

The δ Foundation

Distinction, Arithmetic, and the Limits of Abstraction

Jonathan Washburn

Recognition Physics Research Institute

Fifth edition, version 5.1 · November 2026

Pith:2026.000055v1 [math.LO] 9 Nov 2026

How to read this book

This book follows a construction. We begin with a distinction that can be repeated and ask what is required to preserve its finite records. We build arithmetic and show that each stage of the number tower, from the tallies through the ratios to the line, is the unique solution of its task; for the line this uses countable choice. We examine the rules by which its statements can be proved, identify the additional principles needed for stronger decisions, and compare the resulting rules with Peano arithmetic and Heyting arithmetic. We then price compactness and linear algebra on the line, compute the values of solutions of polynomial differential equations with rational data to certified precision, whenever one computed trajectory passes a finite check, and build a calculus of δ's own objects in whose computable form statements of compactness can be shown not derivable. Part V connects finite descriptions with finite physical records. The last three parts take up the cost of a comparison that Recognition Science places at the base of physics. We prove that its composition law and calibration force one cost, restate the steps of the forcing chain built on it in our own words and price each one, and carry classical mechanics, the second law, linear gravity, the conventions of a classical library and the theory of learning onto the line, saying for each result whether it is free, what it costs, or that it is proved in classical mathematics.

Elementary algebra and experience following a proof are enough to begin. Parts I and II introduce the logical language they use; prior study of category theory is unnecessary. Chapter 9 is a more demanding introduction to a small dependent type theory. Its examples give the main argument, while its final soundness proof requires a slower reading of substitution and reduction. The chapter provides a route through those details rather than assuming they are already familiar.

Part V requires finite-dimensional complex linear algebra: orthonormal bases, eigenvalues, positive matrices and tensor products. Its control arguments also use differentiation, integration and limits of matrix exponentials. Appendix C explains the notation and the specific moves used here; it is a companion to that background, not a replacement for a first course. The central proofs are part of the text, and each physical construction states the operations it assumes.

Parts VI and VII also use real analysis at the level of a first course (Taylor expansions, integrals, finite Markov chains), the elementary theory of the ordinary differential equations of mechanics, and, in Chapters 23 and 27, singular homology and the algebra of symmetric tensors. Each chapter recalls what it uses. Part VIII uses the compactness of Chapter 13, the calculus of Chapter 16, the records of Section 17.7 and the programs of Chapter 20.

The standard of proof is that of ordinary mathematical proof, and every result is proved in the book: where it is stated, or, when the proof is long and routine, in an appendix that the statement's proof names. One heading marks an exception. A proof headed *Standard theorem* reproves a classical result for the book's own calculus and cites the original with its theorem and page; it states the steps of the adaptation, and the appendix it names proves them. The proof for the fixed machine of Theorem 15.33 is of the long and routine kind: Section 15.9 gives the argument, and Appendix I constructs every stage program and proves its bound. No

proof is a sketch. Standard facts of computability theory and general topology are cited where they are used: the unsolvability of the halting problem, the s-m-n and recursion theorems, and the basic properties of $\beta\mathbb{N}$. In Parts VI to VIII a few results rest on stated inputs, and each says so in its scope box: the numbers for the tick of Recognition Science in Chapter 26 assume five numerical inputs, (T1)–(T5), from interval computations the book does not reproduce; the proofs of Theorems 27.32 and 27.33 give their route and the source of each estimate; a few proofs in Chapters 27 and 30 end in finite exact computations that they describe; and the counts of library theorems in Chapter 28, with two size counts of compiled programs there, report an audit and are not theorems. The price table of Appendix Q records the statement and price of every row. In its key, "exact" means that both directions hold; a row with a reference is proved in this book at the result it names, and most rows without one are not proved here. Section 28.12 proves its headline rows, and most rows that this book proves name the result. A few results of the last sections of Chapters 21 to 29 cite a preprint (Zenodo, 2026) for a full treatment, and Theorem 22.73 is cited from one.

Read the examples before the theorem statements. In each theorem, notice the hypotheses: they identify the circumstances in which its conclusion is forced. In each proof, ask where a hypothesis is used. That habit is especially useful when the book compares a construction with a stronger logical principle. An additional principle matters only if it permits a step that was previously unavailable, and that has to be proved. Chapter 6 shows that, with induction on every formula, one of the book's three named rules, in the form the arithmetic language admits, permits nothing new, while another is exactly as strong as excluded middle.

The δ foundation concerns actual mathematics: the structures that distinction and its consistent development require. Making that claim precise requires us to state the finite constructions with which we begin, the reasoning available for studying them, and the conditions attached to each extension. Those details are the argument.

The companion volume, *The Structure of Reality*, develops the physical forcing chain. The present volume gives the mathematical construction room to be taught on its own. Its last three parts restate the steps of that chain in the book's own words, on the line built from ratios, and prove or price each restated step; no result here rests on the companion volume. They then do the same for classical mechanics, the conventions of a classical library and learning. Its physical chapters identify the finite bridge and the resources its constructions use.

The eight parts can be read at different speeds. Part I develops the arithmetic carrier and its extensions. Part II makes proof rules and their logical resources explicit. Part III compares forms of construction and examines the difference between arbitrarily precise approximation and exact decision, including what completeness of the line costs. Part IV prices compactness and linear algebra, counts the work of computing the values of solutions of polynomial differential equations with rational data to a requested precision, whenever one computed trajectory passes a finite check, and builds a calculus of δ's own objects in which the statements of compactness can be written; in its computable form they are not derivable, even from LPO and with excluded middle as a base rule. Part V asks what finite records and specified physical apparatus can preserve. Part VI proves that the composition law and its calibration force the cost J, restates the forcing chain of Recognition Science and prices it step by step, proves that a circle can be linked only in three dimensions, and reads the light cone off the records of a ledger, naming the premises on the way from posting to time, the square law, charges

and space; it also proves that the verdict of Recognition Science holds with no cited premise, prices the moral code, and asks which values the world's records keep real. Part VII carries orbits, rigid bodies, lattice fields, entropy and the second law onto the line and summarizes what the cost and Regge's action say about gravity, through quartic order and the scattering of four gravitons. Part VIII prices the conventions of a classical library and, in a table of 1467 statements, the mathematics that physics uses; it treats the logic of recorded histories and the learning of laws from records, and collects the open problems of the whole book. Read Parts I to V in order on a first pass; each uses distinctions made by the preceding one. Parts VI to VIII rely on Chapters 6, 7 and 12 and on Chapter 21. From Chapter 21 on, a map is read without unique choice (Section 21.1), so prices there are prices of functions, such as the WLPO function. Each of their chapters names the further sections it uses (for example, Chapter 23 uses Chapter 13, Chapter 24 Section 17.7, Chapter 25 Chapter 15, Chapter 28 Chapters 13, 16 and 20, Chapter 29 Chapter 17, and Chapter 30 Chapter 24).

Appendix C explains the proof moves and matrix notation used along the way. Appendix D collects the arithmetic rules, and Appendix N is a notation guide and glossary. Selected exercise solutions are in Appendix O. Appendix P provides routes back to the main arguments, and Appendix Q lists the price table of the mathematics physics uses. Appendices A and B supply the finite-control construction and the count identity used in the physical chapters. Appendices E and F carry the routine verifications behind Chapters 6 and 12, Appendix K the routine cases of the comparison with Peano arithmetic in Chapter 20, Appendix L the proofs of its two extensions, to one universe and to bounded arithmetic, and Appendix M the arguments for Gödel's second theorem and for finite set theory. Appendices G, H, I and J carry the longer proofs of Chapters 14, 15 and 16; Appendix I constructs the machine of Theorem 15.33 stage by stage.

The book brings several kinds of argument together, and it is worth saying which are recovered and which are its own. Recovered: the arithmetic laws, signed differences and ratios reconstruct standard mathematics (Chapters 3 and 4); the initiality and categoricity arguments (Theorems 2.3 and 2.7) have their classical antecedent in Dedekind [2], and the limits of first-order characterization go back to Skolem [3]; Heyting and Peano arithmetic, the omniscience principles and their separations, and Bishop's constructive analysis are the established theories against which the book's rules and prices are measured [4, 5, 14, 20]; and Gödel's incompleteness theorem [8] and Kaye and Wong's reading of finite set theory [13] are proved again for the book's own calculi (Theorems 20.42 and 20.47), as are Löb's theorem [10], the negative and Friedman translations and the no-counterexample interpretation [65, 67, 66, 68, 55] (Theorems 20.49 and 20.57, Lemma 20.62 and Theorem 20.63). Brouwer's fan theorem [29] and its equivalents in constructive reverse mathematics [37, 30, 32] are the background of Chapter 13, Kleene's tree [51] and Lifschitz's realizability [57] supply two of the book's models, and Pouly and Graça's polynomial-time solution of polynomial differential equations [45] is the classical counterpart of Chapter 15, and the classes of questions that finite records confirm, refute and settle are the characterizations of formal learning theory [64] (Theorem 17.13). The paper on the δ-calculus [1] supplies the arithmetic and recognition framework. The book's own contribution is a way of making the strength of each conclusion visible, together with the results that this makes possible: a four-entry ledger that records which additional rules and which kind of induction a derivation uses (Chapter 6); the exact prices of the three rules, which have two strengths with induction on every formula and four without it (Theo-

rem 6.11 and Propositions 6.13 and 6.14); one classification priced under six combinations of supplied information and requested answer, with the six prices separated pairwise by models of intuitionistic second-order arithmetic (Theorems 7.33 and 7.38); the typed separations of iteration, indexed recursion, dependent elimination and the start discriminator (Chapter 9); the exact prices of the decisions and completeness properties of the line built from ratios (Chapter 12 and Table 12.3); the prices of compactness, where Heine–Borel for $[0, 1]$ and for every compact set presented with its nets is exactly the decidable fan theorem, with no choice principle (Theorems 13.6 and 13.25, Table 13.1); the prices of linear algebra on the line, where selecting a basis and rank–nullity, with independence meaning that only the zero combination vanishes, are each exactly WLPO, with unique choice, part of the reading of maps, used in the converse (Theorem 14.16); a solver for polynomial differential equations with rational data whose answers carry a proved error bound and whose arithmetic is counted exactly, and one fixed one-tape binary machine that, for every problem whose region certificate holds at the level given in the request, answers every request for m bits within $K(m + 1)^{\lambda}$ steps, with one exponent λ for all problems and a constant K that depends on the problem, a machine that does not check the certificate and reads a polynomial field through the request of its quadratized problem (Theorems 15.22, 15.23 and 15.33); a calculus of δ's own objects into which second-order intuitionistic arithmetic compiles, and which with all its terms derives every true sentence built from equations over carriers that do not change between its two stages (Proposition 16.15), while its computable form, the computable calculus, has a model, assembled from Kleene's tree and standard counterexamples of recursive analysis [51, 76, 77, 40], in which LPO is true, so that each question is decided, but the answers cannot be collected into a sequence and the decidable fan theorem and Heine–Borel for presented compact sets, written with equations, are false, so that neither is derivable there, even from LPO and with excluded middle as a base rule (Theorems 16.11, 16.27, 16.32 and 16.34); what deciding questions about growing records costs: a question that a given recorder settles, by a proof without excluded middle, is decided with the binary Markov principle, and one settled question costs exactly that principle (Theorem 17.17); the verdict of a law tested at finite precision, which always exists, and the price of exactness: deciding an equation costs WLPO, and turning its failure into a refutation costs the binary Markov principle (Theorem 17.22); models of intuitionistic second-order arithmetic for prices strictly between pairs of the six (Corollary 7.49), where the principles $\mathrm{MP}^{\vee}$, WMP and LLPO_n and several of their separations are known (Section 7.13); the price of converting a classical theorem into one without excluded middle, shape by shape, for the shapes of the book's arithmetic language, in which a quantifier-free formula has no bounded quantifiers (Theorem 20.66); the comparison of a language of functions with the rules and with Peano arithmetic, compiled rule by rule (Theorems 20.26 and 20.39); and a finite bridge to physical records that carries effective names, controls, errors, resources and a conservation bound through one account (Theorems 18.11 and 19.5). With excluded middle, the rules prove exactly the theorems of Peano arithmetic (Theorem 8.11), and with no added principle exactly those of Heyting arithmetic. These claims belong to the specified languages and representations.

The last three parts add results of a different kind: what the physical derivations of Recognition Science, and the classical mathematics that physics uses, cost on the line. On $\mathbb{R}_{\delta}$ the composition law and its calibration force the cost J with no omniscience principle and no choice (Theorem 21.14), and one bit of calibration, $J(2) = \frac{1}{4}$, with any one mild regularity condition

such as a bound on one interval, already does (Theorem 21.42); the verdict of Recognition Science holds with no cited premise, its transcendence input proved in classical mathematics (Theorem 22.58); of the 130 rows of the price table of the forcing chain, as the book restates it, 119 are free, ten have an exact price, one is bracketed and none is open (Chapter 22); three is the only dimension in which a circle can be linked, again with no omniscience and no choice, which is the topological fact behind the dimension step (Chapter 23); the mechanics of Kepler and Newton, the perihelion advance of relativistic test orbits to second order and the second law hold with explicit constants, on the line or, where the chapters say so, in classical mathematics (Chapters 25 and 26); on transverse-traceless strains the long-wave form of J is the Fierz–Pauli form of linearized gravity, and the cost built from J agrees with Regge's action, read as the gravitational action, to second order only, so at the Euclidean background of the lattice J has no post-Newtonian parameter β of its own, while the long-wave vertices of Regge's action on the Kuhn lattice are one half of Einstein's through quartic order on every Euclidean background of the Kuhn family, and its gravitons scatter at tree level and long wavelength as general relativity's on every flat background of the family off its poles (Chapter 27); with the code's hypothesis in its free form, every conclusion of the main theorem of the moral code is free on the line, and only decisions over every conceivable record carry a price (Section 22.14); the conventions of a classical library have prices on the line, most of them exact, and of 1467 statements of the mathematics physics uses 1166 are free, 279 have an exact price in the table's sense and 22 are open (Chapter 28 and Appendix Q); and learning a law from records has exact prices too (Chapter 29). Recovered there are the laws of Kepler, Newton and Lagrange, the Laplace–Runge–Lenz vector, Chetaev's instability argument, Doeblin's contraction, Regge's calculus and the Fierz–Pauli form, Ostrowski's theorem, de Finetti's theorem, Morse and Hedlund's bound, Littlestone's dimension and the learning criteria of Gold and of Case and Smith, each credited where it is used.

The physical chapters use established matrix and dilation theory [22, 23, 24]. Their task is to connect complete finite descriptions to the stated register, controls and error accounting, and to derive a recorder bound for the specified conserved charges. The conservation argument belongs to the line of work begun by Wigner and Araki and Yanase [26, 27].

Contents

How to read this book **ii**

I Distinction becomes arithmetic 1

1 A distinction that can be repeated 2

1.1 Start, then take one more step . . . 2

1.2 Why a proof can follow the construction . . . 3

1.3 A familiar notation, with a proved translation . . . 4

1.4 From an orbit to a realization . . . 5

Questions to work through . . . 5

2 Why a faithful tally has one structure 6

2.1 The data a translation must preserve . . . 6

2.2 There is exactly one reading . . . 7

2.3 What makes the reading faithful? . . . 7

2.4 The exact uniqueness result . . . 9

2.5 Where the force of the argument lies . . . 10

Questions to work through . . . 10

3 Addition, multiplication, and order 11

3.1 Addition follows the second record . . . 11

3.2 Multiplication repeats an addition . . . 12

3.3 Why multiplication distributes . . . 13

3.4 The step determines both operations . . . 14

3.5 Comparing two finite records . . . 15

3.6 Subtraction stops at the start . . . 16

Questions to work through . . . 17

4 Differences and ratios 18

4.1 Keep both sides of a difference . . . 18

4.2 Operations must survive a change of representative . . . 19
4.3 The integer laws follow on the classes . . . 20
4.4 Reading the signed construction . . . 21
4.5 A ratio needs a nonzero denominator . . . 21
4.6 Constructing arithmetic on ratio classes . . . 22
4.7 Why a reciprocal keeps the sign in the numerator . . . 23
4.8 Checking the field laws . . . 24
4.9 Each extension is determined by its task . . . 25
Questions to work through . . . 27

II What a proof is allowed to use 28

5 The language and the rules 29
5.1 Expressions are finite constructions . . . 29
5.2 One complete argument . . . 30
5.3 The permitted inferences . . . 30
5.4 Reading the expressions as arithmetic . . . 32
5.5 Why every permitted step preserves truth . . . 34
5.6 The same induction, now inside the language . . . 35
5.7 What each family of rules supplies . . . 35
Questions to work through . . . 39

6 The price of an additional principle 40
6.1 A witness, a decision, and a double negation . . . 40
6.2 A ledger for a finite argument . . . 42
6.3 No added principle is not the same as an empty ledger . . . 42
6.4 The record agrees with the derivation . . . 43
6.5 Soundness with the prices attached . . . 44
6.6 A route's price and a theorem's exact strength . . . 45
6.7 What the three rules cost . . . 45
Questions to work through . . . 61

7 What can recognize the same record? 62

7.1 Identifications that respect addition 62

7.2 How a period reaches an earlier collision 63

7.3 The constructive classification 64

7.4 What changes when the collision is not supplied 65

7.5 What a decision procedure alone can buy 66

7.6 The unrestricted catalogue and excluded middle 69

7.7 What the classification tells us about a reader 70

7.8 Signed records keep only the pure periods 71

7.9 One classification, six prices 73

7.10 A system and its models of truth values 74

7.11 The six equivalences inside the system 80

7.12 Four models, and why the six prices are different 89

7.13 Prices between the six 94

7.14 One halting question, priced instance by instance 102

Questions to work through 114

8 Changing the representation 116

8.1 An interpretation is fixed by its operations 116

8.2 Arithmetic on lists of one mark 117

8.3 Terms agree under corresponding assignments 118

8.4 Truth is preserved in both directions 119

8.5 What survives a one-way translation 120

8.6 What happens to the proof and its ledger 122

8.7 A second proof language: Peano arithmetic 122

8.8 A quotient represents what a reading retains 124

8.9 The limits are part of the invariance theorem 125

Questions to work through 125

III Construction, iteration and the continuum 127

9 Iteration, recursion and induction 128

9.1 Three ways to proceed from a start 128

9.2 The language in which the question is exact . . . 129
9.3 What indexed recursion immediately gives . . . 131
9.4 An iterator can merge two roots . . . 132
9.5 Recursion can leave a whole chain unreached . . . 133
9.6 Induction does not invent a discriminator . . . 135
9.7 The strict sequence of resources . . . 136
9.8 Why the countermodels cover every rule . . . 137
9.8.1 A total interpretation and its limits . . . 137
9.8.2 Substitution and safe expressions . . . 138
9.8.3 Joining conversions without changing the language . . . 139
9.8.4 Soundness of the complete typing derivations . . . 141
Questions to work through . . . 143

10 Finite construction and effective description 144
10.1 An explicit budget for a tally . . . 144
10.2 Arithmetic carries its own finite bounds . . . 145
10.3 A record with a decoder . . . 146
10.4 A finite rule can describe an unbounded calculation . . . 146
10.5 Storing the program preserves its execution . . . 148
10.6 One program for every requested precision . . . 149
10.7 A finite prefix is a different quantifier . . . 150
Questions to work through . . . 151

11 The continuum through rational enclosures 152
11.1 Why every later interval constrains every earlier one . . . 152
11.2 A unique value lies in every enclosure . . . 153
11.3 When two protocols describe the same point . . . 153
11.4 Rationals and arithmetic act on the intervals . . . 154
11.5 Every classical real has such a display . . . 155
11.6 Why finite prefixes do not provide a uniform name . . . 157
Questions to work through . . . 158

12 Exact comparison has a cost **159**

12.1 Finite stages and exact predicates . . . 159

12.2 Questions about a binary sequence . . . 161

12.3 Encoding an event as a small jump . . . 161

12.4 Exact zero decisions are equivalent to WLPO . . . 163

12.5 Sign trichotomy is equivalent to LPO . . . 163

12.6 One-sided order is equivalent to LLPO . . . 164

12.7 The difference between denying and finding . . . 165

12.8 One line built from ratios . . . 168

12.9 What completeness costs . . . 175

12.9.1 Countable choice decides nothing . . . 175

12.9.2 Limits and located cuts: where countable choice enters . . . 181

12.9.3 The line is determined by its task . . . 189

12.9.4 The table of completeness prices . . . 192

12.9.5 Order tools for the proofs of the table . . . 194

12.9.6 Deciding a point against a ratio . . . 195

12.9.7 Two cuts that bound a set . . . 196

12.9.8 Monotone sequences and sequential compactness cost LPO . . . 197

12.9.9 Least upper bounds cost WLPO, LPO, WEM or EM . . . 200

12.9.10 Roots, maxima and expansions cost LLPO or nothing . . . 203

12.9.11 Without unique choice, excluded middle does not replace countable choice . . . 211

12.9.12 A located cut that unique choice does not make a point . . . 215

12.9.13 Countable choice proves the four assertions without excluded middle . . . 218

12.9.14 LLPO does not replace countable choice . . . 219

12.9.15 The same prices away from the line . . . 225

12.9.16 Monotone convergence splits as LPO does . . . 229

12.10 A price table for the calculus on the line . . . 230

12.11 Integrals over the whole line and space, and what completeness costs . . . 241

Questions to work through . . . 247

IV Compactness, algebra and computation 249

13 What compactness costs 250

13.1 Bars and the fan theorems . . . 251
13.2 Heine–Borel on $[0, 1]$ is the decidable fan theorem . . . 252
13.3 Where the fan theorem sits . . . 255
13.4 Uniform continuity . . . 258
13.5 Compact sets presented with their nets . . . 263
13.6 Heine–Borel, positive bounds and infima for presented sets . . . 265
13.7 Heine–Cantor for presented sets . . . 270
13.8 Bishop compactness . . . 272
Questions to work through . . . 275

14 What linear algebra costs 276

14.1 Selections, bases and three kinds of division . . . 277
14.2 Cancelling without inverting . . . 278
14.3 A basis is a sequence of divisions . . . 279
14.4 Linear algebra costs exactly WLPO . . . 282
14.5 The sign of a nonzero point . . . 284
14.6 Dividing by a nonzero point . . . 286
14.7 Signs as units . . . 290
14.8 The prices . . . 292
14.9 Without unique choice . . . 292
Questions to work through . . . 292

15 Differential equations to certified precision 293

15.1 A number is a program with a proof of its precision . . . 293
15.2 Carrying every rounding error forward . . . 296
15.3 The number of steps at a fixed order . . . 297
15.4 Taylor coefficients of a quadratic field . . . 298
15.5 A fixed number of steps of growing order . . . 301
15.6 Counting the work . . . 303
15.7 Polynomial fields become quadratic . . . 305

15.8 What the certified point is 306
15.9 One machine for every problem and every precision 307
Questions to work through 314

16 A calculus of δ's own objects 315
16.1 Objects that exist at two stages 316
16.2 Terms, formulas and their reading 317
16.3 Derivations and the twenty-nine base rules 319
16.4 Every derivation holds at both stages 321
16.5 Intuitionistic second-order arithmetic inside the calculus 323
16.6 The full calculus has no recursive model 326
16.7 The computable calculus and its model 329
16.8 LPO decides each row but cannot collect the rows 333
16.9 Kleene's tree 335
16.10 Compactness is not derivable in the computable calculus 336
16.11 What remains open 342
Questions to work through 343

V Mathematics and finite physical records 344

17 What a finite record can establish 345
17.1 The target, the record and the reader 345
17.2 Several worlds can remain compatible with one record 346
17.3 A finite capacity bound 347
17.4 Adaptive tests still produce a bounded transcript 348
17.5 An error tolerance gives a separation bound 349
17.6 Finite length without a common budget 350
17.7 What a record can confirm, refute or settle 351
17.8 The physical task left by the mathematics 360
Questions to work through 361

18 The construction of finite operations 362
18.1 An operation with a retained outcome 362

18.2 An explicit reversible extension 364
18.3 A fixed register and a complete reader 365
18.4 An error bound that keeps a reference system 366
18.5 Two descriptions of finite operations 367
18.6 Searching with a margin 368
18.7 Recovering the matrix description 369
18.8 Enacting a sequence with positive energy 370
18.8.1 Retaining the phase 371
18.9 The finite bridge and its physical interfaces 372
Questions to work through 374

19 Conservation and the cost of keeping a record 375
19.1 The event, the interaction and the pointer 375
19.2 Accuracy controls the noise of a sharp event 376
19.3 What exact conservation does to the pulled-back event 377
19.4 The native source and its two charges 378
19.5 A four-entry witness inside the actual source 379
19.6 How much charge spread can the recorder supply? 380
19.7 The two-charge recorder theorem 381
19.8 What the resource is, and what the result establishes 382
Questions to work through 383

20 What has been forced 384
20.1 The arithmetic structure is unique 384
20.2 Identifying records and extending logic are visible choices 385
20.3 Peano arithmetic and a richer language 386
20.3.1 The classical derivations are Peano arithmetic 387
20.3.2 A language of functions 387
20.3.3 Computations described by arithmetic 390
20.3.4 A universal program whose correctness the rules derive 392
20.3.5 Equality that follows the type 397
20.3.6 Compiled terms return admissible values 398
20.3.7 Substitution moves with the environment 400

20.3.8 Every rule becomes a derivation . . . 404
20.3.9 Recovering the arithmetic sentence . . . 407
20.3.10 The comparison with Peano arithmetic . . . 409
20.3.11 Consistency, incompleteness and finite sets . . . 414
20.3.12 Löb's theorem and a tower of consistency statements . . . 418
20.3.13 A small case . . . 423
20.4 What a classical proof gives without excluded middle . . . 424
20.4.1 Translating a classical derivation step by step . . . 424
20.4.2 Two quantifiers: a witness and a program . . . 426
20.4.3 Friedman's translation . . . 428
20.4.4 Three quantifiers: a strategy against counterexamples . . . 429
20.4.5 What each shape of classical theorem costs . . . 432
20.5 Finite access and a uniform construction differ . . . 433
20.6 The physical bridge has a construction and a constraint . . . 434
20.7 What a further claim must show . . . 436
Questions to work through . . . 437

VI The comparison cost and its forcing chain 438

21 The cost of a comparison 440
21.1 Reading maps without unique choice . . . 441
21.2 Logarithm, sine, cosine and π on the line . . . 442
21.3 The composition law forces J . . . 447
21.4 J is the cost of one fair binary distinction . . . 454
21.5 The ledger is $\mathbb{Z}[\varphi]$ . . . 456
21.6 What deciding a cost law costs . . . 460
21.7 A certain change costs a quarter cycle . . . 466
21.8 One binary distinction fixes the cost . . . 470
21.9 Bounds force the curvature . . . 474
21.10 The prices of the branches . . . 475
21.11 Exact values of the cost . . . 477
21.12 The cost on ratio records . . . 479

Questions to work through 481

22 The forcing chain of Recognition Science, priced 483

22.1 Steps, native statements and prices 484
22.2 The meter reads ratios 485
22.3 The apparatus and the native reporting theorem 487
22.4 What deciding the consequences of J costs 491
22.5 The golden hierarchy 492
22.6 Three readings of the first costly join 494
22.7 Eight states, and where three dimensions come from 496
22.8 The second rung 497
22.9 The constants 506
22.10 The numbered chain, step by step 512
22.11 Where the prices sit 520
22.12 The verdict with no cited premise 528
22.13 The brackets of the chain, closed 536
22.14 The moral code, priced 538
Questions to work through 547

23 Why three dimensions 549

23.1 Spheres, simplices and chains built from the line 550
23.2 The theorem: a circle links only in the 3-sphere 552
23.3 Tools that work in every dimension 553
23.4 The suspension ladder 558
23.5 From a linked circle to the third homology of the sphere 561
23.6 Dimensions zero and one 563
23.7 The 3-sphere links: a loop that sweeps four quarter turns 564
23.8 Room in four dimensions: pulling a loop off the equatorial circle 568
23.9 Why three 570
23.10 The winding number as data, and its price 572
Questions to work through 575

24 Ledgers, rates and the light cone **576**

24.1 One account fills a light cone . . . 579
24.2 The cost of a ratio reads the motion . . . 581
24.3 Proper time is superadditive, and the deficit is J . . . 583
24.4 The recognition worldline . . . 584
24.5 Many accounts reach only the taxicab cone . . . 588
24.6 The record cone of one coherent binary distinction . . . 590
24.7 Where the distinction is carried . . . 594
24.8 Least comparison cost is greatest proper time . . . 598
24.9 Clocks: resolution, self-recognition and the action floor . . . 602
24.10 What one balance history can be asked . . . 609
24.11 From records to time, quantum, matter and space . . . 616
24.12 Which values the world keeps real . . . 626
24.13 The eight-tick comparison at its exact scope . . . 632
24.14 What a store with few addresses keeps real . . . 633
24.15 Certifying a law at finite precision . . . 635
24.16 Crossings and postings . . . 637
24.17 Controlled posts and the coherence of a side . . . 638
24.18 Charts of the causal order of postings . . . 641
Questions to work through . . . 646

VII Classical physics on the line **647**

25 Orbits and bodies **648**

25.1 Motions on the line . . . 650
25.2 Kepler from Newton . . . 654
25.3 Other central forces . . . 659
25.4 Bound orbits return . . . 661
25.5 The N-body problem . . . 665
25.6 The relativistic orbit equation . . . 670
25.7 The perihelion advance of actual orbits . . . 675
25.8 What Recognition Science supplies here . . . 684

25.9 Certified runs are motions . . . 689
25.10 The rigid body . . . 696
25.11 Pulsar timing in far time . . . 701
25.12 The point mass from the lattice's own field equations . . . 704
Questions to work through . . . 707

26 Fields, heat and the second law **709**
26.1 Entropy and the Gibbs principle on the line . . . 710
26.2 The second law for a Markov step . . . 716
26.3 What the second law gives Recognition Science's dynamics . . . 719
26.4 Recognition Science's tick: walls, windows and a rate . . . 722
26.5 The Metropolis tick on the line, and a gap for every ball . . . 729
26.6 The pace of the spine . . . 732
26.7 Fields on the lattice . . . 743
26.8 Laws and their prices . . . 749
26.9 Thermal light on the lattice . . . 755
26.10 Kepler's motion for all time . . . 760
Questions to work through . . . 760

27 Gravity: the long-wave form of the cost and Regge's action **762**
27.1 The long-wave form of J is the Fierz–Pauli form on transverse strains . . . 763
27.2 The Lorentzian identity on $\mathbb{R}_\delta$ . . . 766
27.3 Completions, locality and the scale . . . 769
27.4 Holonomies, deficits and the sign of the graviton term . . . 774
27.5 At third order the cost breaks gauge invariance . . . 779
27.6 The Regge table for every background . . . 784
27.7 The near-flat chain on $\mathbb{R}_\delta$ . . . 788
27.8 The background $\kappa = 1$ at its exact scope . . . 796
27.9 The locality clause is a named premise . . . 800
27.10 What reading the cost as the graviton's action needs . . . 805
27.11 Regge's action is one half of Einstein's through quartic order . . . 811
27.12 Among area charges, gauge consistency selects Regge's action through cubic order . . . 817

27.13 Hinge shape costs cannot complete a changed coupling 822
27.14 Four gravitons at tree level 824
27.15 The graviton's dispersion beyond long wavelength 827
27.16 The premises of reading the cost as the graviton's action 829
Questions to work through 831

VIII The library, learning and open problems 833

28 The classical library on the line 834
28.1 Tests and decisions as data 835
28.2 Conventions at excluded points 837
28.3 Choosing representatives 842
28.4 The linear order and the ordered field 845
28.5 Suprema, separation and compactness 848
28.6 The derivative convention 852
28.7 Monotonicity from a pointwise derivative 855
28.8 WLPO and Π^0_2 weak excluded middle 858
28.9 Names, a space-filling curve and tail signs 861
28.10 Moving the library: density, cones and counts 866
28.11 Classical proofs as programs of Gödel's T 871
28.12 What the mathematics of physics costs 883
Questions to work through 897

29 Time, chance and learning 899
29.1 The logic of one recorded history 900
29.2 Chance from exchangeable records 905
29.3 The golden record, the places of ratios and growth 910
29.4 Fixed points on δ's triangle 914
29.5 What identifying a law costs 916
29.6 Mistakes, dimensions and samples 920
29.7 Learners that are programs 923
29.8 Learning with the cost J 926
29.9 Learning on the programs of Gödel's T 929

29.10 Methods, bounds and decision tables . . . 933
29.11 How many revisions a catalogue forces . . . 941
29.12 Catalogues at every level of the arithmetical hierarchy . . . 948
Questions to work through . . . 954

30 Conjectures and refutations **956**
30.1 Calibrating a phase with golden probes . . . 957
30.2 Two recognizers, and the memory that recognition needs . . . 962
30.3 Self-similar ledgers: rates, balance and fluctuation . . . 964
30.4 The price of a certain change . . . 972
30.5 Open problems . . . 987
Questions to work through . . . 995

A Finite controls and approximation **996**
A.1 The matrix directions that suffice . . . 996
A.2 From finite operations to continuous directions . . . 997
A.3 Two native cells supply all matrix directions . . . 998
A.4 Overlapping pairs control a whole register . . . 1000
A.5 A finite alphabet has an infinite-order seed . . . 1002
A.6 Sign masks isolate the seed's edges . . . 1003
A.7 The finite alphabet connects the cells . . . 1004
A.8 A complete record-and-field assembly . . . 1005

B Deriving the conserved native count **1009**
B.1 Separate the spectral gaps of the clock . . . 1009
B.2 The entries needed for the certificate . . . 1009
B.3 A finite expression for the odd-mode projection . . . 1010
B.4 Transport to a conserving recorder . . . 1012

C A toolkit for reading the proofs **1013**
C.1 Read a theorem as a task . . . 1013
C.2 Induction follows a finite construction . . . 1013
C.3 Maps, round trips and quotients . . . 1014
C.4 Quantifiers say what information must be supplied . . . 1015

C.5 Equivalences and countermodels answer different questions 1015
C.6 Reading the finite matrices 1016
C.7 Eigenvalues, positivity and evolution 1017
C.8 Keeping both systems in the calculation 1017

D The rules at a glance **1019**
D.1 Reading the notation 1019
D.2 The arithmetic base 1019
D.3 The logical rules 1020
D.4 Additional schemas and their records 1020
D.5 The three orbit constructions 1021

E The prices of the three rules: supporting proofs **1023**
E.1 Arithmetic inside the rules 1023
E.2 Structures with two stages 1025
E.3 Realizers 1027
E.4 The halting matrix 1029

F Proofs for the continuum chapter **1032**
F.1 Realizers for the rules 1032
F.2 Located cuts and Archimedean ordered fields 1033
F.3 Limits in the recursive model 1038
F.4 Stage estimates on the line 1039

G Proofs for the linear-algebra chapter **1042**
G.1 Matrices 1042
G.2 Unique roots and monotone maps 1044
G.3 Without unique choice 1046

H Proofs for certified computation **1049**
H.1 The size of every operand 1049
H.2 The classical solution value 1051
H.3 Polynomial fields 1054

I The machine of Theorem 15.33 **1055**

I.1 The tape: rows, tracks and marks . . . 1055
I.2 Programs, counted runs and loops . . . 1057
I.3 Moving along the tape . . . 1059
I.4 Counters, columns and rows . . . 1060
I.5 Binary arithmetic in rows . . . 1064
I.6 The machine's own integers . . . 1069
I.7 Reading the request . . . 1071
I.8 The problem's constants: preparing the request and folding the field . . . 1087
I.9 The problem's constants: the schedule and the width . . . 1109
I.10 The first vector . . . 1124
I.11 One coefficient of the Taylor table . . . 1138
I.12 The Taylor sum of a step . . . 1163
I.13 The step loop and the selected word . . . 1183
I.14 Rounding and emitting the answer . . . 1187
I.15 Framing the answer and erasing the workspace . . . 1194
I.16 The whole count . . . 1198

J Proofs for the calculus of δ's objects **1202**

J.1 Equivalence and reindexing . . . 1202
J.2 The remaining cases of the compiler . . . 1203
J.3 The full calculus derives every true equational sentence . . . 1203
J.4 Programs that track the maps . . . 1206
J.5 Reading the sentences in the model . . . 1207
J.6 Addresses and the dead cover . . . 1209
J.7 The cover series as a named map . . . 1209
J.8 The tent function . . . 1209

K The comparison with Peano arithmetic: routine cases **1211**

K.1 Order and pairing . . . 1211
K.2 The computation formula . . . 1212
K.3 The record test . . . 1215
K.4 Closures . . . 1217

K.5 Reading a variable . . . 1217
K.6 The conversion cases . . . 1218
K.7 The type conversion case . . . 1218

L One universe and bounded arithmetic: the proofs 1219

L.1 The checker and the sentence Con(PA) . . . 1219
L.2 The calculus with one universe and its arithmetic . . . 1226
L.3 The interpretation of codes . . . 1227
L.4 Environments and the syntactic lemmas . . . 1228
L.5 The checker is sound inside $\mathscr{F}_U$ . . . 1229
L.6 From the core to the sentence . . . 1231
L.7 The embedding and the strictness . . . 1231
L.8 The set model . . . 1232
L.9 Completeness and the models of bounded arithmetic . . . 1236
L.10 Arithmetic in a model of bounded arithmetic . . . 1237
L.11 The exponentiation graph . . . 1238
L.12 Pair formation from exponentiation . . . 1240

M Incompleteness and finite set theory: the proofs 1242

M.1 Models of Peano arithmetic . . . 1242
M.2 Codes and certificates . . . 1242
M.3 The first, second and fourth conditions . . . 1244
M.4 The bridge to the consistency sentence . . . 1246
M.5 The third condition . . . 1247
M.6 Finite set theory and binary membership . . . 1248

N Notation and glossary 1250

N.1 Arithmetic and maps . . . 1250
N.2 Formulas, judgments and resources . . . 1251
N.3 Typed constructions and finite descriptions . . . 1253
N.4 Physical records and finite matrices . . . 1255
N.5 Glossary . . . 1256
N.6 The cost, physics and learning (Parts VI to VIII) . . . 1260

O Selected solutions **1266**

O.1 Distinction and rigidity . . . 1266
O.2 Arithmetic and its extensions . . . 1267
O.3 Rules and logical resources . . . 1268
O.4 Recognition and representation . . . 1270
O.5 Iteration and effective description . . . 1274
O.6 Enclosures and exact comparison . . . 1274
O.7 Compactness, algebra and computation . . . 1276
O.8 Records and physical construction . . . 1281
O.9 Conservation and the closing argument . . . 1283
O.10 The comparison cost and its forcing chain . . . 1284
O.11 Classical physics on the line . . . 1292
O.12 The library, learning and open problems . . . 1299

P Finding an argument **1304**

P.1 What determines arithmetic? . . . 1304
P.2 What does a proof use? . . . 1304
P.3 Which constructions give stronger eliminations? . . . 1305
P.4 Does a richer language prove more about numbers? . . . 1305
P.5 When does finite information give one uniform procedure? . . . 1306
P.6 What can a calculus of δ's objects derive? . . . 1306
P.7 What can a physical record preserve? . . . 1307
P.8 What does a comparison cost? . . . 1307
P.9 What does the forcing chain of Recognition Science cost? . . . 1308
P.10 Why three dimensions? . . . 1308
P.11 What does one account's ledger contain? . . . 1308
P.12 What does classical physics cost on the line? . . . 1309
P.13 What does the cost say about gravity? . . . 1309
P.14 What do a library's conventions cost? . . . 1310
P.15 What does learning cost? . . . 1310
P.16 Which conjectures are settled, and which stay open? . . . 1310
P.17 Headline results and what their proofs cite . . . 1311

Q The price table of the mathematics physics uses **1314**

How to read a price 1314

Q.1 Conventions at excluded points 1315

Q.2 Limits, continuity and derivatives 1316

Q.3 Exponential, logarithm, trigonometric and complex functions 1328

Q.4 Integrals and measure 1352

Q.5 Differential equations and dynamics 1355

Q.6 Linear algebra and spectra 1360

Q.7 Spectra, lattice models and thermodynamic limits 1400

Q.8 Topology, compactness and fixed points 1402

Q.9 Probability 1413

Q.10 Functional and Fourier analysis 1420

Q.11 A first survey across the areas 1425

Bibliography **1437**

Part I

Distinction becomes arithmetic

Chapter 1

A distinction that can be repeated

Suppose we record arrivals by making a mark. Before anyone arrives, there is no mark. The first arrival adds one. The next arrival adds another. We can change the pen, the shape of the mark, or the name we give the record. What must survive those changes if the record is still to count the same arrivals?

The question is small enough to answer completely. It is also the beginning of a much larger question: how much mathematics is already determined when we insist that distinctions can be made, retained, and repeated faithfully?

We begin with the record itself. We need a starting record and an operation that adds one step. We then ask what their finite repetition constructs. The arithmetic operations will come later. At this point, even the familiar decimal numerals would conceal more than they explain.

1.1 Start, then take one more step

Write 0_δ for the starting record. Write S for the operation that takes one further step. The first few expressions are

$$0_\delta, \qquad S(0_\delta), \qquad S(S(0_\delta)), \qquad S(S(S(0_\delta))).$$

Each expression says how it was built. The last expression, for example, starts at 0_δ and applies the step three times. Nothing in this notation specifies the material of a recorder or the duration of an event. Those are questions about a realization. Here we are constructing the finite records that a faithful realization must distinguish.

Definition 1.1 (The finite distinction orbit)**.** The orbit $\mathbb{N}_\delta$ consists of the finite expressions generated by these two rules:

1. 0_δ is an expression.
2. If a is an expression, then $S(a)$ is an expression.

There are no further ways to construct an expression. Equality is equality of the constructed expressions: the starting expression is different from every step expression, and two step expressions are equal exactly when their inner expressions are equal.

The final sentence matters. A device might return to an earlier state after several updates. Such a device can read the orbit, but its states do not retain every distinction made in the orbit. We will study identifications of records later. First we need the object before any such identification (Figure 1.1).

Two facts can now be read directly from the construction.

$$0_\delta \xrightarrow{\ S\ } S(0_\delta) \xrightarrow{\ S\ } S(S(0_\delta)) \longrightarrow \cdots$$

Figure 1.1. The orbit is generated from a start by finite repetition of a step. The arrows describe construction. No period or identification has been imposed.

Proposition 1.2 (A start is not a step; steps preserve distinction)**.** *For all $a, b \in \mathbb{N}_\delta$,*

$$0_\delta \neq S(a), \qquad S(a) = S(b) \implies a = b.$$

Proof. The expression 0_δ is built by the first rule. An expression $S(a)$ is built by the second. The equality of expressions does not identify these different outer forms, so $0_\delta \neq S(a)$.

If $S(a) = S(b)$, both expressions have the same outer constructor. Equality of these expressions therefore requires equality of the expressions to which that constructor was applied. Hence $a = b$. □

The second property is called *injectivity* of the step: different inputs cannot become the same output after one step. It is the elementary form of preserving a distinction. The first property says that stepping never produces the starting record. Together they tell us how to compare constructed records without assuming arithmetic in advance.

1.2 Why a proof can follow the construction

A finite expression has a finite construction. To establish a property of all expressions, it is enough to establish it at the start and show that one further step preserves it. The proof follows the only two ways an expression can be formed.

Proposition 1.3 (Induction on the orbit)**.** *Let $P(a)$ be a property of orbit expressions. Suppose that*

1. *$P(0_\delta)$ holds;*
2. *for every $a \in \mathbb{N}_\delta$, $P(a)$ implies $P(S(a))$.*

Then $P(a)$ holds for every $a \in \mathbb{N}_\delta$.

Proof. Take an expression together with its finite construction. If it was built by the starting rule, the first premise supplies the property. If it was built by applying the step to an earlier expression, the property of that earlier expression and the second premise supply the property of the new one. Applying this reasoning along the finite construction establishes the property at its final expression. These are all the construction rules, so they cover every expression in the orbit. □

This proof makes a background commitment visible. We are reasoning about finite inductively generated expressions and are allowed to follow their construction. That is the setting in which the orbit has been defined. The resulting induction principle belongs to that setting.

We have not produced a language, a notion of finite construction, or the rules for reasoning about it from an absence of all rules.

Keeping this point explicit helps us ask a sharper question. Given this construction, which further structures are determined? And when a stronger conclusion needs a further principle, where does that principle enter? Much of this book is devoted to making those dependencies visible.

1.3 A familiar notation, with a proved translation

The expressions become cumbersome quickly. We would like to use ordinary natural-number notation without losing the construction it abbreviates. For that purpose, construct two maps.

The first map, $r : \mathbb{N}_\delta \to \mathbb{N}$, reads an orbit expression as its ordinary natural-number index:

$$r(0_\delta) = 0, \qquad r(S(a)) = r(a) + 1. \tag{1.1}$$

The second map, $e : \mathbb{N} \to \mathbb{N}_\delta$, expands an index into an orbit expression:

$$e(0) = 0_\delta, \qquad e(n+1) = S(e(n)). \tag{1.2}$$

Here $\mathbb{N}$ is the usual natural-number system used in the surrounding mathematics. Introducing it here provides a comparison with familiar notation. The orbit itself was already specified by its finite construction.

Theorem 1.4 (Reading and expansion are inverse)**.** *For every $n \in \mathbb{N}$ and $a \in \mathbb{N}_\delta$,*

$$r(e(n)) = n, \qquad e(r(a)) = a.$$

Consequently, reading and expansion give mutually inverse identifications between the two carriers.

Proof. First prove $r(e(n)) = n$ by induction on n. At zero,

$$r(e(0)) = r(0_\delta) = 0.$$

Suppose $r(e(n)) = n$. Then the recursive rules give

$$r(e(n+1)) = r(S(e(n))) = r(e(n)) + 1 = n + 1.$$

This proves the first identity.

For the second identity, use induction on the orbit. At the starting expression,

$$e(r(0_\delta)) = e(0) = 0_\delta.$$

If $e(r(a)) = a$, then

$$e(r(S(a))) = e(r(a) + 1) = S(e(r(a))) = S(a).$$

The orbit induction principle therefore proves $e(r(a)) = a$ for every a. The two identities are exactly the assertion that the maps are inverse. □

The result allows notation to become lighter without becoming ambiguous. When it is useful, we can write 1_δ for $S(0_\delta)$ and 2_δ for $S(S(0_\delta))$. These names refer to particular expressions. They do not add operations or impose equations on the orbit.

The proof also shows what a successful change of representation should provide. A proposed correspondence must do more than assign labels. It must tell us how to go in both directions and establish that no information is lost on either return journey. Here both directions are explicit.

1.4 From an orbit to a realization

The constructed orbit and a physical tally answer different questions. The orbit specifies the finite records. A tally implements a way of representing them. To examine that representation, we need to say what the tally does at the start and what it does after one more step.

This leads to the next chapter's central construction. Let another system have a starting point b and an update s. There is an evident way to read the orbit into it: send 0_δ to b, then send each additional orbit step to one application of s. What is less obvious, until proved, is that these two requirements leave no other reading.

That uniqueness does not yet say that the target retains every distinction. We will prove the unique reading first, then identify the additional conditions under which it is a one-to-one correspondence. This order keeps the force of each premise visible.

Questions to work through

1. In the proof of Proposition 1.2, which part uses the fact that the two constructors have different outer forms, and which part uses equality of their inner expressions?

2. Reproduce the two induction steps in Theorem 1.4 without looking at the proof. Explain why the first induction runs over $\mathbb{N}$ and the second over $\mathbb{N}_\delta$.

3. Which background rule is used to pass from the two construction cases to a statement about every finite orbit expression?

Chapter 2

Why a faithful tally has one structure

A tally need not look like a string of marks. It can use words, positions, or some other family of states. To compare it with the orbit, the relevant data are a starting state and a rule for the next state. We can ask for a translation that respects those data before asking whether the translation is faithful.

The distinction between those two questions is essential. Every pointed step system receives a unique reading of the orbit. Only a system that preserves distinctions and contains no unreachable extra elements is necessarily another presentation of the whole orbit.

These mapping arguments have a classical antecedent in Dedekind's treatment of arithmetic [2]. We give the construction and its proof so that the conditions behind uniqueness remain visible.

2.1 The data a translation must preserve

Definition 2.1 (A pointed step system)**.** A pointed step system is a carrier M, a designated element $b \in M$ and a function $s : M \to M$. We write it as (M, b, s).

At this stage there are no further conditions. In particular, we have not said that s is injective, that b is different from every output of s, or that every element of M is generated from b.

Definition 2.2 (A structure-preserving map)**.** A map $f : \mathbb{N}_\delta \to M$ preserves the starting point and the step when

$$f(0_\delta) = b, \qquad f(S(a)) = s(f(a)) \quad \text{for every } a \in \mathbb{N}_\delta. \tag{2.1}$$

The second equation says that the order of translation and stepping does not matter. We may take a step in the orbit and translate the result, or translate first and take a step in the target. Both routes must agree.

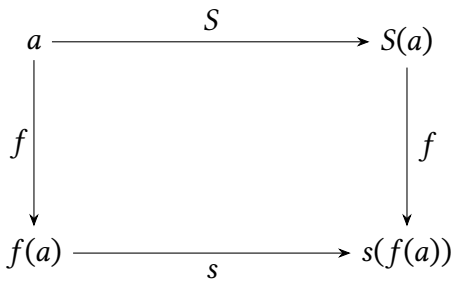


Figure 2.1. Preserving the step means that both routes have the same result. The equation $f(0_\delta) = b$ separately fixes the starting point.

2.2 There is exactly one reading

The required map can be constructed directly. Define $R_M : \mathbb{N}_\delta \to M$ by recursion:

$$R_M(0_\delta) = b, \qquad R_M(S(a)) = s(R_M(a)). \tag{2.2}$$

To evaluate R_M on an expression, follow the construction of the expression. Start with b and apply s once for each outer step. The expression is finite, so this instruction specifies its value through a finite recursive construction.

Theorem 2.3 (The unique reading of the orbit)**.** *For every pointed step system (M, b, s), there is exactly one map $f : \mathbb{N}_\delta \to M$ preserving the starting point and the step.*

Proof. The recursively constructed R_M satisfies both equations in (2.1), so such a map exists.

For uniqueness, let f be any map satisfying those equations. We prove $f(a) = R_M(a)$ by induction on a. At the starting expression,

$$f(0_\delta) = b = R_M(0_\delta).$$

Suppose $f(a) = R_M(a)$. Preservation of the step and the recursive rule give

$$f(S(a)) = s(f(a)) = s(R_M(a)) = R_M(S(a)).$$

Induction proves equality at every orbit expression. Thus every permitted map agrees with R_M, and there is exactly one. □

This property is called *initiality*. The orbit is initial among pointed step systems because every such system receives exactly one structure-preserving map from it. The name abbreviates the theorem; the content is the existence and uniqueness just proved.

Notice what the proof used. It followed the construction of the source orbit. It required no injectivity, separation or induction property in the target M. The map therefore exists even when the target identifies some of the records. Initiality tells us that the reading is determined. It does not tell us that the reading loses nothing.

2.3 What makes the reading faithful?

There are two possible failures to exclude. First, two different orbit expressions might receive the same target value. Second, the target might contain elements that no orbit expression reaches. The first is a failure of injectivity; the second is a failure of surjectivity.

The hypotheses below address both.

Definition 2.4 (A faithful generated step system)**.** A pointed step system (M, b, s) is faithful and generated when:

1. *The step is injective:* $s(x) = s(y)$ implies $x = y$.
2. *The start is not a step:* $s(x) \neq b$ for every $x \in M$.

3. *Induction holds on the target:* for every property P of elements of M, if $P(b)$ holds and $P(x)$ implies $P(s(x))$ for every x, then $P(y)$ holds for every $y \in M$.

The third condition quantifies over properties in the surrounding mathematics. It is the full generation principle used in the argument below, and it cannot be weakened to the properties expressible by formulas of an arithmetic language. Skolem constructed systems with a start, a step, addition and multiplication that satisfy every first-order arithmetic sentence true of the natural numbers, every instance of induction for such formulas included, and that nevertheless contain elements no tally reaches [3]. In such a system the property used in Lemma 2.6, "is reached from the orbit," cannot be written as a formula of that language.

Lemma 2.5 (No two orbit records are merged)**.** *If (M, b, s) is a faithful generated step system, then R_M is injective.*

Proof. We prove that $R_M(a) = R_M(c)$ implies $a = c$, using induction on a.

First let $a = 0_\delta$. If $c = 0_\delta$, the conclusion is immediate. If $c = S(d)$, the assumed equality becomes

$$b = s(R_M(d)),$$

contrary to separation of the start from every step. Thus c must also be 0_δ.

Now suppose the result has been proved for a, and consider an equality $R_M(S(a)) = R_M(c)$. If $c = 0_\delta$, this equality says $s(R_M(a)) = b$, again contradicting separation. The remaining case is $c = S(d)$. The equality then reads

$$s(R_M(a)) = s(R_M(d)).$$

Injectivity of s gives $R_M(a) = R_M(d)$. The induction hypothesis gives $a = d$, and hence $S(a) = S(d) = c$.

These cases exhaust the possible forms of c at each induction step. The reading is therefore injective. □

The argument tells us exactly where each condition works. Separation rules out a comparison between the start and a step. Injectivity removes a common outer step so that the comparison can be reduced to earlier expressions. No assumption about reaching every target element was needed.

Lemma 2.6 (Every target element is reached)**.** *If (M, b, s) is a faithful generated step system, then R_M is surjective.*

Proof. Apply target induction to the property

$$P(y) \quad \Longleftrightarrow \quad \text{there is an } a \in \mathbb{N}_\delta \text{ with } R_M(a) = y.$$

The property holds at b, witnessed by 0_δ, since $R_M(0_\delta) = b$. Suppose it holds at x, witnessed by an expression a. Then

$$R_M(S(a)) = s(R_M(a)) = s(x),$$

so $S(a)$ witnesses the property at $s(x)$. Target induction now gives the property at every $y \in M$. This is surjectivity. □

Here the target's generation principle does all the work. We chose the property "is reached from the orbit," proved it at the start, and showed that it survives a step. The principle then excludes any extra target element outside the reading's image.

2.4 The exact uniqueness result

A bijective structure-preserving map is an isomorphism of pointed step systems. It identifies their elements while preserving the specified start and step. The preceding results now fit together.

Theorem 2.7 (Uniqueness of a faithful generated tally)**.** *For every faithful generated step system (M, b, s), the canonical reading $R_M : \mathbb{N}_\delta \to M$ is bijective. It is the unique bijective map preserving the starting point and the step.*

Proof. By Lemma 2.5, the first two hypotheses make R_M injective. By Lemma 2.6, the third makes it surjective. It is therefore bijective and already preserves the structure by its construction.

Any other bijective structure-preserving map is, in particular, a structure-preserving map. The uniqueness established in Theorem 2.3 forces it to equal R_M. Thus the isomorphism is unique. □

The result is stronger than saying that some convenient labels can be matched. There is only one matching compatible with the starting point and the step. Changing the names or the carrier does not introduce a choice into that matching.

It is also more precise than the statement that every system with a step is arithmetic. A raw step system always receives a reading. To identify its whole carrier with the orbit, we used injectivity, separation and generation. The conclusion carries those conditions with it.

The conditions are not only sufficient. They are exactly what a bijective reading requires.

Proposition 2.8 (The faithfulness conditions are exact)**.** *Let (M, b, s) be a pointed step system.*

1. *Target induction holds exactly when R_M is surjective.*
2. *The system is faithful and generated exactly when R_M is bijective.*

Proof. The proof of Lemma 2.6 used only target induction, so target induction gives surjectivity. Conversely, suppose R_M is surjective, and let P satisfy $P(b)$ and $P(x) \Rightarrow P(s(x))$ for every x. Induction on the orbit proves $P(R_M(a))$ for every a, since $R_M(0_\delta) = b$ and $R_M(S(a)) = s(R_M(a))$. Every $y \in M$ is some $R_M(a)$, so $P(y)$ holds.

Theorem 2.7 proved the forward half of the second assertion. Conversely, let R_M be bijective. The first assertion gives target induction. If $s(x) = s(y)$, write $x = R_M(a)$ and $y = R_M(c)$; then $R_M(S(a)) = R_M(S(c))$, so $S(a) = S(c)$ by injectivity of R_M, and Proposition 1.2 gives $a = c$ and hence $x = y$. If $s(x) = b$ with $x = R_M(a)$, then $R_M(S(a)) = R_M(0_\delta)$, so $S(a) = 0_\delta$, contrary to Proposition 1.2. □

Injectivity of the reading needs less than the first two conditions. The proof of Lemma 2.5 used them only at elements the reading reaches, and there it needs them exactly: the reading is injective precisely when the step is injective on reached elements and never takes a reached element to the start. The converse is the argument just given for the second assertion, applied to reached elements only, so that no surjectivity is needed to write them as $R_M(a)$ and $R_M(c)$. Add to the orbit one extra element $*$ with $s(*) = 0_\delta$. The start is now a step, yet the reading is still injective. What fails is surjectivity: nothing reaches $*$.

2.5 Where the force of the argument lies

There are three distinct achievements here. First, we constructed a map from the orbit into an arbitrary pointed step system. Second, we proved that no other map preserves the specified data. Third, for faithful generated targets, we proved that this map reaches every element and merges none.

The first two give initiality. The third gives categoricity within the stated class of targets. Their combination gives a unique isomorphism. Keeping these achievements separate makes the theorem easier to use: a later construction may need only the unique reading, while a change of arithmetic carrier needs the full correspondence.

This is a familiar kind of mathematical uniqueness theorem. Its role here is foundational: it shows that the arithmetic carrier obtained from finite distinction is fixed up to its unique structure-preserving correspondence. The further work is to construct operations on that carrier and to identify what additional reasoning principles are needed for stronger statements. The number of symbols used to describe the carrier does not answer either question. The constructions and their proofs do.

Questions to work through

1. In Theorem 2.3, identify the exact induction hypothesis and the equation that uses preservation of the target step. Which of the three target conditions in Definition 2.4 appears in this proof?

2. In Lemma 2.5, mark each use of separation and each use of injectivity. Explain why the target induction principle is absent.

3. State the property to which target induction is applied in Lemma 2.6. Give the witness at the start and the witness after a step.

4. Why does uniqueness among all structure-preserving maps imply uniqueness among bijective structure-preserving maps?

5. In the system obtained by adding $*$ with $s(*) = 0_\delta$, decide which conditions of Definition 2.4 hold. Exhibit a property for which target induction fails, and explain which assertion of Proposition 2.8 this illustrates.

Chapter 3

Addition, multiplication, and order

A tally tells us how many steps have been taken. Arithmetic begins when we use one tally to act on another. To add, start at one orbit position and take the steps recorded by the other. To multiply, repeat an addition. To compare, remove matching steps until one of the records reaches its start.

These instructions concern the expressions already constructed. We do not enlarge their carrier or postulate a table of arithmetic laws. We define the operations, then prove what the definitions require. The order of the proofs matters: a familiar identity cannot justify an earlier step merely because we recognize it from ordinary arithmetic.

For the rest of this chapter, write 0 for 0_δ and 1 for $S(0)$. The letters a, b, c denote elements of $\mathbb{N}_\delta$. Products such as ab mean the multiplication defined below.

3.1 Addition follows the second record

Definition 3.1 (Addition)**.** Define addition by the two recursive clauses

$$a + 0 = a, \qquad a + S(b) = S(a + b). \tag{3.1}$$

The second input controls the recursion. If it is the starting record, there is nothing to do. If it has one more step, perform the earlier addition and then take one more step. Each recursive call removes an outer step from that input, so evaluation ends on every constructed expression.

The definition is deliberately asymmetric. It gives $a + 0 = a$ immediately, but it does not give $0 + a = a$ immediately. It tells us how to move a step out of the second input, but not out of the first. Both missing facts have proofs.

Lemma 3.2 (Moving a step through addition)**.** *For all $a, b \in \mathbb{N}_\delta$,*

$$0 + a = a, \qquad S(a) + b = S(a + b).$$

Proof. For the first identity, induct on a. At 0 the identity is $0 + 0 = 0$, which is a defining clause. If $0 + a = a$, then

$$0 + S(a) = S(0 + a) = S(a).$$

For the second identity, fix a and induct on b. At 0, both sides are $S(a)$. If $S(a) + b = S(a + b)$, then

$$S(a) + S(b) = S(S(a) + b) = S(S(a + b)) = S(a + S(b)).$$

This is the required identity at $S(b)$. □

The second proof illustrates a recurring technique. An equation is easy to use when it matches the argument on which the definition recurses. If it does not, first prove a lemma that moves the construction into that position.

Theorem 3.3 (Order and grouping of addition)**.** *Addition is commutative and associative:*

$$a + b = b + a, \qquad (a + b) + c = a + (b + c).$$

Proof. For commutativity, induct on a, keeping b arbitrary. At 0, the two sides agree by the defining zero clause and Lemma 3.2. If $a + b = b + a$, then

$$S(a) + b = S(a + b) = S(b + a) = b + S(a).$$

For associativity, induct on c. At 0, both sides reduce to $a + b$. If $(a + b) + c = a + (b + c)$, then

$$\begin{aligned}(a + b) + S(c) &= S((a + b) + c)\\ &= S(a + (b + c))\\ &= a + S(b + c)\\ &= a + (b + S(c)).\end{aligned}$$

The first, third and fourth equalities use the recursive addition clause; the second uses the induction hypothesis. □

We may now rearrange and regroup finite sums. The permission has been earned by these proofs. It was not part of the definition of addition.

Proposition 3.4 (Cancellation)**.** *If $a + b = a + c$, then $b = c$. If $a + c = b + c$, then $a = b$.*

Proof. For the first assertion, induct on a. At 0, remove $0+$ from both sides by Lemma 3.2. At $S(a)$, the equality becomes $S(a + b) = S(a + c)$. Injectivity of S gives $a + b = a + c$, and the induction hypothesis gives $b = c$. For the second assertion, commute the common right addend to the left and apply the first assertion. □

Cancellation will do substantial work in the next chapter. When two signed descriptions contain the same contribution on both sides of a comparison, it lets us remove that contribution without changing the comparison.

3.2 Multiplication repeats an addition

Definition 3.5 (Multiplication)**.** Define multiplication by

$$a \cdot 0 = 0, \qquad a \cdot S(b) = a \cdot b + a. \tag{3.2}$$

The second input again controls the recursion. The starting case performs no additions. Each further step adds one more copy of a to the result. Thus the operation is built from the earlier addition, just as addition was built from the step.

Lemma 3.6 (The other argument of multiplication)**.** *For all $a, b \in \mathbb{N}_\delta$,*

$$0 \cdot b = 0, \qquad S(a) \cdot b = a \cdot b + b.$$

Proof. Induction on b proves the first assertion. Its base case is a defining clause. In the step, $0 \cdot S(b) = 0 \cdot b + 0 = 0$.

For the second assertion, again induct on b. At 0, both sides are 0. At the next step, the induction hypothesis gives

$$\begin{aligned} S(a) \cdot S(b) &= S(a) \cdot b + S(a) \\ &= (a \cdot b + b) + S(a) \\ &= S((a \cdot b + b) + a) \\ &= S((a \cdot b + a) + b) \\ &= (a \cdot b + a) + S(b) \\ &= a \cdot S(b) + S(b). \end{aligned}$$

The middle rearrangement uses only the addition laws already established. □

Proposition 3.7 (Commutativity and the unit)**.** *For all $a, b \in \mathbb{N}_\delta$,*

$$ab = ba, \qquad a1 = 1a = a.$$

Proof. Induct on a for commutativity. The zero case follows from the two zero laws. In the step,

$$S(a)b = ab + b = ba + b = bS(a).$$

For the unit, $a1 = aS(0) = a0 + a = 0 + a = a$. Commutativity gives $1a = a$. □

3.3 Why multiplication distributes

To multiply by a sum, the second input asks for two successive groups of additions. We should therefore obtain the sum of the two products. The following induction makes that explanation exact.

Theorem 3.8 (Distributivity)**.** *For all $a, b, c \in \mathbb{N}_\delta$,*

$$a(b + c) = ab + ac, \qquad (a + b)c = ac + bc.$$

Proof. For the first identity, fix a, b and induct on c. At 0, both sides reduce to ab. If $a(b + c) = ab + ac$, then

$$\begin{aligned} a(b + S(c)) &= aS(b + c) \\ &= a(b + c) + a \\ &= (ab + ac) + a \\ &= ab + (ac + a) \\ &= ab + aS(c). \end{aligned}$$

The fourth equality is associativity of addition. For the second identity, commute the product, apply the first identity and commute the two resulting products:

$$(a+b)c = c(a+b) = ca + cb = ac + bc. \qquad \square$$

Proposition 3.9 (Grouping of multiplication)**.** *For all $a, b, c \in \mathbb{N}_\delta$, $(ab)c = a(bc)$.*

Proof. Induct on c. Both sides are 0 at the start. At a successor,

$$(ab)S(c) = (ab)c + ab = a(bc) + ab = a(bc + b) = a(bS(c)).$$

The equalities use, in order, the multiplication clause, the induction hypothesis, distributivity and the multiplication clause again. $\square$

This completes the dependency chain for the elementary algebra: recursive addition, addition laws, recursive multiplication, commutativity, distributivity and associativity. Later expansions of products of sums depend on this chain.

3.4 The step determines both operations

We defined addition and multiplication by recursion, then proved their laws. The question can be turned around. Suppose someone proposes other operations $\oplus$ and $\otimes$ on $\mathbb{N}_\delta$ and claims only a few familiar laws for them. Could they differ from ours? Not once adding one is required to be the step.

Theorem 3.10 (A few laws force the operations)**.** *Let $\oplus$ and $\otimes$ be binary operations on $\mathbb{N}_\delta$.*

1. *If $\oplus$ is associative, and $a \oplus 0 = a$ and $a \oplus 1 = S(a)$ for every a, then $a \oplus b = a + b$ for all a, b.*
2. *If $a \otimes 0 = 0$, $a \otimes 1 = a$ and $a \otimes (b+c) = a \otimes b + a \otimes c$ for all a, b, c, then $a \otimes b = ab$ for all a, b.*

Proof. For the first assertion, fix a and induct on b. At 0, the hypothesis gives $a \oplus 0 = a = a + 0$. Suppose $a \oplus b = a + b$. Since $S(b) = b \oplus 1$ by hypothesis,

$$a \oplus S(b) = a \oplus (b \oplus 1) = (a \oplus b) \oplus 1 = S(a \oplus b) = S(a + b) = a + S(b).$$

The equalities use the step hypothesis, associativity, the step hypothesis again, the induction hypothesis and the addition clause.

For the second assertion, again fix a and induct on b. At 0 both sides are 0. The addition clauses give $b + 1 = S(b + 0) = S(b)$, so if $a \otimes b = ab$, then

$$a \otimes S(b) = a \otimes (b + 1) = a \otimes b + a \otimes 1 = ab + a = aS(b). \qquad \square$$

The hypotheses do separate work. Associativity cannot be omitted from the first assertion. The operation that agrees with $+$ when its second input is 0 or 1, and returns its first input otherwise, satisfies the other two conditions. It gives $0 \oplus 2 = 0$, and it is not associative: $(0 \oplus 1) \oplus 1 = 2$, while $0 \oplus (1 \oplus 1) = 0$. The second assertion uses no commutativity.

These laws belong to a familiar list. A *commutative semiring* is a set with two associative and commutative operations, an element neutral for the first and an element neutral for the second, such that multiplying by the first neutral element gives that element and multiplication distributes over addition. The theorems of this chapter show that $\mathbb{N}_\delta$, with 0, 1, + and $\cdot$, is a commutative semiring.

Proposition 3.11 (The orbit maps uniquely into every commutative semiring)**.** *For every commutative semiring A, with neutral elements 0_A and 1_A, there is exactly one map $f : \mathbb{N}_\delta \to A$ with*

$$f(0) = 0_A, \qquad f(1) = 1_A, \qquad f(a+b) = f(a) + f(b), \qquad f(ab) = f(a)f(b).$$

Proof. Let f be the canonical reading of Chapter 2 into the pointed step system $(A, 0_A, x \mapsto x + 1_A)$. Then $f(1) = 0_A + 1_A = 1_A$. For additivity, induct on b; the base case uses the neutral element, and the step is

$$f(a + S(b)) = f(S(a+b)) = f(a+b) + 1_A = f(a) + f(b) + 1_A = f(a) + f(S(b)).$$

Multiplicativity follows in the same way: the base case uses $f(a) \cdot 0_A = 0_A$, and

$$f(aS(b)) = f(ab + a) = f(a)f(b) + f(a) = f(a)\big(f(b) + 1_A\big) = f(a)f(S(b)).$$

For uniqueness, any map g with these properties satisfies $g(0) = 0_A$ and $g(S(n)) = g(n+1) = g(n) + 1_A$. It therefore preserves the start and the step of the same pointed system, and Theorem 2.3 gives $g = f$. □

Corollary 3.12 (The only semiring whose unit step is the tally step)**.** *Let $\oplus$, $\otimes$, e and u make $\mathbb{N}_\delta$ a commutative semiring, with e neutral for $\oplus$ and u neutral for $\otimes$. If $a \oplus u = S(a)$ for every a, then $e = 0$, $u = 1$, $\oplus = +$ and $\otimes = \cdot$.*

Proof. The element e is 0 or a step. Suppose $e = S(w)$. Then $w \oplus u = S(w) = e$, and

$$0 = 0 \oplus e = 0 \oplus (w \oplus u) = (0 \oplus w) \oplus u = S(0 \oplus w),$$

contrary to Proposition 1.2. Hence $e = 0$, and $u = e \oplus u = S(e) = 1$. The first part of Theorem 3.10 now gives $\oplus = +$. The semiring laws give $a \otimes 0 = 0$, $a \otimes 1 = a$ and distributivity over +, so the second part gives $\otimes = \cdot$. □

The arithmetic of the tally is therefore not one convenient choice among several. Ask for a commutative semiring on the orbit in which adding the unit takes one step, and exactly one structure remains. The step carries that content: other commutative semiring structures on $\mathbb{N}_\delta$ keep its start and add differently, as the last question at the end of the chapter shows.

3.5 Comparing two finite records

Comparison need not inspect a decimal numeral. The expressions themselves carry enough information. Compare their outer forms. If both are steps, remove one step from each and compare what remains.

Definition 3.13 (Finite comparison)**.** Let $L(a, b)$ have the values true or false, with recursive clauses

$$\begin{aligned} L(0, b) &= \text{true}, \\ L(S(a), 0) &= \text{false}, \\ L(S(a), S(b)) &= L(a, b). \end{aligned}$$

Write $a \leq b$ when $L(a, b) = \text{true}$.

The cases are exhaustive. The first input reaches 0, or the second does, or both have a step to remove. On each recursive call, an outer step is removed from both inputs. Every finite comparison therefore ends with an answer. This fact will later distinguish finite arithmetic from comparison of arbitrary continuum descriptions.

Theorem 3.14 (Agreement with the natural-number reading)**.** *Let $r : \mathbb{N}_\delta \to \mathbb{N}$ be the reading of Chapter 1. Then*

$$r(a + b) = r(a) + r(b), \qquad r(ab) = r(a)r(b),$$

and

$$L(a, b) = \text{true} \quad \Longleftrightarrow \quad r(a) \leq r(b).$$

Proof. For addition, induct on b. At 0 the assertion reduces to $r(a) = r(a)$. For the step, use $r(S(x)) = r(x) + 1$ and the induction hypothesis:

$$r(a + S(b)) = r(S(a + b)) = r(a) + r(b) + 1 = r(a) + r(S(b)).$$

For multiplication, induct on b again. The zero case is immediate. In the step, the addition identity just proved gives

$$r(aS(b)) = r(ab + a) = r(a)r(b) + r(a) = r(a)r(S(b)).$$

For comparison, follow its three recursive cases. At $(0, b)$, the left side is true and $0 \leq r(b)$. At $(S(a), 0)$, the left side is false and $r(a) + 1 \leq 0$ is false. At $(S(a), S(b))$, remove the common successor: $r(a) + 1 \leq r(b) + 1$ is equivalent to $r(a) \leq r(b)$. The induction hypothesis identifies this last comparison with $L(a, b) = \text{true}$. □

The theorem connects the constructed arithmetic with its familiar notation. It does not supply the definitions retrospectively: the recursive operations and comparison were specified before their readings were proved.

3.6 Subtraction stops at the start

Removing matching steps also supplies a finite subtraction procedure. Define the *truncated difference* $a \dot{-} b$ by

$$\begin{aligned} a \dot{-} 0 &= a, \\ 0 \dot{-} S(b) &= 0, \\ S(a) \dot{-} S(b) &= a \dot{-} b. \end{aligned} \tag{3.3}$$

There is no expression before 0 in this carrier. If the first input runs out before the second, the procedure returns 0. This is a useful operation, but it does not retain a negative difference.

Proposition 3.15 (Reading a truncated difference)**.** *The natural-number reading of $a \dot{-} b$ is the ordinary truncated natural difference $r(a) \dot{-} r(b)$.*

Proof. Induct on a, treating b as arbitrary. If $a = 0$, the result is 0 whether b is 0 or a successor. If $a = S(c)$ and $b = 0$, both sides give $r(S(c))$. If $b = S(d)$, both truncated subtractions remove their common successor, reducing the assertion to the induction hypothesis at (c, d). □

The next chapter keeps both sides of a difference instead of stopping when one side runs out. That change in representation is enough to make signed arithmetic possible.

Questions to work through

1. In the proof of commutativity of addition, identify the exact point where the lemma for a step in the first argument is needed.

2. Reconstruct the cancellation proof without using the map r. Which property of the step permits the induction to continue?

3. In the distributivity proof, mark each equality by its justification: definition, induction hypothesis or an earlier theorem.

4. Explain why the comparison procedure terminates on the finite expressions. Which part of that explanation would need replacement for an input given by an unending approximation procedure?

5. Verify the operation described after Theorem 3.10: check $a \oplus 0 = a$ and $a \oplus 1 = S(a)$, then locate the first step of the theorem's proof that fails for it.

6. Exchange 1 and 2 in the natural-number reading and transport the ordinary operations along the resulting bijection. Show that the result is a commutative semiring on $\mathbb{N}_\delta$ whose neutral element for addition is 0, compute $1 \oplus 1$, and identify the hypothesis of Corollary 3.12 that fails.

Chapter 4

Differences and ratios

Suppose one record counts contributions and another counts removals. Keeping only their truncated difference loses information when removals exceed contributions. Keeping the two records separately avoids that loss. It also creates a new problem: different pairs can describe the same difference.

The same problem appears with ratios. Changing the unit of comparison can change both parts of a fraction while leaving its value unchanged. Arithmetic must respect these changes of description. Otherwise the result of a calculation would depend on how we happened to write its inputs.

This chapter constructs the two extensions in that order. First we identify balanced pairs of orbit records. Then we identify ratios by cross multiplication. At each stage we must prove that the identification is consistent and that the operations preserve it.

4.1 Keep both sides of a difference

Definition 4.1 (A signed record and its balance)**.** A signed record is a pair (a, b) of elements of $\mathbb{N}_\delta$. Its first entry records the positive side and its second entry the negative side. Define

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a + d = c + b. \tag{4.1}$$

The right side uses only orbit addition. It does not subtract anything. It compares the first positive side together with the second negative side against the second positive side together with the first negative side. Equality of those totals is the condition for balance.

The notation (a, b) still names a particular pair. The notation $[a, b]$ will name its *class*: all pairs balanced with it. Before using classes, we must show that balance is an equivalence relation. This means that a pair balances with itself, balance can be reversed, and two successive balances compose.

Proposition 4.2 (Balance is an equivalence relation)**.** *The relation $\sim$ is reflexive, symmetric and transitive.*

Proof. Reflexivity is the identity $a + b = a + b$. Symmetry reverses the equality $a + d = c + b$.

For transitivity, suppose

$$a + d = c + b, \qquad c + f = e + d.$$

We must prove $a + f = e + b$. Add f to the first equality, then use the second equality and the addition laws:

$$(a + f) + d = (a + d) + f = (c + b) + f = (c + f) + b = (e + d) + b = (e + b) + d.$$

Cancel the common addend d by Proposition 3.4. The result is the required balance. □

Definition 4.3 (Constructed integers)**.** The carrier $\mathbb{Z}_\delta$ consists of signed records identified by balance. Thus $[a,b] = [c,d]$ exactly when $a + d = c + b$.

This operation of forming classes is called taking a *quotient*. Here the word has no connection with division. It means that an explicitly specified equivalence relation determines when descriptions count as the same object. The equivalence proof is what makes the classes coherent.

4.2 Operations must survive a change of representative

The natural candidate operations on pairs are

$$\begin{aligned} (a,b) + (c,d) &= (a + c, b + d), \\ -(a,b) &= (b,a), \\ (a,b)(c,d) &= (ac + bd, ad + bc). \end{aligned} \tag{4.2}$$

The product places equal-sign contributions on the positive side and opposite-sign contributions on the negative side. Every entry is an orbit expression, so all three formulas construct signed records.

To turn these formulas into operations on $\mathbb{Z}_\delta$, we must prove more: replacing an input by a balanced pair must produce a balanced output. This requirement is called *representative independence*.

Theorem 4.4 (Signed arithmetic is representative independent)**.** *Addition, negation and multiplication in* (4.2) *preserve balance. They therefore define operations on* $\mathbb{Z}_\delta$.

Proof. Suppose $(a,b) \sim (a',b')$, so $a + b' = a' + b$.

For addition of a fixed pair (c,d), rearranging terms gives

$$(a + c) + (b' + d) = (a + b') + (c + d) = (a' + b) + (c + d) = (a' + c) + (b + d).$$

This is exactly the balance of the two sums. The same argument with the two summands exchanged handles replacement of the other input.

For negation, the required equation is $b + a' = b' + a$. It is the assumed balance with its sides reversed and its summands commuted.

For multiplication by (c,d), the two sides of the required balance are

$$(ac + bd) + (a'd + b'c), \qquad (a'c + b'd) + (ad + bc).$$

Group the terms in the first expression by their factors c and d:

$$\begin{aligned} (ac + bd) + (a'd + b'c) &= (a + b')c + (b + a')d \\ &= (a' + b)c + (b' + a)d \\ &= (a'c + b'd) + (ad + bc). \end{aligned}$$

The middle equality uses the given balance and its commuted form. The other equalities use distributivity and the addition laws from Chapter 3. Exchanging the two factors handles replacement of the second input. Replacing both inputs is now justified in two successive steps, using transitivity of balance. □

We may consequently write the same operation symbols on classes. In particular, set

$$0 = [0, 0], \qquad 1 = [1, 0], \qquad -[a, b] = [b, a].$$

The role of the brackets matters. A pair and its negative usually differ as records, but their sum has class 0 because its two entries balance.

4.3 The integer laws follow on the classes

Theorem 4.5 (The laws of signed arithmetic)**.** *Addition on $\mathbb{Z}_\delta$ is commutative and associative, has identity 0, and satisfies $x + (-x) = 0$. Multiplication is commutative and associative, has identity 1, and distributes over addition.*

Proof. Choose representatives $x = [a, b]$, $y = [c, d]$, $z = [e, f]$.

Commutativity and associativity of addition hold in each entry by the corresponding orbit laws. Adding $[0, 0]$ leaves both entries unchanged. Also

$$x + (-x) = [a + b, b + a] = [0, 0],$$

because $a + b = b + a$ is the balance condition with $[0, 0]$.

For multiplication, commutativity of the orbit product makes the two entries of xy equal to the two entries of yx. Multiplying by $[1, 0]$ gives $[a, b]$ using the unit and zero laws.

To check associativity, expand both $(xy)z$ and $x(yz)$. After regrouping, each has positive entry

$$ace + bde + adf + bcf$$

and negative entry

$$acf + bdf + ade + bce.$$

Every regrouping is justified by the associative and distributive orbit laws. Thus the representatives themselves agree.

For left distributivity, $x(y + z)$ has entries

$$\big(a(c + e) + b(d + f),\ a(d + f) + b(c + e)\big).$$

Expansion gives

$$\big(ac + ae + bd + bf,\ ad + af + bc + be\big),$$

which, after regrouping, are the entries of $xy + xz$. Right distributivity follows by commuting the product and applying left distributivity. □

These are the laws of a commutative ring. The name packages the identities we have just proved. Subtraction is now defined by $x - y = x + (-y)$; unlike truncated subtraction, it retains a negative result.

4.4 Reading the signed construction

We can compare the constructed integers with familiar integer notation. Let r be the natural-number reading from Chapter 1, and define the reading of a pair by

$$I(a,b) = r(a) - r(b) \quad \text{in } \mathbb{Z}. \tag{4.3}$$

Here the right side is a comparison with ordinary signed arithmetic. It was not used to define the pairs, balance, or their operations.

Proposition 4.6 (The signed reading is faithful)**.** *Two signed records balance exactly when their readings under I agree. The induced map $I : \mathbb{Z}_\delta \to \mathbb{Z}$ is bijective and preserves zero, one, addition, negation and multiplication.*

Proof. The addition reading of Theorem 3.14 gives

$$\begin{aligned} (a,b) \sim (c,d) \quad &\iff \quad r(a) + r(d) = r(c) + r(b) \\ &\iff \quad r(a) - r(b) = r(c) - r(d). \end{aligned}$$

The first equivalence also uses faithfulness of r. The second is ordinary integer rearrangement. Thus I is well defined on classes and injective.

Every nonnegative integer n is the reading of $[e(n), 0]$, where e is the expansion map from Chapter 1. Every negative integer $-n$, for $n > 0$, is the reading of $[0, e(n)]$. This proves surjectivity.

Preservation of zero and one follows by reading their representatives. For addition and negation, expand the definitions:

$$\begin{aligned} I([a+c, b+d]) &= (r(a) - r(b)) + (r(c) - r(d)), \\ I([b,a]) &= -(r(a) - r(b)). \end{aligned}$$

For multiplication, the two orbit reading identities give

$$\begin{aligned} I([ac+bd, ad+bc]) &= r(a)r(c) + r(b)r(d) - r(a)r(d) - r(b)r(c) \\ &= (r(a) - r(b))(r(c) - r(d)). \end{aligned}$$

Thus the operations agree under the reading. □

The proof distinguishes construction from comparison. The carrier and its laws were built from orbit records; the bijection then identifies that construction with the usual integers. We can use this faithful reading to check an integer identity without changing the constructed objects.

4.5 A ratio needs a nonzero denominator

A ratio record consists of a signed pair and a nonzero orbit denominator. Keeping the denominator in $\mathbb{N}_\delta$ places its sign in the numerator. We will write u, v, w for the integer classes of the chosen numerator pairs, and abbreviate a ratio record by (u, d). The record retains

its actual pair; the abbreviation saves us from writing both of its entries in every formula. The letters d, e, f denote nonzero orbit denominators, also read as the corresponding positive integers when they occur in signed products.

Representative independence of signed arithmetic ensures that replacing a numerator pair by a balanced pair does not change any cross comparison below. This is why the abbreviated notation is legitimate.

Definition 4.7 (Cross equality)**.** Two ratio records (u, d) and (v, e) are cross equal when

$$ue = vd. \tag{4.4}$$

The equation is equality of constructed integers. Equivalently, before passing to signed classes, the two scaled signed records balance.

The denominator restriction does real work. It ensures that a common denominator can be cancelled in the proof that cross equality is transitive.

Proposition 4.8 (Cross equality is an equivalence relation)**.** *Cross equality on ratio records is reflexive, symmetric and transitive.*

Proof. Reflexivity is $ud = ud$; symmetry reverses a cross equality. For transitivity, suppose $ue = vd$ and $vf = we$. Multiply the first equality by f and the second by d. The integer laws give

$$(uf)e = (ue)f = (vd)f = (vf)d = (we)d = (wd)e.$$

The common factor e is nonzero. Apply the faithful integer reading of Proposition 4.6, cancel that nonzero integer factor, and use faithfulness to return to constructed integers. The result is $uf = wd$, as required. □

The cancellation used here is ordinary integer cancellation. Its use is explicit: the preceding proposition supplied an operation-preserving bijection, so an integer equality proved after reading is also an equality of the constructed classes. No rational arithmetic has entered this proof.

Definition 4.9 (Constructed rationals)**.** The carrier $\mathbb{Q}_\delta$ consists of ratio records identified by cross equality. Write $\langle u; d\rangle$ for the class of (u, d).

The angle brackets keep this second quotient visually separate from the signed class $[a, b]$. Once the construction is understood, the familiar notation u/d can be used for $\langle u; d\rangle$.

4.6 Constructing arithmetic on ratio classes

Define

$$\begin{aligned} \langle u; d\rangle + \langle v; e\rangle &= \langle ue + vd; de\rangle, \\ -\langle u; d\rangle &= \langle -u; d\rangle, \\ \langle u; d\rangle\langle v; e\rangle &= \langle uv; de\rangle. \end{aligned} \tag{4.5}$$

The new denominator is allowed: a product of two nonzero orbit positions is nonzero. This follows either from their faithful natural-number readings or directly from repeated addition of a nonzero position.

Theorem 4.10 (Rational arithmetic is representative independent)**.** *The operations in* (4.5) *preserve cross equality.*

Proof. Suppose $ud' = u'd$ and $ve' = v'e$. These are the hypotheses for replacing (u, d) by (u', d') and (v, e) by (v', e').

For addition, cross multiplication of the proposed outputs gives

$$\begin{aligned}(ue + vd)(d'e') &= (ud')ee' + (ve')dd' \\ &= (u'd)ee' + (v'e)dd' \\ &= (u'e' + v'd')(de).\end{aligned}$$

Thus the two sums are cross equal.

For multiplication,

$$(uv)(d'e') = (ud')(ve') = (u'd)(v'e) = (u'v')(de).$$

For negation, $ud' = u'd$ gives $(-u)d' = (-u')d$ by negating both sides. All three operations are therefore well defined on ratio classes. □

The formulas were specified as constructions. The theorem proves that they respect the identification, which is the exact condition needed to lift them to classes. It does not by itself show that no other construction could serve. Section 4.9 proves that uniqueness.

4.7 Why a reciprocal keeps the sign in the numerator

For a nonzero numerator u, let $|u|$ denote its positive magnitude. This magnitude has a finite construction. For a signed representative (a, b), remove matching steps and add the two remaining truncated differences:

$$|(a, b)| = (a \dot{-} b) + (b \dot{-} a). \tag{4.6}$$

Its natural-number reading is $|r(a) - r(b)|$. To verify this, compare $r(a)$ and $r(b)$. In the first order, the second truncated difference is zero and the first is $r(a) - r(b)$; in the other order their roles are reversed. The truncated-difference reading from the previous chapter therefore gives the stated absolute value. In particular, the magnitude is nonzero exactly when the signed class is nonzero.

The sign is also decided by finite comparison: (a, b) is nonnegative exactly when $L(b, a)$ is true. This agrees with $0 \le r(a) - r(b)$ by Theorem 3.14. No search through an infinite description is involved.

Define the reciprocal on a ratio with nonzero numerator by

$$\langle u; d\rangle^{-1} = \begin{cases} \langle d; |u|\rangle, & u > 0, \\ \langle -d; |u|\rangle, & u < 0. \end{cases} \tag{4.7}$$

Both cases have a nonzero orbit denominator. To make the operation defined on every class, set $0^{-1} = 0$. This last clause is a convention for the value of a total operation; the inverse cancellation theorem retains its nonzero hypothesis.

Proposition 4.11 (The reciprocal is well defined and cancels)**.** *Cross-equal ratio records have cross-equal reciprocals. For every nonzero $q \in \mathbb{Q}_\delta$,*

$$qq^{-1} = q^{-1}q = 1, \qquad 1 = \langle 1; 1\rangle.$$

Proof. Suppose $ue = vd$ with d, e positive. If $u = 0$, cancellation of the nonzero factor d gives $v = 0$; both reciprocal records have zero class. The same argument in reverse shows that nonzero numerators remain nonzero.

In the nonzero case, u and v have the same sign. Indeed, multiplication by a positive integer preserves sign, so opposite signs would contradict $ue = vd$. If both are positive, the reciprocal cross equation is $dv = eu$, the original equality reversed. If both are negative, it is $(-d)(-v) = (-e)(-u)$, again the same equality. This proves representative independence in all cases.

Now let $q = \langle u; d\rangle$ be nonzero. For $u > 0$, its product with the reciprocal is $\langle ud; du\rangle$, cross equal to $\langle 1; 1\rangle$. For $u < 0$, the product is $\langle u(-d); d(-u)\rangle$, whose numerator and denominator also agree. Commutativity gives cancellation in the opposite order. □

4.8 Checking the field laws

Theorem 4.12 (The rational field laws)**.** *The constructed rationals have commutative, associative addition with zero and additive inverses; commutative, associative multiplication with one; distributivity; distinct zero and one; and multiplicative inverses for nonzero elements.*

Proof. Write three representatives as $(u, d), (v, e), (w, f)$. For addition, commutativity exchanges $ue + vd$ with $vd + ue$ and de with ed. The two groupings of a three-term sum, after expansion, have the common denominator def and numerator $uef + vdf + wde$. The two numerators agree by the integer laws, and the denominators agree by the orbit laws. This proves associativity.

Adding $\langle 0; 1\rangle$ leaves $\langle u; d\rangle$ unchanged. Adding its negative produces $\langle 0; d^2\rangle$, which is cross equal to $\langle 0; 1\rangle$.

For multiplication, the two orders give uv over de, and the two groupings of three factors give uvw over def. The integer and orbit laws justify all rearrangements. Multiplying by $\langle 1; 1\rangle$ leaves both parts unchanged.

For left distributivity, the representative of the left side is

$$\langle u(vf + we); def\rangle.$$

The representative obtained by adding the two products on the right is

$$\langle uvdf + uwde; d^2ef\rangle.$$

The second numerator and denominator are each the first multiplied by d, so cross multiplication proves equality. Commutativity supplies right distributivity.

Zero and one are distinct because their cross equality would assert $0 = 1$ in the integer construction, contradicting its faithful reading. Proposition 4.11 supplies the inverses of nonzero elements. □

The construction has now reached a field using finite records, balanced identifications, cross multiplication and explicit operations. It has not introduced a procedure for deciding every possible mathematical question. The finite comparisons used here terminate because they act on finite constructed data.

4.9 Each extension is determined by its task

The constructions of this chapter specified pairs, relations and formulas. A reader may ask whether different choices would have served as well. Each extension answers a definite question, and each is the unique answer.

Signed records make subtraction possible while keeping the orbit's addition. Write $\iota(n) = [n, 0]$. A *group* is a set with an associative operation, written $+$, a neutral element and inverses; it is a *commutative group* when the operation is also commutative. A map f from $\mathbb{N}_\delta$ into a group is *additive* when $f(a + b) = f(a) + f(b)$ for all a, b.

Theorem 4.13 (Signed records complete the orbit's addition)**.** *The map ι is additive. For every group G, commutative or not, and every additive map $f : \mathbb{N}_\delta \to G$, there is exactly one additive map $g : \mathbb{Z}_\delta \to G$ with $g(\iota(n)) = f(n)$ for every n. It is*

$$g([a, b]) = f(a) - f(b),$$

and its values lie in the cyclic subgroup of G generated by $f(1)$.

Proof. Additivity of ι is the addition formula $[a, 0] + [b, 0] = [a + b, 0 + 0] = [a + b, 0]$. An additive map sends 0 to the neutral element, since $f(0) = f(0 + 0) = f(0) + f(0)$, and then $f(a + 1) = f(a) + f(1)$ shows, by induction on a, that $f(a)$ is the sum of a copies of $f(1)$. So every value of f, and every negative of a value, is a sum of copies of $f(1)$ or of $-f(1)$, and any two such sums commute. The computations below involve only these elements, so they may be rearranged as in a commutative group, whether or not G is one. The formula for g is representative independent: if $a + d = c + b$, then $f(a) + f(d) = f(c) + f(b)$, so $f(a) - f(b) = f(c) - f(d)$. It is additive, because $g([a + c, b + d]) = f(a + c) - f(b + d) = \big(f(a) - f(b)\big) + \big(f(c) - f(d)\big)$. Since $f(0)$ is the neutral element, $g(\iota(n)) = f(n) - f(0) = f(n)$. Each value $f(a) - f(b)$ is the sum of a copies of $f(1)$ and b copies of $-f(1)$, so it lies in the cyclic subgroup generated by $f(1)$. For uniqueness, every class satisfies $[a, b] = \iota(a) + \big(-\iota(b)\big)$, and an additive map h on $\mathbb{Z}_\delta$ preserves negatives, because $h(x) + h(-x) = h(0)$ is the neutral element. So any such map takes $[a, b]$ to $f(a) - f(b)$. □

The pair $(\mathbb{Z}_\delta, \iota)$ is determined by this property up to exactly one isomorphism compatible with the maps from the orbit: two pairs with the property map into each other, and uniqueness forces both composites to be identities. The map matters, not only the carrier. The rationals with $n \mapsto \langle \iota(n); 1 \rangle$, and the integers with $n \mapsto 2n$, both fail. In each, the element assigned to 1 has a half, and an additive map into the integers extending the natural-number reading would have to send that half to an integer h with $h + h = 1$.

Corollary 4.14 (The constructed integers map uniquely into every commutative ring)**.** *For every commutative ring R there is exactly one map $\mathbb{Z}_\delta \to R$ preserving* 1*, addition and multiplication.*

Proof. Let $f : \mathbb{N}_\delta \to R$ be the map of Proposition 3.11, and let g be its extension given by Theorem 4.13. It sends $1 = [1,0]$ to 1_R, and it preserves products:

$$g\big([ac+bd, ad+bc]\big) = f(a)f(c)+f(b)f(d)-f(a)f(d)-f(b)f(c) = \big(f(a)-f(b)\big)\big(f(c)-f(d)\big).$$

Conversely, if h preserves 1, addition and multiplication, then so does $h \circ \iota$, because ι preserves them, and $h \circ \iota$ sends 0 to 0_R because it is additive. Proposition 3.11 gives $h \circ \iota = f$, and the uniqueness in Theorem 4.13 gives $h = g$. □

A field K has *characteristic zero* when $1_K + \cdots + 1_K$, with one or more terms, is never 0_K. Equivalently, the map $f : \mathbb{N}_\delta \to K$ of Proposition 3.11 sends every nonzero tally to a nonzero element.

Theorem 4.15 (The constructed rationals come first among fields of characteristic zero)**.** *For every field K of characteristic zero there is exactly one map $\mathbb{Q}_\delta \to K$ preserving* 1*, addition and multiplication. It sends $\langle u; d\rangle$ to $g(u)/f(d)$, where f and g are the maps of Proposition 3.11 and Corollary 4.14.*

Proof. The denominator $f(d)$ is nonzero because d is nonzero. If $ue = vd$, the ring map g gives $g(u)f(e) = g(v)f(d)$, so $g(u)/f(d) = g(v)/f(e)$: the formula is representative independent. It preserves sums, since

$$\frac{g(ue+vd)}{f(de)} = \frac{g(u)f(e)+g(v)f(d)}{f(d)f(e)} = \frac{g(u)}{f(d)} + \frac{g(v)}{f(e)},$$

it preserves products in the same way, and it sends $\langle 1;1\rangle$ to 1_K. Conversely, let k preserve 1, addition and multiplication. On the classes $\langle u;1\rangle$ it preserves these operations as a function of u, so Corollary 4.14 gives $k(\langle u;1\rangle) = g(u)$. The cross equality $\langle u;d\rangle\langle \iota(d);1\rangle = \langle u;1\rangle$ then gives $k(\langle u;d\rangle)\, f(d) = g(u)$. □

Two fields with this property are isomorphic by exactly one isomorphism of fields. The ordinary rationals are one of them, so exactly one isomorphism of fields connects $\mathbb{Q}_\delta$ with them. The real numbers have characteristic zero but do not come first: a map from them into the rationals preserving the operations would send $\sqrt{2}$ to a rational whose square is 2.

The order used in this chapter's sign test is also forced. Call a subset P of a commutative ring a *positive cone* when it is closed under addition and multiplication, every element x satisfies $x \in P$, $x = 0$ or $-x \in P$, and no element has both $x \in P$ and $-x \in P$.

Proposition 4.16 (Only one order fits the arithmetic)**.** *The classes of $\mathbb{Z}_\delta$ read as positive integers form a positive cone, and they form the only one. Likewise, the classes $\langle u; d\rangle$ with positive numerator form the only positive cone of $\mathbb{Q}_\delta$.*

Proof. The displayed sets are cones: their readings are closed under addition and multiplication, and the sign of an integer satisfies the other two conditions. For ratios, the sign of the numerator is representative independent by the argument in Proposition 4.11.

Let P be any positive cone. The last condition with $x = 0$ shows $0 \notin P$. If $x \neq 0$, then x or $-x$ lies in P, and in either case $x \cdot x = (-x)(-x) \in P$. Hence $1 = 1 \cdot 1 \in P$, and by addition every $\iota(n)$ with $n \neq 0$ lies in P. In $\mathbb{Z}_\delta$ these are exactly the classes read as positive integers. A class read as negative has its negative in P, so it is not in P. Thus P is the displayed cone. In $\mathbb{Q}_\delta$, also $\langle 1; d\rangle = \langle \iota(d); 1\rangle\langle 1; d\rangle^2 \in P$. A ratio with positive numerator, $\langle u; d\rangle = \langle u; 1\rangle\langle 1; d\rangle$, therefore lies in P, and a ratio with negative numerator does not, because its negative does. □

The order is therefore part of the arithmetic rather than an additional choice. Any ordering compatible with addition and multiplication in this sense is the one read from the finite comparison. The line built from these ratios in Chapter 12 is determined by its task from the other side: every Archimedean ordered field maps into it in exactly one way that preserves the arithmetic and the order, once countable choice is granted (Theorem 12.38). Which further conclusions require additional principles is the question of the next part.

Questions to work through

1. In the transitivity proof for balance, explain why equality of $a + d$ and $c + b$ is more useful than trying to subtract in $\mathbb{N}_\delta$.

2. Identify the two different tasks in defining an operation on classes: constructing its output and proving representative independence. Locate both tasks for signed multiplication.

3. Explain precisely where a nonzero denominator is used in the transitivity proof for cross equality.

4. Reproduce the two sign cases of the reciprocal proof. Explain why assigning $0^{-1} = 0$ does not establish inverse cancellation at zero.

5. Show directly that the integers with $n \mapsto 2n$ fail Theorem 4.13: take f to be the natural-number reading and find the equation that has no integer solution.

6. Using only the three cone conditions, show that every positive cone of $\mathbb{Q}_\delta$ contains $\langle 1; d\rangle$ for every nonzero d. Which condition excludes a cone containing -1?

Part II

What a proof is allowed to use

Chapter 5

The language and the rules

The equation $1 + 1 = 2$ can be read at two levels. We can calculate with the constructed numbers and establish the equality. We can also write a finite argument whose permitted steps are completely specified. The first activity concerns numbers. The second makes the argument itself an object we can study.

This distinction matters when we ask what a result requires. A calculation may be valid, yet conceal the use of a stronger principle in one of its steps. To expose that use, we need to say which expressions are allowed, which inferences count as steps, and how those expressions are interpreted.

We begin with arithmetic. This choice gives us a domain already constructed in Part I and a small language whose rules can be stated in full. It also keeps two questions separate: what the arithmetic structure is, and what a particular formal argument can establish about it.

5.1 Expressions are finite constructions

An arithmetic expression such as $S(0)+S(0)$ names a number. A statement such as $S(0)+S(0) = S(S(0))$ says something that can be proved. We call the first kind of expression a *term* and the second a *formula*.

Definition 5.1 (The arithmetic language)**.** Terms are built from variables $x_0, x_1, \ldots$, the symbol 0, and the operations S, $+$ and $\cdot$. Formulas are built from equalities of terms, the false formula $\bot$, conjunction $\wedge$, disjunction $\vee$, implication $\to$, and the quantifiers $\forall$ and $\exists$. Negation abbreviates implication to falsehood:

$$\neg\varphi \;:=\; \varphi \to \bot.$$

Each term and formula is a finite expression generated by these rules.

The symbols 1 and 2 remain abbreviations for $S(0)$ and $S(S(0))$. An equality between two terms is a formula; it does not assert that the two written expressions have the same shape. Thus $1 + 1$ and 2 are different terms, and $1 + 1 = 2$ is a statement about their values.

We use positions for variables so that the treatment of quantifiers is exact. Immediately inside a quantifier, x_0 denotes its newly bound variable. The earlier variables move up one position. Inside two nested quantifiers, x_0 refers to the inner one and x_1 to the outer one. Ordinary notation such as $\forall x\, \varphi(x)$ will be used when this binding convention is unambiguous.

Two operations on expressions are needed. *Lifting* inserts space for a new variable. *Substitution* replaces a variable by a term. They must respect the distinction between a free variable and one bound by a quantifier.

Definition 5.2 (Lifting and substitution)**.** To lift an expression by d at cutoff c, leave a variable x_n unchanged when $n < c$, and replace it by x_{n+d} otherwise. Apply the operation to each constituent expression; on entering a quantifier, increase c by one. Write $\varphi\uparrow$ for lifting by one at cutoff zero.

To substitute a term t at position k, replace x_k by t, lower an index $n > k$ to $n - 1$, and leave an index $n < k$ unchanged. Again act on each constituent expression. Under a quantifier, substitute at $k + 1$ and first lift the free variables of t by one. Write $\varphi(t)$ for substitution at position zero.

The lowering in substitution has a purpose: instantiating a quantifier removes its variable position. The lifting under a quantifier prevents a free variable in the substituted term from becoming bound accidentally. For example, let φ be $\forall(x_0 = x_1)$: every number equals the value of the free variable, which is written x_1 inside the quantifier. Substitute the term x_0 at position zero. Under the quantifier the rule substitutes at position one and first lifts the term to x_1, so $\varphi(x_0)$ is $\forall(x_0 = x_1)$ again: every number equals the value of x_0, which is false in $\mathbb{N}$. Replacing x_1 by x_0 without the lift would give $\forall(x_0 = x_0)$, which is true: the quantifier would have captured the substituted variable and changed the meaning. These operations are determined by the finite expression. Neither asks whether a formula is true.

5.2 One complete argument

A *context* Γ is a finite list of formulas admitted temporarily as hypotheses. A judgment

$$\Gamma \vdash \varphi$$

says that a finite derivation of φ from those hypotheses has been given. A blank context is written simply $\vdash \varphi$.

Here is the promised proof of $1 + 1 = 2$. Two arithmetic rules supply

$$1 + 0 = 1, \qquad 1 + S(0) = S(1 + 0).$$

Equality substitution replaces $1 + 0$ by 1 in the second formula. The whole derivation is therefore

$$\frac{\vdash 1 + 0 = 1 \qquad \vdash 1 + S(0) = S(1 + 0)}{\vdash 1 + S(0) = S(1)} \quad \text{equality substitution.} \tag{5.1}$$

The concluding formula is $1 + 1 = 2$ after expanding the numeral abbreviations. There is no induction in this argument, and there is no decision about an arbitrary formula. Its two leaves are defining equations of addition.

For complete precision, the formula with a substitution position is $1 + S(0) = S(x_0)$. Filling that position with $1 + 0$ produces the second premise; filling it with 1 produces the conclusion. This is the general equality rule in a small case.

5.3 The permitted inferences

We now give the rules behind judgments. Each displayed premise must itself have a finite derivation. Temporary hypotheses appear to the left of $\vdash$. A rule with no premises starts a branch of the derivation tree.

Hypotheses and equality

Any formula listed in Γ may be used. Equality has reflexivity and substitution:

$$\Gamma \vdash t = t, \qquad \frac{\Gamma \vdash t = s \qquad \Gamma \vdash \varphi(t)}{\Gamma \vdash \varphi(s)}.$$

The terms and the formula with its substitution position are specified as part of the application of the rule. A resemblance between formulas is not enough; the premise must be the indicated instance.

The arithmetic structure

The start and the step satisfy

$$\Gamma \vdash \neg(S(t) = 0), \qquad \frac{\Gamma \vdash S(t) = S(u)}{\Gamma \vdash t = u}.$$

The recursive operations supply four further formulas:

$$\Gamma \vdash t + 0 = t, \qquad \Gamma \vdash t + S(u) = S(t + u),$$
$$\Gamma \vdash t \cdot 0 = 0, \qquad \Gamma \vdash t \cdot S(u) = t \cdot u + t.$$

These are rules of the arithmetic language. Their validity in the constructed carrier was established in Part I. Writing them as rules does not erase the construction that licenses their interpretation.

Induction has the form

$$\frac{\Gamma \vdash \varphi(0) \qquad \Gamma \vdash \forall x\big(\varphi(x) \to \varphi(S(x))\big)}{\Gamma \vdash \forall x\, \varphi(x)}. \tag{5.2}$$

Here the bound induction variable is fresh relative to the hypotheses. In the second premise, $\varphi(S(x))$ denotes φ with x_0 replaced by $S(x_0)$ and every other position unchanged, that is, the substitution of $S(x_0)$ at position one in $\varphi\uparrow$. Other parameters of φ may remain free. The language permits this rule for every formula; the next chapter records separately whether the formula itself contains a quantifier.

Combining and using propositions

Implication is introduced by discharging a temporary hypothesis and used by applying it:

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi}, \qquad \frac{\Gamma \vdash \varphi \to \psi \qquad \Gamma \vdash \varphi}{\Gamma \vdash \psi}.$$

The comma means that the indicated hypothesis has been added to the context. For conjunction, proofs of both components give a proof of their conjunction; either component can then be recovered:

$$\frac{\Gamma \vdash \varphi \qquad \Gamma \vdash \psi}{\Gamma \vdash \varphi \wedge \psi}, \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \varphi}, \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \psi}.$$

For disjunction, a proof of either component gives the disjunction. To use a disjunction, both cases must lead to the same conclusion:

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \varphi \vee \psi}, \qquad \frac{\Gamma \vdash \psi}{\Gamma \vdash \varphi \vee \psi},$$

$$\frac{\Gamma \vdash \varphi \vee \psi \qquad \Gamma, \varphi \vdash \chi \qquad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi}.$$

Finally, from a derivation of $\bot$ one may infer any specified formula. This is the rule for an impossible case. None of these rules supplies a proof of $\varphi \vee \neg\varphi$ for an arbitrary φ.

Quantifiers and a genuinely arbitrary value

Universal introduction proves a formula at a fresh variable; universal elimination substitutes a term:

$$\frac{\Gamma\uparrow \vdash \varphi}{\Gamma \vdash \forall\varphi}, \qquad \frac{\Gamma \vdash \forall\varphi}{\Gamma \vdash \varphi(t)}.$$

Every formula in $\Gamma \uparrow$ has been lifted over the fresh variable. The hypotheses therefore cannot constrain that variable. This is the exact force of calling it arbitrary.

Existential introduction exhibits a witness. Existential elimination uses a fresh witness without letting it escape into the conclusion:

$$\frac{\Gamma \vdash \varphi(t)}{\Gamma \vdash \exists\varphi}, \qquad \frac{\Gamma \vdash \exists\varphi \qquad \Gamma\uparrow, \varphi \vdash \psi\uparrow}{\Gamma \vdash \psi}.$$

The last condition matters. The conclusion must be the lifted form of a formula from the original context, so it cannot depend on the particular witness chosen inside the branch.

Together these are the base arithmetic rules. A derivation is a finite tree assembled from them, with matching premises at every junction. Its validity as a derivation is a question about expressions and rules. To connect it to arithmetic truth, we still need an interpretation theorem.

5.4 Reading the expressions as arithmetic

We use the ordinary natural-number reading proved faithful in Chapter 1. An *environment* ρ assigns a natural number to each free variable position. Write $[n, \rho]$ for the extended environment with

$$[n, \rho](0) = n, \qquad [n, \rho](k+1) = \rho(k).$$

The new variable receives n and the earlier assignments are retained.

Definition 5.3 (Evaluation and satisfaction)**.** The value $E_\rho(t)$ of a term is defined recursively:

$$E_\rho(x_k) = \rho(k), \qquad E_\rho(0) = 0,$$
$$E_\rho(S(t)) = E_\rho(t) + 1, \qquad E_\rho(t + u) = E_\rho(t) + E_\rho(u),$$
$$E_\rho(t \cdot u) = E_\rho(t)E_\rho(u).$$

Write $\rho \vDash \varphi$ when the formula is satisfied. Equality is interpreted as equality of term values and $\bot$ as falsehood. Conjunction, disjunction and implication are interpreted by their corresponding constructive logical operations. Quantifiers are read by

$$\rho \vDash \forall\varphi \iff \text{for every } n \in \mathbb{N},\ [n, \rho] \vDash \varphi,$$
$$\rho \vDash \exists\varphi \iff \text{there is } n \in \mathbb{N} \text{ with } [n, \rho] \vDash \varphi.$$

An environment satisfies a context when it satisfies every formula in it.

The formulas of the language are finite objects. The assertions that they are satisfied are made in the surrounding mathematics. That surrounding reasoning is called the *metatheory*. It supplies the finite inductive constructions, equality and logical reasoning being used to study the language. We have made this background visible; we have not eliminated it.

Three identities ensure that manipulating an expression corresponds to manipulating its interpretation.

Lemma 5.4 (Binding respects interpretation)**.** *For every formula φ, term t, environment ρ and natural number n,*

$$[n, \rho] \vDash \varphi\uparrow \iff \rho \vDash \varphi, \tag{5.3}$$
$$\rho \vDash \varphi(t) \iff [E_\rho(t), \rho] \vDash \varphi. \tag{5.4}$$

Replacing the induction variable by its successor correspondingly replaces its assigned value n by $n + 1$.

Proof. First check terms. Lifting a variable moves its position past the newly inserted entry, so its value remains its previous assignment. Substitution at position k has three cases: the variable is x_k, is below k, or is above k. Its value is respectively the substituted term's value, the unchanged earlier assignment, or the assignment shifted down one position. These are exactly the environment changes specified by the operations. Zero is unchanged; the successor, sum and product cases follow by applying their evaluation clauses to the already established identities for their constituent terms. This proves the term identities by structural induction.

Now induct on formulas. An equality uses the term identities on both sides. Falsehood is unchanged. Each propositional connective applies to the corresponding already established interpretations of its components. Under a quantifier, the new bound position receives its own value and the cutoff increases by one. For substitution, lifting the inserted term keeps its value independent of this newly bound position. Thus the extended environments agree at position zero and at every later position. For a universal formula this argument applies to every assigned value; for an existential formula it applies to its witness, in both directions. These are all the formula constructors, proving the two equivalences.

For the final assertion, the inserted term $S(x_0)$ has value $n+1$ when x_0 has value n. The other free positions remain fixed by the lifting and substitution convention. Apply the substitution identity. □

This is the bookkeeping behind the freshness conditions. Its purpose is to make the soundness proof about the same assignments before and after a rule is applied.

5.5 Why every permitted step preserves truth

Theorem 5.5 (Soundness of the base arithmetic rules)**.** *Suppose a finite derivation of* $\Gamma \vdash \varphi$ *uses the base rules listed in this chapter. If* ρ *satisfies every hypothesis in* Γ*, then* $\rho \vDash \varphi$*.*

Proof. Induct on the derivation tree, keeping the context and environment arbitrary. For each final rule, the induction hypotheses say that its premise derivations preserve satisfaction in their respective contexts. We check every kind of rule.

Hypothesis and equality. A selected hypothesis is true because the whole context is satisfied. Reflexivity evaluates to equality of a value with itself. For substitution, the first premise gives $E_\rho(t) = E_\rho(s)$. The second, by Lemma 5.4, gives satisfaction of the hole formula with $E_\rho(t)$ in its first position. Replace that equal value by $E_\rho(s)$ and use the same lemma in reverse.

The arithmetic rules. A successor value is not zero, and equality of successor values implies equality of the preceding values. The four operation equations evaluate to the recursive arithmetic identities. Thus these rule conclusions are satisfied. In the injectivity rule the premise supplies the required equality of successors.

Implication. To introduce an implication, assume the antecedent is satisfied. The augmented context is then satisfied, so the induction hypothesis for the branch gives the consequent. This proves the interpreted implication. To eliminate one, the premise interpretations give an implication and its antecedent; applying the former to the latter gives the consequent.

Conjunction, disjunction and falsehood. Conjunction introduction combines the two satisfied components; either elimination selects the required component. A disjunction introduction supplies its proved side. In disjunction elimination, the interpreted disjunction gives one of its two alternatives. Add that alternative to the satisfied context and apply the induction hypothesis for the corresponding branch. Both branches have the same concluding formula. If the premise is falsehood, its interpreted truth is impossible, and the falsehood rule is sound.

Universal quantification. For introduction, choose an arbitrary natural number n and extend ρ to $[n, \rho]$. Every lifted hypothesis is satisfied by (5.3). The induction hypothesis therefore gives the body of the universal formula at n. Since n was arbitrary, the universal statement is satisfied. For elimination, apply the interpreted universal statement to $E_\rho(t)$, then use (5.4) to obtain the desired instance.

Existential quantification. Introduction uses $E_\rho(t)$ as its witness; the substitution identity supplies the body at that witness. For elimination, the first premise supplies a witness n. The extended environment satisfies the lifted original hypotheses and the existential body. The branch induction hypothesis gives $\psi\uparrow$ in this environment. Equation (5.3) then returns ψ

under the original ρ. This is where the restriction on the conclusion of existential elimination is used.

Induction. The first premise gives the body at zero. The second gives, at every n, an implication from the body at n to the body at $n+1$; Lemma 5.4 identifies the successor instance. Induction in the surrounding natural-number mathematics now gives the body at every n. This is precisely satisfaction of the universal conclusion. The context remains satisfied throughout because its formulas do not acquire the fresh induction variable.

Every base rule has been covered. Since the derivation tree is finite, the induction reaches its conclusion. □

The last case shows the scope of the result plainly. Soundness of the object-language induction rule was proved using induction in the metatheory. This is a proof about the stated language in its stated arithmetic interpretation. It is not an internal proof, written in that language, of its own unrestricted soundness.

The proof also shows which facts about the natural numbers it used. The arithmetic rules needed the recursive identities, separation of the start from every step, and injectivity of the step. The induction rule needed induction. The remaining cases used only that quantifiers range over the numbers. Section 5.7 uses this observation to show what each family of rules supplies.

5.6 The same induction, now inside the language

Chapter 3 proved $0 + a = a$ about the constructed operation. We can now display an object-language proof of its universal formula. Set $\varphi(x)$ to $0 + x = x$. The base is the rule $0 + 0 = 0$. For the step, temporarily assume $0 + x = x$. The recursive rule gives

$$0 + S(x) = S(0 + x).$$

Equality substitution changes the right side to $S(x)$. Discharge the temporary hypothesis and generalize the fresh variable to obtain

$$\forall x\big(0 + x = x \;\rightarrow\; 0 + S(x) = S(x)\big).$$

Apply (5.2). The conclusion is

$$\vdash \forall x(0 + x = x). \tag{5.5}$$

Soundness now reads this finite derivation as a true arithmetic statement. The formula on which we inducted was the equality $0 + x = x$, with no quantifier inside it. The quantifier in the conclusion does not change that fact. This distinction will matter for the induction record in the next chapter.

5.7 What each family of rules supplies

Besides the four recursive equations, which define the operations, the arithmetic rules fall into three families. *Separation* is the rule $\neg(S(t) = 0)$: a step never returns to the start. *Injectivity*

is the rule from $S(t) = S(u)$ to $t = u$: equal successors have equal predecessors. *Induction* is rule (5.2). They are the object-language counterparts of the three conditions of Definition 2.4. We can now ask what each family contributes that the others cannot.

Begin with what needs no induction at all. The defining equations already settle every closed calculation.

Proposition 5.6 (Closed calculations need no induction)**.** *Let φ be a quantifier-free formula with no free variables. If φ is true, then $\vdash \varphi$ has a derivation that uses no induction rule. If φ is false, then $\vdash \neg\varphi$ has such a derivation. Neither derivation uses an additional principle.*

Proof. First, every closed term is provably equal to the numeral of its value, with no induction. Proceed by recursion on the term in the metatheory. A numeral needs nothing; for $S(t)$, substitution replaces t by its numeral $\bar{n}$, and $S(\bar{n})$ is the numeral $\overline{n+1}$. For a sum, equality substitution replaces the summands by their numerals $\bar{m}$ and $\bar{n}$; then n uses of the successor equation for addition and one use of its zero equation turn $\bar{m} + \bar{n}$ into $\overline{m+n}$. Products are treated the same way with the multiplication equations, using the case of sums. Symmetry and transitivity of equality, each one use of equality substitution together with reflexivity, combine these steps.

The values of closed terms can be computed and compared, so each closed quantifier-free formula is either true or false. We prove both claims together by induction on φ. Take an equation $t = s$ with values m and n. If $m = n$, both sides are provably equal to the same numeral, so $t = s$ follows. If $m < n$, assume $t = s$, hence $\bar{m} = \bar{n}$. Injectivity, applied m times, gives $0 = \overline{n-m}$. The right side is a successor, so separation refutes it; discharging the assumption proves $\neg(t = s)$. The case $m > n$ is symmetric. Falsehood is refuted by $\bot \to \bot$, which discharges the hypothesis $\bot$. A true conjunction combines two derivations, and a false one refutes the false component. A true disjunction introduces its true side, and a false one refutes both sides by disjunction elimination. A true implication has a true consequent or a false antecedent, and either gives the implication; a false one has a true antecedent and a false consequent, which together refute it. No step introduces an induction rule. □

So every closed instance of a familiar law, such as $0 + \bar{n} = \bar{n}$, $\bar{m} + \bar{n} = \bar{n} + \bar{m}$ or $\neg(S(\bar{n}) = \bar{n})$, has a derivation with no induction. The universal statements are a different matter. To show that something cannot be derived, we need an interpretation in which the permitted rules still preserve truth while the statement fails. The proof of Theorem 5.5 supplies one almost at once, because the particular facts about the natural numbers entered only at the arithmetic rules and at induction.

Definition 5.7 (A structure for the arithmetic language)**.** A *structure* M consists of a carrier, an element 0_M, an operation S_M of one argument, and operations $+_M$ and $\cdot_M$ of two. Environments assign elements of the carrier to variable positions. Evaluation and satisfaction are given by the clauses of Definition 5.3 with M in place of $\mathbb{N}$, and quantifiers range over the carrier. Write $M, \rho \vDash \varphi$.

Theorem 5.8 (Soundness in other structures)**.** *Let a finite derivation of $\Gamma \vdash \varphi$ use the base rules, and let ρ be an environment in a structure M that satisfies every formula of Γ. Then $M, \rho \vDash \varphi$ in each of the following cases.*

1. *The derivation contains no induction rule, and M satisfies the six arithmetic rules: $S_M(a) \neq 0_M$, S_M is injective, and the four recursive equations hold for all elements.*
2. *The derivation contains no separation rule, and M has exactly one element.*
3. *The derivation contains no injectivity rule, and M is the* Boolean structure *on $\{0, 1\}$ with $S(a) = 1$, $a + b = \max(a, b)$ and $a \cdot b = \min(a, b)$.*

Proof. Lemma 5.4 holds in every structure, with the same proof, because that proof used only the evaluation clauses. Now repeat the proof of Theorem 5.5. The hypothesis, equality, propositional and quantifier cases go through unchanged, with quantifiers ranging over the carrier of M. Only the arithmetic rules and induction need attention.

In the first case each arithmetic rule is true in M by hypothesis, and the induction case does not arise.

In the second case every equation is satisfied, so reflexivity, substitution, injectivity and the four recursive equations all hold. Induction holds as well, since its base already gives the body at the only element. Separation is the one rule that fails, and it does not occur.

In the third case check the equations directly:

$$a + 0 = a, \qquad a + S(b) = \max(a, 1) = 1 = S(a + b),$$
$$a \cdot 0 = 0, \qquad a \cdot S(b) = \min(a, 1) = a = \max(\min(a, b), a) = a \cdot b + a.$$

Separation holds because $S(a) = 1 \neq 0$. Induction holds because the base gives the body at 0, and one use of the step gives it at $1 = S(0)$, the only other element. Injectivity fails, since $S(0) = S(1)$, and it does not occur. □

The first case is the interesting one. Let W consist of the natural numbers together with two further elements a and b. On natural numbers the operations are the usual ones. In addition,

$$S(a) = a, \qquad S(b) = b, \qquad a + n = a, \qquad b + n = b, \qquad x + a = x + b = b$$

for every natural number n and every element x. A product is 0 when either factor is 0, the usual product when both factors are natural numbers, and b in every other case.

The six arithmetic rules hold in W. No element steps to 0, and S is injective because a and b are fixed points that no natural number reaches. For $x + 0 = x$ there is nothing new to check. For $x + S(y) = S(x + y)$, the new cases are of two kinds: if x is a or b and y is a number, both sides are x; if y is a or b, then $S(y) = y$ and both sides are b. The rule $x \cdot 0 = 0$ is built into the table. For $x \cdot S(y) = x \cdot y + x$, both sides are 0 when $x = 0$; when $x \neq 0$ and x or y is a or b, both sides are b. In W, three familiar laws fail:

$$S(a) = a, \qquad 0 + a = b \neq a, \qquad a + 0 = a \neq b = 0 + a.$$

Corollary 5.9 (What only induction supplies)**.** *None of the formulas*

$$\forall x \neg(S(x) = x), \qquad \forall x(0 + x = x), \qquad \forall x \forall y(x + y = y + x)$$

has a derivation from the base rules without an induction rule. Each has a derivation whose induction formulas are all quantifier-free, and each of their closed instances has a derivation with no induction.

Proof. An induction-free derivation of any of them would make it true in W, by the first case of Theorem 5.8; the displayed values refute each one. The derivation of $\forall x(0+x = x)$ is (5.5). For $\forall x \neg(S(x) = x)$, induct on the quantifier-free formula $\neg(S(x) = x)$: separation gives the base, and in the step injectivity turns $S(S(x)) = S(x)$ into $S(x) = x$, contradicting the hypothesis. Commutativity follows the proofs of Lemma 3.2 and Theorem 3.3, with each induction on an equation in which the other variable is a parameter. The closed instances are covered by Proposition 5.6. □

The other two cases of Theorem 5.8 complete the picture. Every equation holds in the one-point structure, so no derivation from no hypotheses that avoids separation proves any formula $\neg(t = s)$, not even $\neg(S(0) = 0)$. In the Boolean structure, $S(S(0))$ and $S(0)$ both have value 1, so no derivation from no hypotheses that avoids injectivity proves $\neg(S(S(0)) = S(0))$. With all three families present, each of these statements has a derivation with no additional principle and only quantifier-free induction. Each family therefore supplies a statement that the other two cannot:

Family removed	A structure satisfying the rest	A statement lost
separation	one element	$\neg(S(0) = 0)$
injectivity	the Boolean structure	$\neg(S(S(0)) = S(0))$
induction	W	$\forall x \neg(S(x) = x)$

The same method shows that the statement "every number is zero or a successor", $\forall x(x = 0 \vee \exists y(x = S(y)))$, needs induction. Let C consist of the natural numbers together with a second chain $c_0, c_1, c_2, \ldots$, with $S(c_k) = c_{k+1}$. Set $x + c_0 = c_0$ and $x \cdot c_0 = c_0$ for every element x, and let the recursive equations $x + 0 = x$, $x + S(y) = S(x + y)$, $x \cdot 0 = 0$ and $x \cdot S(y) = x \cdot y + x$ determine every other sum and product. Every element is reached from 0 or from c_0 by steps, so this defines both operations, and the four recursive equations hold by construction. Separation and injectivity hold because S moves one place along each chain. The element c_0 is neither zero nor a successor, so the statement fails in C, and by the first case of Theorem 5.8 it has no derivation without induction. It has one by induction on the formula $x = 0 \vee \exists y(x = S(y))$, which contains a quantifier. The structure W, by contrast, satisfies the statement, since $a = S(a)$ and $b = S(b)$. Corollary 5.9 therefore survives when the statement is added as a hypothesis. In this language, then, induction supplies the passage from all the instances of a law to the universal law itself.

Remark 5.10. Chapter 6 adds rules for excluded middle and two related principles. Every formula has a decided truth value in the one-point and Boolean structures, so those rules are true there, and the second and third cases of Theorem 5.8 extend to derivations that use them. Corollary 5.9 also extends, with or without the hypothesis that every number is zero or a successor. With excluded middle in the metatheory the added rules are true in W. Without it, one reads disjunction and existence in W through double negation. Induction-free derivations remain sound under that reading, with or without the added rules, and because equality in W is decidable the reading agrees with ordinary truth on the three universal formulas of the corollary. With that hypothesis and the excluded-middle rule, the induction-free rules present Robinson's arithmetic, so Robinson's arithmetic proves none of the three formulas.

We now have both sides of the connection. A derivation gives a finite record of permitted reasoning; an interpretation gives arithmetic content to its expressions; soundness proves that the first preserves the second. The next task is to retain this connection when the rules are enlarged.

Questions to work through

1. In (5.1), identify the hole formula and the two terms substituted into it. Which rule supplies each premise?

2. Explain why a fresh variable may receive an arbitrary value in the universal-introduction case of the soundness proof. Where is the same fact used for existential elimination?

3. Locate the two inductions in the soundness argument: induction on the derivation and the induction used to interpret the induction rule. What is the object of each?

4. In (5.5), distinguish the formula used for induction from the formula finally proved. Does the first contain a quantifier?

5. Check the six arithmetic rules in the structure W. Once $S(a) = a$, which values can $x + a$ take, given the rule $x + S(y) = S(x + y)$ at $y = a$? Why would a structure with only the one new element a fail to refute $\forall x(0 + x = x)$?

6. Derive $0 + S(S(0)) = S(S(0))$ without induction. Explain why no induction-free derivation of $\forall x(0 + x = x)$ exists, although every instance has one.

7. Which rule fails in the Boolean structure? Explain why no derivation from no hypotheses that avoids that rule can prove $\neg(S(S(0)) = S(0))$, and give a derivation that uses it.

Chapter 6

The price of an additional principle

Consider the formula

$$(0 = 0) \vee \neg(0 = 0).$$

One way to obtain it is to invoke excluded middle: every formula holds or its negation holds. Another way is to prove $0 = 0$ by reflexivity and then introduce the left side of the disjunction. The second argument never asks for a decision about an arbitrary formula.

The conclusion is identical. The resources used by the two arguments are different. If we record the first argument faithfully, excluded middle appears in its record. It would be a mistake to infer that the conclusion therefore requires excluded middle. The second argument has already answered that question.

This chapter makes resource accounting precise. We will distinguish the principles a derivation uses, the conditions under which its conclusion is true, and the separate work needed to prove that a principle is unavoidable for a whole mathematical assertion. We then carry out that work for the three additional rules themselves. With induction on every formula they generate exactly two theories, which we name; without induction, or with induction on quantifier-free formulas only, they mark four strengths.

6.1 A witness, a decision, and a double negation

To establish $\exists n\, P(n)$ constructively, one supplies a value of n together with the required evidence for $P(n)$. To establish $\neg\neg\exists n\, P(n)$, one gives a contradiction from the assumption that there is no such value. The second form does not itself display a witness. A rule that converts it to the first is therefore a substantive addition to the permitted steps.

Similarly, deciding each individual instance $P(n)$ does not include a specified bound on where a successful instance might occur. Finite inspection of initial values and a decision about the entire unbounded predicate are different tasks. This difference motivates the following principles.

Definition 6.1 (Three named additional principles)**.** *Excluded middle*, abbreviated EM, states

$$P \vee \neg P \qquad \text{for every proposition } P.$$

The *limited principle of omniscience*, in the predicate form used here and abbreviated LPO, states

$$\big[\forall n\,(P(n) \vee \neg P(n))\big] \longrightarrow \big[(\exists n\, P(n)) \vee (\forall n\, \neg P(n))\big]. \tag{6.1}$$

Markov's principle, abbreviated MP, states

$$\big[\forall n\,(P(n) \vee \neg P(n))\big] \longrightarrow \big[\neg\neg\exists n\, P(n) \longrightarrow \exists n\, P(n)\big]. \tag{6.2}$$

Here n ranges over the natural numbers and P is an arbitrary predicate in the surrounding mathematics.

EM concerns every proposition. LPO concerns a family with an individual decision for every member and returns a global alternative. MP uses such individual decisions together with a double-negated existence assertion to obtain existence. Stating these principles separately lets a theorem say which one it actually uses.

There is also an object-language version of each principle. To the arithmetic rules of Chapter 5, we may add the schemas

$$\Gamma \vdash \varphi \vee \neg\varphi, \qquad \text{EM}, \tag{6.3}$$

$$\Gamma \vdash (\forall x(\varphi(x) \vee \neg\varphi(x))) \to ((\exists x\,\varphi(x)) \vee (\forall x\,\neg\varphi(x))), \qquad \text{LPO}, \tag{6.4}$$

$$\Gamma \vdash \neg\neg\exists x\,\varphi(x) \to \exists x\,\varphi(x), \qquad \text{MP}, \tag{6.5}$$

where the MP rule is permitted only when φ is quantifier-free. These are schemas over the finite formulas of the arithmetic language. Definition 6.1 concerns propositions and predicates in the metatheory. The latter principles suffice to interpret the former rules; the two levels should not be silently identified. Section 6.7 shows how far apart they are. With induction on every formula, the MP schema permits nothing new and the LPO schema proves every instance of the EM schema. As principles about arbitrary predicates in the metatheory, the three keep the separate strengths they have in constructive mathematics [4].

Why does the object-language MP rule omit an explicit premise giving individual decisions? Its restriction supplies that premise.

Lemma 6.2 (Quantifier-free formulas can be decided)**.** *If φ contains no quantifier, then for every environment ρ,*

$$(\rho \vDash \varphi) \vee \neg(\rho \vDash \varphi).$$

Proof. Induct on the formula. For equality, evaluate the two terms and compare their natural-number values. Falsehood has the negative answer. For a conjunction, decide its two components: both positive answers give the conjunction, and a negative answer to either refutes it. For a disjunction, a positive answer to either supplies a side; two negative answers refute both possible sides.

For an implication $A \to B$, first decide B. If B holds, the implication follows. If B does not hold, decide A. A proof of A then refutes the implication, since applying it would prove B; a refutation of A proves the implication by eliminating the impossible antecedent. These are all cases because the formula has no quantifier. □

The lemma decides a fixed finite formula at a fixed environment. By itself it says nothing about an existential formula, which has infinitely many instances. The restriction on the MP schema is part of its mathematical statement, and replacing it by an unrestricted schema would change the rule being studied. The restriction also has a consequence that is not visible here. In this language an existential formula over a quantifier-free body can be decided by a finite search, and Section 6.7 shows that the restricted rule therefore permits nothing new once induction on every formula is allowed.

6.2 A ledger for a finite argument

The word *cost* in this part of the book means a recorded logical resource. It is neither elapsed time nor an energy quantity. The later physical chapters introduce their own physical resources with their units and hypotheses.

Definition 6.3 (The four-entry ledger)**.** A ledger is a tuple

$$O = (e, \ell, m; i), \qquad e, \ell, m, i \in \{\text{false}, \text{true}\}.$$

The first three entries record use of EM, LPO and MP. The final entry records whether an induction formula with a quantifier was used. The empty ledger has all four entries false. Ledgers combine by taking the Boolean inclusive-or separately in each entry; write this combination as $O_1 \sqcup O_2$.

A base rule contributes no entry of its own, with one stated exception: an induction rule contributes the final entry when its induction formula contains a quantifier. An EM, LPO or MP rule contributes its corresponding entry. At a rule with premises, combine all their ledgers and the rule's own contribution.

This defines the ledger of a finite derivation by following its tree. An entry says that the resource occurred at least once. It does not count how often it occurred. In particular, using the same resource twice does not change its Boolean entry after the first use.

Proposition 6.4 (Resources cannot disappear in combination)**.** *A named principle present in either input ledger is present in their combination. Moreover,*

$$O_1 \sqcup O_2 = \emptyset \quad \Longleftrightarrow \quad O_1 = \emptyset \text{ and } O_2 = \emptyset,$$

where $\emptyset$ denotes the empty four-entry ledger.

Proof. Each output entry is the inclusive-or of the corresponding input entries. It is true whenever either input is true, and false exactly when both inputs are false. Apply this observation to all four entries. □

No rearrangement of the surrounding base rules erases an occurrence already recorded in a premise. Removing a resource requires changing the derivation itself, for example by replacing a branch with a proof that avoids the additional rule.

6.3 No added principle is not the same as an empty ledger

We need two descriptions. A derivation is *posit-free* when it uses none of the three additional principles. Its first three entries are then false. A derivation has an *empty ledger* only when its fourth entry is also false. Thus empty-ledger proofs restrict the induction formulas to the quantifier-free tier, while posit-free proofs may use induction on quantified formulas.

In this accounting convention, the verdict "forced" tests the first three entries. It is accompanied by the induction tier. The verdict is relative to the stated base rules and to this alphabet of

additional principles. Reading it as "no background assumptions of any kind" would discard the language, the base rules and the metatheory that made the argument possible.

The distinction can be seen directly in examples already available. The proof of $1 + 1 = 2$ has no induction and an empty ledger. The proof of $\forall x(0 + x = x)$ uses induction, but its induction formula is the quantifier-free equality $0 + x = x$. Its ledger is also empty.

For a contrasting induction, let φ be $\forall y(y = y)$, independent of the induction variable x. Its base follows by reflexivity and universal introduction. Its step is the identity implication $\varphi \to \varphi$, generalized over x. Induction gives $\forall x \forall y(y = y)$. This derivation posts the final entry because φ contains a quantifier, even though no EM, LPO or MP rule occurs. The record describes this particular use of induction; it is not a claim that this trivial conclusion needs that use.

6.4 The record agrees with the derivation

There are two ways to obtain the resource record. One is to carry it through each valid inference while verifying the premises and conclusion. The other is to ignore those formulas and simply inspect the rule labels and induction formulas in the whole tree. The latter inspection by itself does not establish that the tree is a valid derivation. On a valid tree, the two records agree exactly.

Theorem 6.5 (Agreement of resource accounting)**.** *For every accepted finite derivation, its threaded ledger equals the ledger obtained by inspecting its rule occurrences. Consequently its first three entries are false if and only if the tree contains no EM, LPO or MP rule. Its induction entry is true if and only if it contains induction on a formula with a quantifier.*

Proof. Induct on the tree. At a hypothesis, equality-reflexivity or arithmetic leaf, both records are empty. At an additional-principle leaf, both contain exactly its named entry. A one-premise base rule carries its premise ledger unchanged in both accounts. A two-premise rule combines the two records by $\sqcup$; disjunction elimination combines its three records. The induction hypotheses identify every premise record, so the combined records agree.

At an induction node, both accounts combine the base and step records. They then inspect the same finite formula: if it is quantifier-free, neither adds an entry; if it has a quantifier, both add the final entry. This covers all rules and proves equality of the two accounts.

For the consequences, an inclusive-or over finitely many occurrences is false exactly when all its inputs are false, by repeated application of Proposition 6.4. Therefore absence of the three posit entries is equivalent to absence of their three rule kinds. The same reasoning, with the induction formula test, gives the characterization of the final entry. □

The acceptance hypothesis cannot be dropped when interpreting the record as a proof's resources. A tree with mismatched premise formulas can have a perfectly clear list of rule labels. That list does not repair the invalid inference.

6.5 Soundness with the prices attached

A ledger has a semantic use as well as a syntactic one. It tells us which additional metatheoretic principles suffice to justify a derivation.

Theorem 6.6 (Conditional soundness)**.** *Let a derivation of $\Gamma \vdash \varphi$ have ledger O. Assume the metatheoretic EM principle if its EM entry is true, LPO if its LPO entry is true, and MP if its MP entry is true. Every environment satisfying Γ then satisfies φ.*

Proof. Use the derivation induction of Theorem 5.5. All base-rule cases are unchanged. The principle assumptions required for a premise are available because a true premise entry remains true in the conclusion ledger, by Proposition 6.4. It remains to check the three new leaves.

At an EM leaf, apply the supplied EM principle to the proposition $\rho \Vdash \varphi$. This gives precisely the interpretation of $\varphi \vee \neg\varphi$.

At an LPO leaf, suppose the interpreted premise says that each instance of the body is decidable. Apply the supplied LPO principle to

$$P(n) \quad \text{defined by} \quad [n, \rho] \Vdash \varphi.$$

Its existential alternative is the interpretation of $\exists\varphi$; its universal negative alternative is the interpretation of $\forall\neg\varphi$. This is the required implication.

At an MP leaf, the rule requires the body to be quantifier-free. Lemma 6.2 supplies a decision for each instance of the same predicate $P(n)$. The supplied MP principle then converts the double-negated existential premise into the existential conclusion.

The induction entry requires no further principle in this semantic theorem: the surrounding natural-number mathematics already supplies induction on the interpreted formula, as explicitly used in the base soundness proof. Its use inside the derivation remains recorded. □

Corollary 6.7 (Soundness without the three additional principles)**.** *For a posit-free derivation, conditional soundness requires none of EM, LPO or MP. This holds at either recorded induction tier.*

Proof. All three principle entries are false, so none of the three conditional assumptions in Theorem 6.6 is required. □

This establishes sufficiency, with the source of each permission visible. It does not say that every conclusion with an EM-labelled proof is unprovable without EM. The opening example proves the contrary. The assumptions in Theorem 6.6 can also be weakened: Corollary 6.15 shows that an MP entry requires no principle at all, and that an EM or LPO entry requires only the metatheoretic LPO for the predicates that formulas define, which is the least assumption under which every derivation with such an entry is sound.

6.6 A route's price and a theorem's exact strength

The two proofs of $(0 = 0) \vee \neg(0 = 0)$ give a useful complete comparison. The EM-rule proof has ledger

$$(\text{true}, \text{false}, \text{false}; \text{false}).$$

The reflexivity-and-disjunction proof has the empty ledger. Both have the same conclusion. The record is exact about each tree, but the first record overstates what is necessary to prove that conclusion.

The rule occurrence also retains more information than its Boolean entry: it specifies the particular formula to which the principle was applied. In this case that formula is $0 = 0$, whose positive side is already provable. Forgetting the formula and retaining only the name EM loses the reason that this occurrence can be discharged.

There is a second limitation. The named alphabet in this chapter contains three principles. It is a useful way to report their use, but three names need not mark three strengths, and no fixed alphabet can distinguish every intermediate principle someone might formulate. Over the base rules with induction on every formula, the three names mark exactly two strengths (Theorem 6.11). A conclusion's intrinsic logical strength is a question about derivability and reductions over a specified base, not merely about which names a chosen record can print.

To prove an *exact* logical cost, two directions are needed. First, derive the mathematical assertion using the proposed principle. Second, show that the assertion itself yields that principle over the same retained base. The second direction is a *reversal*. It supplies the lower bound that inspection of one proof cannot supply.

6.7 What the three rules cost

We now apply the two-direction test to the three rules themselves. Say that a rule *reduces* to a set of rules when every instance of its schema has a derivation, from no hypotheses, whose additional rules all belong to that set. A rule that reduces to a set can be removed from any derivation. Replace each of its leaves by the derivation of the same instance, with the leaf's hypotheses added to its context. A derivation stays valid when unused hypotheses are added, each lifted wherever a quantifier rule lifts the context; an induction on the derivation proves this, because no rule requires a hypothesis to be absent. The conclusion and the context of the whole derivation do not change. Its ledger can change in the induction entry, since some replacement derivations below use induction on formulas with quantifiers.

The answers depend on which induction rules the replacement derivations may use. We first allow induction on every formula.

The restricted Markov rule permits nothing new

The MP rule applies only to a quantifier-free body. In this language such a body cannot describe a search whose outcome is unknown. The reason is a fact about polynomials.

Lemma 6.8 (Quantifier-free formulas are eventually constant)**.** *Let φ be quantifier-free and let ρ be an environment. There is a number B, computed from ρ by evaluating terms, such that*

$$[n, \rho] \vDash \varphi \quad \Longleftrightarrow \quad [B, \rho] \vDash \varphi \qquad \text{for every } n \geq B.$$

Proof. Fix ρ. Each term t has the value

$$p_t(n) = c_0 + c_1 n + \cdots + c_k n^k$$

under $[n, \rho]$, for natural-number coefficients c_i that do not depend on n. Induct on t: the variable x_0 gives the polynomial n; another variable and 0 give constants; S adds one to c_0; sums and products of such polynomials are again such polynomials. Each coefficient is the value of a term in the other variables, built by the same recursion.

Let B be one more than the sum of all coefficients of both sides of every equation in φ. If every coefficient is 0, every equation holds at every n and there is nothing to prove. Otherwise $B \geq 2$. For $n \geq B$ every coefficient is less than n, so $c_0 + c_1 n + \cdots + c_k n^k$ is the base-n expansion with digits $c_0, \ldots, c_k$. Base-n expansions are unique. Hence an equation $t = s$ holds at such an n exactly when t and s have the same coefficients, after the shorter list is padded with zeros. That condition does not mention n. Each equation in φ therefore has one truth value for all $n \geq B$, and so does every combination of equations by $\wedge$, $\vee$, $\to$ and $\bot$. □

Take, for example, $\varphi(x)$ to be $x + x = y$, with x the searched variable and y a parameter. The left side has coefficients $(0, 2)$ and the right side has (y), so $B = y + 3$. When $n \geq y + 3$ the digit lists $(0, 2)$ and $(y, 0)$ differ, and the equation fails. Whether some n satisfies it is settled by checking $n = 0, 1, \ldots, y + 3$.

Theorem 6.9 (The restricted Markov rule costs nothing)**.** *Let φ be quantifier-free, let ρ be an environment, and let B be the bound of Lemma 6.8.*

1. *$\rho \vDash \exists \varphi$ holds exactly when $[n, \rho] \vDash \varphi$ for some $n \leq B$. It is therefore decided by finitely many applications of Lemma 6.2.*
2. *Consequently $\rho \vDash \neg\neg\exists \varphi \to \exists \varphi$ holds with none of EM, LPO or MP in the metatheory.*
3. *Every instance of the MP schema has a derivation from the base rules, with induction on every formula permitted, and with no EM, LPO or MP rule. The derivation constructed below uses induction on formulas that contain quantifiers.*

Conditions and scope.

Base and premises

The arithmetic language of Chapter 5, a quantifier-free formula φ with one searched variable, an environment ρ and the bound B of Lemma 6.8. Parts 1 and 2 concern the interpretation, with no added principle in the metatheory. Part 3 concerns the object-language schema (6.5), which applies only to quantifier-free bodies, over the base rules with induction on every formula permitted.

Conclusion

$\exists\varphi$ at ρ is decided by a search up to B; its interpreted Markov instance holds with none of EM, LPO or MP; and every instance of the restricted schema (6.5) has a derivation with no EM, LPO or MP rule, which uses induction on formulas with quantifiers.

Not claimed

Markov's principle for arbitrary decidable predicates, the principle of Definition 6.1, which is a different principle.

Proof. For the first assertion, a witness $n > B$ can be replaced by B, by the lemma; then examine $n = 0, 1, \dots, B$, deciding each instance by Lemma 6.2. For the second, run that search. If it finds an instance, the existential holds. If it finds none, the first assertion gives $\neg\exists\varphi$, which contradicts the double-negated premise, and the impossible case gives the conclusion.

The third assertion carries out the same argument inside the language. Every derivation below is from no hypotheses unless a context is shown, uses no EM, LPO or MP rule, and keeps the parameters of φ free; we write x for the searched variable. We use the abbreviations

$$u \le w \text{ for } \exists j(u{+}j = w), \qquad u < w \text{ for } \exists j(u{+}S(j) = w), \qquad A \leftrightarrow B \text{ for } (A \to B)\wedge(B \to A).$$

Appendix E.1 derives the arithmetic used along the way: the laws of addition and multiplication, each by induction on an equation; the decision $\psi \vee \neg\psi$ of every quantifier-free ψ; and five facts proved by induction on formulas with quantifiers, namely that every number is 0 or a successor, the decision of equality, the uniqueness of remainders, cancellation of a nonzero factor, and the comparison $x < y \vee y \le x$ (Lemma E.1).

Coefficients and the bound. The proof of Lemma 6.8 assigns to every term s a list of coefficients, lowest degree first. We use the same recursion on terms, keeping each coefficient as a term in the parameters: x has the list $(0, S(0))$, a parameter p has (p), and 0 has the empty list; $S(s)$ replaces the first coefficient c of s by $S(c)$, or the empty list by $(S(0))$; $s + r$ adds the two lists entry by entry, keeping the longer tail; and $s{\cdot}r$ is computed by $(c, c_1, \dots){\cdot}\vec{d} = (0, (c_1, \dots){\cdot}\vec{d}) + (c{\cdot}\vec{d})$, where $c \cdot \vec{d}$ multiplies every entry of $\vec{d}$ by c, and the empty list times $\vec{d}$ is empty. Let Σ be the sum $c_0 + (c_1 + (\cdots + (c_N + 0)))$ of all the coefficients of both sides of every equation of φ, and let T be $S(\Sigma)$. Its value is the bound B of Lemma 6.8. For a term X, the *polynomial form* of a list is

$$h_X() = 0, \qquad h_X(c, c_1, \dots, c_k) = h_X(c_1, \dots, c_k) \cdot X + c,$$

so that $h_X(c_0, \dots, c_k)$ is $c_0 + c_1X + \cdots + c_kX^k$ arranged by Horner's rule.

Every term equals its polynomial form. For every term s and every term X,

$$\vdash s(X) = h_X(\text{coefficients of } s),$$

where $s(X)$ substitutes X for x. This is one derivation for each s, built by recursion on s in the metatheory, and it uses only the laws of addition and multiplication. For instance, x itself is $(0 \cdot X + S(0)) \cdot X + 0$ because $0 \cdot X = 0$ and $S(0) \cdot X = X$; for a sum, $(a \cdot X + c) + (b \cdot X + d) = (a + b) \cdot X + (c + d)$; for a product, $(a \cdot X + c) \cdot h = (a \cdot h) \cdot X + c \cdot h$. The remaining cases are in Lemma E.2.

Digits below the base are unique. The heart of the argument is one digit:

$$\vdash c < X \to d < X \to U \cdot X + c = V \cdot X + d \to (c = d \wedge U = V). \tag{6.6}$$

Write $c + S(k) = X$ and $d + S(l) = X$. Uniqueness of remainders, with divisor X, gives $c = d$. Cancelling c leaves $U \cdot X = V \cdot X$, and X is the successor $S(c + k)$, so cancellation of a nonzero factor gives $U = V$. Conversely, $c = d$ and $U = V$ give the equation by substitution. Now say that two lists have *equal digits* when the formula $D(\vec{c}, \vec{d})$ holds, where D compares the lists entry by entry after padding the shorter one with zeros: $D((c, \vec{c}), (d, \vec{d}))$ is $c = d \wedge D(\vec{c}, \vec{d})$, a missing entry is compared with 0, and $D((), ())$ is $0 = 0$. Recursion on the two lists, applying (6.6) at each entry and writing the empty form 0 as $0 \cdot X + 0$ where one list runs out, gives

$$0 < X,\ \text{every } c_i < X,\ \text{every } d_i < X \ \vdash\ h_X(\vec{c}) = h_X(\vec{d}) \leftrightarrow D(\vec{c}, \vec{d}). \tag{6.7}$$

The hypothesis $0 < X$ is needed for the padding digit 0.

Beyond the bound, φ does not depend on x. Let D_φ be the formula obtained from φ by replacing each equation $s = r$ with $D(\text{coefficients of } s, \text{coefficients of } r)$. It contains neither x nor any new variable. Suppose $T \le X$. Every coefficient c of φ is one of the summands of Σ, so $c \le \Sigma$, and $S(\Sigma) \le X$ then gives $c < X$; likewise $0 \le \Sigma$ gives $0 < X$. By the normal forms and (6.7), each equation of $\varphi(X)$ is equivalent to its digit comparison, and these equivalences pass through $\wedge$, $\vee$, $\to$ and $\bot$. So

$$T \le X \ \vdash\ \varphi(X) \leftrightarrow D_\varphi.$$

Apply this at $X = T + z$, using $T \le T + z$ with witness z, and at $X = T$, using $T \le T$ with witness 0. Both sides are equivalent to the same D_φ, so universal introduction gives

$$\vdash \forall z \big(\varphi(T + z) \leftrightarrow \varphi(T)\big). \tag{6.8}$$

This is the derivable form of Lemma 6.8.

The bounded search. Let $Q(n)$ be the formula

$$\exists x\, \varphi(x)\ \vee\ \forall y \big(y < n \to \neg\varphi(y)\big).$$

We prove $\vdash \forall n\, Q(n)$ by induction on n; the induction formula contains quantifiers. At $n = 0$ take the right side: $y < 0$ is impossible, since $y + S(j) = 0$ gives $S(y + j) = 0$ against separation. For the step assume $Q(n)$. If its left side holds, so does that of $Q(S(n))$. Otherwise every $y < n$ has $\neg\varphi(y)$. Decide the quantifier-free $\varphi(n)$. If it holds, n witnesses $\exists x\, \varphi(x)$. If not, let $y < S(n)$. Then $y < n$ or $y = n$ (Lemma E.1); in the first case $\neg\varphi(y)$ is assumed, and in the second it follows from $\neg\varphi(n)$ by equality substitution. So the right side of $Q(S(n))$ holds.

The alternative. We now derive

$$\forall z \big(\varphi(T + z) \leftrightarrow \varphi(T)\big)\ \vdash\ \exists x\, \varphi(x) \vee \forall x \neg\varphi(x).$$

Instantiate the bounded search at $n = T$. If $\exists x\, \varphi(x)$, take the left side. Otherwise every $y < T$ has $\neg\varphi(y)$; decide $\varphi(T)$. If it holds, T is a witness. If not, take any x and compare it with T: $x < T$ or $T \le x$ (Lemma E.1). In the first case $\neg\varphi(x)$ is assumed. In the second, open $T \le x$ as $T + z = x$; the hypothesis at z gives $\varphi(T + z) \to \varphi(T)$, so $\neg\varphi(T)$ gives $\neg\varphi(T + z)$, and equality substitution along $T + z = x$ gives $\neg\varphi(x)$. Universal introduction yields $\forall x \neg\varphi(x)$.

The Markov instance. Discharge the hypothesis of the alternative with (6.8), and assume $\neg\neg\exists x\, \varphi(x)$. In the first case of the alternative the conclusion $\exists x\, \varphi(x)$ is at hand. In the second,

$\forall x\,\neg\varphi(x)$ gives $\neg\exists x\,\varphi(x)$: a witness y with $\varphi(y)$ meets $\neg\varphi(y)$, and existential elimination closes. Applying the premise yields $\bot$, and the rule for an impossible case gives $\exists x\,\varphi(x)$. Implication introduction completes the derivation of $\neg\neg\exists x\,\varphi \to \exists x\,\varphi$.

No EM, LPO or MP rule occurs. The induction formulas of the bounded search, of the comparison of numbers, of the uniqueness of remainders, of the cancellation of a nonzero factor, of the decision of equality and of the split into zero and successor contain quantifiers, so the induction entry of the ledger is true. □

The theorem explains why, once induction on every formula is allowed, the restriction leaves nothing for Markov's principle to do in this language. Markov's principle concerns a search that is decidable at each step but whose stopping point is unknown, such as the search for the step at which a computation halts. A quantifier-free body comes with a bound, computed from its parameters, beyond which nothing changes, so its search is bounded in advance. No quantifier-free formula $\theta(e, n)$ can therefore say that n codes a halting computation of the program with code e on input e: a search up to the bound would then decide whether that program halts, and no program decides that [21]. Programs and their numerical codes are treated in Chapter 10. Describing a computation with only 0, S, $+$ and $\cdot$ requires quantifiers. For the same reason, a version of LPO restricted to quantifier-free bodies would decide nothing new in the interpretation: its conclusion is settled by the same finite search. The principle of Definition 6.1, which applies to every decidable predicate of the metatheory, is a different matter.

The LPO rule proves excluded middle

The LPO rule has no such restriction. Its premise asks for a decision of every instance, and the rule itself supplies decisions for quantified formulas, one quantifier at a time.

Theorem 6.10 (The LPO and EM rules prove the same formulas)**.** *Permit induction on every formula. For every formula φ and every context Γ, the judgment $\Gamma \vdash \varphi \vee \neg\varphi$ has a derivation from the base rules and the LPO rule, with no EM or MP rule. Conversely, every instance of the LPO schema has a derivation from the base rules and the EM rule, with no induction.*

Proof. Start with equality. Induction on x, with the induction formula $\forall y(x = y \vee \neg(x = y))$, proves

$$\vdash \forall x\,\forall y\big(x = y \vee \neg(x = y)\big). \tag{6.9}$$

The base and the step each split into cases according to whether y is 0 or a successor. An induction on y with a quantifier-free formula performs that split, since its step does not use its own hypothesis. In the base, $0 = 0$ holds by reflexivity and $\neg(0 = S(y))$ by separation. In the step, $\neg(S(x) = 0)$ holds by separation; for $S(x)$ and $S(y)$, the hypothesis decides $x = y$, and its two answers give $S(x) = S(y)$ by substitution and $\neg(S(x) = S(y))$ by injectivity.

Now prove $\Gamma \vdash \varphi \vee \neg\varphi$ for every formula φ and every context Γ together, by induction on φ. Allowing every context matters, because the quantifier cases pass to a lifted context.

Equations are instances of (6.9). *Falsehood* takes the right alternative, $\bot \to \bot$. For a *conjunction*, *disjunction* or *implication*, apply disjunction elimination to the decisions of its two components and argue by cases, exactly as in the proof of Lemma 6.2.

An existential formula $\exists\psi$. The induction hypothesis in the lifted context, followed by universal introduction, gives $\forall x(\psi \vee \neg\psi)$. This is the premise of the LPO rule at ψ, which returns $(\exists x\,\psi) \vee (\forall x\,\neg\psi)$. In the second case existential elimination turns $\forall x\,\neg\psi$ into $\neg\exists x\,\psi$.

A universal formula $\forall\psi$. From $\forall x(\psi \vee \neg\psi)$ obtain $\forall x(\neg\psi \vee \neg\neg\psi)$, since ψ gives $\neg\neg\psi$. The LPO rule at $\neg\psi$ returns $(\exists x\,\neg\psi) \vee (\forall x\,\neg\neg\psi)$. A witness of $\neg\psi$ refutes $\forall x\,\psi$. In the other case decide ψ at an arbitrary x: the negative answer contradicts $\neg\neg\psi$, so ψ holds, and universal introduction gives $\forall x\,\psi$.

Only the LPO rule introduces a principle in these derivations. For the converse, the EM rule at $\exists x\,\varphi$ gives $(\exists x\,\varphi) \vee \neg\exists x\,\varphi$, and $\neg\exists x\,\varphi$ gives $\forall x\,\neg\varphi$: assume φ at a fresh x, introduce the existential, and reach a contradiction. The premise of the LPO instance is not even used. □

The only inductions in these derivations are those in the proof of (6.9), and they serve only the equations. The MP rule can do that job instead. For an equation $t = s$, let β be $t = s \vee \neg(t = s)$, and let z be a variable that does not occur in it. The formula $\neg\neg\beta$ has a derivation with no added rule: from $\neg\beta$ we obtain $\neg(t = s)$, hence β, a contradiction. So $\neg\neg\exists z\,\beta$ is derivable, and the MP rule at the quantifier-free β returns $\exists z\,\beta$, hence β. Thus excluded middle also follows from the LPO and MP rules together, with no induction at all.

Two strengths, or four

Theorem 6.11 (The price of each rule when induction is unrestricted)**.** *Allow induction on every formula. Then:*

1. *MP reduces to the empty set of rules;*
2. *EM reduces to LPO, and LPO reduces to EM;*
3. *neither LPO nor EM reduces to MP, and in particular neither reduces to the empty set.*

Give MP level 0 *and LPO and EM level* 1*, and give a set of rules the largest level of its members, or* 0 *if it is empty. A rule reduces to a set exactly when its level is at most the level of the set.*

Conditions and scope.

Base and premises
The base rules of Chapter 5, with induction on every formula permitted in the replacement derivations; the rules EM, LPO and MP of (6.3)–(6.5), the MP rule restricted to quantifier-free bodies; reduction as defined at the start of Section 6.7.

Conclusion
The restricted MP rule reduces to the empty set; EM and LPO reduce to each other; neither LPO nor EM reduces to MP. Two levels, 0 for MP and 1 for LPO and EM, decide every reduction.

Proof. The proof of the third assertion uses material from later in the book: the programs of Definition 10.5 with the pairing (10.3) and the codes (10.4), the arithmetic description of

computations and the universal program of Sections 20.3.3 and 20.3.4, and Theorem E.9 with Lemmas E.6–E.11. On a first reading it may be deferred until after Section 20.3.4.

The first two assertions are Theorems 6.9 and 6.10. For the third it suffices to treat LPO. If EM reduced to MP, so would LPO: in the derivation of an LPO instance from one EM rule given by Theorem 6.10, replace the EM leaf by a derivation of the same instance from MP rules, as in Section 6.7. And a rule that reduces to the empty set reduces to {MP}. We show that LPO does not reduce to MP, using Kleene's realizability interpretation [50, 4] and a classical metatheory.

Programs. The programs are those of Definition 10.5, with the codes of (10.4). Decoding never fails, so every number is the code of exactly one program; write $\{r\}(a)\downarrow v$ when the program with code r returns v on input a. Every number is a pair $\langle r_1, r_2\rangle$ in exactly one way, by the round trip for (10.3). Besides the constructors of Definition 10.5 we use two standard facts of computability [19, Chapter XII]. A program can run a code it has computed: there is a universal program, which returns v on $\langle r, a\rangle$ exactly when $\{r\}(a)\downarrow v$; the program U of Section 20.3.4 is one. And a program can fix part of an input. Let K_a be the program that composes the successor a times with the zero program, so that it returns a on every input. On input b the program $r \circ \mathrm{pair}(K_a, \mathrm{id})$, where id, the pairing of the two projections, returns its input by the round trip for (10.3), runs r on $\langle a, b\rangle$, and its code is computed from r and a by (10.4).

Realizers. Let ρ be an environment. A number r *realizes* a formula at ρ according to the following clauses.

- r realizes $t = s$ when t and s have the same value under ρ; nothing realizes $\bot$.
- $\langle r_1, r_2\rangle$ realizes $\varphi \wedge \psi$ when r_1 realizes φ and r_2 realizes ψ.
- $\langle i, r'\rangle$ realizes $\varphi \vee \psi$ when $i = 0$ and r' realizes φ, or $i \neq 0$ and r' realizes ψ.
- r realizes $\varphi \to \psi$ when, for every a that realizes φ at ρ, $\{r\}(a)\downarrow v$ for some v that realizes ψ at ρ.
- r realizes $\forall\varphi$ when, for every number n, $\{r\}(n)\downarrow v$ for some v that realizes φ at $[n, \rho]$.
- $\langle n, r'\rangle$ realizes $\exists\varphi$ when r' realizes φ at $[n, \rho]$.

So a realizer of an existential statement carries its witness, and a realizer of a disjunction says which side holds. Since nothing realizes $\bot$, a number realizes $\neg\varphi$ exactly when no number realizes φ, and then every number does. A formula with no free variables is realized or not independently of ρ.

Derivations are realized. Realizers of formulas with free variables must be computed uniformly from the values of those variables. For a finite list $l = (l_0, \ldots, l_{m-1})$ let ρ_l be the environment with $\rho_l(k) = l_k$ for $k < m$ and $\rho_l(k) = 0$ otherwise, and code lists by $\ulcorner()\urcorner = 0$ and $\ulcorner(n, l)\urcorner = \langle n, \ulcorner l\urcorner\rangle + 1$. Then $\rho_{(n,l)} = [n, \rho_l]$, and the value of a term under ρ_l is computed from $\ulcorner l\urcorner$. Theorem E.9 proves the following. *For every derivation of $\gamma_1, \ldots, \gamma_k \vdash \varphi$ whose additional rules are all MP rules, with any induction rules, there is a program c such that for every list l and all numbers $a_1, \ldots, a_k$ with a_i realizing γ_i at ρ_l, c returns on input $\langle \ulcorner l\urcorner, \ulcorner(a_1, \ldots, a_k)\urcorner\rangle$ a number that realizes φ at ρ_l.* The program is built along the derivation. Three cases show why the statement has this form.

Universal introduction derives $\forall\varphi$ from a derivation of φ in the lifted context. A realizer of $\forall\varphi$ must be one program that works for every value n of the new variable. Given l and the a_i, it is the program that, on n, runs the program of the premise on $\langle\ulcorner(n,l)\urcorner,\ulcorner(a_1,\dots,a_k)\urcorner\rangle$; its code is computed by fixing part of an input. The same a_i realize the lifted hypotheses at $\rho_{(n,l)}=[n,\rho_l]$, because realizability respects lifting as satisfaction does (Lemma E.6).

Induction derives $\forall x\,\varphi$ from $\varphi(0)$ and $\forall x(\varphi(x)\to\varphi(S(x)))$. Let r_0 and s be the realizers that the premises' programs return. The realizer of the conclusion is the program that, on n, computes v_n by primitive recursion: $v_0=r_0$, and v_{y+1} is obtained by running s on y and then running the program so obtained on v_y. Its code is computed from r_0 and s, which the premises' programs return on the given input, by fixing part of an input. Induction on n shows that v_n realizes φ at $[n,\rho_l]$. The step uses the final assertion of Lemma E.6: a realizer of $\varphi(S(x))$ at $[y,\rho_l]$ realizes φ at $[y+1,\rho_l]$. Nothing here depends on the shape of φ, so induction on every formula is realized.

An MP rule concludes $\neg\neg\exists x\,\varphi\to\exists x\,\varphi$ with φ quantifier-free. Truth of a quantifier-free formula under ρ_l is computed from $\ulcorner l\urcorner$ by evaluating terms, and a true quantifier-free formula has a realizer computed in the same way, while a realized one is true (Lemma E.7). The realizer returned for the MP instance is the code of the program that ignores its input and searches $n=0,1,2,\dots$ for the first n with $[n,\rho_l]\vDash\varphi$, then returns $\langle n,q\rangle$ with q the computed realizer of φ at $[n,\rho_l]$. Let b realize $\neg\neg\exists x\,\varphi$ at ρ_l. If no n satisfied φ, no number would realize $\exists x\,\varphi$, since its second component would realize, hence make true, an instance of φ; then 0 would realize $\neg\exists x\,\varphi$, and $\{b\}(0)$ would return a realizer of $\bot$, which does not exist. So, by the classical metatheory, some n satisfies φ, and the search stops. The realizer b is never run; it only guarantees that the search ends.

A derivation of a formula with no free variables, from no hypotheses, therefore yields a realizer: run its program on $\langle\ulcorner()\urcorner,\ulcorner()\urcorner\rangle$.

An LPO instance with no realizer. Section 20.3.3 reads a program step by step within an *allowance*. By Lemma E.11 there is a formula $\theta(e,x)$, with the two free variables shown, such that

(i) θ is built from equations by $\wedge$, $\vee$, $\forall$ and $\exists$, with no implication;

(ii) $\theta(e,x)$ is true exactly when program e returns a value on input 0 within allowance x, and program e returns a value on input 0 exactly when it does so within some allowance;

(iii) a program decides, from e and x, whether $\theta(e,x)$ is true, and when it is, computes a number that realizes it.

By (i), a number that realizes $\theta(e,x)$ makes it true: each clause passes realizability of the parts to truth of the whole (Lemma E.8). Consequently, when $\theta(e,x)$ is false every number realizes $\neg\theta(e,x)$.

Let Λ be the closed formula

$$\forall e\Big[\forall x\big(\theta(e,x)\vee\neg\theta(e,x)\big)\to\Big(\exists x\,\theta(e,x)\vee\forall x\,\neg\theta(e,x)\Big)\Big].$$

The formula in brackets is the LPO instance at θ, with e free. If LPO reduced to MP, that instance would have a derivation from no hypotheses whose additional rules are MP rules,

and universal introduction, which adds no rule, would turn it into such a derivation of Λ. Then some number r would realize Λ.

For each e, let $p(e)$ be the code of the program that, on x, returns $\langle 0, q\rangle$ with q a realizer of $\theta(e,x)$ when $\theta(e,x)$ is true, and $\langle 1, 0\rangle$ when it is false. By (iii) and the consequence of (i), $p(e)$ realizes $\forall x(\theta(e,x) \vee \neg\theta(e,x))$, and $p(e)$ is computed from e by fixing part of an input. Now let F be the program that, on e, runs r on e, runs the result on $p(e)$, and returns the first component of what it obtains. Since r realizes Λ, F returns a value on every e, and the pair it inspects realizes $\exists x\,\theta(e,x) \vee \forall x\neg\theta(e,x)$. If the first component is 0, the second realizes $\exists x\,\theta(e,x)$; it carries a witness x and a realizer of $\theta(e,x)$, so $\theta(e,x)$ is true and program e returns a value on input 0. If the first component is not 0, the second realizes $\forall x\neg\theta(e,x)$. Were $\theta(e,x)$ true for some x, (iii) would supply a realizer of it, and the realizer of $\neg\theta(e,x)$ would turn that into a realizer of $\bot$. So $\theta(e,x)$ is false for every x, and by (ii) program e returns no value on input 0. Thus F decides, for every e, whether program e halts on input 0.

No program does that. If one did, compose it with the computation of the code of $e \circ K_e$, the program that ignores its input and runs e on e. The result would decide whether program e halts on input e, and no program decides that [21]. Hence Λ has no realizer, the LPO instance at θ has no derivation from MP rules, and LPO does not reduce to MP.

For the final statement, a rule reduces to every set that contains it, and a rule that reduces to a set reduces to every larger set. MP has level 0 and reduces to the empty set by the first assertion, hence to every set. LPO and EM have level 1. A set of level 1 contains LPO or EM, and each of these reduces to the other by the second assertion, so both reduce to such a set. A set of level 0 is contained in {MP}, and by the third assertion neither LPO nor EM reduces to it. □

In constructive mathematics the principles of Definition 6.1 form a chain of strictly increasing strength, MP below LPO below EM [4]. For the rules over this language, with induction on every formula, the chain collapses in two places: the restricted Markov rule (the quantifier-free schema (6.5)) costs nothing, and LPO costs as much as EM. The theorem also leaves no freedom in assigning prices. Suppose each rule receives a degree in an ordered system with a least element and joins, so that a rule reducing to a set never has a degree above the join of the set's degrees, the empty set having the least element as its join. Then MP must receive the least degree and LPO must receive the degree of EM. The two-level chart is the finest such assignment: every other one is obtained from it by exactly one map that preserves the least element, joins and the degrees of the rules.

Remark 6.12 (Heyting arithmetic and Peano arithmetic)**.** The two strengths have standard names, and both identifications are proved against presentations written with no ledger. The base rules of Chapter 5, with induction for every formula, have the language, the successor, addition and multiplication laws and the induction of *Heyting arithmetic*, the intuitionistic first-order theory of 0, S, $+$ and $\cdot$ [4], and we use that name for the posit-free derivations. Chapter 8 shows that they are exactly the translations of derivations in the usual presentation of Heyting arithmetic, up to the way the injectivity of the successor is written (Corollary 8.12). In particular, from every finite context, the posit-free derivations prove exactly what Heyting arithmetic proves. With the EM rule added they correspond to *Peano arithmetic*, the classical theory. Theorem 8.11 in Chapter 8 proves this second identification: from every finite context,

the derivations whose LPO and MP entries are false prove exactly what Peano arithmetic proves. Together with Theorem 6.11 this covers every combination of the three rules, with induction on every formula permitted. If MP may be used but neither EM nor LPO, removing the MP leaves of a derivation gives a posit-free derivation of the same judgment, so the derivable judgments are those of Heyting arithmetic. If EM or LPO may be used, removing the MP leaves and exchanging LPO leaves for EM derivations, or EM leaves for LPO derivations when only LPO is permitted, shows that the derivable judgments are exactly those of Peano arithmetic. The two theories differ, by the third assertion of Theorem 6.11.

Without any induction rule, the picture changes.

Proposition 6.13 (The price of each rule without induction)**.** *Consider only derivations that contain no induction rule. Then EM reduces to LPO and MP together, and each of LPO and MP reduces to EM. LPO does not reduce to MP, MP does not reduce to LPO, and EM reduces to neither of them alone. The three rules therefore mark four strengths: no rule; MP; LPO; and EM, which is equivalent to LPO and MP together.*

Conditions and scope.

Base and premises
Derivations that contain no induction rule, over the base rules of Chapter 5 and the rules EM, LPO and MP of (6.3)–(6.5), the MP rule restricted to quantifier-free bodies.

Conclusion
EM reduces to LPO and MP together, and each of LPO and MP reduces to EM; LPO does not reduce to MP, MP does not reduce to LPO, and EM reduces to neither alone. The rules mark four strengths: no rule, MP, LPO, and EM.

Proof. We first give the three positive reductions, then the four separations, and finally read off the four strengths. Every derivation in the first part is free of induction rules.

The positive reductions. The converse half of Theorem 6.10 derives every LPO instance from one EM rule, at $\exists x\,\varphi$, without induction. For MP, let φ be quantifier-free and assume $\neg\neg\exists x\,\varphi$. The EM rule at $\exists x\,\varphi$ gives $(\exists x\,\varphi) \vee \neg\exists x\,\varphi$. In the first case the conclusion is at hand. In the second, implication elimination applies the hypothesis $\neg\neg\exists x\,\varphi$ to $\neg\exists x\,\varphi$ and yields $\bot$, and the rule for an impossible case gives $\exists x\,\varphi$. Disjunction elimination and implication introduction give the MP instance, with one EM rule and no induction. For EM from LPO and MP, recall that the proof of Theorem 6.10 derives $\Gamma \vdash \varphi \vee \neg\varphi$ by induction on φ, and that induction rules enter it only where an equation $t = s$ is decided through (6.9). Replace each such decision by the derivation of the remark that follows the theorem: with β the formula $t = s \vee \neg(t = s)$ and z a variable that does not occur in it, that derivation proves $\neg\neg\exists z\,\beta$ with no added rule, applies the MP rule at the quantifier-free body β, and recovers β by existential elimination. Hypotheses that it does not use may be added to its context, as in Section 6.7. The result derives every EM instance from LPO and MP rules with no induction rule.

The separations from MP. The third assertion of Theorem 6.11 shows that neither LPO nor EM reduces to MP even when induction on every formula is permitted. A derivation without induction is in particular a derivation, so neither reduces to MP here, and neither reduces to the empty set.

Two stages. The separations from LPO use a model with two stages, and from here on the argument uses a classical metatheory. A *two-stage structure* consists of two structures M_0 and M_1 in the sense of Definition 5.7, each satisfying the six arithmetic rules, and a map h from M_0 *onto* M_1 that preserves 0, S, $+$ and $\cdot$. We call M_0 the earlier stage, M_1 the later stage and h the passage. For an environment ρ in M_0 write $h\rho$ for the environment $k \mapsto h(\rho(k))$ in M_1. A formula is *forced* at ρ, written $\rho \Vdash \varphi$, according to the following clauses:

- $\rho \Vdash t = s$ when t and s have the same value in M_0 under ρ; $\bot$ is never forced;
- $\rho \Vdash \varphi \wedge \psi$ when both are forced, and $\rho \Vdash \varphi \vee \psi$ when one of them is forced;
- $\rho \Vdash \varphi \to \psi$ when $\rho \Vdash \varphi$ implies $\rho \Vdash \psi$, and also $M_1, h\rho \vDash \varphi$ implies $M_1, h\rho \vDash \psi$;
- $\rho \Vdash \forall\varphi$ when $[a, \rho] \Vdash \varphi$ for every element a of M_0, and also $M_1, [b, h\rho] \vDash \varphi$ for every element b of M_1;
- $\rho \Vdash \exists\varphi$ when $[a, \rho] \Vdash \varphi$ for some element a of M_0.

Equations, disjunctions and existential statements are thus read at the earlier stage alone, while an implication or a universal statement must also hold at the later stage. In particular $\rho \Vdash \neg\varphi$ holds exactly when φ is not forced at ρ and is false in M_1 under $h\rho$.

Appendix E proves three facts about these clauses. Forcing *persists*: if $\rho \Vdash \varphi$, then $M_1, h\rho \vDash \varphi$ (Lemma E.3). Forcing respects lifting and substitution in the way Lemma 5.4 describes for satisfaction (the same lemma). And every derivation with no induction rule whose additional rules are all LPO rules is *sound* for forcing: if ρ forces every formula of its context, ρ forces its conclusion (Theorem E.5). The proof of soundness is a rule-by-rule check. Its one case that is not routine is the LPO rule, and it explains why the passage must be onto. Suppose $\rho \Vdash \forall x(\varphi \vee \neg\varphi)$. If some element a of M_0 has $[a, \rho] \Vdash \varphi$, the left alternative $\exists x\varphi$ is forced. Otherwise every a forces $\neg\varphi$, so φ is false in M_1 at every $h(a)$. Because h is onto, these are all the elements of M_1, so $\forall x\neg\varphi$ is forced as well. At the later stage alone the LPO instance is true by classical reasoning.

The model. The later stage is $\mathbb{N}$. The earlier stage C has the elements $0, 1, 2, \ldots$ and a second copy $0', 1', 2', \ldots$, all distinct; its zero is the unprimed 0. Write $|a|$ for the number obtained by erasing a prime. Set $S(n) = n + 1$ and $S(n') = (n + 1)'$. The sum $a + b$ has erased value $|a| + |b|$ and carries a prime exactly when a or b does. The product $a \cdot b$ has erased value $|a|\,|b|$ and carries a prime exactly when a or b does, except that $a \cdot 0 = 0$ for every a. Thus $3' \cdot 0 = 0$ while $0 \cdot 0' = 0'$; the operations of C are not commutative.

C satisfies the six arithmetic rules. A successor is never the unprimed 0: $S(n) = n + 1$, and $S(n')$ is primed. The successor adds one to the erased value and keeps the prime, so it is injective. The equation $a + 0 = a$ holds because adding the unprimed 0 changes neither the value nor the prime, and $a \cdot 0 = 0$ is the exception. In $a + S(b) = S(a + b)$ both sides have erased value $|a| + |b| + 1$ and carry a prime exactly when a or b does. In $a \cdot S(b) = a \cdot b + a$ both sides have erased value $|a|\,|b| + |a|$. Since $S(b)$ is never the unprimed 0, the left side carries a prime exactly when a or b does. On the right, if b is the unprimed 0 then $a \cdot b = 0$ and the sum is primed exactly when a is, which agrees with the left side; otherwise $a \cdot b$ is primed exactly when a or b is, and so is the sum. The passage $h(a) = |a|$ preserves 0 and S, and it preserves sums

and products because they were defined through erased values; the exception has the value $0 = |a| \cdot 0$. It is onto, since $h(n) = n$. So C, $\mathbb{N}$ and h form a two-stage structure.

Excluded middle fails. Let ρ send x to $0'$. The equation $x = 0$ is not forced, because $0'$ and 0 are different elements of C. Its negation is not forced either, because $h(0') = 0$ makes $x = 0$ true in $\mathbb{N}$ under $h\rho$. So the EM instance $x = 0 \vee \neg(x = 0)$ is not forced at ρ, and neither is $\forall x(x = 0 \vee \neg(x = 0))$, whose earlier clause fails at the element $0'$. A derivation from no hypotheses is forced at every environment, since the empty context is forced. By the soundness theorem, neither formula has a derivation without induction whose additional rules are LPO rules. So EM does not reduce to LPO.

Markov's principle fails. Let β be $x = 0 \vee \neg(x = 0)$ and let z be a variable that does not occur in it, so that $\neg\neg\exists z\,\beta \to \exists z\,\beta$ is an instance of the MP schema at the quantifier-free body β. It is the instance the remark after Theorem 6.10 uses to decide $x = 0$. Take ρ with $x \mapsto 0'$ again. In $\mathbb{N}$ under $h\rho$ the formula β is true, since x has the value 0; so $\exists z\,\beta$ is true there and $\neg\exists z\,\beta$ is false. Hence $\neg\exists z\,\beta$ is not forced at ρ (its later clause fails), and $\neg\neg\exists z\,\beta$ is forced: its earlier clause holds because $\neg\exists z\,\beta$ is not forced, and its later clause because $\exists z\,\beta$ is true in $\mathbb{N}$. But $\exists z\,\beta$ is not forced. Written with positions, its body is the lifted formula $\beta\uparrow$, so a witness c in C would give $[c, \rho] \Vdash \beta\uparrow$, which by the lifting half of Lemma E.3 says $\rho \Vdash \beta$; we have just seen that this fails. The first clause of the implication therefore fails, and the MP instance is not forced at ρ. By the soundness theorem it has no derivation without induction whose additional rules are LPO rules, and MP does not reduce to LPO.

Four strengths. Give EM the degree $(1, 1)$, LPO the degree $(1, 0)$ and MP the degree $(0, 1)$, and give a set of rules the coordinatewise maximum of the degrees of its members, or $(0, 0)$ if it is empty. A rule reduces without induction to a set exactly when its degree is at most the degree of the set in both coordinates. If the degree is at most that of the set, then either the rule belongs to the set, or it is LPO or MP and the set contains EM, or it is EM and the set contains both LPO and MP; the positive reductions cover each case. If the degree is not at most that of the set, then either the first coordinate fails, and the rule is LPO or EM while the set is contained in {MP}, or the second fails, and the rule is MP or EM while the set is contained in {LPO}. A rule that reduced to a subset would reduce to the larger set, and the four separations exclude each of these cases. The four degrees are four strengths: no rule; MP; LPO; and EM, which has the degree of LPO and MP together. □

Theorem 6.11 and Proposition 6.13 do not conflict: they concern different sets of permitted derivations. Induction on quantified formulas decides equality, the job that MP did in the remark after Theorem 6.10, and it derives every MP instance. Once the restricted rule costs nothing, the reduction of EM to LPO and MP becomes a reduction of EM to LPO, and the four strengths collapse to two. This is why the ledger records induction in an entry of its own: the price of a principle can depend on the induction available.

The empty ledger permits induction on quantifier-free formulas only. That much induction decides $x = 0$, by splitting into the zero and successor cases, but it does not decide equality in general, and the four strengths persist.

Proposition 6.14 (The price of each rule with quantifier-free induction)**.** *Consider only derivations whose induction formulas are all quantifier-free, that is, derivations whose ledger has a false induction entry. Then the reductions are exactly those of Proposition 6.13: EM reduces to LPO and MP together, each of LPO and MP reduces to EM, LPO does not reduce to MP, MP does not reduce to LPO, and EM reduces to neither of them alone. Two explicit formulas witness the separations from LPO: the MP instance*

$$\neg\neg\exists x(x + y = z) \to \exists x(x + y = z)$$

has no derivation whose ledger has false EM, MP and induction entries, even with LPO permitted, and therefore none with an empty ledger; nor has $\forall x\,\forall y(x = y \vee \neg(x = y))$*.*

Conditions and scope.

Base and premises

Derivations whose induction formulas are all quantifier-free, over the base rules of Chapter 5 and the rules EM, LPO and MP of (6.3)–(6.5), the MP rule restricted to quantifier-free bodies.

Conclusion

The reductions are exactly those of Proposition 6.13. With LPO permitted but no EM or MP rule and no induction on a formula with quantifiers, neither the MP instance $\neg\neg\exists x(x + y = z) \to \exists x(x + y = z)$ nor $\forall x\,\forall y(x = y \vee \neg(x = y))$ has a derivation.

Proof. The positive reductions are those of Proposition 6.13. Their derivations use no induction at all, so their ledgers have a false induction entry. The third assertion of Theorem 6.11 shows that neither LPO nor EM reduces to MP even when induction on every formula is permitted, so neither reduces to MP here. It remains to show that neither MP nor EM reduces to LPO, through the two displayed formulas. The structure C of Proposition 6.13 no longer serves, because quantifier-free induction proves $\forall x(x = 0 \vee \neg(x = 0))$, by the split into the zero and successor cases. We build another two-stage structure, again with a classical metatheory, in which the earlier stage also forces quantifier-free induction, and apply Theorem E.5 in its second form.

The later stage. Let P consist of the polynomials in t with integer coefficients whose leading coefficient is nonnegative; these are 0 and the polynomials with a positive leading coefficient. Its operations are 0, $p \mapsto p + 1$, and the sum and product of polynomials. For nonzero p, q in P the leading terms never cancel. If p has the larger degree, $p + q$ has the leading coefficient of p; if the degrees are equal, the leading coefficients add to a positive number. So $p + q$ lies in P and has the larger of the two degrees. The leading coefficient of pq is the product of the leading coefficients, which is positive, so pq lies in P and has degree $\deg p + \deg q$. Adding 0 or multiplying by 0 stays in P, and the successor is the sum with 1. The constant elements of P are the natural numbers. If p has positive degree, then for every natural number k the polynomial $p - k$ has the same degree and leading coefficient, so it lies in P and has positive degree; the elements $p - k$ are pairwise distinct, and the successor of $p - (k + 1)$ is $p - k$.

P satisfies the six arithmetic rules. If $p + 1 = 0$, then $p = -1$, whose leading coefficient is negative, so a successor is never 0. Adding 1 is injective. The four recursive equations, such as $p(q + 1) = pq + p$, are identities of polynomials.

Eventual constancy in P. Fix a quantifier-free formula ψ and an environment σ in P; the variable $x = x_0$ is singled out, and the others are parameters with the values σ gives them. For each term s let $s_\sigma(X)$ be the polynomial in a new variable X, with coefficients in the integer polynomials in t, obtained by writing X for x_0, the polynomial $\sigma(k)$ for x_{k+1}, and computing with 0, S, $+$ and $\cdot$ as polynomials. Induction on s shows that the value of s under $[w, \sigma]$ is $s_\sigma(w)$ for every w in P. So an equation $s = r$ holds under $[w, \sigma]$ exactly when w is a root of $s_\sigma - r_\sigma$.

Now let $w_0, w_1, w_2, \ldots$ be pairwise distinct elements of P. We claim that for each equation $s = r$ there is an N such that, for every $k \geq N$, the equation holds under $[w_k, \sigma]$ exactly when $s_\sigma = r_\sigma$ as polynomials in X. If $s_\sigma = r_\sigma$, it holds at every w. Otherwise $s_\sigma - r_\sigma$ is a nonzero polynomial whose coefficients lie in the integer polynomials in t, which have no zero divisors. Such a polynomial has at most as many roots as its degree. A root c splits off a factor: $s_\sigma - r_\sigma = (X - c) \cdot g$ with g of degree one less, by division by $X - c$; and every other root c' is a root of g, because $(c' - c)\, g(c') = 0$ and there are no zero divisors. Induction on the degree gives the bound. Since the w_k are distinct, only finitely many of them are roots; take N beyond those. Call the truth value obtained by giving each equation $s = r$ of ψ the value of $s_\sigma = r_\sigma$, and combining these by the truth tables of $\wedge$, $\vee$, $\to$ and $\bot$, the *generic value* of ψ at σ. Satisfaction of a quantifier-free formula is computed from its equations by the same truth tables, and ψ has finitely many equations. Taking the largest of their thresholds, we obtain an N such that ψ holds under $[w_k, \sigma]$, for every $k \geq N$, exactly when its generic value at σ is true.

Quantifier-free induction holds in P. Suppose $\psi(0)$ and $\forall x(\psi(x) \to \psi(S(x)))$ hold under σ. By Lemma 5.4, ψ holds under $[0, \sigma]$, and whenever it holds under $[w, \sigma]$ it holds under $[w + 1, \sigma]$. Ordinary induction gives ψ under $[n, \sigma]$ for every natural number n. The natural numbers are pairwise distinct elements of P, so the generic value of ψ at σ is true. Let w be any element of P. If w is constant, it is a natural number. Otherwise w has positive degree, and $w, w - 1, w - 2, \ldots$ are pairwise distinct elements of P; for large N, ψ holds under $[w - N, \sigma]$, because the generic value is true. Applying the step N times carries ψ up to $[w, \sigma]$. Thus $\forall x \psi(x)$ holds under σ, and P satisfies every quantifier-free induction instance.

The earlier stage. Let Q consist of the pairs (u, v) of elements of P that are equal or both of positive degree, with $0 = (0, 0)$ and the successor, sum and product taken in each coordinate. Every element of Q is either (n, n) for a natural number n or has both coordinates of positive degree. Q is closed under the operations. The successor keeps equal pairs equal and does not change degrees. A sum of two equal pairs is equal; otherwise one summand has both coordinates of positive degree, and since a sum in P has the larger degree of its summands, so does the sum. A product of two equal pairs is equal. Otherwise one factor has both coordinates of positive degree. If the other factor is $(0, 0)$, the product is $(0, 0)$. If not, both coordinates of the other factor are nonzero, because an element of Q with a zero coordinate is $(0, 0)$; degrees add, so both coordinates of the product have positive degree. The six arithmetic rules hold in each coordinate, hence in Q; in particular a successor $(u + 1, v + 1)$ is never $(0, 0)$. The passage $h(u, v) = u$ preserves 0, S, $+$ and $\cdot$, and it is onto, since $h(u, u) = u$. So Q, P and h form a two-stage structure.

Quantifier-free induction is forced at the earlier stage. For an environment ρ in Q write ρ^1 and ρ^2 for its two coordinate environments in P; thus $h\rho = \rho^1$. Since the operations act in each coordinate, the value of a term under ρ is the pair of its values under ρ^1 and ρ^2, and $\rho \Vdash s = r$ holds exactly when $s = r$ holds in P under ρ^1 and under ρ^2. For a quantifier-free

formula, induction on the formula shows that whether it is forced at ρ is determined by two assignments of truth values to its equations: the *earlier* assignment, which records which equations are forced at ρ, and the *later* one, which records which hold in P under $h\rho$. Equations are read from the earlier assignment, $\bot$ is false, $\wedge$ and $\vee$ combine their components, and an implication $\varphi \to \chi$ requires that forcing of φ give forcing of χ and that its truth table, read from the later assignment, be true.

Let ψ be quantifier-free and suppose $\rho \Vdash \psi(0)$ and $\rho \Vdash \forall x(\psi(x) \to \psi(S(x)))$. We show $\rho \Vdash \forall x\psi$, which has two parts. The later part asks that ψ hold in P under $[b, \rho^1]$ for every b in P. By persistence (Lemma E.3) both premises hold in P under ρ^1, and quantifier-free induction in P gives $\forall x\psi$ there. The earlier part asks that $[a, \rho] \Vdash \psi$ for every a in Q. By the binding lemma, $[(0,0), \rho] \Vdash \psi$, and the earlier clauses of the second premise show that $[a, \rho] \Vdash \psi$ implies $[S(a), \rho] \Vdash \psi$ for every a. So $[(n,n), \rho] \Vdash \psi$ for every natural number n. Now let $a = (u, v)$ have both coordinates of positive degree, and set $a_k = (u-k, v-k)$. These pairs lie in Q, $a_0 = a$, and $S(a_{k+1}) = a_k$. Along both sequences (k,k) and a_k the first coordinates are pairwise distinct, and so are the second coordinates. Apply eventual constancy to each equation $s = r$ of ψ, once with the parameters ρ^1 and once with ρ^2. For all large k, along either sequence, the equation is forced exactly when $s_{\rho^1} = r_{\rho^1}$ and $s_{\rho^2} = r_{\rho^2}$, and it holds after the passage exactly when $s_{\rho^1} = r_{\rho^1}$. These limiting assignments depend on ρ and ψ alone, not on the sequence. So there is an N such that, for every $k \geq N$, forcing of ψ at $[(k,k), \rho]$ and at $[a_k, \rho]$ have the same value, the one that the two-stage reading gives to the limiting assignments. Along (k,k) that value is true, so $[a_N, \rho] \Vdash \psi$, and N applications of the step carry it up to $[a_0, \rho] = [a, \rho]$.

Soundness. The two-stage structure (Q, P, h) satisfies the hypotheses of the second form of Theorem E.5. Hence every derivation whose ledger has false EM, MP and induction entries, and therefore whose only additional rules are LPO rules and whose induction formulas are quantifier-free, is forced at every environment that forces its context. In particular a formula derived from no hypotheses in this way is forced at every environment.

The MP instance fails. Let ρ send y to (t, t^2) and z to (t^2, t); both pairs lie in Q. Under $h\rho$ the variables have the values t and t^2, and $x = t^2 - t$, which lies in P with leading coefficient 1, satisfies $x + t = t^2$. So $\exists x(x + y = z)$ holds in P under $h\rho$, and $\neg\exists x(x + y = z)$ is false there. Consequently $\neg\exists x(x + y = z)$ is not forced at ρ, its later clause failing, and $\neg\neg\exists x(x + y = z)$ is forced: its earlier clause holds because $\neg\exists x(x + y = z)$ is not forced, and its later clause because $\exists x(x + y = z)$ holds in P. But no pair (a, b) in Q satisfies $(a, b) + (t, t^2) = (t^2, t)$, since the second coordinate would force $b = t - t^2$, whose leading coefficient is -1. So $\exists x(x + y = z)$ is not forced, the earlier clause of the implication fails, and the MP instance is not forced at ρ. It therefore has no derivation whose ledger has false EM, MP and induction entries, even with LPO permitted, and in particular none with an empty ledger.

Excluded middle fails. Let ρ send x to (t, t) and y to (t, t^2). The two pairs differ in the second coordinate, so $x = y$ is not forced. They have the same first coordinate, so $x = y$ holds in P under $h\rho$ and $\neg(x = y)$ is not forced either. So $x = y \vee \neg(x = y)$ is not forced at ρ, and neither is $\forall x \forall y(x = y \vee \neg(x = y))$, whose earlier clauses lead to this environment. Neither formula has a derivation whose ledger has false EM, MP and induction entries.

Thus MP and EM do not reduce to LPO at this level. With the positive reductions and the separations from MP, the comparison of degrees at the end of the proof of Proposition 6.13 applies word for word, and the reductions with quantifier-free induction are exactly those

without induction. □

The collapse to two strengths therefore needs induction on quantified formulas, the resource that the fourth entry of the ledger records.

What soundness needs

These theorems sharpen Theorem 6.6. That theorem assumes the metatheoretic MP for an MP entry and the metatheoretic EM for an EM entry. Neither assumption is needed in that form.

Corollary 6.15 (The metatheoretic price of the three rules)**.** *Let a derivation of $\Gamma \vdash \varphi$ have ledger O. If the EM or the LPO entry of O is true, assume the metatheoretic LPO of Definition 6.1 for the predicates $n \mapsto \big([n, \rho] \vDash \psi\big)$, one for each formula ψ and environment ρ; otherwise assume nothing. Every environment satisfying Γ then satisfies φ. This assumption is the least possible. If every derivation whose additional rules are all LPO rules is sound, or every derivation whose additional rules are all EM rules is sound, then LPO holds for these predicates; and for them it is equivalent to excluded middle for the propositions $\rho \vDash \psi$.*

Proof. Follow the proof of Theorem 6.6. The base cases and the LPO leaves are unchanged, and only the EM and MP leaves need new arguments. At an MP leaf, the second assertion of Theorem 6.9 gives the interpreted instance with no principle. At an EM leaf we need $(\rho \vDash \varphi) \vee \neg(\rho \vDash \varphi)$. The metatheoretic LPO supplies this for every formula, by the argument of Theorem 6.10 read in the metatheory, by induction on the formula and for every environment at once. Equations are decided by comparing values, and the connectives preserve decisions. For $\exists\psi$, apply LPO to the predicate $[n, \rho] \vDash \psi$, which the induction hypothesis decides at every n. For $\forall\psi$, apply it to $[n, \rho] \vDash \neg\psi$, and use the decision of each instance to pass from $\forall n\, \neg\neg\psi$ to $\forall n\, \psi$. Only predicates of the stated form occur, so LPO for them suffices, and the same argument shows that it gives excluded middle for every proposition $\rho \vDash \psi$.

For the converse, the LPO instance at ψ is a derivation from no hypotheses whose only additional rule is LPO. Its soundness at ρ is exactly LPO for the predicate $[n, \rho] \vDash \psi$. The EM instance at ψ gives, in the same way, $(\rho \vDash \psi) \vee \neg(\rho \vDash \psi)$; applied to $\exists\psi$, this excluded middle decides whether some n satisfies $[n, \rho] \vDash \psi$ or none does, which is LPO for that predicate. □

For the formulas of this language, then, interpreting a derivation never needs excluded middle for arbitrary propositions. What the derivations with an EM or LPO entry need, taken together, is exactly LPO for the predicates that formulas define, which is the same as excluded middle for the propositions that formulas express. Excluded middle for arbitrary propositions is not implied by LPO [4]; formulas of the arithmetic language are not arbitrary propositions.

The next chapter carries the two-direction test to a concrete structural question: which identifications can a reader of repeated records make? Supplying an explicit identified pair, asserting only that some pair must exist, supplying only a decision procedure and asking for a yes-or-no answer, the full alternative or the parity of the index, and asking for a classification with no decision procedure lead to different theorem statements, and their costs are proved in both directions. Those costs are stated with metatheoretic principles: those of Definition 6.1, applied to arbitrary decidable predicates, and the weaker omniscience principles of Definition 7.8. At that level Markov's principle is an assumption that the classification

genuinely needs: the classification implies it, and a model in Section 7.12 shows that intuitionistic second-order arithmetic proves neither. The restricted Markov rule of this chapter, the quantifier-free schema (6.5), by contrast costs nothing once induction on every formula is allowed. Theorem 7.38 shows that no two of the principles priced in the next chapter are equivalent.

Questions to work through

1. Write the two derivations of $(0 = 0) \vee \neg(0 = 0)$ and annotate each node with its ledger. What does their comparison establish, and what would examining only the EM proof fail to establish?

2. Reproduce the implication case of Lemma 6.2. Explain why deciding its finite components does not decide an arbitrary existential formula.

3. Why does the quantified-induction example have a nonempty ledger while remaining posit-free? Identify the tested formula rather than looking only at the conclusion.

4. In the proof of conditional soundness, identify the use of the MP rule's quantifier-free restriction. What premise of the metatheoretic principle does it supply?

5. For $\varphi(x)$ given by $x \cdot x = y + y$, find the coefficient lists and the bound B of Lemma 6.8. Carry out the search for $y = 8$. Why does the lemma fail for the formula $\exists z\,(x = z + z)$, which contains a quantifier?

6. Derive an instance of the LPO schema from the EM rule. Where is existential introduction used, and what is the ledger of the result?

7. Locate the inductions in the proof of Theorem 6.10. Show how one MP instance, at a formula under a quantifier whose variable does not occur in it, replaces them for the equation $x = 0$. What is the ledger of the result?

8. Explain why Theorem 6.11 and Proposition 6.13 are consistent. What does induction on quantified formulas supply that separates them, and why does Proposition 6.14 show that induction on quantifier-free formulas does not supply it?

Chapter 7

What can recognize the same record?

A tally distinguishes successive counts. A reader of a tally may retain less. Two different counts can receive the same reading. If the reader also respects addition, that loss of distinction cannot occur arbitrarily: an identification must remain valid after the same further count is added to both sides.

This requirement is strong enough to classify every possible pattern of identification. The answer has a simple shape. Either every count remains distinct, or an initial segment is followed by a repeating part. The mathematical content lies in proving that no other shape is possible.

There is also a logical question. What information have we been given about the reader, and what do we ask the classification to return? An actual pair of identified counts provides more than the assertion that the reader cannot distinguish everything. A way to decide each pair provides more than an arbitrary relation. Asking only whether the reader is faithful asks for less than its index and period. Keeping these differences visible leads to six precise versions of the classification. One needs no additional principle, and each of the other five is equivalent to exactly one named logical principle. The six principles, counting the empty one, are pairwise different over intuitionistic second-order arithmetic (Theorem 7.38), so no version sits at two of them.

We use natural-number coordinates for the orbit in this chapter. The round-trip maps of Chapter 1 preserve addition, so the results describe the same constructed arithmetic. No new carrier is being substituted for it.

7.1 Identifications that respect addition

Definition 7.1 (An additive congruence)**.** A relation $x \sim y$ on the natural numbers is an *additive congruence* if it is reflexive, symmetric and transitive, and satisfies

$$w \sim x \text{ and } y \sim z \quad \Longrightarrow \quad w + y \sim x + z.$$

In particular, $x \sim y$ implies $x + t \sim y + t$ for every t.

The final implication follows by combining $x \sim y$ with $t \sim t$. Its direction matters. A congruence permits us to add the same count to both sides. It need not permit cancellation. A reader can lose distinctions as the count advances. Section 7.8 identifies the congruences that do permit cancellation: they are exactly the ones that survive the passage to signed records.

For an addition-preserving reading r, define

$$x \sim_r y \quad \Longleftrightarrow \quad r(x) = r(y).$$

Equality of readings is an equivalence relation. Moreover, $r(w + y) = r(w) + r(y)$, so equal readings remain equal under addition. Thus the relation is an additive congruence. It is often called the *kernel* of the reading: it records exactly which inputs the reading identifies.

Definition 7.2 (The index–period relation)**.** Let i be a natural number and let $p > 0$. Define

$$x \sim_{i,p} y \quad \Longleftrightarrow \quad x = y \text{ or } \big(i \le x,\ i \le y,\ x \bmod p = y \bmod p\big). \tag{7.1}$$

The number i is the *index*; p is the *period*.

Before i, a count is identified only with itself. From i onward, counts with the same remainder modulo p are identified. For example, the definition gives the repeating chain

$$i \sim_{i,p} i + p \sim_{i,p} i + 2p \sim_{i,p} \cdots.$$

The displayed continuation describes repetitions of the same finite rule. It does not require a completed physical record of all those counts.

Proposition 7.3 (The family is well defined)**.** *For each i and $p > 0$, relation* (7.1) *is an additive congruence. Furthermore,*

$$n \sim_{i,p} n + p \quad \Longleftrightarrow \quad i \le n. \tag{7.2}$$

Proof. Reflexivity uses the equality alternative. Symmetry follows by reversing equality or equality of remainders. For transitivity, an equality alternative lets us replace one endpoint directly. If both relations use the second alternative, all necessary endpoints lie at or beyond i, and equality of remainders is transitive.

For addition, if both input relations are equalities, their sums are equal. If at least one uses the second alternative, both corresponding sums are at least i. Equality of remainders of the inputs gives equality of remainders of their sums, since the remainder of a sum is determined by the remainders of its summands. This proves compatibility with addition.

Finally n and $n + p$ are distinct because $p > 0$, and they have the same remainder modulo p. Their relation therefore holds exactly when n is at least i. □

7.2 How a period reaches an earlier collision

The difficult part of classification is to rule out irregular behaviour. We cannot simply cancel a large common addition in a congruence. Instead, we use an existing repetition to bring two chains together.

Say that a positive number d is a *period at x* when $x \sim x + d$. Addition compatibility and transitivity immediately give

$$x \sim x + kd \qquad (k \in \mathbb{N}).$$

The same period also holds at every later point: add $n - x$ to the relation when $n \ge x$.

Lemma 7.4 (Moving a period back to a collision)**.** *Suppose $d > 0$, $p > 0$, $x \sim x + d$, and $y \sim y + p$ with $y \geq x$. Then $x \sim x + p$.*

Proof. Choose a natural number k with $M = x+kd \geq y$. One explicit choice is $k = \lfloor (y-x)/d \rfloor + 1$. Repeating the first relation gives $x \sim M$. Translating the second relation from y to M gives $M \sim M + p$. Translating $x \sim M$ by p gives $x + p \sim M + p$. Symmetry and transitivity now combine the three relations:

$$x \sim M \sim M + p \sim x + p.$$

No cancellation of a congruence was used. □

Lemma 7.5 (A least period controls every collision)**.** *Suppose $p > 0$, $a \sim a + p$, and p is the least positive period at a. Every positive period at any point is a multiple of p. Also, if $x < y$ and $x \sim y$, then $x \sim x + p$.*

Proof. Let $e > 0$ be a period at some x. Translate that period to $a + x$, which lies at or beyond a. Apply Lemma 7.4, using the existing period p at a. The result is $a \sim a + e$.

Divide e by p: write $e = qp + r$ with $0 \leq r < p$. Repeating the p-period gives $a \sim a + qp$. After adding r, this becomes $a + r \sim a + e$. Together with $a \sim a + e$, it gives $a \sim a + r$. If $r > 0$, this contradicts minimality of p. Therefore $r = 0$, so p divides e.

For the final assertion, $y - x > 0$ is a period at x. Translate the p-period from a to $a + x$, and apply Lemma 7.4 with the existing positive period $y - x$ at x. It yields $x \sim x + p$. □

These two lemmas explain why an apparently local collision determines the shape of every later one. Their hypotheses contain the periods they use; neither proof searches for a period without evidence that one exists.

7.3 The constructive classification

Theorem 7.6 (The classification with an explicit collision)**.** *Let $\sim$ be an additive congruence. Suppose we have a decision procedure for each assertion $x \sim y$, and are given $a \sim b$ with $a \neq b$. Then there are unique natural numbers i and $p > 0$ such that $\sim$ is exactly $\sim_{i,p}$.*

The information in the hypothesis is concrete. The procedure decides individual pairs, and the supplied pair shows that at least one identification occurs. We will use two finite searches whose endpoints are supplied by that pair.

Construction and classification. Reverse a and b if necessary so that $a < b$. Search the integers $1, \ldots, b - a$ for the least p such that $a \sim a + p$. Every test is decidable, and the last candidate succeeds because $a + (b - a) = b$. Thus p exists, is positive and is the least positive period at a.

Next search $0, \ldots, a$ for the least i such that $i \sim i + p$. Again every test is decidable, and a succeeds. This i is also the least index among all natural numbers with that property: any smaller index would have occurred in the same search.

We prove equality of the two relations. Suppose first that $x \sim y$. If $x = y$, the equality alternative of (7.1) applies. If they differ, arrange $x < y$ by symmetry. Lemma 7.5 gives $x \sim x + p$, so minimality of i gives $i \le x \le y$. The positive gap $y - x$ is a period at x, hence a multiple of p. Consequently x and y have the same remainder modulo p. This is the second alternative of (7.1).

Conversely, equality is always a congruence relation. For the second alternative, arrange $x \le y$. Equal remainders mean $y - x = kp$ for some natural number k. The relation $i \sim i + p$ translates to a p-period at x because $i \le x$. Repeating it k times gives $x \sim x + kp = y$. Symmetry handles the opposite ordering. Thus the relations agree at every pair. □

Uniqueness. Suppose the same relation has two descriptions, with parameters (i, p) and (i', p'). The relation $i \sim i + p$, read in the second description, implies that p' divides p: its two unequal endpoints have equal remainders modulo p'. Reversing the descriptions shows that p divides p'. Positivity gives $p = p'$.

With the period fixed, equation (7.2) identifies the index. Since $i' \sim i' + p$, the first description gives $i \le i'$. The second description similarly gives $i' \le i$. Hence $i = i'$. □

The classification proof separates the search from the structural argument. Once a least positive period at a and the least index for that period are supplied, Lemma 7.5 and the two relation inclusions establish the answer. The decision procedure is used to find those values from the explicit collision.

7.4 What changes when the collision is not supplied

Now replace the displayed collision by a negative statement:

$$\neg\big[\forall x, y\,(x \sim y \implies x = y)\big]. \tag{7.3}$$

This says that the congruence is not equality. Constructively, it does not yet hand us a pair to start the two searches.

From here on, *decidable* means that a procedure giving a positive or negative decision is supplied. We use Markov's principle in the corresponding form: for a predicate with such pointwise decisions, $\neg\neg\exists n\, P(n)$ implies $\exists n\, P(n)$. This makes explicit the decision data behind the instance decisions discussed in Chapter 6.

This is the principle of Definition 6.1, applied to an arbitrary decidable predicate. It is not the arithmetic MP rule of Chapter 6. That rule applies only to quantifier-free formulas of the arithmetic language, and Theorem 6.9 shows that each of its instances is settled by a bounded search and, with induction on every formula permitted, is derivable from the base rules. A decidable predicate here is any predicate that comes with decisions, and at this level Markov's principle is genuinely needed. The reversal below shows that the classification implies it, and Section 7.12 gives a model of intuitionistic second-order arithmetic in which Markov's principle fails and the classification of the next theorem fails with it. That system therefore proves neither.

Theorem 7.7 (The classification from negative nontriviality)**.** *The following assertion is equivalent to Markov's principle, over the constructive mathematics retained here:*

For every decidable additive congruence satisfying (7.3)*, there exist i and $p > 0$ for which the congruence is $\sim_{i,p}$.*

Markov's principle suffices. Let $B(k)$ assert that some $x < k$ satisfies $x \sim k$. This is a finite search, so the given pairwise decisions decide $B(k)$. There cannot be no such k. Indeed, if none existed, any related pair x, y would have to be equal: if $x < y$, the count x would witness $B(y)$, and if $y < x$, symmetry would make y witness $B(x)$. This would contradict (7.3). We have therefore established $\neg\neg\exists k\, B(k)$.

Markov's principle gives a successful k, and hence an explicit distinct related pair. Theorem 7.6 applies. □

The converse is more informative. It shows how an arbitrary search question can be stored inside a congruence. Any procedure or principle that grants the stated classification must then be able to answer that question in the required sense.

The classification implies Markov's principle. Let $\theta(n)$ be any decidable predicate, and assume $\neg\neg\exists n\, \theta(n)$. Define

$$x \sim_\theta y \quad \Longleftrightarrow \quad x = y \text{ or } \exists k \le \min(x, y)\ \theta(k). \tag{7.4}$$

This relation is decidable by bounded search. It is reflexive and symmetric. For transitivity, equality alternatives are immediate. If $x \sim_\theta y$ and $y \sim_\theta z$ are both witnessed by successful indices, compare x and z. If $x \le z$, the first witness lies below both x and z; if $z < x$, the second does. Thus $x \sim_\theta z$. For addition compatibility, a successful witness below a related pair remains below the corresponding pair of sums. If both relations are equalities, the sums are equal. So this is an additive congruence.

It cannot be equality. If it were, any successful n would give $n \sim_\theta n+1$ and hence the impossible equation $n = n + 1$. That would refute $\exists n\, \theta(n)$, contradicting its double negation.

Apply the asserted classification to obtain i, p with $p > 0$. The period relation gives $i \sim_\theta i+p$. Its equality alternative is impossible. Its other alternative supplies a k with $\theta(k)$. We have obtained $\exists k\, \theta(k)$ from its double negation, for an arbitrary decidable predicate. That is Markov's principle. □

The theorem concerns the assertion for *every* congruence of the specified kind. It does not say that each particular classification needs MP. When an explicit collision is already available, the constructive proof applies directly.

7.5 What a decision procedure alone can buy

Keep the pairwise decision procedure, but supply neither a collision nor the assertion that the congruence is not equality. What may we then ask of the classification? Three questions are natural. Is the congruence equality? Which of the two alternatives holds: equality, or an

index–period description? And, more modestly, if there is a description, is its index odd or even?

Each question has its own exact price. The prices are principles about binary sequences, and no two of them are equivalent over intuitionistic second-order arithmetic (Theorem 7.38). We state them here; Chapter 12 meets them again on the continuum.

Definition 7.8 (Omniscience principles for binary sequences)**.** Let $\alpha : \mathbb{N} \to \{0,1\}$ be a binary sequence; its value at each specified index is decidable. The *weak limited principle of omniscience*, WLPO, states that every binary sequence satisfies

$$(\forall n,\ \alpha(n) = 0) \quad \text{or} \quad \neg(\forall n,\ \alpha(n) = 0).$$

The *limited principle of omniscience*, LPO, states that every binary sequence satisfies

$$(\forall n,\ \alpha(n) = 0) \quad \text{or} \quad (\exists n,\ \alpha(n) = 1).$$

The *lesser limited principle of omniscience*, LLPO, concerns sequences with at most one value equal to 1. For each such sequence it states

$$(\forall k,\ \alpha(2k) = 0) \quad \text{or} \quad (\forall k,\ \alpha(2k+1) = 0).$$

The binary form of Markov's principle states, for every binary sequence,

$$\mathrm{MP}_{\mathrm{bin}} : \qquad \neg(\forall n, \alpha(n) = 0) \implies \exists n, \alpha(n) = 1.$$

A decision procedure for a predicate on the natural numbers is the same information as the binary sequence of its answers. So the binary LPO is the predicate form of Definition 6.1 for predicates that come with decisions, and $\mathrm{MP}_{\mathrm{bin}}$ is equivalent to the form of Markov's principle used in Theorem 7.7, since $\neg(\forall n,\ \alpha(n) = 0)$ and $\neg\neg\exists n,\ \alpha(n) = 1$ imply each other. WLPO decides whether the all-zero assertion holds, without producing an index. LLPO, given at most one event, names a parity class free of events; it need not say whether an event occurs in the other class.

For a decidable congruence $\sim$, all three questions can be put to one sequence. Let

$$\beta_{\sim}(n) = 1 \quad \iff \quad \text{some } x < n \text{ satisfies } x \sim n. \tag{7.5}$$

Each value is a finite search, decided by the supplied procedure. The sequence is all zero exactly when $\sim$ is equality. If $\sim$ is equality, no $x < n$ satisfies $x \sim n$, because $x < n$ gives $x \neq n$. Conversely, let the sequence be all zero and $x \sim y$, and decide whether $x = y$. If $x \neq y$, then $x < y$ or $y < x$; by symmetry the smaller count is related to the larger, so $\beta_{\sim}(\max(x,y)) = 1$, which is excluded. Hence $x = y$. Only the decidability of equality and order between two counts is used. A value $\beta_{\sim}(n) = 1$ is itself an explicit collision: the search at n returns an $x < n$ with $x \sim n$.

In the other direction we reuse the congruence $\sim_\theta$ of (7.4), whose construction did not use the double-negation hypothesis. Given a binary sequence α, take $\theta(n)$ to be $\alpha(n) = 1$. Then $\sim_\theta$ is a decidable additive congruence, and

$$\sim_\theta \text{ is equality} \quad \iff \quad \forall n,\ \alpha(n) = 0. \tag{7.6}$$

An event at n relates n and $n+1$. Without events, only the equality alternative of (7.4) remains.

Theorem 7.9 (Three questions with a decision procedure)**.** *Over the retained constructive base:*

1. *The assertion that every decidable additive congruence is equality or is not equality is equivalent to WLPO.*
2. *The assertion that every decidable additive congruence is equality or equals some $\sim_{i,p}$ with $p > 0$ is equivalent to LPO.*
3. *The assertion that, for every decidable additive congruence, either every index–period description of it has odd index or every such description has even index, is equivalent to LLPO.*

Moreover, with no added principle, the assertion in (2) is equivalent to the conjunction of the assertion in (1) with the assertion of Theorem 7.7.

Equality has no index–period description, so both alternatives in (3) hold for it vacuously. The uniqueness argument of Theorem 7.6 uses neither the decision procedure nor the collision, so a congruence has at most one description, and (3) asks for one parity bit and nothing more.

The equality test is WLPO. Assume WLPO and apply it to $\beta_\sim$. If the sequence is all zero, $\sim$ is equality. If it is not all zero, $\sim$ cannot be equality, since equality would make it all zero. Conversely, given a binary sequence α, apply the assertion to $\sim_\theta$. By (7.6) its two alternatives are exactly the two alternatives of WLPO for α. □

The full alternative is LPO. Assume LPO and apply it to $\beta_\sim$. If the sequence is all zero, $\sim$ is equality. Otherwise LPO supplies k with $\beta_\sim(k) = 1$. The bounded search at k now produces an explicit pair $x \neq y$ with $x \sim y$, and Theorem 7.6 gives the description.

Conversely, apply the assertion to $\sim_\theta$. The equality alternative gives $\forall n,\ \alpha(n) = 0$ by (7.6). In the other alternative, $i \sim_\theta i + p$ with $p > 0$. These counts are unequal, so the definition of $\sim_\theta$ supplies $k \leq i$ with $\alpha(k) = 1$. Either way we have the conclusion of LPO.

For the final assertion, the equality test in (1) leaves two cases. The positive case is the first alternative of (2). In the negative case the congruence satisfies (7.3), and the assertion of Theorem 7.7 gives a description. Conversely, (2) gives (1) at once, because an index–period congruence is not equality. It also gives the assertion of Theorem 7.7, because for a congruence that is not equality its equality alternative is excluded. □

The last part is the catalogue's copy of Proposition 12.9: deciding the alternative is deciding whether there is anything to find, followed by finding it.

The parity of the index is LLPO. The forward direction needs a sequence with at most one event that records the parity. Call a pair $x < y$ the *first collision* of $\sim$ when $x \sim y$, no related pair $x' < y'$ has $y' < y$, and no $x' < x$ satisfies $x' \sim y$. Two first collisions must agree: each would otherwise violate a leastness condition of the other. Whether a given pair is the first collision is a finite check. For $b \in \{0, 1\}$ define

$$\gamma(2y + b) = 1 \quad \Longleftrightarrow \quad \text{some } x < y \text{ with } x \bmod 2 = b \text{ forms the first collision } (x, y).$$

Each value is decided by finite search, and uniqueness of the first collision gives at most one event.

Suppose $\sim$ equals $\sim_{i,p}$. A related pair $x < y$ satisfies $i \le x$ and $x + p \le y$, because the gap $y - x$ is then a positive multiple of p. Hence the first collision is $(i, i + p)$, and γ has its event at $2(i + p) + (i \bmod 2)$. Now apply LLPO to γ. If every even entry is zero, no description has even index, so every description has odd index. If every odd entry is zero, every description has even index.

Conversely, let α have at most one event. If $\alpha(k) = 1$, the congruence $\sim_\theta$ identifies two counts exactly when they are equal or both at least k; that is $\sim_{k,1}$. Apply the assertion to $\sim_\theta$. In its first alternative, an event at an even place $2j$ would give the description $\sim_{2j,1}$ with even index; so $\alpha(2j) \neq 1$, and since α is binary, $\alpha(2j) = 0$ for every j. The second alternative gives $\alpha(2j + 1) = 0$ for every j in the same way. These are the two alternatives of LLPO. □

The larger element of the first collision is the half y of the position $2y + b$, and the smaller element is then the least count related to y. That is why one parity bit is all the sequence needs to carry.

7.6 The unrestricted catalogue and excluded middle

Finally suppose no pairwise decision procedure is supplied. Ask for a catalogue that always returns one of two mathematical alternatives: either the congruence is equality, or it has an index–period description.

Theorem 7.10 (The full catalogue has exactly the strength of EM)**.** *The assertion*

$$\textit{every additive congruence is equality or some } \sim_{i,p} \textit{ with } p > 0$$

is equivalent to excluded middle over the retained constructive base.

Excluded middle suffices. Use EM to decide whether the congruence is equality. If so, take that alternative. Otherwise EM gives a distinct related pair: denying the existence of any such pair would make every related pair equal. Reverse the pair if needed to obtain $a < b$.

We need least numbers for predicates that may not have supplied decision procedures. EM supplies the following least-number argument. If $P(n)$ is known, use strong induction on n. EM either gives a $k < n$ with $P(k)$, in which case the induction hypothesis gives a least witness, or says no earlier witness exists, in which case n itself is least. Thus every inhabited natural-number predicate has a least witness under EM.

Apply this argument first to positive periods at a, using $b - a$ as a witness, and then to indices carrying that least period, using a as a witness. The structural argument in Theorem 7.6 now identifies the congruence as $\sim_{i,p}$. This proves the catalogue. □

The catalogue implies excluded middle. Let P be any proposition. Define

$$x \sim_P y \quad \Longleftrightarrow \quad x = y \text{ or } P. \tag{7.7}$$

This is an additive congruence without first deciding P. Reflexivity uses equality; symmetry and transitivity preserve an available proof of P or reduce to equality. For addition, if both input relations are equalities, their sums are equal; otherwise a proof of P already supplies the conclusion.

Apply the catalogue. In its equality branch, P would imply $0 \sim_P 1$, and hence $0 = 1$. Therefore $\neg P$. In its index–period branch, $i \sim_P i + p$ with $p > 0$. These numbers are unequal, so definition (7.7) gives P. Either branch supplies a side of $P \vee \neg P$. Since P was arbitrary, excluded middle follows. □

This proof does not turn EM into an executable decision procedure for every proposition. It establishes an equivalence between mathematical assertions in the metatheory. The forward proof keeps its least-witness claims at that level. The version with supplied pairwise procedures and an explicit collision has the more concrete bounded searches. Between the two lies the version of Section 7.5: the same alternative, asked of a congruence that comes with a decision procedure but no collision, costs exactly LPO.

7.7 What the classification tells us about a reader

Return to an addition-preserving reading r from the orbit into an additive commutative monoid V. Its kernel, $x \sim_r y$ exactly when $r(x) = r(y)$, is an additive congruence, and r is injective exactly when its kernel is equality. A question about the reader is therefore a question about a congruence, and the prices follow it.

The only new ingredient is the information carried by the target. If equality of values in V can be decided, the kernel is a decidable congruence. Conversely, every additive congruence $\sim$ is the kernel of a reading: send x to its class in the quotient $\mathbb{N}/\sim$, whose addition $[x] + [y] = [x + y]$ is well defined because $\sim$ respects addition. Two classes are equal exactly when their representatives are related, so equality of classes is decidable when $\sim$ is.

Theorem 7.11 (The recognition dichotomy and its prices)**.** *Consider readings that preserve zero and addition from the orbit into additive commutative monoids.*

1. *If equality in the target of r is decidable and $r(a) = r(b)$ for given $a \neq b$, then the kernel of r is $\sim_{i,p}$ for unique i and $p > 0$. No added principle is used.*
2. *For readings into targets with decidable equality, each of the following assertions about every such reading is equivalent to the principle named after it:*
   (a) *the reading is injective or not injective: WLPO;*
   (b) *the reading is injective or its kernel is some $\sim_{i,p}$ with $p > 0$: LPO;*
   (c) *if the reading is not injective, its kernel is some $\sim_{i,p}$ with $p > 0$: MP;*
   (d) *every index–period description of the kernel has odd index, or every one has even index: LLPO.*
3. *The assertion that every such reading, into an arbitrary additive commutative monoid, is injective or has an index–period kernel is equivalent to EM.*

Proof. For (1), the kernel is a decidable congruence with the exhibited collision $a \sim_r b$; apply Theorem 7.6.

For (2) and (3), compare each assertion with the catalogue assertion of the same price. A reading supplies its kernel, which is decidable when equality in the target is, and injectivity of the reading corresponds to the kernel being equality. A congruence supplies its quotient reading $x \mapsto [x]$, whose kernel is the congruence itself, and whose target has decidable equality when the congruence is decidable. These two translations carry each reading assertion to the corresponding catalogue assertion and back, with no added principle. Theorems 7.7, 7.9 and 7.10 now give the prices. □

The classical statement of the dichotomy, for arbitrary targets, therefore costs exactly excluded middle. A reader whose outputs can be compared pays only LPO for the same dichotomy, and nothing at all once a single collision has been observed.

The result concerns the reading's identifications. It does not assert that every target object is reached, that every physical device has this form, or that addition preservation can be omitted. Those are different hypotheses. Chapter 8 will make the relationship between a quotient and the range of a reading precise.

7.8 Signed records keep only the pure periods

A congruence permits adding the same count to both sides; it need not permit removing one. The signed records of Chapter 4 show exactly when it does. An *additive congruence* on $\mathbb{Z}_\delta$ is defined as in Definition 7.1, with signed records in place of counts. Reading a count n as the signed record $[n, 0]$, such a congruence $\approx$ restricts to the relation $x \sim y$ defined by $[x, 0] \approx [y, 0]$. Call an additive congruence on the natural numbers *cancellative* when $x + z \sim y + z$ implies $x \sim y$.

Theorem 7.12 (Congruences of the signed records)**.** *Restriction is a bijection from the additive congruences of $\mathbb{Z}_\delta$ onto the cancellative additive congruences of the natural numbers. A cancellative congruence $\sim$ extends by*

$$[a, b] \approx [c, d] \quad \Longleftrightarrow \quad a + d \sim c + b, \tag{7.8}$$

and this is its only extension. Among the index–period congruences, $\sim_{i,p}$ is cancellative exactly when $i = 0$.

Proof. A restricted relation is an additive congruence because $[x + y, 0] = [x, 0] + [y, 0]$. It is cancellative because $\mathbb{Z}_\delta$ has negatives: adding $[0, z]$ to both sides of $[x + z, 0] \approx [y + z, 0]$ gives $[x, 0] \approx [y, 0]$.

Now let $\sim$ be cancellative. Relation (7.8) does not depend on representatives. If (a, b) is replaced by a balanced pair (a', b'), so that $a + b' = a' + b$, add b' to both sides of $a + d \sim c + b$. The left side becomes $(a' + d) + b$, so $(a' + d) + b \sim (c + b') + b$, and cancellation removes b. The same argument handles the second pair and the reverse direction. Reflexivity and symmetry are immediate. For transitivity, add f to $a + d \sim c + b$ and b to $c + f \sim e + d$; the two results chain to $(a + f) + d \sim (e + b) + d$, and cancellation removes d. Adding two instances of (7.8) gives the instance for the sums, so the extension respects addition.

The extension restricts to $\sim$, since $[x,0] \approx [y,0]$ means $x+0 \sim y+0$. Conversely, let $\approx'$ be any congruence on $\mathbb{Z}_\delta$ restricting to $\sim$. Adding $[b+d,0]$ to both sides turns $[a,b] \approx' [c,d]$ into $[a+d,0] \approx' [c+b,0]$, and adding $[0,b+d]$ reverses the step. So $\approx'$ is relation (7.8). This proves the bijection and the uniqueness of the extension.

Finally, if $i > 0$ then $0 + i \sim_{i,p} p + i$, since both sums are at least i and have the same remainder modulo p, while 0 and p are different and $0 < i$, so they are not related. If $i = 0$, equal remainders of $x + z$ and $y + z$ give equal remainders of x and y. □

Passing to signed records therefore erases the index. A congruence with index $i > 0$, which keeps the counts below i apart, has no counterpart on the signed records; only the pure periods survive. The prices do not change, provided the parity question is asked of the period: every description now has index zero, so the parity of the index costs nothing, while the parity of the period costs exactly LLPO.

Corollary 7.13 (The signed catalogue)**.** *Let $\approx$ be a decidable additive congruence on $\mathbb{Z}_\delta$ with an exhibited pair of distinct related records. Then there is exactly one $p > 0$ such that*

$$[a,b] \approx [c,d] \quad \Longleftrightarrow \quad (a+d) \bmod p = (c+b) \bmod p.$$

No added principle is used. For congruences on $\mathbb{Z}_\delta$, each of the following assertions is equivalent to the principle named after it:

(a) every decidable congruence that is only known not to be equality has this form: MP;

(b) for every decidable congruence, either every period p describing it in this form is odd, or every such period is even: LLPO;

(c) every decidable congruence is equality or is not equality: WLPO;

(d) every decidable congruence is equality or has this form: LPO;

(e) every congruence, with no decision procedure, is equality or has this form: EM.

The parity of the index, asked as in Theorem 7.9, costs nothing here: every index–period description of the restriction of a congruence on $\mathbb{Z}_\delta$ has index 0.

Proof. Restriction carries a decision procedure for $\approx$ to one for $\sim$, and (7.8) carries one back. Distinct related records $[a,b] \approx [c,d]$ give the collision $a+d \sim c+b$, with $a+d \neq c+b$ because the records are distinct. Theorem 7.6 describes $\sim$ as $\sim_{i,p}$, cancellation forces $i = 0$, and (7.8) gives the displayed relation. The period is unique because the description is.

For assertions (a), (c), (d) and (e), the forward directions follow in the same way from Theorems 7.7, 7.9 and 7.10 applied to the restriction. By (7.8), $\approx$ is equality exactly when $\sim$ is: if $\sim$ is equality, $[a,b] \approx [c,d]$ says $a + d = c + b$, which is equality of the two records, and if $\approx$ is equality, so is its restriction.

The reversals cannot simply reuse $\sim_\theta$: when the first witness of θ is positive, $\sim_\theta$ has a positive index and is not cancellative. Read a signed record through the faithful integer reading I of Proposition 4.6, and for a decidable θ put

$$u \approx_\theta v \quad \Longleftrightarrow \quad u = v, \text{ or } \theta \text{ has a least witness } n \text{ and } (n+1)! \text{ divides } I(v) - I(u).$$

This relation is decidable: when $u \neq v$ the difference $I(v) - I(u)$ is nonzero, and a witness n with $(n+1)!$ dividing it satisfies $n < (n+1)! \leq |I(v) - I(u)|$, so the search is bounded. A least witness is unique, so any two uses of the second alternative involve the same factorial; this makes $\approx_\theta$ reflexive, symmetric and transitive, and it respects addition. If $\approx_\theta$ is equality, θ has no witness, since a witness gives a least one n by bounded search and then relates $[0,0]$ to the distinct record $[(n+1)!, 0]$. If $\approx_\theta$ has a period $p > 0$, it relates $[0,0]$ to $[p,0]$, which needs a witness. These two facts give MP from (a) and LPO from (d), exactly as in the proofs of Theorems 7.7 and 7.9. Without a witness only the alternative $u = v$ remains, so $\approx_\theta$ is equality exactly when θ has no witness. For $\theta(n)$ given by $\alpha(n) = 1$, the two alternatives of (c) at $\approx_\theta$ are therefore the two alternatives of WLPO for α. For EM, take any proposition P and let $u \approx_P v$ hold when $u = v$, or when P holds and 2 divides $I(v) - I(u)$. Its equality alternative refutes P, because P would relate $[0,0]$ and $[2,0]$; a period $p > 0$ relates $[0,0]$ and $[p,0]$, which requires P.

The parity question needs its own sequence, because the index of a restriction carries no information. For (b), let $\gamma(2y + b) = 1$ exactly when y is the least positive count with $[0,0] \approx [y,0]$ and b is the remainder of y modulo 2. Each value is a bounded search, and the least such count is unique, so γ has at most one event. If $\approx$ is described by the period p, the least such count is p, and the event sits at $2p + (p \bmod 2)$, which has the parity of p. Apply LLPO to γ. If every even entry is zero, no describing period is even; if every odd entry is zero, none is odd.

Conversely, let α have at most one event. Put $q(n) = 2n + 1$ when n is even and $q(n) = 2n + 2$ when n is odd, so that $q(n) > n$ and $q(n)$ has the parity of $n + 1$. Let $u \approx_\alpha v$ hold when $u = v$, or when $\alpha(n) = 1$ for some n such that $q(n)$ divides $I(v) - I(u)$. A witness n for a nonzero difference d satisfies $n < q(n) \leq |d|$, so each pair is decided by bounded search. Since there is at most one event, any two uses of the second alternative involve the same n; hence $\approx_\alpha$ is an additive congruence. It is equality when α has no event, and it is described by the period $q(n)$ when $\alpha(n) = 1$. An event at an even place gives an odd period, and an event at an odd place an even one. So the first alternative of (b), every period odd, gives $\alpha(2j + 1) = 0$ for every j, and the second gives $\alpha(2j) = 0$ for every j. These are the two alternatives of LLPO.

Finally, the restriction of a congruence on $\mathbb{Z}_\delta$ is cancellative, so by Theorem 7.12 every index–period description of it has index 0. The alternative "every description has even index" therefore holds for every congruence, with no principle at all. □

7.9 One classification, six prices

We have also learned how to read a logical price correctly. One classification, of one kind of object, has six exact prices. The price is fixed by what is supplied and by what is asked for:

Supplied	Asked for	Exact price
decisions and a collision	the description	none
decisions, and "not equality"	the description	MP
decisions	the parity of the index	LLPO
decisions	equality or not	WLPO
decisions	equality or the description	LPO
nothing	equality or the description	EM

Theorems 7.6, 7.7, 7.9 and 7.10 prove the rows. Theorem 7.11 gives the same table for readers, and Corollary 7.13 all six rows for signed records, the parity row read on the period. The reversals, rather than the presence of a label on one proof, establish these exact equivalences. Each reversal stores an arbitrary sequence or proposition inside a congruence, so any principle that grants the classification must answer that question.

The equivalences place each form exactly at its principle, and the principles themselves are different. No two of the six are equivalent, so no version of the classification is equivalent to two of them. Each separation is a model of intuitionistic second-order arithmetic in which one principle holds and the other fails. Three take their truth values in the open sets of a space: the one-point compactification of an uncountable discrete set (Markov's principle holds, LLPO fails), the remainder $\beta\mathbb{N} \setminus \mathbb{N}$ (LLPO holds, WLPO fails), and $\beta\mathbb{N}$ itself (WLPO holds, MP and LPO fail). The fourth takes its truth values in a chain (LPO holds, EM fails). Each equivalence of the table is also a derivation in that system, so in every model, these four included, each version of the classification has exactly the truth value of its principle. Section 7.10 states the system, Section 7.11 derives the six equivalences inside it, Section 7.12 builds the models, and Section 7.13 shows that the six prices are not all the prices.

7.10 A system and its models of truth values

An equivalence is proved by two constructions. A difference between two principles needs something of another kind: a precise system of reasoning, and an interpretation of that system in which every derivable sentence is true while one principle holds and the other fails. We supply both. We also write each of the six versions of the classification as a sentence of the system and derive, inside the system, its equivalence with its principle (Theorem 7.33). An interpretation of the system then gives each version exactly the value of its principle, for the same reason that it makes every derivable sentence true: no separate computation inside the model is needed. Interpretations of this kind are standard in constructive mathematics [4, 5]; the work particular to this chapter is the derivation of the six equivalences. Chapter 16 builds a system of the same kind for δ's own objects, points, functions and compact sets included, and shows that the system of this section translates into it.

Definition 7.14 (Intuitionistic second-order arithmetic)**.** The language has two kinds of variable: $x, y, \ldots$ for numbers and $X, Y, \ldots$ for sets of numbers. Terms are built from number variables with 0, S, $+$ and $\cdot$; the atomic formulas are equations $t = u$ and memberships $t \in X$; formulas are built from these with $\bot$, $\wedge$, $\vee$, $\to$ and quantifiers over both kinds of variable. We write $\neg\varphi$ for $\varphi \to \bot$ and $\varphi \leftrightarrow \psi$ for $(\varphi \to \psi) \wedge (\psi \to \varphi)$.

A derivation is a finite tree of judgments $\Gamma \vdash \varphi$, where Γ is a finite list of formulas, built from the following rules.

(i) The rules of Chapter 5 for hypotheses, for reflexivity and substitution of equals, for the connectives and for the number quantifiers, and its six arithmetic rules as axioms: $\neg(S(t) = 0)$, $S(t) = S(u) \to t = u$, $t + 0 = t$, $t + S(u) = S(t + u)$, $t \cdot 0 = 0$ and $t \cdot S(u) = t \cdot u + t$. The substitution rule may substitute into any formula of the present language. In Chapter 5 variables are positions and a fresh variable is made by lifting. With the named variables of this chapter the lifting conditions are the usual eigenvariable conditions: universal introduction derives $\varphi(x)$ from Γ for a number variable x not free in Γ and concludes $\forall x\, \varphi(x)$, and existential elimination concludes ψ from $\exists x\, \varphi(x)$ and a derivation of ψ from $\Gamma, \varphi(x)$, where x is free neither in Γ nor in ψ.

(ii) The same four rules for the set quantifiers. Universal introduction proves $\varphi(X)$ for a set variable X not free in Γ and concludes $\forall X\, \varphi(X)$; universal elimination passes from $\forall X\, \varphi(X)$ to $\varphi(Y)$ for any set variable Y; existential introduction passes from $\varphi(Y)$ to $\exists X\, \varphi(X)$; existential elimination concludes ψ from $\exists X\, \varphi(X)$ and a derivation of ψ from $\Gamma, \varphi(X)$, where X is free neither in Γ nor in ψ.

(iii) The induction axiom

$$\forall X \Big(0 \in X \to \forall x \big(x \in X \to S(x) \in X \big) \to \forall x (x \in X) \Big).$$

(iv) The comprehension axiom $\exists X \forall x (x \in X \leftrightarrow \varphi)$ for every formula φ in which X is not free. The formula φ may contain other free variables of both kinds; they are the parameters of the set it describes.

This system is usually called HAS [4, 5]. A *derivation in HAS* of a sentence is a derivation of it from the empty list of hypotheses.

A derivation of $\Gamma \vdash \varphi$ gives a derivation of $\Gamma' \vdash \varphi$ whenever every formula of Γ occurs in Γ': by induction on the derivation, after renaming each eigenvariable that is free in Γ' to a fresh variable. We use this without comment whenever a derived statement is applied inside a longer argument.

The induction rule (5.2) of Chapter 5 is not among the rules. It is derived.

Lemma 7.15 (Induction for every formula)**.** *For every formula $\varphi(x)$, possibly with parameters of both kinds, HAS derives*

$$\varphi(0) \to \forall x \big(\varphi(x) \to \varphi(S(x)) \big) \to \forall x\, \varphi(x).$$

Consequently the induction rule (5.2) *is a derived rule of HAS, for formulas of the present language.*

Proof. Assume $\varphi(0)$ and $\forall x (\varphi(x) \to \varphi(S(x)))$. Comprehension gives a set X with $x \in X \leftrightarrow \varphi(x)$ for every x; we use it by existential elimination, with a set variable X that occurs neither in the hypotheses nor in the conclusion $\forall x\, \varphi(x)$. Then $0 \in X$, and $x \in X$ gives $\varphi(x)$, hence $\varphi(S(x))$, hence $S(x) \in X$. The induction axiom at X gives $\forall x (x \in X)$, and so $\forall x\, \varphi(x)$. Discharging the two hypotheses gives the displayed sentence. Moving it into the context Γ of the two premises of (5.2) and applying it to them gives the rule. □

Lemma 7.16 (Heyting arithmetic inside HAS)**.** *Every posit-free derivation of Chapter 6, that is, every derivation from the base rules of Chapter 5 with induction on every formula and with no EM, LPO or MP rule, becomes a derivation in HAS of the same judgment, when its formulas are read as formulas of the present language without set variables. In particular, every theorem of Heyting arithmetic is derivable in HAS.*

Proof. By induction on the derivation. Every base rule other than induction and injectivity is a rule of HAS by clause (i) of Definition 7.14, with the same premises and conclusion; the injectivity rule, from $S(t) = S(u)$ to $t = u$, is the axiom $S(t) = S(u) \to t = u$ followed by implication elimination. The lifting conditions of Chapter 5 become the eigenvariable conditions of clause (i). An induction rule is replaced by the derived rule of Lemma 7.15. The second sentence is the first applied to the posit-free derivations, which prove exactly the theorems of Heyting arithmetic (Remark 6.12). □

Order is written with existential quantifiers:

$$x \le y \;:\equiv\; \exists d\,(x + d = y), \qquad x < y \;:\equiv\; \exists d\,\big(x + S(d) = y\big). \tag{7.9}$$

We use the pairing

$$J(x, y) = (x + y) \cdot (x + y) + x. \tag{7.10}$$

Lemma 7.17 (Arithmetic used in the derivations)**.** *HAS derives the universal closures of the following formulas.*

(a) $x = y \lor \neg(x = y)$.

(b) $x \le y \lor y < x$; $x < y \lor x = y \lor y < x$.

(c) $\neg(x < 0)$; $x < S(y) \leftrightarrow x \le y$; $x \le y \land \neg(x = y) \to x < y$; $x < y \to \neg(y \le x)$; $x \le y \to y \le z \to x \le z$; $x < y \to y \le z \to x < z$; $x \le x + y$; $\neg(p = 0) \to x < x + p$; $x < y \to x \le y$; $x \le y \to y < z \to x < z$; $x \le y \to x + z \le y + z$.

(d) $\neg(d = 0) \to y \le x + d \cdot y$.

(e) $\neg(p = 0) \to \exists q\, \exists r\big(r < p \land e = q \cdot p + r\big)$.

(f) $\exists h\,(x = h + h) \lor \exists h\,\big(x = S(h + h)\big)$; $\neg(h + h = S(k + k))$; $h + h = k + k \to h = k$.

(g) $x + z = y + z \to x = y$, *together with the commutative semiring laws for* $+$ *and* $\cdot$.

(h) $J(x, y) = J(x', y') \to x = x' \land y = y'$.

Proof. By Lemma 7.16 it is enough that each formula is a theorem of Heyting arithmetic, and each is proved there by induction from the six arithmetic rules. The semiring laws and cancellation in (g) are the derivations of Chapter 3 written in the language, as in Section 5.7. For (a), induct on x for the formula $\forall y(x = y \lor \neg(x = y))$: every y is 0 or a successor (a one-step induction), $S(x) \neq 0$, and $S(x) = S(y)$ holds exactly when $x = y$. For (b), induct on x for $\forall y(x \le y \lor y < x)$, using that $y = 0$ or $y = S(y')$; the second form follows with (a). The

formulas in (c) follow from the definitions (7.9) by cancellation, the rule $\neg(S(t) = 0)$ and the fact that every number is 0 or a successor. For (d), write $d = S(d')$; then $x+d\cdot y = (x+d'\cdot y)+y$. For (e), induct on e: for $e = 0$ take $q = r = 0$, and $0 < p$ since $p \neq 0$; from $e = qp + r$ with $r < p$, either $S(r) < p$, and $S(e) = qp + S(r)$, or $S(r) = p$, and $S(e) = (q + 1)p + 0$. For the first part of (f), induct on x: $0 = 0 + 0$; $x = h + h$ gives $S(x) = S(h + h)$; and $x = S(h + h)$ gives $S(x) = S(h) + S(h)$. For the second part, induct on h for $\forall k\neg(h + h = S(k + k))$: $0 \neq S(k + k)$, and $S(h) + S(h) = S(k + k)$ forces $k = S(k')$ and then $h + h = S(k' + k')$. For the third part, $h < k$ would give $h + h < k + k$, and $k < h$ symmetrically, so (b) gives $h = k$. For (h), put $s = x + y$ and $s' = x' + y'$. If $s < s'$ then $J(x, y) \leq s\cdot s + s < (s + 1)\cdot(s + 1) \leq s'\cdot s' \leq J(x', y')$, using that $u \leq v$ gives $u\cdot u \leq v\cdot v$, and symmetrically if $s' < s$; so (b) gives $s = s'$, cancellation in $J(x, y) = J(x', y')$ gives $x = x'$, and cancellation in $x + y = x + y'$ gives $y = y'$. □

The objects of this chapter now become sets. A binary sequence is a set X with $n \in X \vee \neg\, n \in X$ for every n, read as $\alpha(n) = 1$ exactly when $n \in X$. A relation on counts is a set R read through the pairing: we write $x \sim y$, or $x \sim_R y$ when R must be named, for $J(x, y) \in R$. By Lemma 7.17(h) distinct pairs have distinct codes, so every relation on counts arises in this way from a set of numbers.

Definition 7.18 (The principles and the versions as sentences)**.** For a set X and a set R write

$$\begin{aligned}
\mathrm{Dec}(X) &:\equiv \forall n\,(n \in X \vee \neg\, n \in X), & \mathrm{Dec}_2(R) &:\equiv \forall x\,\forall y(x \sim y \vee \neg\, x \sim y),\\
\mathrm{Z}(X) &:\equiv \forall n\,\neg\, n \in X, & \mathrm{E}(X) &:\equiv \exists n\,(n \in X),\\
\mathrm{A}(X) &:\equiv \forall m\,\forall n\,(m \in X \to n \in X \to m = n), & \mathrm{Eq}(R) &:\equiv \forall x\,\forall y(x \sim y \to x = y),\\
\mathrm{Ev}(X) &:\equiv \forall k\,\neg\,(k + k) \in X, & \mathrm{Od}(X) &:\equiv \forall k\,\neg\, S(k + k) \in X.
\end{aligned}$$

$\mathrm{Cong}(R)$ is the conjunction of

$$\forall x(x \sim x), \qquad \forall x\,\forall y(x \sim y \to y \sim x), \qquad \forall x\,\forall y\,\forall z(x \sim y \to y \sim z \to x \sim z),$$
$$\forall w\,\forall x\,\forall y\,\forall z(w \sim x \to y \sim z \to w + y \sim x + z).$$

The relation (7.1) and a description are

$$\begin{aligned}
x \equiv y\,(\mathrm{mod}\ p) &:\equiv \exists d\big(x = y + p\cdot d \ \vee\ y = x + p\cdot d\big),\\
x \sim_{i,p} y &:\equiv x = y \ \vee\ \big(i \leq x \wedge i \leq y \wedge x \equiv y\,(\mathrm{mod}\ p)\big),\\
\mathrm{Desc}_R(i, p) &:\equiv \forall x\,\forall y\big(x \sim y \leftrightarrow x \sim_{i,p} y\big),\\
\mathrm{D}(R) &:\equiv \exists i\,\exists p\,\big(\neg(p = 0) \wedge \mathrm{Desc}_R(i, p)\big),
\end{aligned}$$

and the two parity assertions are

$$\begin{aligned}
\mathrm{Odd}(R) &:\equiv \forall i\,\forall p\,\big(\neg(p = 0) \to \mathrm{Desc}_R(i, p) \to \exists h\,(i = S(h + h))\big),\\
\mathrm{Even}(R) &:\equiv \forall i\,\forall p\,\big(\neg(p = 0) \to \mathrm{Desc}_R(i, p) \to \exists h\,(i = h + h)\big).
\end{aligned}$$

The principles of Definition 7.8 and excluded middle are the sentences

$$\begin{aligned}
\mathrm{MP} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \neg\mathrm{Z}(X) \to \mathrm{E}(X)\big),\\
\mathrm{LLPO} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{A}(X) \to \mathrm{Ev}(X) \vee \mathrm{Od}(X)\big),\\
\mathrm{WLPO} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{Z}(X) \vee \neg\mathrm{Z}(X)\big),\\
\mathrm{LPO} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{Z}(X) \vee \mathrm{E}(X)\big),\\
\mathrm{EM} &:\equiv \forall X(0 \in X \vee \neg\, 0 \in X),
\end{aligned}$$

and the empty principle is the sentence $\bot \to \bot$. The six versions of the classification are

$$\begin{aligned}
\mathsf{V}_{\mathrm{none}} &:\equiv \forall R\,\big(\mathrm{Dec}_2(R) \to \mathrm{Cong}(R) \to \forall x \forall y(\neg(x = y) \to x \sim y \to \mathrm{D}(R))\big),\\
\mathsf{V}_{\mathrm{MP}} &:\equiv \forall R\,\big(\mathrm{Dec}_2(R) \to \mathrm{Cong}(R) \to \neg\mathrm{Eq}(R) \to \mathrm{D}(R)\big),\\
\mathsf{V}_{\mathrm{LLPO}} &:\equiv \forall R\,\big(\mathrm{Dec}_2(R) \to \mathrm{Cong}(R) \to \mathrm{Odd}(R) \vee \mathrm{Even}(R)\big),\\
\mathsf{V}_{\mathrm{WLPO}} &:\equiv \forall R\,\big(\mathrm{Dec}_2(R) \to \mathrm{Cong}(R) \to \mathrm{Eq}(R) \vee \neg\mathrm{Eq}(R)\big),\\
\mathsf{V}_{\mathrm{LPO}} &:\equiv \forall R\,\big(\mathrm{Dec}_2(R) \to \mathrm{Cong}(R) \to \mathrm{Eq}(R) \vee \mathrm{D}(R)\big),\\
\mathsf{V}_{\mathrm{EM}} &:\equiv \forall R\,\big(\mathrm{Cong}(R) \to \mathrm{Eq}(R) \vee \mathrm{D}(R)\big).
\end{aligned} \tag{7.11}$$

The six sentences (7.11) state in this language exactly what the six rows of the table in Section 7.9 assert: $\mathsf{V}_{\mathrm{none}}$ is the existence assertion of Theorem 7.6, V_{MP} the assertion of Theorem 7.7, $\mathsf{V}_{\mathrm{WLPO}}$, $\mathsf{V}_{\mathrm{LPO}}$ and $\mathsf{V}_{\mathrm{LLPO}}$ the assertions (1), (2) and (3) of Theorem 7.9, and V_{EM} the assertion of Theorem 7.10. A decision procedure for a relation is Dec_2, and a description by i and $p > 0$ is $\neg(p = 0) \wedge \mathrm{Desc}_R(i, p)$. Excluded middle is stated for the single proposition $0 \in X$; through comprehension it gives every instance $\varphi \vee \neg\varphi$, since a set X with $\forall x(x \in X \leftrightarrow \varphi)$, where x is not free in φ, has $0 \in X \leftrightarrow \varphi$.

Definition 7.19 (Models of truth values)**.** A *system of truth values* is a set H ordered by strength, with a meet and a join for every family of elements and an implication $\Rightarrow$ satisfying

$$a \wedge c \le b \quad \Longleftrightarrow \quad c \le (a \Rightarrow b).$$

Its least and greatest elements are $\bot$ and $\top$, and $\neg a$ is $a \Rightarrow \bot$. A formula of Definition 7.14 receives a value in H under an assignment to its free variables. Numbers range over $\mathbb{N}$, and an equation has the value $\top$ when it holds there and $\bot$ otherwise. Sets range over all assignments $A : \mathbb{N} \to H$, and $t \in X$ has the value $A(t)$. The connectives $\wedge$, $\vee$ and $\to$ are meet, join and $\Rightarrow$; a universal quantifier is the meet of the values of its instances, and an existential quantifier their join.

We write $[\![\varphi]\!]_{\rho,\sigma}$ for the value of φ under the assignment ρ of numbers to number variables and σ of maps $\mathbb{N} \to H$ to set variables, and $\rho[x \mapsto n]$, $\sigma[X \mapsto A]$ for the assignments changed at one variable. A number term t has the value $t^{\rho} \in \mathbb{N}$ computed in $\mathbb{N}$. In such a model a relation read through the pairing is an assignment of a truth value to each pair of counts, and every such assignment arises, since J is injective.

The open sets of a topological space form such a system. Meet and join of two open sets are intersection and union, the join of a family is its union, the meet of a family is the interior of its intersection, and $U \Rightarrow V$ is the interior of the union of V with the complement of U. So $\neg U$ is the largest open set disjoint from U. A chain with a meet and a join for every family is another example. For $0 < 1 < 2 < \cdots < \infty$, the implication $a \Rightarrow b$ is $\top$ when $a \le b$ and b otherwise.

Three properties of a system of truth values carry the proof below. First, $a \wedge (a \Rightarrow b) \le b$, which is the defining equivalence at $c = a \Rightarrow b$. Second, meeting with a fixed element distributes over every join,

$$a \wedge \bigvee_j b_j = \bigvee_j (a \wedge b_j), \tag{7.12}$$

because $\bigvee_j (a \wedge b_j) \le c$ means $b_j \le (a \Rightarrow c)$ for every j, hence $\bigvee_j b_j \le (a \Rightarrow c)$, hence $a \wedge \bigvee_j b_j \le c$. Taking $c = \bigvee_j (a \wedge b_j)$ gives one inequality, and the other holds because each $a \wedge b_j$ lies below

$a \wedge \bigvee_j b_j$. Third, $(a \wedge b) \Rightarrow c$ and $a \Rightarrow (b \Rightarrow c)$ are equal, since an element lies below either exactly when its meet with $a \wedge b$ lies below c.

Lemma 7.20 (Values respect substitution)**.** *For every formula φ, number variable x, term t and set variables X, Y: $[\![\varphi(t)]\!]_{\rho,\sigma} = [\![\varphi(x)]\!]_{\rho[x \mapsto t^{\rho}],\sigma}$ and $[\![\varphi(Y)]\!]_{\rho,\sigma} = [\![\varphi(X)]\!]_{\rho,\sigma[X \mapsto \sigma(Y)]}$, where the substitutions rename bound variables as usual; and the value of φ depends only on the values of its free variables.*

Proof. By induction on φ. For an atomic formula the value of a term after substitution is the value of the term at the changed assignment, by induction on the term. The connectives commute with substitution and with the operations of H. At a quantifier, the substitution passes under the binder with its bound variable renamed away from t (or Y), and the induction hypothesis applies to each instance before the meet or join is taken. The last assertion is proved in the same induction. These are the identities of Lemma 5.4, with values in H in place of satisfaction and with set variables carried along; the proof is the same. □

Lemma 7.21 (Derivable sentences are true in every model)**.** *If a sentence of Definition 7.14 has a derivation, its value in every system of truth values is $\top$.*

Proof. For a list $\Gamma = \varphi_1, \dots, \varphi_k$ write $[\![\Gamma]\!]_{\rho,\sigma} = [\![\varphi_1]\!]_{\rho,\sigma} \wedge \cdots \wedge [\![\varphi_k]\!]_{\rho,\sigma}$, which is $\top$ for the empty list. We prove, by induction on derivations, that a derivation of $\Gamma \vdash \varphi$ gives

$$[\![\Gamma]\!]_{\rho,\sigma} \leq [\![\varphi]\!]_{\rho,\sigma} \qquad \text{for every assignment } \rho, \sigma.$$

For a sentence derived from no hypotheses this says $\top \leq [\![\varphi]\!]$, which is the lemma. We fix ρ, σ in each case, abbreviate $g = [\![\Gamma]\!]_{\rho,\sigma}$ and drop the subscripts when they are unchanged.

Hypotheses and absurdity. A member of Γ lies above the meet g. If $g \leq [\![\bot]\!] = \bot$, then $g \leq [\![\varphi]\!]$ for every φ.

Conjunction and disjunction. The rules for $\wedge$ are the facts that $g \leq a \wedge b$ exactly when $g \leq a$ and $g \leq b$. Disjunction introduction uses $a \leq a \vee b$ and $b \leq a \vee b$. For disjunction elimination, the premises give $g \leq a \vee b$, $g \wedge a \leq c$ and $g \wedge b \leq c$, where a, b, c are the values of the two cases and of the conclusion. By (7.12), $g = g \wedge (a \vee b) = (g \wedge a) \vee (g \wedge b) \leq c$.

Implication. Introduction turns $g \wedge a \leq b$ into $g \leq (a \Rightarrow b)$ by the defining equivalence. Elimination turns $g \leq (a \Rightarrow b)$ and $g \leq a$ into $g \leq a \wedge (a \Rightarrow b) \leq b$.

Number quantifiers. Universal introduction has the premise $\Gamma \vdash \varphi(x)$ with x not free in Γ. For each $n \in \mathbb{N}$ the induction hypothesis at $\rho[x \mapsto n]$ gives $[\![\Gamma]\!]_{\rho[x \mapsto n],\sigma} \leq [\![\varphi(x)]\!]_{\rho[x \mapsto n],\sigma}$, and the left side is g by Lemma 7.20, because x is not free in Γ. Hence g lies below the meet of these values, which is $[\![\forall x \varphi(x)]\!]$. Universal elimination uses that this meet lies below its instance at $n = t^{\rho}$, which is $[\![\varphi(t)]\!]$ by Lemma 7.20. Existential introduction uses that $[\![\varphi(t)]\!]$ is one of the values whose join is $[\![\exists x \varphi(x)]\!]$. For existential elimination, the second premise derives ψ from $\Gamma, \varphi(x)$ with x free neither in Γ nor in ψ. At $\rho[x \mapsto n]$ it gives $g \wedge [\![\varphi(x)]\!]_{\rho[x \mapsto n],\sigma} \leq [\![\psi]\!]$ for every n, and the first premise gives $g \leq \bigvee_n [\![\varphi(x)]\!]_{\rho[x \mapsto n],\sigma}$. By (7.12),

$$g = g \wedge \bigvee_n [\![\varphi(x)]\!]_{\rho[x \mapsto n],\sigma} = \bigvee_n \big(g \wedge [\![\varphi(x)]\!]_{\rho[x \mapsto n],\sigma}\big) \leq [\![\psi]\!].$$

This is the step at which joins of arbitrary families, and their distributivity, are needed.

Set quantifiers. The four cases are the same, with the assignments $\sigma[X \mapsto A]$, for all $A : \mathbb{N} \to H$, in place of $\rho[x \mapsto n]$. Universal elimination at a set variable Y uses the instance $A = \sigma(Y)$, and existential introduction from $\varphi(Y)$ the same instance, both through Lemma 7.20.

Equality. An equation $t = t$ has the value $\top$. For substitution of equals, the premises give $g \le [\![t = s]\!]$ and $g \le [\![\varphi(t)]\!]$. If $t^\rho = s^\rho$, then $[\![\varphi(t)]\!] = [\![\varphi(x)]\!]_{\rho[x \mapsto t^\rho],\sigma} = [\![\varphi(s)]\!]$ by Lemma 7.20. Otherwise $[\![t = s]\!] = \bot$, so $g = \bot$.

Arithmetic axioms. The four equations hold in $\mathbb{N}$ at every assignment, so they have the value $\top$. The equation $S(t) = 0$ fails in $\mathbb{N}$, so its negation has the value $\bot \Rightarrow \bot = \top$. If $S(t) = S(u)$ holds in $\mathbb{N}$ then so does $t = u$, and the implication has the value $\top \Rightarrow \top$; otherwise its antecedent has the value $\bot$. Either way the value is $\top$.

Induction axiom. Fix $A : \mathbb{N} \to H$ and put $c = A(0) \wedge \bigwedge_n \big(A(n) \Rightarrow A(n+1)\big)$. By induction on m, $c \le A(m)$: for $m = 0$ this is a meetand of c, and $c \le A(m)$ together with $c \le (A(m) \Rightarrow A(m+1))$ gives $c \le A(m+1)$. So $c \le \bigwedge_m A(m)$, and by the third property above $A(0) \Rightarrow \big(\bigwedge_n (A(n) \Rightarrow A(n+1)) \Rightarrow \bigwedge_m A(m)\big) = c \Rightarrow \bigwedge_m A(m) = \top$. The axiom is the meet of these values over all A.

Comprehension. Let X not be free in φ and put $A(n) = [\![\varphi]\!]_{\rho[x \mapsto n],\sigma}$. At the assignment $\sigma[X \mapsto A]$, the formula $x \in X$ has the value $A(n)$ at $\rho[x \mapsto n]$, and so does φ, because X is not free in it. So each instance of $x \in X \leftrightarrow \varphi$ has the value $(A(n) \Rightarrow A(n)) \wedge (A(n) \Rightarrow A(n)) = \top$, their meet is $\top$, and the join over all set assignments, which is the value of the axiom, is $\top$. $\square$

7.11 The six equivalences inside the system

We now derive in HAS each version's equivalence with its principle. The arguments are those of Lemmas 7.4 and 7.5 and of the proofs of Theorems 7.6, 7.7, 7.9 and 7.10, written so that every step is a rule of Definition 7.14 or one of the derived rules proved here. Three features of those arguments need care when they are carried out inside a formal system. Every finite search becomes a set given by comprehension, and its decidability is derived by induction. Every least witness comes from the least-number principle for decidable sets, derived in the same way. Every encoding of a sequence or a proposition in a congruence is a set given by comprehension, and its defining property is derived through the injectivity of the pairing.

We reason informally, under three conventions that make each step an inference of HAS. A case distinction is made only on a disjunction that has already been derived: usually a decision from Lemma 7.17, or from a hypothesis $\mathrm{Dec}(X)$ or $\mathrm{Dec}_2(R)$. A set is introduced only by a comprehension instance, displayed with its defining formula; its parameters are the other free variables of that formula. A least element is taken only of a decidable set, by Lemma 7.22. A case that has been refuted closes by the rule for an impossible case, so its conclusion is whatever the argument needs. Throughout, R is a set variable and $x \sim y$ means $J(x, y) \in R$; Lemma 7.17 is used without further mention for order, cancellation, division with remainder and parity.

Lemma 7.22 (Bounded search and least numbers)**.** *For every set X, HAS derives from* $\mathrm{Dec}(X)$*:*

(a) for every n: either some $k < n$ has $k \in X$ and $\neg\, j \in X$ for all $j < k$, or $\neg\, k \in X$ for all $k < n$;

(b) for every n: $\exists k\,(k < n \wedge k \in X) \vee \neg\exists k\,(k < n \wedge k \in X)$;

(c) if $n \in X$, then some $k \in X$ has $\neg\, j \in X$ for all $j < k$.

Proof. For (a), induct on n (Lemma 7.15, with X as a parameter). For $n = 0$ the second alternative holds, since no k satisfies $k < 0$. Suppose the assertion holds at n. In its first alternative, the same k serves at $S(n)$, since $k < n$ gives $k < S(n)$. In its second alternative, decide $n \in X$. If $n \in X$, then $k = n$ gives the first alternative at $S(n)$: $n < S(n)$, and every $j < n$ has $\neg\, j \in X$. If $\neg\, n \in X$, the second alternative holds at $S(n)$: from $k < S(n)$ we get $k \leq n$, so either $k < n$ or $k = n$, and both are excluded.

Part (b) is a weakening of (a). For (c), apply (a) at $S(n)$. Its second alternative is refuted by $n < S(n)$ and $n \in X$, and its first alternative is the conclusion. □

Part (c) is the least-number principle for decidable sets. Every search of this chapter is an application of it to a set given by comprehension.

Lemma 7.23 (Congruence laws)**.** *HAS derives from* $\mathrm{Cong}(R)$*, for all numbers:*

(a) $x \sim y \to x + t \sim y + t$;

(b) $x \sim x + d \to x \sim x + d \cdot k$;

(c) if $z \sim z + p$ and $z \leq w$, then $w \sim w + p$;

(d) (period transfer) if $\neg(d = 0)$, $x \sim x + d$, $x \leq y$ and $y \sim y + p$, then $x \sim x + p$.

Proof. For (a), combine $x \sim y$ with $t \sim t$ by addition compatibility. For (b), induct on k. At $k = 0$, $x \sim x$ by reflexivity, and $x + d \cdot 0 = x$. If $x \sim x + d \cdot k$, translate $x \sim x + d$ by $d \cdot k$ using (a): $x + d \cdot k \sim x + d + d \cdot k = x + d \cdot S(k)$. Transitivity gives $x \sim x + d \cdot S(k)$. For (c), write $w = z + c$ and translate by c.

For (d), put $M = x + d \cdot y$. Since $d \neq 0$, Lemma 7.17(d) gives $y \leq M$; write $M = y + c$. By (b), $x \sim M$. Translating $y \sim y + p$ by c gives $M \sim M + p$, and translating $x \sim M$ by p gives $x + p \sim M + p$. Symmetry and transitivity give $x \sim M \sim M + p \sim x + p$. This is Lemma 7.4, with the explicit choice $k = y$ in place of $\lfloor (y - x)/d \rfloor + 1$. With this choice the hypothesis $x \leq y$ is not used; we keep it so that the statement is that of Lemma 7.4. □

Lemma 7.24 (The structure with a least period and a least index)**.** *HAS derives the following from* $\mathrm{Cong}(R)$*,* $\neg(p = 0)$*,* $a \sim a + p$*,*

$$\forall d\big(d < p \to \neg(d = 0) \to \neg\,(a \sim a + d)\big), \qquad i \sim i + p, \qquad \forall z\,\big(z < i \to \neg\,(z \sim z + p)\big).$$

(a) If $x < y$ and $x \sim y$, then $i \leq x$, $i \leq y$ and $x \equiv y \pmod{p}$.

(b) $\mathrm{Desc}_R(i, p)$*.*

Proof. (a) Write $y = x + e$ with $e = S(e')$, so $e \neq 0$. Translating $a \sim a + p$ by x gives $(a + x) \sim (a + x) + p$. Since $x \leq a + x$ and $x \sim x + e$, period transfer (Lemma 7.23(d)) gives $x \sim x + p$. Decide $i \leq x$ or $x < i$; the second case contradicts the leastness of i at $z = x$. So $i \leq x$, and $i \leq y$ because $x \leq y$.

Next, translating $x \sim x + e$ by a gives $(a + x) \sim (a + x) + e$, and period transfer from the period $p \neq 0$ at a, with $a \leq a + x$, gives $a \sim a + e$. Divide: $e = q \cdot p + r$ with $r < p$. By Lemma 7.23(b), $a \sim a + p \cdot q$; translating by r gives $a + r \sim a + e$. With $a \sim a + e$, symmetry and transitivity give $a \sim a + r$. Decide $r = 0$. If $r \neq 0$, the leastness of p at $d = r$ refutes $a \sim a + r$. So $r = 0$, $y = x + p \cdot q$, and $x \equiv y \pmod{p}$ with $d = q$ in its second alternative.

(b) Let $x \sim y$. Decide $x = y$; equality is the first alternative of $x \sim_{i,p} y$. Otherwise $x < y$ or $y < x$. If $x < y$, part (a) gives the second alternative. If $y < x$, symmetry gives $y \sim x$, part (a) gives $i \leq y$, $i \leq x$ and $y \equiv x \pmod{p}$, and exchanging the two alternatives in the definition of the congruence gives $x \equiv y \pmod{p}$.

Conversely let $x \sim_{i,p} y$. If $x = y$, reflexivity gives $x \sim y$. Otherwise $i \leq x$, $i \leq y$, and $y = x + p \cdot d$ or $x = y + p \cdot d$. In the first case, $i \sim i + p$ and $i \leq x$ give $x \sim x + p$ by Lemma 7.23(c), and then $x \sim x + p \cdot d = y$ by (b) of that lemma. The second case gives $y \sim x$ in the same way, and symmetry gives $x \sim y$. □

Part (a) combines the two assertions of Lemma 7.5 at the pair (x, y) with the leastness of i, and part (b) is the structural half of Theorem 7.6. Neither uses a decision procedure for $\sim$.

Proposition 7.25 (The classification with an explicit collision)**.** *HAS derives* V_{none}.

Proof. Let R satisfy $\mathrm{Dec}_2(R)$ and $\mathrm{Cong}(R)$, and let $\neg(a = b)$ and $a \sim b$. Then $a < b$ or $b < a$. In the second case, symmetry gives $b \sim a$, and the argument below applies with a and b exchanged; so let $a < b$, and write $b = a + e$ with $e \neq 0$.

The least period. Comprehension gives a set P_a with

$$\forall d \big(d \in P_a \leftrightarrow \neg(d = 0) \wedge a \sim a + d \big),$$

with parameters a and R. It is decidable: $d = 0$ is decided by Lemma 7.17(a), $a \sim a + d$ by $\mathrm{Dec}_2(R)$, and a conjunction of two decided formulas is decided. Since $e \in P_a$, Lemma 7.22(c) gives $p \in P_a$ with no member of P_a below it: $p \neq 0$, $a \sim a + p$, and $\neg(a \sim a + d)$ for every $d < p$ with $d \neq 0$.

The least index. Comprehension gives a set I_p with $\forall z (z \in I_p \leftrightarrow z \sim z + p)$. It is decidable by $\mathrm{Dec}_2(R)$, and $a \in I_p$. Lemma 7.22(c) gives $i \sim i + p$ with $\neg(z \sim z + p)$ for every $z < i$.

Lemma 7.24(b) now gives $\mathrm{Desc}_R(i, p)$, and with $p \neq 0$ this is $\mathrm{D}(R)$. □

The two searches of the proof of Theorem 7.6 ran over $1, \ldots, b - a$ and $0, \ldots, a$. Here they are the least members of the sets P_a and I_p, which contain $e = b - a$ and a; the bounds are the same.

Lemma 7.26 (The collision search as a set)**.** *Let R satisfy* $\mathrm{Dec}_2(R)$ *and* $\mathrm{Cong}(R)$*. Comprehension gives a set B with*

$$\forall n \big(n \in B \leftrightarrow \exists x (x < n \wedge x \sim n)\big),$$

the set of the collision search $\beta_\sim$ of (7.5)*. HAS derives:*

(a) $\mathrm{Dec}(B)$*;*

(b) $\mathrm{Z}(B) \to \mathrm{Eq}(R)$*;*

(c) $\mathrm{Eq}(R) \to \mathrm{Z}(B)$*;*

(d) $\mathrm{E}(B) \to \mathrm{D}(R)$*.*

Proof. (a) Fix n. Comprehension gives a set C with $\forall x (x \in C \leftrightarrow x \sim n)$, with parameter n; it is decidable by $\mathrm{Dec}_2(R)$. Lemma 7.22(b) at the bound n decides $\exists x (x < n \wedge x \in C)$, which is $n \in B$.

(b) Assume $\mathrm{Z}(B)$ and $x \sim y$. Decide $x = y$. Otherwise $x < y$ or $y < x$. If $x < y$, then x witnesses $y \in B$, against $\mathrm{Z}(B)$. If $y < x$, symmetry gives $y \sim x$, and y witnesses $x \in B$. Both cases are refuted, so $x = y$.

(c) Assume $\mathrm{Eq}(R)$ and $n \in B$, with $x < n$ and $x \sim n$. Then $x = n$, which contradicts $x < n$. So $\neg\, n \in B$ for every n.

(d) From $n \in B$ we obtain $x < n$ with $x \sim n$, so $x \neq n$, and Proposition 7.25 gives $\mathrm{D}(R)$. □

Proposition 7.27 (The forward directions)**.** *HAS derives* $\mathrm{MP} \to \mathsf{V}_{\mathrm{MP}}$, $\mathrm{WLPO} \to \mathsf{V}_{\mathrm{WLPO}}$, $\mathrm{LPO} \to \mathsf{V}_{\mathrm{LPO}}$ *and* $\mathrm{EM} \to \mathsf{V}_{\mathrm{EM}}$.

Proof. For the first three, let R satisfy $\mathrm{Dec}_2(R)$ and $\mathrm{Cong}(R)$, and let B be its collision set (Lemma 7.26), which is decidable.

MP. Assume $\neg\mathrm{Eq}(R)$. By part (b) of the lemma, $\mathrm{Z}(B)$ would give $\mathrm{Eq}(R)$, so $\neg\mathrm{Z}(B)$. MP at B gives $\mathrm{E}(B)$, and part (d) gives $\mathrm{D}(R)$.

WLPO. WLPO at B gives $\mathrm{Z}(B) \vee \neg\mathrm{Z}(B)$. The first case gives $\mathrm{Eq}(R)$ by part (b). In the second, $\mathrm{Eq}(R)$ would give $\mathrm{Z}(B)$ by part (c), so $\neg\mathrm{Eq}(R)$.

LPO. LPO at B gives $\mathrm{Z}(B) \vee \mathrm{E}(B)$, hence $\mathrm{Eq}(R)$ by part (b) or $\mathrm{D}(R)$ by part (d).

EM. Now R satisfies only $\mathrm{Cong}(R)$. First, EM decides the relation. For numbers x and y, comprehension gives a set Y with $\forall z (z \in Y \leftrightarrow x \sim y)$, where z is a variable not occurring in $x \sim y$; EM at Y gives $0 \in Y \vee \neg\, 0 \in Y$, that is, $x \sim y \vee \neg\, x \sim y$. So $\mathrm{Dec}_2(R)$ holds, and the collision set B exists and is decidable. Comprehension gives a set Y' with $\forall z (z \in Y' \leftrightarrow \mathrm{E}(B))$, and EM at Y' gives $\mathrm{E}(B) \vee \neg\mathrm{E}(B)$. The first case gives $\mathrm{D}(R)$. In the second, $n \in B$ would give $\mathrm{E}(B)$, so $\mathrm{Z}(B)$, and $\mathrm{Eq}(R)$ follows. □

The proof of Theorem 7.10 took least witnesses for arbitrary predicates from EM. The derivation above uses EM only twice: to decide the relation, after which the least witnesses come

from the decidable searches, and to decide whether the collision set is inhabited. Either argument proves the forward direction; the second is the shorter one inside HAS.

For the reversals, a set is stored inside a congruence. For a set X, put

$$\theta_X(x,y) \;:\equiv\; x = y \;\vee\; \exists k\,(k \le x \wedge k \le y \wedge k \in X),$$

the relation (7.4) with $\theta(k)$ read as $k \in X$.

Lemma 7.28 (The congruence of a set)**.** *For every set X, comprehension gives a set T_X with*

$$\forall z\Big(z \in T_X \leftrightarrow \exists x\,\exists y\big(z = J(x,y) \wedge \theta_X(x,y)\big)\Big).$$

Write $x \sim_\theta y$ for $J(x,y) \in T_X$. HAS derives:

(a) $x \sim_\theta y \leftrightarrow \theta_X(x,y)$*;*

(b) $\mathrm{Cong}(T_X)$*;*

(c) $\mathrm{Dec}(X) \to \mathrm{Dec}_2(T_X)$*;*

(d) $\mathrm{Eq}(T_X) \to \mathrm{Z}(X)$ *and* $\mathrm{Z}(X) \to \mathrm{Eq}(T_X)$*;*

(e) $\mathrm{D}(T_X) \to \mathrm{E}(X)$*;*

(f) if $\mathrm{A}(X)$ *and* $k \in X$*, then* $\mathrm{Desc}_{T_X}(k,1)$*.*

Proof. (a) From right to left, x and y themselves witness the existential. From left to right, $J(x,y) = J(x',y')$ with $\theta_X(x',y')$ gives $x = x'$ and $y = y'$ by Lemma 7.17(h), and substitution of equals gives $\theta_X(x,y)$. By (a), the remaining parts are statements about θ_X.

(b) Reflexivity uses $x = x$, and symmetry exchanges the roles of x and y in each alternative. For transitivity, let $\theta_X(x,y)$ and $\theta_X(y,z)$. If either uses the equation, substitute it into the other. Otherwise there are $k \le x$, $k \le y$ with $k \in X$, and $k' \le y$, $k' \le z$ with $k' \in X$. Decide $x \le z$ or $z < x$. If $x \le z$, then $k \le z$, and k witnesses $\theta_X(x,z)$; if $z < x$, then $k' \le x$, and k' witnesses it. For addition, let $\theta_X(w,x)$ and $\theta_X(y,z)$. If both use equations, $w + y = x + z$. If the first uses a witness $k \le w$, $k \le x$, then $k \le w + y$ and $k \le x + z$, so k witnesses $\theta_X(w+y, x+z)$; if only the second uses a witness $k \le y$, $k \le z$, then again $k \le w + y$ and $k \le x + z$.

(c) Given x and y, decide $x = y$. For the second alternative, decide $x \le y$ or $y < x$ and let m be the smaller number. Then $\exists k\,(k \le x \wedge k \le y \wedge k \in X)$ is equivalent to $\exists k\,(k < S(m) \wedge k \in X)$, which Lemma 7.22(b) decides. A disjunction of two decided formulas is decided, so (a) gives $x \sim_\theta y \vee \neg\, x \sim_\theta y$.

(d) Assume $\mathrm{Eq}(T_X)$ and $k \in X$. Since $k \le k$ and $k \le S(k)$, we have $\theta_X(k, S(k))$, so $k \sim_\theta S(k)$ and $k = S(k)$, which is impossible. So $\neg\, k \in X$ for every k. Conversely, assume $\mathrm{Z}(X)$ and $x \sim_\theta y$. The witness alternative of $\theta_X(x,y)$ would give a member of X, so $x = y$.

(e) Let $p \neq 0$ and $\mathrm{Desc}_{T_X}(i,p)$. We have $i \sim_{i,p} i + p$: indeed $i \le i$, $i \le i + p$, and $i + p = i + p \cdot 1$. So $i \sim_\theta i + p$, that is, $\theta_X(i, i+p)$. Its first alternative $i = i + p$ is impossible because $i < i + p$, and its second gives a member of X.

(f) Let $\mathrm{A}(X)$ and $k \in X$. If $x \sim_\theta y$, either $x = y$, the first alternative of $x \sim_{k,1} y$, or some $k' \in X$ has $k' \le x$ and $k' \le y$; then $k' = k$ by $\mathrm{A}(X)$, so $k \le x$ and $k \le y$. Moreover $x \equiv y \pmod 1$: decide $x \le y$ or $y < x$, and write $y = x + 1 \cdot d$ or $x = y + 1 \cdot d$. Conversely, $x \sim_{k,1} y$ gives either $x = y$ or $k \le x$, $k \le y$ with $k \in X$, and both give $\theta_X(x, y)$. □

Lemma 7.29 (The congruence of a proposition)**.** *For every set X, write P for $0 \in X$. Comprehension gives a set Q_X with*

$$\forall z \Big(z \in Q_X \leftrightarrow \exists x \exists y \big(z = J(x, y) \wedge (x = y \vee P) \big) \Big).$$

Write $x \sim_P y$ for $J(x, y) \in Q_X$. HAS derives $x \sim_P y \leftrightarrow (x = y \vee P)$, $\mathrm{Cong}(Q_X)$, $\mathrm{Eq}(Q_X) \to \neg P$ and $\mathrm{D}(Q_X) \to P$.

Proof. The first assertion follows from injectivity of the pairing, as in Lemma 7.28(a). This is the relation (7.7). Reflexivity and symmetry are immediate. For transitivity and addition, if any premise uses the alternative P, the conclusion uses it too; otherwise the premises are equations and so is the conclusion. If $\mathrm{Eq}(Q_X)$ and P, then $0 \sim_P 1$, so $0 = 1$, which contradicts $\neg(S(0) = 0)$. If $p \neq 0$ and $\mathrm{Desc}_{Q_X}(i, p)$, then $i \sim_P i + p$ as in Lemma 7.28(e), and $i \neq i + p$, so P. □

Proposition 7.30 (The reversals)**.** *HAS derives $\mathsf{V}_{\mathrm{MP}} \to \mathrm{MP}$, $\mathsf{V}_{\mathrm{WLPO}} \to \mathrm{WLPO}$, $\mathsf{V}_{\mathrm{LPO}} \to \mathrm{LPO}$, $\mathsf{V}_{\mathrm{LLPO}} \to \mathrm{LLPO}$ and $\mathsf{V}_{\mathrm{EM}} \to \mathrm{EM}$.*

Proof. For the first four, let X satisfy $\mathrm{Dec}(X)$ and let T_X be the set of Lemma 7.28. By parts (b) and (c), T_X is a decidable congruence, so each version applies to it.

MP. Assume $\neg\mathrm{Z}(X)$. Then $\neg\mathrm{Eq}(T_X)$, since $\mathrm{Eq}(T_X)$ would give $\mathrm{Z}(X)$ by part (d). V_{MP} gives $\mathrm{D}(T_X)$, and part (e) gives $\mathrm{E}(X)$.

WLPO. $\mathsf{V}_{\mathrm{WLPO}}$ gives $\mathrm{Eq}(T_X) \vee \neg\mathrm{Eq}(T_X)$. By part (d), the first case gives $\mathrm{Z}(X)$, and in the second, $\mathrm{Z}(X)$ would give $\mathrm{Eq}(T_X)$, so $\neg\mathrm{Z}(X)$.

LPO. $\mathsf{V}_{\mathrm{LPO}}$ gives $\mathrm{Eq}(T_X) \vee \mathrm{D}(T_X)$, hence $\mathrm{Z}(X)$ by part (d) or $\mathrm{E}(X)$ by part (e).

LLPO. Assume also $\mathrm{A}(X)$. $\mathsf{V}_{\mathrm{LLPO}}$ gives $\mathrm{Odd}(T_X) \vee \mathrm{Even}(T_X)$. Suppose $\mathrm{Odd}(T_X)$, and let $k + k \in X$. Part (f) gives $\mathrm{Desc}_{T_X}(k + k, 1)$, and $1 \neq 0$, so $\mathrm{Odd}(T_X)$ gives h with $k + k = S(h + h)$, which Lemma 7.17(f) refutes. So $\mathrm{Ev}(X)$. Suppose $\mathrm{Even}(T_X)$, and let $S(k + k) \in X$. Then $\mathrm{Desc}_{T_X}(S(k + k), 1)$ gives h with $S(k + k) = h + h$, which is refuted in the same way. So $\mathrm{Od}(X)$.

EM. Let X be any set and let Q_X be the set of Lemma 7.29, a congruence. V_{EM} gives $\mathrm{Eq}(Q_X) \vee \mathrm{D}(Q_X)$, hence $\neg P$ or P, where P is $0 \in X$. □

The forward direction for LLPO needs the first collision. For a set B and a relation R put

$$\begin{aligned} \mathrm{LB}(c) &:\equiv c \in B \wedge \forall c' \, (c' < c \to \neg\, c' \in B), \\ \mathrm{LP}(x, c) &:\equiv x < c \wedge x \sim c \wedge \forall x' \, (x' < x \to \neg\, x' \sim c). \end{aligned}$$

When B is the collision set of R, $\mathrm{LB}(c)$ says that c is the least count with a smaller partner, and $\mathrm{LP}(x, c)$ that x is its least partner. The pair (x, c) is then the first collision of the proof of Theorem 7.9(3).

Lemma 7.31 (The first collision as a sequence). *Let R satisfy $\mathrm{Dec}_2(R)$ and $\mathrm{Cong}(R)$, and let B be its collision set. Comprehension gives a set G with*

$$\forall m \Big(m \in G \leftrightarrow \exists c \exists x \big(\mathrm{LB}(c) \wedge \mathrm{LP}(x,c) \wedge \big((m = c + c \wedge \exists h\,(x = h + h)) \\ \vee\, (m = S(c+c) \wedge \exists h\,(x = S(h+h))) \big) \big) \Big),$$

the set of the sequence γ with $\gamma(2c + b) = 1$ exactly when (x, c) is the first collision and b is the remainder of x modulo 2. HAS derives:

(a) $\mathrm{LB}(c) \wedge \mathrm{LB}(c') \to c = c'$ *and* $\mathrm{LP}(x,c) \wedge \mathrm{LP}(x',c) \to x = x'$*;*

(b) $\mathrm{A}(G)$*;*

(c) $\mathrm{Dec}(G)$*;*

(d) if $p \neq 0$ and $\mathrm{Desc}_R(i,p)$, then $\mathrm{LB}(i+p)$ and $\mathrm{LP}(i, i+p)$; hence $(i+p)+(i+p) \in G$ when i is even, and $S((i+p)+(i+p)) \in G$ when i is odd.

Proof. (a) Decide $c = c'$. Otherwise $c < c'$ or $c' < c$, and the leastness in $\mathrm{LB}(c')$, respectively $\mathrm{LB}(c)$, refutes $c \in B$, respectively $c' \in B$. The same argument, with the leastness clause of LP, gives $x = x'$.

(b) Let $m, m' \in G$, through (c, x) and (c', x'). By (a), $c = c'$ and $x = x'$. If both use the same alternative, $m = m'$. If they use different alternatives, x is both even and odd, which Lemma 7.17(f) refutes.

(c) Fix m. Lemma 7.17(f) gives c with $m = c + c$ or $m = S(c + c)$. If $m \in G$ through a pair (c'', x''), then $c'' = c$: the equations $c + c = c'' + c''$ and $S(c + c) = S(c'' + c'')$ each give $c = c''$, by injectivity of the step and Lemma 7.17(f), and the mixed equations $c + c = S(c'' + c'')$ and $S(c + c) = c'' + c''$ are refuted by Lemma 7.17(f).

Decide $\mathrm{LB}(c)$: $c \in B$ is decided by Lemma 7.26(a), and $\exists c'\,(c' < c \wedge c' \in B)$ by Lemma 7.22(b). If $\neg\mathrm{LB}(c)$, then $\neg\, m \in G$, since a pair (c'', x'') through which $m \in G$ has $c'' = c$ and $\mathrm{LB}(c'')$.

If $\mathrm{LB}(c)$, then $c \in B$ gives $x_0 < c$ with $x_0 \sim c$. The set $\{x : x \sim c\}$, given by comprehension, is decidable and contains x_0, so Lemma 7.22(c) gives $x \sim c$ with $\neg(x' \sim c)$ for all $x' < x$. Then $x \leq x_0$, since $x_0 < x$ would contradict leastness, so $x < c$ and $\mathrm{LP}(x, c)$. Decide whether x is even or odd. If $m = c + c$ and x is even, or $m = S(c + c)$ and x is odd, then (c, x) witnesses $m \in G$. In the two other combinations $\neg\, m \in G$: a witness (c'', x'') would have $c'' = c$, $x'' = x$ by (a), and its alternative would make x both even and odd, or $c + c = S(c + c)$.

(d) Let $p \neq 0$ and $\mathrm{Desc}_R(i, p)$. As in Lemma 7.28(e), $i \sim i + p$, and $i < i + p$, so $i + p \in B$. Suppose $c' < i + p$ and $c' \in B$, with $x' < c'$ and $x' \sim c'$. The description gives $x' \sim_{i,p} c'$. Its equality alternative is refuted by $x' < c'$, so $i \leq x'$ and $x' \equiv c' \pmod p$. The form $x' = c' + p \cdot d$ is refuted by $x' < c'$. In the form $c' = x' + p \cdot d$ we have $d \neq 0$, since $d = 0$ would give $c' = x'$; writing $d = S(d')$ gives $c' = x' + p + p \cdot d' \geq i + p$, which contradicts $c' < i + p$. So $\mathrm{LB}(i + p)$. Next, $i < i + p$ and $i \sim i + p$; if $x' < i$ and $x' \sim i + p$, the description gives $x' \sim_{i,p} i + p$, whose equality alternative is refuted by $x' < i < i + p$ and whose other alternative gives $i \leq x'$, against $x' < i$. So $\mathrm{LP}(i, i + p)$. The last assertion is the definition of G at $c = i + p$, $x = i$. □

Proposition 7.32 (The forward direction for LLPO)**.** *HAS derives* $\mathrm{LLPO} \to \mathsf{V}_{\mathrm{LLPO}}$.

Proof. Let R satisfy $\mathrm{Dec}_2(R)$ and $\mathrm{Cong}(R)$, with collision set B and first-collision set G. By Lemma 7.31(b) and (c), LLPO applies to G and gives $\mathrm{Ev}(G) \vee \mathrm{Od}(G)$.

Suppose $\mathrm{Ev}(G)$, and let $p \neq 0$ and $\mathrm{Desc}_R(i, p)$. Decide whether i is even or odd. If i is even, part (d) of the lemma puts $(i + p) + (i + p)$ in G, against $\mathrm{Ev}(G)$; so i is odd. This is $\mathrm{Odd}(R)$. Suppose $\mathrm{Od}(G)$. If i is odd, part (d) puts $S((i + p) + (i + p))$ in G, against $\mathrm{Od}(G)$; so i is even. This is $\mathrm{Even}(R)$. □

Theorem 7.33 (The six equivalences are derivable)**.** *HAS derives* $\mathsf{V}_{\mathrm{none}}$ *and the five biconditionals*

$$\mathsf{V}_{\mathrm{MP}} \leftrightarrow \mathrm{MP}, \quad \mathsf{V}_{\mathrm{LLPO}} \leftrightarrow \mathrm{LLPO}, \quad \mathsf{V}_{\mathrm{WLPO}} \leftrightarrow \mathrm{WLPO}, \quad \mathsf{V}_{\mathrm{LPO}} \leftrightarrow \mathrm{LPO}, \quad \mathsf{V}_{\mathrm{EM}} \leftrightarrow \mathrm{EM}.$$

Equivalently, for each of the six versions V *and its principle* P*, with* $\bot \to \bot$ *as the empty principle, HAS derives* $\mathsf{V} \leftrightarrow P$.

Conditions and scope.

Base and premises
HAS (Definition 7.14), intuitionistic second-order arithmetic with the induction axiom and comprehension for every formula; no further axiom. The versions are the sentences (7.11) and the principles those of Definition 7.18.

Conclusion
Derivations in HAS of $\mathsf{V}_{\mathrm{none}}$ and of $\mathsf{V}_P \leftrightarrow P$ for each of MP, LLPO, WLPO, LPO and EM.

Proof. Proposition 7.25 derives $\mathsf{V}_{\mathrm{none}}$, and so $\mathsf{V}_{\mathrm{none}} \leftrightarrow (\bot \to \bot)$, since both sides are derivable. The forward directions are Propositions 7.27 and 7.32, the reversals Proposition 7.30. □

Corollary 7.34 (Each version has the value of its principle)**.** *In every system of truth values, each of the six sentences* $\mathsf{V}_{\mathrm{none}}, \ldots, \mathsf{V}_{\mathrm{EM}}$ *has the same value as its principle; in particular* $\mathsf{V}_{\mathrm{none}}$ *has the value* $\top$.

Proof. Let v and q be the values of a version and of its principle; both are sentences, so the values do not depend on an assignment. By Theorem 7.33 and Lemma 7.21, $(v \Rightarrow q) \wedge (q \Rightarrow v) = \top$. So $\top \leq (v \Rightarrow q)$, which gives $v \leq q$ by the defining property of implication, and likewise $q \leq v$. Hence $v = q$. The empty principle $\bot \to \bot$ has the value $\top$. □

No argument of the chapter has to be rerun inside a model. The arguments run once, inside HAS, and soundness carries their conclusions into every model, including models in which a relation takes intermediate truth values and the searches of the proofs are not decided at every point.

Table 7.1 lists every encoding the chapter uses: each finite search, each way of storing a sequence or a proposition inside a congruence, and the translations to readers and to signed records. For the six versions of the classification, every encoding is a set given by

Encoding	Where it is defined	What it serves	Interpreted in HAS by
least positive period at a; least index for that period	proof of Theorem 7.6	the version with a collision (no principle)	the sets P_a and I_p of Proposition 7.25; decided by $\mathrm{Dec}_2(R)$; least members by Lemma 7.22(c); the description by Lemma 7.24
collision search $\beta_\sim$	(7.5); the search $B(k)$ in the proof of Theorem 7.7	forward directions for MP, WLPO, LPO and EM	the set B of Lemma 7.26: decidable; empty exactly when the relation is equality; a member is a collision and gives a description
first-collision sequence γ	proof of Theorem 7.9(3)	forward direction for LLPO	the set G of Lemma 7.31: decidable; at most one member; for a description by (i, p) its member is $2(i + p) + (i \bmod 2)$
$\sim_\theta$	(7.4), (7.6)	reversals to MP, WLPO, LPO and LLPO	the set T_X of Lemma 7.28: a congruence; decidable when the sequence is; equality exactly when the sequence is all zero; a description yields an event; with at most one event, at k, it is $\sim_{k,1}$
$\sim_P$	(7.7)	reversal to EM	the set Q_X of Lemma 7.29: a congruence; equality refutes P; a description proves P
least witnesses under EM	proof of Theorem 7.10	forward direction for EM	not needed: in Proposition 7.27 EM decides the relation and whether the collision set B is inhabited, and the decidable searches supply the least witnesses
kernel $\sim_r$ of a reading; quotient reading $x \mapsto [x]$	Section 7.7	Theorem 7.11	not used by Theorem 7.38
restriction to counts; extension (7.8)	Theorem 7.12	Corollary 7.13	not used by Theorem 7.38
$\approx_\theta$, $\approx_P$, $\approx_\alpha$; least positive count related to $[0, 0]$	proof of Corollary 7.13	the signed rows	not used by Theorem 7.38

Table 7.1. Every encoding of the chapter and where it is interpreted. The rows above the middle rule are the encodings of the six versions; each is a set given by comprehension, and the properties listed are derived in HAS. The rows below it serve the readers and the signed records.

comprehension, and every property the proofs use is derived in HAS; soundness then interprets the encoding, with all those properties, in every model at once. The last three rows serve Theorem 7.11 and Corollary 7.13. Their equivalences are proved in the constructive base of this chapter, and Theorem 7.38 makes no claim about them in the models, so none of them needs an interpretation there.

7.12 Four models, and why the six prices are different

A principle whose value is not $\top$ in some model has no derivation, by Lemma 7.21. If one sentence has the value $\top$ and another does not, the implication from the first to the second has the value of the second and has no derivation either. One model in which P holds and Q fails therefore shows that P and Q are not equivalent.

The principles decrease in strength along EM, LPO, WLPO, LLPO, and LPO is WLPO and MP together. These are derivations in HAS, so they hold in every model, and they let one check in a model settle several entries of the table below.

Proposition 7.35 (The order of strength)**.** *HAS derives* EM $\to$ LPO, LPO $\to$ WLPO, LPO $\to$ MP, WLPO $\to$ LLPO *and* WLPO $\to$ (MP $\to$ LPO).

Proof. Let X satisfy $\mathrm{Dec}(X)$. Each principle is applied at X, or at a set given by comprehension from X.

EM gives LPO. Comprehension gives a set Y with $\forall z\,(z \in Y \leftrightarrow \mathrm{E}(X))$, and EM at Y gives $\mathrm{E}(X) \vee \neg\mathrm{E}(X)$. In the second case a member of X would give $\mathrm{E}(X)$, so $\mathrm{Z}(X)$.

LPO gives WLPO and MP. LPO gives $\mathrm{Z}(X) \vee \mathrm{E}(X)$. A member of X refutes $\mathrm{Z}(X)$, so the second alternative gives $\neg\mathrm{Z}(X)$, and $\mathrm{Z}(X) \vee \neg\mathrm{Z}(X)$ follows. Under the hypothesis $\neg\mathrm{Z}(X)$ of MP, the first alternative is refuted and the second is the conclusion $\mathrm{E}(X)$.

WLPO and MP give LPO. WLPO gives $\mathrm{Z}(X) \vee \neg\mathrm{Z}(X)$, and in the second case MP gives $\mathrm{E}(X)$.

WLPO gives LLPO. Let also $\mathrm{A}(X)$ hold. Comprehension gives Y with $\forall k\,(k \in Y \leftrightarrow k + k \in X)$, which is decidable because X is, and $\mathrm{Z}(Y)$ is $\mathrm{Ev}(X)$. WLPO at Y gives $\mathrm{Z}(Y) \vee \neg\mathrm{Z}(Y)$. In the second case we derive $\mathrm{Od}(X)$: if $S(k + k) \in X$, then every j with $j \in Y$ gives $j + j \in X$, and $\mathrm{A}(X)$ gives $S(k + k) = j + j$, which Lemma 7.17(f) refutes; so $\mathrm{Z}(Y)$, against $\neg\mathrm{Z}(Y)$. Hence $\neg\, S(k + k) \in X$ for every k. □

Corollary 7.36 (Implications valid in every model)**.** *In every system of truth values,* $[\![\mathrm{EM}]\!] \le [\![\mathrm{LPO}]\!]$, $[\![\mathrm{LPO}]\!] \le [\![\mathrm{MP}]\!]$, $[\![\mathrm{LPO}]\!] \le [\![\mathrm{WLPO}]\!]$ *and* $[\![\mathrm{WLPO}]\!] \le [\![\mathrm{LLPO}]\!]$.

Proof. By Lemma 7.21, a derivable implication $P \to Q$ between sentences has the value $\top$, that is, $\top \le ([\![P]\!] \Rightarrow [\![Q]\!])$, which says $[\![P]\!] \le [\![Q]\!]$. Apply this to Proposition 7.35. □

In the open sets of a space, a binary sequence is a sequence of sets $A_0, A_1, \ldots$ that are both open and closed, since $A_n \vee \neg A_n = \top$ says exactly that the complement of A_n is open. At a point x the sequence reads $\alpha(n) = 1$ when $x \in A_n$, so it varies continuously with the point. Write O

for the union of the sets A_n. The value of $\exists n\, \alpha(n) = 1$ is O, and the value of $\forall n\, \alpha(n) = 0$ is $\neg O$. At this sequence the principles say:

LPO:	O is closed;	
WLPO:	the closure of O is open;	
MP:	O is the interior of its closure;	(7.13)
LLPO:	if the A_n are disjoint, the union of the even-indexed ones and the union of the odd-indexed ones have disjoint closures.	

Excluded middle says that every open set is closed. To refute a principle it is enough to find one sequence at which its condition fails. To verify one, the condition must hold for every sequence, and also, inside each open part of the space, for the sequences decided only on that part. When the space has a base of sets that are both open and closed, the second task reduces to the first.

Lemma 7.37 (Reduction to sequences of open and closed sets)**.** *Let T be a topological space with a base of sets that are both open and closed, and let its open sets be the truth values. Then each of LPO, WLPO and MP has the value T exactly when its condition in* (7.13) *holds for every sequence of sets that are both open and closed, and LLPO has the value T exactly when its condition holds for every such sequence of pairwise disjoint sets.*

Proof. First let every A_n be both open and closed. Then the value d of $\mathrm{Dec}(X)$ is T, the value e of $\mathrm{E}(X)$ is O, the value of $\mathrm{Z}(X)$ is $T \setminus \overline{O}$, its negation is the interior of $\overline{O}$, the value of $\mathrm{A}(X)$ is T exactly when the A_n are pairwise disjoint, and the values of $\mathrm{Ev}(X)$ and $\mathrm{Od}(X)$ are the complements of the closures of the even and of the odd union. So the instance of LPO at this sequence is T exactly when $(T \setminus \overline{O}) \cup O = T$, that is, when O is closed. The instance of WLPO is T exactly when $(T \setminus \overline{O}) \cup \operatorname{int} \overline{O} = T$, that is, when $\overline{O}$ is open. The instance of MP is T exactly when $\operatorname{int} \overline{O} \subseteq O$; since O is open and so lies in $\operatorname{int} \overline{O}$, this says $O = \operatorname{int} \overline{O}$. For a disjoint sequence the instance of LLPO is T exactly when the two closures are disjoint. A principle is the meet of its instances, so if its value is T, each condition holds.

Conversely, assume the conditions, let $A : \mathbb{N} \to H$ be any sequence of open sets, and let W be the value of the hypotheses of an instance: d, met with the value of $\neg\mathrm{Z}(X)$ for MP and with the value of $\mathrm{A}(X)$ for LLPO. Let $x \in W$, and choose V, both open and closed, with $x \in V \subseteq W$. Put $A'_n = A_n \cap V$. Each A'_n is open, and it is closed: its complement is $(T \setminus V) \cup (V \cap \neg A_n)$, because $V \subseteq d \subseteq A_n \cup \neg A_n$, and both pieces are open. Let $O' = \bigcup_n A'_n = O \cap V$. Two facts are used repeatedly. An open set disjoint from every A_n lies in the value of $\mathrm{Z}(X)$, which is the interior of $\bigcap_n (T \setminus A_n)$. And a point of V that lies in some A_n lies in O'; so an open subset of V that misses O' misses every A_n. We show that the conclusion of the instance contains V, hence x.

For LPO, the condition at A' makes O' closed, so $V \setminus O'$ is open; it misses every A_n and lies in the value of $\mathrm{Z}(X)$, while $V \cap O'$ lies in O, the value of $\mathrm{E}(X)$. For WLPO, the condition makes $\overline{O'}$ open, so $V \setminus \overline{O'}$ is open and lies in the value of $\mathrm{Z}(X)$; and $V \cap \overline{O'}$ is an open set that misses the value of $\mathrm{Z}(X)$, since that value is an open set disjoint from O' and hence from $\overline{O'}$, so $V \cap \overline{O'}$ lies in the value of $\neg\mathrm{Z}(X)$. For MP, every point y of V lies in $\overline{O'}$: otherwise some open $U \ni y$ misses O', the open set $U \cap V$ misses every A_n and lies in the value of $\mathrm{Z}(X)$; it also lies in V, inside the value of $\neg\mathrm{Z}(X)$; these two values are disjoint, and $U \cap V$ contains y. So the open

set V lies in $\operatorname{int}\overline{O'}$, which the condition puts inside $O' \subseteq O$, the value of $\mathrm{E}(X)$. For LLPO, the sets A'_n are pairwise disjoint, since V lies in the value of $\mathrm{A}(X)$, which is disjoint from $A_m \cap A_n$ when $m \neq n$. The condition makes the closures of $O'_{\mathrm{ev}} = \bigcup_k A'_{2k}$ and $O'_{\mathrm{od}} = \bigcup_k A'_{2k+1}$ disjoint, so the open sets $V \setminus \overline{O'_{\mathrm{ev}}}$ and $V \setminus \overline{O'_{\mathrm{od}}}$ cover V. The first misses every A_{2k}, so it lies in the value of $\mathrm{Ev}(X)$; the second lies in the value of $\mathrm{Od}(X)$ in the same way.

In each case the hypotheses of the instance at A lie below its conclusion, so the instance has the value T, and so does the principle. □

The models are built in ordinary classical mathematics, with the axiom of choice supplying the ultrafilters. That is the usual setting of an independence proof. The question is what a specified system derives, and the answer is established about that system from outside it.

The one-point compactification of an uncountable set

Let S be an uncountable set, for example the set of all sets of natural numbers, and add one point ∞. The open sets are all subsets of S, together with the sets that contain ∞ and omit only finitely many points of S. The sets both open and closed are the finite subsets of S and their complements, and they form a base: each point of S is isolated, and the complements of finite subsets of S are a base at ∞.

Markov's principle holds. By Lemma 7.37 it is enough to check that a countable union O of sets both open and closed is the interior of its closure. If one of them contains ∞, the complement of O is a finite subset of S, which is open, so O is closed and equals the interior of its closure. Otherwise O is a countable subset of S. Each point of S is isolated, so the closure of O adds at most ∞, and ∞ is not interior to it: every neighbourhood of ∞ contains uncountably many points of S, hence points outside O and outside its closure. Again O is the interior of its closure. LLPO fails. Choose distinct points $s_0, s_1, \ldots$ of S and put $A_n = \{s_n\}$. These sets are disjoint, the even-indexed and the odd-indexed points form two infinite subsets of S, and ∞ lies in the closure of both.

The remainder $\beta\mathbb{N} \setminus \mathbb{N}$

The points are the non-principal ultrafilters on $\mathbb{N}$. For $B \subseteq \mathbb{N}$ let B^* be the set of those that contain B. These sets are open and closed and form a base, and by compactness every set both open and closed has this form; Walker [18] develops these facts about ultrafilters. Moreover $B^* \cap C^* = (B \cap C)^*$, and B^* is empty exactly when B is finite.

LLPO holds. By Lemma 7.37 it is enough to separate the closures for a disjoint sequence of sets both open and closed. Let the sets be $A_n = B_n^*$, so that $B_m \cap B_n$ is finite when $m \neq n$. Let B be the union, over k, of B_{2k} with its finitely many points in $B_1 \cup B_3 \cup \cdots \cup B_{2k+1}$ removed. Then only finitely many points of each B_{2k} lie outside B, and each B_{2m+1} meets B only in its finite overlaps with $B_0, B_2, \ldots, B_{2m-2}$. Hence every even-indexed A_n lies in B^* and every odd-indexed one in $(\mathbb{N} \setminus B)^*$. These two sets are closed and disjoint, so they separate the closures of the two unions.

WLPO fails. Split $\mathbb{N}$ into infinitely many infinite sets $B_0, B_1, \ldots$, put $A_n = B_n^*$, and let O be their union. If the closure of O were open, it would be open and closed, hence equal to B^* for some

B, and each B_n would lie in B apart from finitely many points. Choose one point b_n of $B \cap B_n$ for each n. The set $C = \{b_0, b_1, \ldots\}$ is infinite, lies in B, and meets each B_n in a single point. So C^* is a nonempty open subset of the closure of O that meets no A_n, and so misses O. That is impossible.

The space $\beta\mathbb{N}$

Now take all ultrafilters on $\mathbb{N}$. The principal ultrafilter at n is an isolated point, which we write n. The sets $\overline{B}$ of ultrafilters containing B are open and closed, they form a base, and $\overline{B}$ is the closure of B. An open set U meets $\mathbb{N}$ in a set B that is dense in U, so the closure of U is $\overline{B}$, which is open. WLPO holds, by Lemma 7.37. For Markov's principle and LPO, take $A_n = \{n\}$. Then $O = \mathbb{N}$, whose closure is the whole space. So O is neither closed nor the interior of its closure, and both principles fail.

The chain $0 < 1 < 2 < \cdots < \infty$

In a chain, $\neg a = \bot$ for every $a \neq \bot$, so $a \vee \neg a = \top$ only when a is $\bot$ or $\top$. For a map A, let d be the value of $\mathrm{Dec}(X)$. If $d = \bot$ the instance of LPO at A is $\top$. Otherwise $A(n) \vee \neg A(n) \geq d$ gives, at each n, either $A(n) = \bot$ or $A(n) \geq d$. The classical mathematics in which the model is built either finds an n with $A(n) \neq \bot$, and then the value of $\mathrm{E}(X)$ is at least d, or shows that $A(n) = \bot$ for every n, and then the value of $\mathrm{Z}(X)$ is $\top$. Either way the instance has the value $\top$. LPO holds. Excluded middle fails at the value 1, since $1 \vee \neg 1 = 1$.

Theorem 7.38 (The six principles are different)**.** *Over intuitionistic second-order arithmetic:*

1. *none of MP, LLPO, WLPO, LPO and EM is derivable, and no two of the six principles none, MP, LLPO, WLPO, LPO and EM are equivalent;*
2. *each of the six versions of the classification in the table of Section 7.9, written as the sentences* (7.11)*, is derivably equivalent to its principle as written in Definition 7.18, so in every system of truth values it has the same value as its principle;*
3. *consequently no version of the classification is equivalent to any principle other than its own.*

Conditions and scope.

Base and premises

HAS (Definition 7.14), with the principles and versions written as the sentences of Definition 7.18 and (7.11). The non-derivations are read off from four systems of truth values, built in classical mathematics with the axiom of choice supplying ultrafilters.

Conclusion

HAS derives none of MP, LLPO, WLPO, LPO and EM and no equivalence between two of the six principles. It derives each version's equivalence with its own principle, and no equivalence of a version with another principle.

Proof. (1) Lemma 7.21 turns a model in which a sentence does not have the value $\top$ into the fact that the sentence has no derivation. The one-point compactification refutes LLPO, WLPO, LPO and EM, and $\beta\mathbb{N}$ refutes MP (Table 7.2), so none of the five priced principles is derivable.

For two different principles P and Q, a model in which P has the value $\top$ and Q does not gives $P \leftrightarrow Q$ the value $(\top \Rightarrow q) \wedge (q \Rightarrow \top) = q \neq \top$, where q is the value of Q; so $P \leftrightarrow Q$ has no derivation. Table 7.2 supplies such a model for each of the fifteen pairs. The empty principle has the value $\top$ everywhere, so the five pairs containing it are separated by the models just named. The one-point compactification separates MP from LLPO, WLPO, LPO and EM; the remainder $\beta\mathbb{N} \setminus \mathbb{N}$ separates LLPO from WLPO, LPO and EM; $\beta\mathbb{N}$ separates WLPO from LPO and EM; and the chain separates LPO from EM. That is $5 + 4 + 3 + 2 + 1 = 15$ pairs.

(2) Theorem 7.33 derives each version's equivalence with its principle, and Corollary 7.34 gives the equality of values in every system of truth values. Table 7.3 lists the resulting value of each version in each of the four models.

(3) Let a version V have the principle P, and let Q be a different principle. If $\mathsf{V} \leftrightarrow Q$ had a derivation, then, with the derivation of $\mathsf{V} \leftrightarrow P$ from (2), the two implications $P \to \mathsf{V} \to Q$ and $Q \to \mathsf{V} \to P$ would give a derivation of $P \leftrightarrow Q$, which (1) excludes. Equivalently: in a model that separates P from Q, the version has the value of P, so $\mathsf{V} \leftrightarrow Q$ does not have the value $\top$ there. □

Model	Holds	Fails
one-point compactification	MP	LLPO, WLPO, LPO, EM
$\beta\mathbb{N} \setminus \mathbb{N}$	LLPO, MP (Theorem 7.42)	WLPO, LPO, EM
$\beta\mathbb{N}$	WLPO, LLPO	MP, LPO, EM
chain $0 < 1 < \cdots < \infty$	LPO, MP, WLPO, LLPO	EM

Table 7.2. Four models of intuitionistic second-order arithmetic. In the first three the truth values are the open sets of the space named. Each entry not checked in the text follows from Corollary 7.36: in every model EM implies LPO, LPO implies MP and WLPO, and WLPO implies LLPO.

Version	one-point	$\beta\mathbb{N} \setminus \mathbb{N}$	$\beta\mathbb{N}$	chain
V_{none}	$\top$	$\top$	$\top$	$\top$
V_{MP}	$\top^{\text{t}}$	$\top^{\text{p}}$	$\neq \top^{\text{t}}$	$\top^{\text{c}}$
V_{LLPO}	$\neq \top^{\text{t}}$	$\top^{\text{t}}$	$\top^{\text{c}}$	$\top^{\text{c}}$
V_{WLPO}	$\neq \top^{\text{c}}$	$\neq \top^{\text{t}}$	$\top^{\text{t}}$	$\top^{\text{c}}$
V_{LPO}	$\neq \top^{\text{c}}$	$\neq \top^{\text{c}}$	$\neq \top^{\text{t}}$	$\top^{\text{t}}$
V_{EM}	$\neq \top^{\text{c}}$	$\neq \top^{\text{c}}$	$\neq \top^{\text{c}}$	$\neq \top^{\text{t}}$

Table 7.3. The value of each version (7.11) in each model of Table 7.2: $\top$, or a value other than $\top$. By Corollary 7.34 a version has the value of its principle. A mark t means that the principle's value is computed in this section, and c that it follows from those computations by Corollary 7.36; the empty principle has the value $\top$ in every model. The mark p means that the value is computed in Theorem 7.42; Theorem 7.38 does not use it.

Two consequences follow. Markov's principle is genuinely needed for the classification from negative nontriviality: the system proves neither the principle nor that version. And the six prices do not form a single ladder. EM, LPO, WLPO and LLPO decrease strictly in strength: each derives the next by Proposition 7.35, and no converse is derivable by (1). LPO is WLPO and MP together, by the same proposition; with Theorem 7.33, HAS also derives $\mathsf{V}_{\mathrm{LPO}} \leftrightarrow (\mathsf{V}_{\mathrm{WLPO}} \wedge \mathsf{V}_{\mathrm{MP}})$, the form inside the system of the last assertion of Theorem 7.9. But Markov's principle is comparable with neither WLPO nor LLPO. The one-point compactification satisfies MP and refutes LLPO, and $\beta\mathbb{N}$ satisfies WLPO and refutes MP.

7.13 Prices between the six

Take a binary sequence with at most one entry 1, known not to be all zero. Which parity class is free of entries? Markov's principle answers, since the entry it finds has a parity; so does LLPO, without the extra knowledge. Is this a seventh price? It is, and not the only one: every gap of the chart of Section 7.12 contains further prices (Figure 7.1). As there, derivations are derivations in HAS, and the models are built in classical mathematics, with the axiom of choice supplying ultrafilters and chosen points.

Definition 7.39 (Further principles as sentences)**.** With the abbreviations of Definition 7.18, and with $\bar{n}$ the numeral of n, put

$$\begin{aligned}
\mathrm{MP}^{\vee} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{A}(X) \to \neg \mathrm{Z}(X) \to \mathrm{Ev}(X) \vee \mathrm{Od}(X)\big),\\
\mathrm{W}(X) &:\equiv \forall Y\Big(\mathrm{Dec}(Y) \to \neg\neg \mathrm{E}(Y) \vee \neg\neg \exists n\,\big(n \in X \wedge \neg\, n \in Y\big)\Big),\\
\mathrm{WMP} &:\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{W}(X) \to \mathrm{E}(X)\big),\\
\mathrm{WEM} &:\equiv \forall X\big(\neg\, 0 \in X \vee \neg\neg\, 0 \in X\big).
\end{aligned}$$

For $n \geq 1$ and $i < n$ let $\mathrm{C}_i^n(X) :\equiv \forall j \neg\, (\bar{n} \cdot j + \bar{\imath}) \in X$, and put

$$\mathrm{LLPO}_n :\equiv \forall X\big(\mathrm{Dec}(X) \to \mathrm{A}(X) \to \mathrm{C}_0^n(X) \vee \mathrm{C}_1^n(X) \vee \cdots \vee \mathrm{C}_{n-1}^n(X)\big).$$

For binary sequences, $\mathrm{MP}^{\vee}$ is LLPO restricted to sequences known not to be all zero: the disjunctive Markov principle of Definition 14.18, which Chapter 14 meets as the price of the weak sign of a nonzero point. WMP is Ishihara's weak Markov principle [60, 43]: if, for every decidable Y, either Y cannot be empty or the part of X outside Y cannot be empty, then X has a member. WEM, weak excluded middle, decides of every proposition whether it is refuted. LLPO_n says that a decidable set with at most one member misses one of the n residue classes modulo n; for $n = 2$ its alternatives say what $\mathrm{Ev}(X)$ and $\mathrm{Od}(X)$ say. These weakenings of LLPO were introduced by Richman [73, Section 8.1]. For $n = 1$ the only class is all of $\mathbb{N}$, so LLPO_1 says that a decidable set with at most one member is empty, which the set $\{0\}$ refutes; the family therefore starts at $n = 2$.

Remark 7.40 (The two halves of Markov's principle)**.** Read for binary sequences, WMP says that $\exists n\, \alpha(n) = 1$ whenever every binary β has $\neg\neg \exists n\, \beta(n) = 1$ or $\neg\neg \exists n\, (\alpha(n) = 1 \wedge \beta(n) = 0)$. In this form $\mathrm{MP}_{\mathrm{bin}}$ is equivalent to WMP and $\mathrm{MP}^{\vee}$ together, with no choice principle [43]. MP gives both, as in Lemma 7.41 below. Conversely, if α is not all zero, so is the sequence α' that keeps only its first entry 1, and for a binary β so is the sequence with $\alpha'(n)\beta(n)$ at $2n$ and

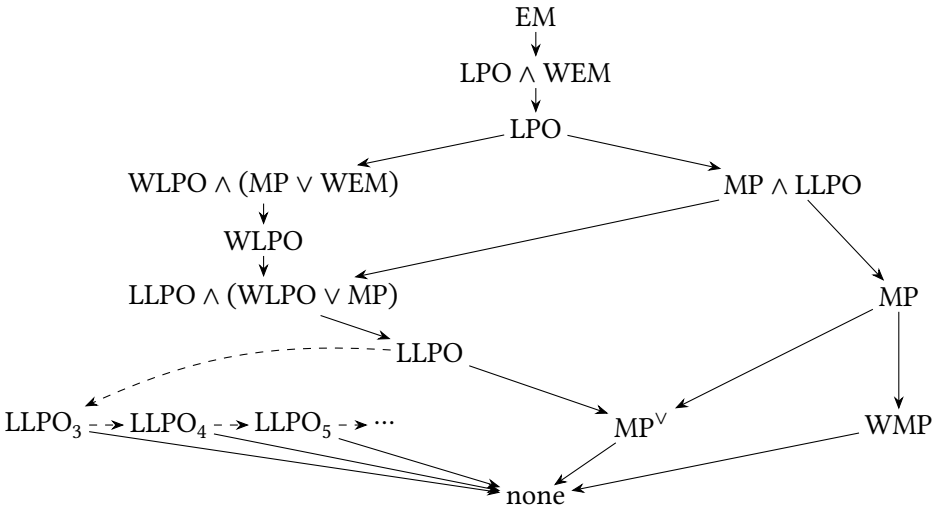


Figure 7.1. The chart of prices over intuitionistic second-order arithmetic. A solid arrow is a derivation; a dashed arrow is an inequality of values in every system of truth values. No arrow can be reversed. The principles $\mathrm{LLPO}, \mathrm{LLPO}_3, \mathrm{LLPO}_4, \ldots$ form a strictly descending chain: each value lies below the next, and for $m \geq 3$, LLPO_m derives neither LLPO nor LLPO_n for $2 \leq n < m$. Each LLPO_n with $n \geq 3$, and WMP, is incomparable with $\mathrm{MP}^\vee$; WMP is also incomparable with LLPO and WLPO, and $\mathrm{MP} \wedge \mathrm{LLPO}$ with WLPO.

$\alpha'(n)(1 - \beta(n))$ at $2n + 1$, which has at most one entry 1. $\mathrm{MP}^\vee$ empties its even places, giving $\neg\neg\exists n\,(\alpha'(n) = 1 \wedge \beta(n) = 0)$, or its odd places, giving $\neg\neg\exists n\,\beta(n) = 1$; so WMP finds an entry 1 of α', hence of α. Over HAS neither half derives the other or MP (Theorem 7.45). Kohlenbach showed that WMP is not derivable even in intuitionistic arithmetic in all finite types with the axiom of choice [75].

Lemma 7.41 (What the six principles derive)**.**

1. *HAS derives* $\mathrm{MP} \to \mathrm{MP}^\vee$, $\mathrm{MP} \to \mathrm{WMP}$, $\mathrm{LLPO} \to \mathrm{MP}^\vee$ *and* $\mathrm{LPO} \to \mathrm{MP} \wedge \mathrm{LLPO}$.
2. *HAS derives* $\mathrm{WLPO} \to \mathrm{LLPO} \wedge (\mathrm{WLPO} \vee \mathrm{MP})$, $\mathrm{LPO} \to \mathrm{WLPO} \wedge (\mathrm{MP} \vee \mathrm{WEM})$ *and* $\mathrm{EM} \to \mathrm{LPO} \wedge \mathrm{WEM}$*, and each of these three conjunctions derives its first conjunct. It also derives* $\mathrm{MP} \wedge \mathrm{LLPO} \to \mathrm{MP}$ *and* $\mathrm{MP} \wedge \mathrm{LLPO} \to \mathrm{LLPO} \wedge (\mathrm{WLPO} \vee \mathrm{MP})$.
3. *For* $2 \leq n \leq m$*, in every system of truth values* $[\![\mathrm{LLPO}]\!] \leq [\![\mathrm{LLPO}_n]\!] \leq [\![\mathrm{LLPO}_m]\!]$.

Proof. (1) Let $\mathrm{Dec}(X)$. For $\mathrm{MP} \to \mathrm{MP}^\vee$ assume also $\mathrm{A}(X)$ and $\neg\mathrm{Z}(X)$. MP gives some $n \in X$, and by Lemma 7.17(f), $n = h + h$ or $n = S(h + h)$. In the first case $\mathrm{Od}(X)$ holds: a member $S(k + k)$ of X would give $h + h = S(k + k)$ by $\mathrm{A}(X)$, which Lemma 7.17(f) refutes. The second case gives $\mathrm{Ev}(X)$ in the same way. For $\mathrm{MP} \to \mathrm{WMP}$ assume $\mathrm{W}(X)$ and apply it at $Y := X$. Its second alternative, $\neg\neg\exists n\,(n \in X \wedge \neg\, n \in X)$, is absurd; its first, $\neg\neg\mathrm{E}(X)$, gives $\neg\mathrm{Z}(X)$, since $\mathrm{Z}(X)$ refutes $\mathrm{E}(X)$. So MP gives $\mathrm{E}(X)$. LLPO gives $\mathrm{MP}^\vee$ directly. Proposition 7.35 gives $\mathrm{LPO} \to \mathrm{MP}$ and $\mathrm{LPO} \to \mathrm{WLPO} \to \mathrm{LLPO}$.

(2) Use Proposition 7.35 and disjunction introduction; EM gives WEM, since $0 \in X$ gives $\neg\neg\, 0 \in X$. That each of the three conjunctions derives its first conjunct is conjunction

elimination. MP $\wedge$ LLPO gives MP by conjunction elimination, and LLPO $\wedge$ (WLPO $\vee$ MP) by conjunction elimination and disjunction introduction.

(3) Fix $2 \le n \le m$ and $A : \mathbb{N} \to H$. Put $g(j) = m\lfloor j/n \rfloor + (j \bmod n)$ and $B_j = A_{g(j)}$; the map g is injective, since $j \bmod n < n \le m$, and sends $nk + i$ to $mk + i$ for $i < n$. The values of $\mathrm{Dec}(X)$ and $\mathrm{A}(X)$ at B are meets over some of the meetands of their values at A, hence at least those values: for $j \neq j'$ the meetand $B_j \Rightarrow (B_{j'} \Rightarrow \bot)$ is that of A at $g(j) \neq g(j')$. For $i < n$ the value of $\mathrm{C}_i^n(X)$ at B is $\bigwedge_k \neg A_{mk+i}$, the value of $\mathrm{C}_i^m(X)$ at A; for $n = 2$ the values of $\mathrm{Ev}(X)$ and $\mathrm{Od}(X)$ at B are those of $\mathrm{C}_0^2(X)$ and $\mathrm{C}_1^2(X)$, since $k + k$ and $2k$ have the same value. So the conclusion of LLPO_n at B, and for $n = 2$ that of LLPO, a join of the values of $\mathrm{C}_i^m(X)$ at A over $i < n$, lies below the conclusion of LLPO_m at A, the join over all $i < m$. Since $a \Rightarrow b$ grows when a shrinks or b grows, the instance of LLPO_n at B, and for $n = 2$ that of LLPO, lies below the instance of LLPO_m at A. The value of a principle lies below each of its instances, and the value of LLPO_m is the meet of its instances over all A. This gives $[\![\mathrm{LLPO}_n]\!] \le [\![\mathrm{LLPO}_m]\!]$, and, with $n = 2$, $[\![\mathrm{LLPO}]\!] \le [\![\mathrm{LLPO}_m]\!]$ for every $m \ge 2$. □

Three models for the halves of Markov's principle

Section 7.12 computes every entry of Table 7.3 but one, which we compute first.

Theorem 7.42 (Markov's principle in the remainder)**.** *In the open sets of $\beta\mathbb{N} \setminus \mathbb{N}$, Markov's principle has the value $\top$. Hence MP and LLPO both hold there, while WLPO and LPO fail.*

Proof. We use the sets B^* of Section 7.12. By Lemma 7.37 it is enough to show that a union O of sets $A_n = B_n^*$ is the interior of its closure. It lies in that interior, being open. Suppose a point p of the interior is not in O, and choose C^* with $p \in C^* \subseteq \overline{O}$. For each n, $p \notin B_n^*$, so $B_n \notin p$, and since an ultrafilter contains a finite union only if it contains one of its members, $D_n = C \setminus (B_0 \cup \cdots \cup B_n)$ belongs to p. Each D_n is infinite, because p contains no finite set. Choose $e_0 \in D_0$, and $e_{k+1} \in D_{k+1}$ with $e_{k+1} > e_k$. The set $E = \{e_0, e_1, \ldots\}$ is infinite and lies in C, and it meets B_n only among $e_0, \ldots, e_{n-1}$, because $e_k \in D_k$ misses B_n for $k \ge n$. So E^* is nonempty, lies in $C^* \subseteq \overline{O}$, and misses every A_n, since $E^* \cap B_n^* = (E \cap B_n)^*$ is empty. A nonempty open subset of $\overline{O}$ that misses O is impossible. LLPO holds and WLPO and LPO fail by Section 7.12. □

With S^+ the one-point compactification of an uncountable set of Section 7.12, the next model compactifies $\mathbb{N}$: let $\widehat{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, and call a set open when it omits ∞ or contains ∞ and all but finitely many natural numbers, so that $0, 1, 2, \ldots$ converges to ∞. Each natural number is isolated, so for an open set U and a natural number m we have $m \in \neg U$ exactly when $m \notin U$, and $m \in \neg\neg U$ exactly when $m \in U$.

Proposition 7.43 (The convergent sequence)**.** *In the open sets of $\widehat{\mathbb{N}}$,* WMP *has the value $\top$ and* $\mathrm{MP}^\vee$ *does not. Consequently MP, LLPO, WLPO and LPO fail there.*

Proof. WMP holds. Let $A : \mathbb{N} \to H$, let $O = \bigcup_n A_n$, the value of $\mathrm{E}(X)$, and let x lie in the values of $\mathrm{Dec}(X)$ and $\mathrm{W}(X)$. We show $x \in O$. The value of $\mathrm{W}(X)$ lies below its instance at the constant map $\bot$, whose value is $\neg\neg O$; so if x is a natural number, it lies in O. Let $x = \infty$ and suppose $\infty \notin O$. Since ∞ lies in $A_n \cup \neg A_n$, it lies in $\neg A_n$ for every n. Let B_n be the

set of even natural numbers in A_n; it is open. The value of $\mathrm{Dec}(Y)$ at B is the whole space: $\infty \in \neg A_n \subseteq \neg B_n$, and each natural number lies in B_n or, being isolated, in $\neg B_n$. So the instance of $\mathrm{W}(X)$ at B puts ∞ in $\neg\neg U$ or in $\neg\neg V$, where $U = \bigcup_n B_n$ and $V = \bigcup_n (A_n \cap \neg B_n)$. Both are impossible. The open set $\neg\neg U$ would contain ∞, hence every large natural number, among them an odd number $2N+1$; but $2N+1$ is outside U, hence in $\neg U$ and not in $\neg\neg U$. Likewise $\neg\neg V$ would contain a large even number $2N+2$; but $2N+2$ lies in no $A_n \cap \neg B_n$, since an even member of A_n is in B_n, so it is in $\neg V$ and not in $\neg\neg V$.

$\mathrm{MP}^\vee$ *fails.* Take $A_n = \{n\}$. Its complement is open, so the value of $\mathrm{Dec}(X)$ is the whole space, and the sets are disjoint, so that of $\mathrm{A}(X)$ is too. An open set disjoint from every $\{n\}$ is empty; so $\neg\mathrm{Z}(X)$ has the value of the whole space. The instance of $\mathrm{MP}^\vee$ at A therefore has the value of $\mathrm{Ev}(X) \vee \mathrm{Od}(X)$. It misses ∞: an open set containing ∞ contains an even number $2k$, so it is not inside $\neg\{2k\}$, and it contains an odd number, so the same holds for the odd places. By Lemmas 7.41 and 7.21, the values of MP and LLPO lie below that of $\mathrm{MP}^\vee$, and by Corollary 7.36 those of WLPO and LPO lie below that of LLPO. □

Proposition 7.44 (Weak Markov fails in $\beta\mathbb{N}$)**.** *In the open sets of* $\beta\mathbb{N}$, WMP *does not have the value* $\top$.

Proof. Use the notation of Section 7.12 and take $A_n = \{n\}$, the isolated principal points. Then $\mathrm{Dec}(X)$ has the value $\top$ and $\mathrm{E}(X)$ the value $\mathbb{N}$, which is not the whole space. We show that $\mathrm{W}(X)$ has the value $\top$; then the instance of WMP at A has the value $\mathbb{N}$. First, if every point n of a set $E \subseteq \mathbb{N}$ lies in an open set U, then $\overline{E} \subseteq \neg\neg U$. Indeed $\overline{E}$ is open, and it misses $\neg U$: a point v of $\overline{E}$ in $\neg U$ would have a basic neighbourhood $\bar{t} \subseteq \neg U$, and $\bar{t}$ meets E, because $t \cap E \in v$, at a point that lies in both U and $\neg U$. Now let $B : \mathbb{N} \to H$ and let u lie in the value of $\mathrm{Dec}(Y)$; choose $s \in u$ with $\bar{s}$ inside that value. For $n \in s$, the point n lies in B_n or in $\neg B_n$. Let E be the set of $n \in s$ with $n \in B_n$. If $E \in u$, then $u \in \overline{E} \subseteq \neg\neg \bigcup_n B_n$. Otherwise $s \setminus E \in u$, and each $n \in s \setminus E$ lies in $A_n \cap \neg B_n$, so $u \in \neg\neg \bigcup_n (A_n \cap \neg B_n)$. Either way u lies in the value of the disjunction in $\mathrm{W}(X)$. So the instance of $\mathrm{W}(X)$ at B has the value $\top$ for every B, and so has $\mathrm{W}(X)$. □

Theorem 7.45 (The two halves of Markov's principle, and MP with LLPO)**.** *Over intuitionistic second-order arithmetic:*

1. $\mathrm{MP} \to \mathrm{MP}^\vee$ *is derivable, while* $\mathrm{MP}^\vee$ *and* $\mathrm{MP}^\vee \to \mathrm{MP}$ *are not;*
2. $\mathrm{MP} \to \mathrm{WMP}$ *is derivable, while* WMP *and* $\mathrm{WMP} \to \mathrm{MP}$ *are not;*
3. $\mathrm{LLPO} \to \mathrm{MP}^\vee$ *is derivable, while* $\mathrm{MP}^\vee \to \mathrm{LLPO}$ *is not;*
4. *none of the six implications between* WMP *and each of* $\mathrm{MP}^\vee$*, LLPO and WLPO, in either direction, is derivable;*
5. $\mathrm{LPO} \to \mathrm{MP} \wedge \mathrm{LLPO}$ *is derivable, while none of* $\mathrm{MP} \wedge \mathrm{LLPO} \to \mathrm{LPO}$, $\mathrm{MP} \to \mathrm{MP} \wedge \mathrm{LLPO}$, $\mathrm{LLPO} \to \mathrm{MP} \wedge \mathrm{LLPO}$, $\mathrm{MP} \wedge \mathrm{LLPO} \to \mathrm{WLPO}$ *and* $\mathrm{WLPO} \to \mathrm{MP} \wedge \mathrm{LLPO}$ *is derivable.*

So LPO is WLPO and MP together (Proposition 7.35), but not LLPO and MP.

Model	$\mathrm{MP}^\vee$	WMP	MP	LLPO	WLPO	LPO
convergent sequence $\widehat{\mathbb{N}}$	$\neq\top$	$\top$	$\neq\top$	$\neq\top$	$\neq\top$	$\neq\top$
one-point compactification S^+	$\top$	$\top$	$\top$	$\neq\top$	$\neq\top$	$\neq\top$
remainder $\beta\mathbb{N}\setminus\mathbb{N}$	$\top$	$\top$	$\top$	$\top$	$\neq\top$	$\neq\top$
$\beta\mathbb{N}$	$\top$	$\neq\top$	$\neq\top$	$\top$	$\top$	$\neq\top$
chain $0<1<\cdots<\infty$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$

Table 7.4. The two halves of Markov's principle and four of the six principles in five models.

Proof. The derivations are Lemma 7.41. The non-derivations are read off from models as in Section 7.12, a conjunction having the meet of the two values; Table 7.4 collects the values. The rows come from Proposition 7.43, Section 7.12 (S^+, $\beta\mathbb{N}$, the chain), Theorem 7.42 and Proposition 7.44, the other entries from Lemma 7.41 with Lemma 7.21, and from Corollary 7.36. Now read off: (1) $\widehat{\mathbb{N}}$ refutes $\mathrm{MP}^\vee$, and $\beta\mathbb{N}$ has $\mathrm{MP}^\vee$ without MP; (2) $\beta\mathbb{N}$ refutes WMP, and $\widehat{\mathbb{N}}$ has WMP without MP; (3) S^+ has $\mathrm{MP}^\vee$ without LLPO; (4) $\widehat{\mathbb{N}}$ has WMP without $\mathrm{MP}^\vee$, LLPO or WLPO, and $\beta\mathbb{N}$ has all three without WMP; (5) $\beta\mathbb{N}\setminus\mathbb{N}$ has MP $\wedge$ LLPO without LPO or WLPO, S^+ has MP without LLPO, and $\beta\mathbb{N}$ has LLPO and WLPO without MP. □

Hendtlass and Lubarsky separate many fragments of weak excluded middle, LPO and Markov's principle, WMP among them, by Kripke and topological models [74]; the models used here are topological.

Infinitely many prices below LLPO

For the family LLPO_n we need models in which the neighbourhoods of one extra point are prescribed by finitely many ultrafilters. For $k\geq 1$ and $l<k$, let $U_{k,l}$ be a nonprincipal ultrafilter on $\mathbb{N}$ that contains the residue class of l modulo k, and let F_k be the family of sets that belong to all of $U_{k,0},\dots,U_{k,k-1}$. Let $X_k=\mathbb{N}\cup\{\infty\}$, with a set W open when $\infty\in W$ implies that the natural numbers in W form a member of F_k. Natural numbers are isolated, and the complement of $\{m\}$ is open, because a nonprincipal ultrafilter contains every cofinite set.

Proposition 7.46 (A family of principles below LLPO)**.**

1. *For $n\geq 2$, the value of LLPO lies below that of LLPO_n in every system of truth values.*
2. *No LLPO_n with $n\geq 2$ is derivable in HAS.*
3. *For $n\geq 3$, neither $\mathrm{LLPO}_n\to\mathrm{LLPO}$ nor $\mathrm{LLPO}_n\to\mathrm{MP}^\vee$ is derivable, and for $n\geq 2$ neither is $\mathrm{MP}^\vee\to\mathrm{LLPO}_n$.*
4. *For $2\leq n<m$, $\mathrm{LLPO}_m\to\mathrm{LLPO}_n$ is not derivable.*

Proof. Part 1 is Lemma 7.41(3). In each model below, an open set disjoint from every A_{nj+i} lies in the value of $\mathrm{C}_i^n(X)$.

LLPO_k *fails in X_k.* Take $A_m=\{m\}$; the values of $\mathrm{Dec}(X)$ and $\mathrm{A}(X)$ are the whole space. If ∞ lay in the value of $\mathrm{C}_l^k(X)$, an open set inside it would contain ∞ and a member T of $F_k\subseteq U_{k,l}$

of natural numbers. The set T meets the residue class of l, at some $m = kj + l$, and m would lie in $\neg\{m\}$.

LLPO_m *holds in* X_k *for* $m > k$. Let $A : \mathbb{N} \to H$ and let x lie in the values of $\mathrm{Dec}(X)$ and of $\mathrm{A}(X)$; call the latter M. We find $i < m$ with x in the value of $\mathrm{C}_i^m(X)$. If $x \in A_a$ for some a, choose $i < m$ different from a mod m. A point of the open set $A_a \cap M$ that lay in some A_{mj+i} would give $a = mj + i$ by $\mathrm{A}(X)$, which is impossible; so $A_a \cap M$ is disjoint from every A_{mj+i} and contains x. If x is a natural number in no A_a, the open set $\{x\}$ serves for $i = 0$. Finally let $x = \infty$ lie in no A_a. The natural numbers of M form a set $T \in F_k$. For $i < m$ let O_i be the set of natural numbers lying in some A_{mj+i}. For $i \neq i'$, the set $O_i \cap O_{i'} \cap T$ is empty, by $\mathrm{A}(X)$. If every O_i belonged to one of the k ultrafilters, two of the $m > k$ classes would belong to the same $U_{k,l}$, together with T, and their empty intersection would belong to it. So some O_i belongs to none of them, its complement belongs to all, and $\{\infty\} \cup (\mathbb{N} \setminus O_i)$ is an open set containing ∞ and disjoint from every A_{mj+i}.

$\mathrm{MP}^\vee$ *fails in* X_2. With $A_m = \{m\}$, an open set containing ∞ contains a member of $U_{2,0}$, hence an even number p, which is not in $\neg\{p\}$; so ∞ is not in the value of $\mathrm{Ev}(X)$, nor of $\mathrm{Z}(X)$, and by $U_{2,1}$ not in that of $\mathrm{Od}(X)$. Natural numbers are not in the value of $\mathrm{Z}(X)$ either, so $\neg\mathrm{Z}(X)$ has the value $\top$, and the instance of $\mathrm{MP}^\vee$ at A misses ∞.

Every LLPO_n *fails in* S^+. With distinct points $s_0, s_1, \ldots$ of S and $A_m = \{s_m\}$, every open set containing ∞ contains some s_{nj+i}, so ∞ lies in no value $\mathrm{C}_i^n(X)$. MP, and with it $\mathrm{MP}^\vee$, holds in S^+.

Now read off: S^+ refutes LLPO_n and validates $\mathrm{MP}^\vee$, which gives part 2 and the last claim of part 3. For $n \geq 3$, X_2 validates LLPO_n and refutes $\mathrm{MP}^\vee$ and LLPO_2, hence LLPO by part 1. For $n < m$, X_n validates LLPO_m and refutes LLPO_n. □

With Lemma 7.41(3), parts 3 and 4 make LLPO, LLPO_3, LLPO_4, … a strictly descending chain. Diener shows that each LLPO_n fails in the topological model over any metric space with an accumulation point [73, Proposition 8.1.5]; for separations of this kind see also [74].

Between LLPO and WLPO, and between WLPO and LPO

The candidates are $\mathrm{LLPO} \wedge (\mathrm{WLPO} \vee \mathrm{MP})$ and $\mathrm{WLPO} \wedge (\mathrm{MP} \vee \mathrm{WEM})$. That each lies between its neighbours is Lemma 7.41(2), and the upper steps are strict by models we have: $\beta\mathbb{N} \setminus \mathbb{N}$ validates LLPO and MP but not WLPO, and $\beta\mathbb{N}$ validates WLPO and WEM but not LPO. WEM holds in $\beta\mathbb{N}$ because the closure of an open set U is open (Section 7.12): then $\neg U$ is the complement of $\overline{U}$, $\neg\neg U$ is $\overline{U}$, and the two cover the space. The lower steps need a model in which the lower principle holds and both alternatives of the disjunction fail at one point. For a set Y and a family $\mathscr{B}$ of its subsets we give a space of ultrafilters on Y the topology with basic open sets $\langle s\rangle$, the ultrafilters containing s, for $s \in \mathscr{B}$. Then $\langle s\rangle \cap \langle t\rangle = \langle s \cap t\rangle$ and the complement of $\langle s\rangle$ is $\langle Y \setminus s\rangle$, so whenever $\mathscr{B}$ is closed under finite intersections and complements the basic sets are open and closed and form a base, and every closed set is an intersection of basic sets.

The model K. Let $Y = \mathbb{N} \times \mathbb{N}$. A point (n, k) lies in column n, and we read k through a fixed bijection $k \mapsto (r(k), q(k))$ of $\mathbb{N}$ onto $\mathbb{N} \times \mathbb{N}$ with $r(k) \leq k$ as a row $r(k)$ and a place $q(k)$. A set is *column-finite* when it meets each column in finitely many points. Let K be the space of

ultrafilters on Y that contain the complement of every column-finite set, with $\mathscr{B}$ all subsets of Y. A basic set $\langle s\rangle$ is empty exactly when s is column-finite: otherwise s meets a column infinitely often, so s and the complements of the column-finite sets have the finite intersection property and lie in a common ultrafilter. The same argument shows that K is compact: if finitely many of the sets $\langle s_i\rangle$ always meet, then no finite intersection of the s_i is column-finite, and one ultrafilter in K contains all of them. So every set both open and closed, being compact, is a finite union of basic sets, hence basic.

The model L. Let $Z = \{0,1\} \times \mathscr{P}(\mathbb{N})$ and $Y = \mathbb{N} \times Z$. Each half $\{b\} \times \mathscr{P}(\mathbb{N})$ of Z is uncountable. Let $\mathscr{B}$ consist of the sets $s \subseteq Y$ whose column sections $s_n = \{z : (n,z) \in s\}$ are each countable or have countable complement in Z. This family is closed under complements and countable unions: the section at n of a union is the union of the sections at n, which is countable when all of them are and otherwise contains a set with countable complement. Let L be the space of all ultrafilters on Y, with this $\mathscr{B}$. The principal ultrafilter x_p at $p \in Y$ lies in $\langle s\rangle$ exactly when $p \in s$, so a basic set $\langle s\rangle$ is nonempty exactly when s is. L is compact by the argument used for K, and every set both open and closed is $\langle b\rangle$ for some $b \in \mathscr{B}$.

Theorem 7.47 (Two further intermediate prices)**.** *Over intuitionistic second-order arithmetic, the four implications of the chains*

$$\mathrm{WLPO} \to \mathrm{LLPO} \wedge (\mathrm{WLPO} \vee \mathrm{MP}) \to \mathrm{LLPO} \quad \textit{and} \quad \mathrm{LPO} \to \mathrm{WLPO} \wedge (\mathrm{MP} \vee \mathrm{WEM}) \to \mathrm{WLPO}$$

are derivable, and none of the four reverse implications is. More precisely, LLPO has the value $\top$ in K while $\mathrm{WLPO} \vee \mathrm{MP}$ *does not, and WLPO has the value $\top$ in L while* $\mathrm{MP} \vee \mathrm{WEM}$ *does not.*

Proof. The derivations and the upper steps were shown above. A model in which P has the value $\top$ and D does not gives $P \to P \wedge D$ the value of D; so it remains to check the two models, where Lemma 7.37 applies.

LLPO holds in K. Let $A_n = \langle b_n\rangle$ be pairwise disjoint, so that $b_m \cap b_n$ is column-finite for $m \neq n$. Put

$$s = \bigcup_k \big(b_{2k+1} \setminus (b_0 \cup b_2 \cup \cdots \cup b_{2k})\big).$$

Then $s \cap b_{2m}$ lies in the union of the sets $b_{2k+1} \cap b_{2m}$ with $k < m$, and $b_{2k+1} \setminus s$ in the union of the sets $b_{2k+1} \cap b_{2j}$ with $j \leq k$; both are column-finite. So every even-indexed A_n lies in $\langle Y \setminus s\rangle$ and every odd-indexed one in $\langle s\rangle$, two disjoint closed sets that separate the closures of the two unions.

The point x_1 of K. For $N \in \mathbb{N}$, $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, the tail $T(N,f,g)$ is the set of (n,k) with $n \geq N$, $r(k) \geq f(n)$ and $q(k) \geq g(n, r(k))$. Two tails contain a third, and no tail is empty, so some ultrafilter x_1 contains every tail. It lies in K: if c is column-finite, choose $f(n)$ above every position of c in column n; since $r(k) \leq k$, the tail $T(0,f,0)$ misses c.

MP fails at x_1. Let $A_n = \langle \{n\} \times \mathbb{N}\rangle$, the columns. A nonempty basic $\langle s\rangle$ has s meeting some column n infinitely often, so it meets A_n; hence $\mathrm{Z}(X)$ has the value $\varnothing$ and $\neg\mathrm{Z}(X)$ the value K. But x_1 contains no column, since $T(n+1,0,0)$ misses column n, so x_1 is not in the value $\bigcup_n A_n$ of $\mathrm{E}(X)$.

WLPO fails at x_1. Let $A_m = \langle R_m\rangle$, where R_m is the set of points in row m, across all columns. Suppose a basic $\langle s\rangle \ni x_1$ lies in the value of $\mathrm{Z}(X)$. If s met each cell (column n, row m) in

finitely many points, with places below $G(n,m)$, then s would miss the tail $T(0,0,G)$; so s meets some cell infinitely often, $s \cap R_m$ is not column-finite, and $\langle s\rangle$ meets A_m, which is impossible. Suppose instead that $\langle s\rangle \ni x_1$ lies in the value of $\neg\mathrm{Z}(X)$. If s met each column in boundedly many rows, with rows below $F(n)$, it would miss $T(0,F,0)$; so in some column n it meets infinitely many rows. Let s' be the set of points of s that have the least position k among the points of s in their cell. Then s' meets each row in at most one point per column, so $\langle s'\rangle \cap A_m = \langle s' \cap R_m\rangle$ is empty for every m, and $\langle s'\rangle$ lies in the value of $\mathrm{Z}(X)$. But s' has infinitely many points in column n, so $\langle s'\rangle$ is a nonempty open subset of $\langle s\rangle$, inside the values of both $\mathrm{Z}(X)$ and its negation. So x_1 lies in neither value, the instance of WLPO at A misses x_1, and x_1 is in neither $[\![\mathrm{WLPO}]\!]$ nor $[\![\mathrm{MP}]\!]$.

WLPO holds in L. Let $A_n = \langle b_n\rangle$ with $b_n \in \mathscr{B}$, and let $t = \bigcup_n b_n$, a member of $\mathscr{B}$. The set $\langle t\rangle$ is closed and contains every A_n. Each of its points x lies in the closure of their union: for a basic $\langle c\rangle \ni x$, the set $c \cap t \in x$ has a point p, in some b_n, and $x_p \in \langle c\rangle \cap A_n$. So the closure of the union is $\langle t\rangle$, which is open.

The point x_0 of L. The tails $\{(n,z) : n \geq N,\ z \notin C\}$, for $N \in \mathbb{N}$ and countable $C \subseteq Z$, are nonempty and directed, so some ultrafilter x_0 contains all of them. It contains no column $\{n\} \times Z$, since the tail for $N = n+1$ misses it. Every $s \in \mathscr{B}$ with $s \in x_0$ meets both halves $\mathbb{N} \times \{b\} \times \mathscr{P}(\mathbb{N})$: if s missed the half of b, no section of s could have countable complement, since that half of Z is uncountable; so all sections would be countable, their union C would be countable, and the tail for $N = 0$ and C would miss s.

MP fails at x_0. The columns lie in $\mathscr{B}$. With A_n the basic set of column n, the value of $\mathrm{Dec}(X)$ is L, and every nonempty basic set contains a principal point, which lies in some A_n; so $\neg\mathrm{Z}(X)$ has the value L, while x_0 is not in $\bigcup_n A_n$.

WEM fails at x_0. For $z \in Z$ the row $\mathbb{N} \times \{z\}$ lies in $\mathscr{B}$. Let P be the union of the basic sets of the rows with z in the half of 1, and Q that of the rows with z in the half of 0. Distinct rows are disjoint, so $Q \subseteq \neg P$. If a basic $\langle s\rangle \ni x_0$ lay in $\neg P$, then s would meet the half of 1 at a point p, and x_p would lie in $\langle s\rangle$ and in P. If it lay in $\neg\neg P$, then s would meet the half of 0 at p, and x_p would lie in $\langle s\rangle$ and in $Q \subseteq \neg P$. So x_0 is not in $\neg P \cup \neg\neg P$, the instance of WEM at the constant map P, and x_0 is in neither $[\![\mathrm{MP}]\!]$ nor $[\![\mathrm{WEM}]\!]$. □

Above LPO

Proposition 7.48 (Excluded middle for decidable sets and its levels)**.**

1. *Let φ be a formula without set quantifiers, possibly with number parameters. If HAS derives, from a list Γ, LPO and $\mathrm{Dec}(X)$ for each set variable X free in φ, then it derives $\varphi \vee \neg\varphi$ from Γ.*

2. *Consequently HAS derives $\mathrm{LPO} \to \forall X\big(\mathrm{Dec}(X) \to \varphi \vee \neg\varphi\big)$ for every such φ whose only set variable is X, whatever its quantifier prefix; and it derives $\forall X\big(\mathrm{Dec}(X) \to \mathrm{E}(X) \vee \neg\mathrm{E}(X)\big) \to \mathrm{LPO}$.*

3. *$\mathrm{EM} \to \mathrm{LPO} \wedge \mathrm{WEM}$ and $\mathrm{LPO} \wedge \mathrm{WEM} \to \mathrm{LPO}$ are derivable, and none of $\mathrm{LPO} \wedge \mathrm{WEM} \to \mathrm{EM}$, $\mathrm{LPO} \to \mathrm{LPO} \wedge \mathrm{WEM}$, $\mathrm{LPO} \to \mathrm{WEM}$ and $\mathrm{WEM} \to \mathrm{LPO}$ is.*

Proof. (1) Induction on φ. Equations are decided by Lemma 7.17(a), memberships $t \in X$ by the hypothesis $\mathrm{Dec}(X)$ at t, and $\bot$ is refuted. Decisions of two formulas decide their conjunction, disjunction and implication. For $\exists n\,\psi$, the induction hypothesis, applied with n fresh, gives $\forall n\,(\psi \vee \neg\psi)$. Comprehension gives a set Y with $\forall n\,(n \in Y \leftrightarrow \psi)$, and Y is decidable. LPO at Y gives $\mathrm{Z}(Y)$, hence $\neg\exists n\,\psi$, or $\mathrm{E}(Y)$, hence $\exists n\,\psi$. For $\forall n\,\psi$, LPO at the set of n with $\neg\psi$ gives $\forall n\,\neg\neg\psi$, hence $\forall n\,\psi$ by the decisions, or a counterexample, hence $\neg\forall n\,\psi$. This is the argument of Theorem 6.10, with comprehension supplying the sets.

(2) The first claim is part 1 with Γ the hypotheses LPO and $\mathrm{Dec}(X)$. For the second, let $\mathrm{Dec}(X)$; the hypothesis gives $\mathrm{E}(X)$ or $\neg\mathrm{E}(X)$, and the second gives $\mathrm{Z}(X)$.

(3) The derivations are Lemma 7.41(2). The chain $0 < 1 < 2 < \cdots < \infty$ validates LPO (Section 7.12) and WEM, because in a chain $\neg a$ is $\top$ for $a = \bot$ and $\bot$ otherwise, so $\neg a \vee \neg\neg a = \top$; EM fails there. The *cone* has three points, two open points a, b and a focal point c whose only open neighbourhood is the whole space. Let $A : \mathbb{N} \to H$. The open point a lies in some A_n, hence in the value of $\mathrm{E}(X)$, or in none, and then the open set $\{a\}$ lies in the value of $\mathrm{Z}(X)$; the same holds for b. If c lies in the value of $\mathrm{Dec}(X)$, then for each n it lies in A_n, which is then the whole space, or in $\neg A_n$, which is then the whole space, so that A_n is empty; so c lies in the value of $\mathrm{E}(X)$, or $\mathrm{Z}(X)$ has the value $\top$. Hence LPO has the value $\top$. WEM fails at c: for $A_0 = \{a\}$, $\neg A_0 = \{b\}$ and $\neg\neg A_0 = \{a\}$, and their union misses c. So the cone refutes $\mathrm{LPO} \to \mathrm{WEM}$ and $\mathrm{LPO} \to \mathrm{LPO} \wedge \mathrm{WEM}$. Finally $\beta\mathbb{N}$ validates WEM and refutes LPO. □

Corollary 7.49 (The six prices are not all the prices)**.** *Between each of the following pairs of principles, HAS has a sentence that is derivable from the stronger, derives the weaker, and is equivalent to neither: none and MP (*$\mathrm{MP}^{\vee}$ *and* WMP*); none and LLPO (*$\mathrm{MP}^{\vee}$*); MP and LPO, and also LLPO and LPO (*$\mathrm{MP} \wedge \mathrm{LLPO}$*); LLPO and WLPO (*$\mathrm{LLPO} \wedge (\mathrm{WLPO} \vee \mathrm{MP})$*); WLPO and LPO (*$\mathrm{WLPO} \wedge (\mathrm{MP} \vee \mathrm{WEM})$*); LPO and EM (*$\mathrm{LPO} \wedge \mathrm{WEM}$*).*

Proof. Lemma 7.41 and conjunction elimination give the derivations (every sentence derives $\bot \to \bot$); Theorems 7.45 and 7.47 and Proposition 7.48 give the rest. □

The chart of prices is therefore neither a ladder nor a finite list. Not every candidate is new: Markov's principle for quantifier-free formulas of arithmetic is derivable in HAS (Theorem 6.9(3) and Lemma 7.16).

7.14 One halting question, priced instance by instance

Fix one program question: does program e return a value on input 0? The halting matrix $\theta(e, x)$ of Lemma E.11 says that it does so within allowance x. Each principle of this chapter then has one instance at θ, a single sentence of first-order arithmetic. The WLPO instance says that for every e the search for a halting time either fails everywhere or does not fail everywhere; the LPO instance says that it fails everywhere or finds a time. Does the first sentence derive the second in the calculus of Chapter 6?

The question is finer than the ones answered so far. Theorem 6.11 prices the three rules of Chapter 6 as schemas, and Theorems 7.38, 7.45 and 7.47 separate the principles of this chapter as schemas over all sets of intuitionistic second-order arithmetic. A separation of schemas

may use a different set at each step. Here the instance is fixed, the sentences have no set variables, and the derivations are those of Chapter 6, with their ledgers and every tier of induction. The answer is that the familiar separations survive instance by instance: WLPO gives neither LPO nor Markov's principle at θ, the bit principles give neither, LLPO gives neither WLPO nor LPO, and Markov's principle proper, unlike the quantifier-free rule of Theorem 6.9, has a price. Every separation is a model and a check; at the halting matrix the models are realizability interpretations, and each check ends with a program that cannot exist: one that decides whether programs halt [21], one that separates the programs returning 0 from those returning 1, or one that picks a halting program from every pair of which one halts. The section closes with the order of the weak arithmetic principles over HAS.

As in Section 7.12, the separations are theorems about a specified calculus, proved from outside it in classical mathematics.

Instances and letters

Definition 7.50 (Instances at a matrix)**.** Let $\varphi(e, x)$ be a formula of the arithmetic language whose free variables are among e and x. Its instances are the sentences

$$\begin{aligned}
\mathrm{MP}_\varphi &:\equiv \forall e\,\big(\neg\neg\exists x\,\varphi \to \exists x\,\varphi\big),\\
\mathrm{LPO}_\varphi &:\equiv \forall e\,\big(\forall x(\varphi \vee \neg\varphi) \to \exists x\,\varphi \vee \forall x\,\neg\varphi\big),\\
\mathrm{WLPO}_\varphi &:\equiv \forall e\,\big(\forall x(\varphi \vee \neg\varphi) \to \forall x\,\neg\varphi \vee \neg\forall x\,\neg\varphi\big),\\
\mathrm{LLPO}_\varphi &:\equiv \forall e\Big(\forall x(\varphi \vee \neg\varphi) \to \forall x\,\forall y\big(\varphi(e,x) \wedge \varphi(e,y) \to x = y\big)\\
&\qquad\qquad \to \forall x\,\neg\varphi(e, x + x) \vee \forall x\,\neg\varphi(e, S(x + x))\Big),\\
\mathrm{MP}^{\vee}_\varphi &:\equiv \forall e_1\,\forall e_2\,\Big(\neg\big(\neg\exists x\,\varphi(e_1,x) \wedge \neg\exists x\,\varphi(e_2,x)\big)\\
&\qquad\qquad \to \neg\neg\exists x\,\varphi(e_1,x) \vee \neg\neg\exists x\,\varphi(e_2,x)\Big).
\end{aligned}$$

The LPO instance is the universal closure of the LPO rule (6.4) at φ; $\mathrm{MP}^{\vee}_\varphi$ is the disjunctive Markov principle of Definition 7.39, read for two programs.

Two matrices carry the section. The first is the halting matrix θ. The second records the output. Let $\mathrm{out}_b(e, K)$ say, as $\theta(e, K)$ does, that some justified table claims a value at the cell $(K, e, 0)$, and that the value is b. Put

$$\begin{aligned}
\lambda(e,x) :\equiv \exists K\Big[&\big(K = 0 \vee \exists K'\,(K = S(K') \wedge \neg\theta(e,K'))\big)\\
&\wedge \big((x = K + K \wedge \mathrm{out}_0(e,K)) \vee (x = S(K + K) \wedge \mathrm{out}_1(e,K))\big)\Big].
\end{aligned}$$

By the proof of Lemma E.11, $\lambda(e, x)$ is true exactly when $x = 2K + b$, where K is the least allowance within which program e returns a value on input 0 and $b \in \{0, 1\}$ is that value. So $\lambda(e, \cdot)$ has at most one witness; it has an even one exactly when program e returns 0, and an odd one exactly when it returns 1. Its truth at (e, x) is decided by a program, which also computes a realizer when it is true. These are the two sets A_0, A_1 of programs that no program separates (proof of Theorem 12.22).

Definition 7.51 (Letters and Σ_1 formulas)**.** A *letter* of a principle P among MP, LPO, WLPO and LLPO is a sentence $\forall z_1 \cdots \forall z_n\, \psi$ where ψ is the body of the instance P_φ (the formula under $\forall e$, or under $\forall e_1 \forall e_2$) at some formula φ; the variables z_i close all its parameters. A Σ_1 *formula* is one built from equations and $\bot$ by $\wedge$, $\vee$, $\exists$ and bounded universal quantifiers $\forall y\big(\exists z\,(b + z = y) \vee \chi\big)$, with b a term free of y and z; such a formula contains no implication, hence no negation. The halting matrix θ is a Σ_1 formula: its bounded universal statements have exactly this form (Lemma E.11). A derivation *from letters* is a derivation of Chapter 6 whose context is a finite list of letters.

So MP_θ is an MP letter at a Σ_1 formula, and an LLPO or WLPO letter may sit at any formula at all. The rules of Chapter 6 add three facts that every proof below uses.

Lemma 7.52 (Three tools)**.**

1. *A derivation from letters whose EM and LPO entries are false can be rebuilt, with the same conclusion and context, as a derivation from the same letters with no added rule; the rebuilt derivation may use induction on formulas with quantifiers.*
2. *A derivation with no added rule is, read in the two-sorted language, a derivation in the theory of Theorem 12.22, from the same hypotheses.*
3. *Kleene realizability, modified realizability and Lifschitz realizability each carry realizers of the hypotheses of such a derivation to a realizer of its conclusion, by one program built along the derivation.*

Proof. (1) Each MP rule is at a quantifier-free body, and by Theorem 6.9(3) it has a derivation with no added rule; replace each MP leaf by that derivation, with the leaf's hypotheses added, as in Section 6.7. (2) The rules of Chapter 5, with induction on every formula, are rules of that theory for formulas without sequence variables, and its arithmetic axioms are the six arithmetic rules. (3) For Kleene realizability this is Theorem E.9; for modified realizability it is the construction in the proof of Theorem 12.22; for Lifschitz realizability it is Lemma 12.69. □

Markov's principle proper has a price

Theorem 6.9 showed that the Markov rule of Chapter 6, restricted to quantifier-free bodies, costs nothing once induction on quantified formulas is allowed: such a body is eventually constant, so its search is bounded in advance. At the halting matrix the search is not bounded, and the price returns.

Theorem 7.53 (Markov's principle at the halting matrix)**.**

1. MP_θ *has no derivation whose EM and LPO entries are false, whatever its induction; in particular none with false EM, LPO and MP entries.*
2. *Read in the two-sorted language,* MP_θ *is not a theorem of the theory of Theorem 12.22, intuitionistic arithmetic with sequences and countable choice for numbers and for sequences.*
3. *Every MP letter at a* Σ_1 *formula,* MP_θ *among them, has a derivation whose ledger is exactly* {EM}*: one EM rule and no induction on a formula with quantifiers.*

Conditions and scope.

Base and premises

The calculus of Chapter 6 with its ledger; the halting matrix θ; for part 2, the theory of Theorem 12.22. Parts 1 and 2 are proved in classical mathematics about these systems.

Conclusion

In the levels of Theorem 6.11, MP_θ sits at level 1 exactly: every derivation of it has a true EM or LPO entry, whatever its induction; one EM rule suffices; countable choice does not supply it. As a hypothesis it stays strictly below EM (Theorem 7.54).

Proof. (3) Let the letter be $\forall \vec{z}\,(\neg\neg\exists x\,\varphi \to \exists x\,\varphi)$. Assume $\neg\neg\exists x\,\varphi$. The EM rule at $\exists x\,\varphi$ gives $\exists x\,\varphi$ or $\neg\exists x\,\varphi$; the second meets the assumption and yields $\bot$, hence $\exists x\,\varphi$. Implication and universal introduction add no rule.

(1) and (2). By Lemma 7.52(1) and (2), a derivation as in (1) gives a theorem of the theory of Theorem 12.22, so it suffices to prove (2): every theorem of that theory has a modified realizer, and we show that MP_θ has none. The argument is the one that refutes $\mathrm{MP}_{\mathrm{bin}}$ there, at θ. Suppose r realizes MP_θ. Then $\{r\}(e)$ lies in the type of $\neg\neg\exists x\,\theta \to \exists x\,\theta$, and 0 lies in every type, so $m(e) = (\{\{r\}(e)\}(0))_0$ is a total computable function of e. If program e halts on input 0, then $\exists x\,\theta(e,x)$ is true; a realizer of θ makes it true, since θ has no implication (Lemma E.8), so $\neg\exists x\,\theta(e,x)$ has no realizer and every element of its type, 0 included, realizes $\neg\neg\exists x\,\theta(e,x)$. Then $\{\{r\}(e)\}(0)$ realizes $\exists x\,\theta(e,x)$, and $\theta(e,m(e))$ is true. So program e halts exactly when $\theta(e,m(e))$ holds, which a program decides by property (iii) of the halting matrix. No program decides halting. □

Kleene's reading realizes MP_θ by search, as in the proof of Theorem 6.11; the separation needs the modified reading, whose realizers must answer on every input of their type.

Theorem 7.54 (Σ_1 Markov letters reach no bit principle)**.** *Let Γ be a finite list of MP letters at Σ_1 formulas. No derivation from Γ whose EM and LPO entries are false concludes LPO_θ, WLPO_θ or LLPO_λ; quantifier-free MP rules and induction on every formula are allowed. In contrast, for every quantifier-free φ the instance LPO_φ has a derivation with no added rule.*

Conditions and scope.

Base and premises

The calculus of Chapter 6; hypotheses that are MP letters at Σ_1 formulas. Proved in classical mathematics about the calculus.

Conclusion

With Theorem 7.53, the price of MP_θ lies strictly between free and EM, and no finite list of Σ_1 Markov letters buys any of LPO_θ, WLPO_θ or LLPO_λ. At a quantifier-free matrix even LPO is free.

Proof. Every Σ_1 Markov letter has a Kleene realizer. For a true Σ_1 formula a program finds a realizer from the values of its free variables: it searches, in parallel, over witnesses for the existential quantifiers and over the steps of the subcomputations; equations are checked, a

disjunction returns the first side found true, and a bounded universal statement returns the program that, for y below the bound, runs the search for its matrix at y, which halts because the matrix is true there. The search halts exactly when the formula is true. So the letter is realized by the program that ignores the realizer of $\neg\neg\exists x\varphi$ and searches for an x at which φ is true together with its realizer: a realizer of $\neg\neg\exists x\varphi$ exists only if $\exists x\varphi$ is realized, hence true (a Σ_1 formula has no implication, so its realizers make it true), and then the search halts.

The three instances have no Kleene realizer. For LPO_θ this is the proof of Theorem 6.11: a realizer, applied to the computable realizers of $\forall x(\theta \vee \neg\theta)$, gives a tag that decides halting. For WLPO_θ the tag decides halting in the same way, as in the proof of Theorem 12.22: a tag for $\neg\forall x\neg\theta(e,x)$ comes with a realizer of that negation, so $\forall x\neg\theta(e,x)$ is false and program e halts. For LLPO_λ, the premise that $\lambda(e,\cdot)$ has at most one witness is true, and it has a realizer that does not depend on e, since a realizer of $\lambda(e,x)$ makes it true (its only negation is of θ, which a program decides); applied to e, to the computable realizer of $\forall x(\lambda \vee \neg\lambda)$ and to that realizer, a realizer of LLPO_λ gives a tag that is a total computable function $d(e)$, and if program e returns 0 the first side is false, so $d(e) \neq 0$, while if it returns 1 the second side is false, so $d(e) = 0$. That separates A_0 from A_1, which no program does.

By Lemma 7.52(3), Theorem E.9 with the letters as hypotheses, a derivation as in the theorem would give a Kleene realizer of its conclusion. For the last assertion: the proof of Theorem 6.9(3) derives $\exists x\varphi \vee \forall x\neg\varphi$ for quantifier-free φ with no added rule (the step called the alternative), and implication and universal introduction give LPO_φ. □

WLPO gives neither LPO nor Markov's principle

HAS derives LPO from WLPO and MP (Proposition 7.35), and the same derivation works at one matrix: WLPO_θ gives $\forall x\neg\theta$ or $\neg\forall x\neg\theta$, the second is $\neg\neg\exists x\theta$, and MP_θ turns it into $\exists x\theta$. So a model that separates WLPO_θ from LPO_θ must also refute MP_θ, and it must supply the halting bit, which WLPO_θ asks for, without supplying the search, which MP_θ asks for. Kleene's reading supplies the search; the modified reading of Theorem 12.22 refutes the bit. The model that works keeps the bit and bounds what can be done with it.

Theorem 7.55 (WLPO gives neither LPO nor MP at the halting matrix)**.** *Neither* $\mathrm{WLPO}_\theta \to \mathrm{LPO}_\theta$ *nor* $\mathrm{WLPO}_\theta \to \mathrm{MP}_\theta$ *has a derivation whose EM and LPO entries are false, whatever its induction; quantifier-free MP rules are allowed. Both* $\mathrm{LPO}_\theta \to \mathrm{WLPO}_\theta$ *and* $\mathrm{WLPO}_\theta \wedge \mathrm{MP}_\theta \to \mathrm{LPO}_\theta$ *have derivations with no added rule.*

Conditions and scope.

Base and premises
The calculus of Chapter 6, at every tier of induction. The non-derivations are proved in classical mathematics, with modified realizability by terms of Gödel's system T and Howard's majorizability.

Conclusion
At the halting matrix, WLPO lies strictly below LPO, and it does not reach Markov's principle.

Proof. The two derivations: a witness of $\exists x\,\theta$ refutes $\forall x\,\neg\theta$, which gives the first; the second is the argument above the theorem.

By Lemma 7.52(1), a derivation whose EM and LPO entries are false can be rebuilt with no added rule, so it suffices to treat derivations with no added rule.

The model. Kreisel's modified realizability can take its realizers among all functionals of finite type over the natural numbers, with the closed terms of Gödel's system T [70] as the programs. Every derivation of Heyting arithmetic from hypotheses gives a closed term of T that turns realizers of the hypotheses into a realizer of the conclusion, whatever functionals the realizers of the hypotheses are [6]. By Remark 6.12 a derivation with no added rule is a derivation of Heyting arithmetic, at any tier of induction. The second ingredient is Howard's *majorizability*, in Bezem's strong form [207, 208, 56]: a number a^* majorizes a when $a^* \geq a$, and a functional f^* majorizes f when, for all arguments, a^* majorizing a gives $f^*(a^*)$ majorizing $f(a)$ and $f^*(a)$. Every closed term of T is majorized by a closed term of T, and a functional majorized by a computable one applied to majorized arguments gives a value majorized by the computable value.

The check for LPO. Suppose a derivation gives a closed term R realizing $\mathrm{WLPO}_\theta \to \mathrm{LPO}_\theta$. Let $h(e)$ be 1 if program e halts on input 0 and 0 otherwise; h is not computable. The functional W that, at e and any decision realizer, returns the tag $h(e)$ with the zero realizer, realizes WLPO_θ: if $h(e) = 0$, every $\theta(e, x)$ is false, hence unrealized, so the zero realizes $\forall x\,\neg\theta(e, x)$; if $h(e) = 1$, that formula is false and unrealized, so the zero realizes its negation. And W is majorized by the computable functional that returns the tag 1. Property (iii) of the halting matrix gives a computable functional D with $D(e)$ realizing $\forall x(\theta(e, x) \vee \neg\theta(e, x))$, with a computable majorant. Now $R(W)(e)(D(e))$ realizes $\exists x\,\theta(e, x) \vee \forall x\,\neg\theta(e, x)$. If program e halts, the second side is false and has no realizer, so the realizer takes the first side, and its witness $x(e)$ satisfies $\theta(e, x(e))$. By majorizability $x(e) \leq g(e)$ for a computable function g, obtained by applying a majorant of R to the majorants of W and D. So program e halts exactly when $\theta(e, x)$ holds for some $x \leq g(e)$, which a program decides. That is impossible.

The check for MP. With a closed term R realizing $\mathrm{WLPO}_\theta \to \mathrm{MP}_\theta$, apply $R(W)(e)$ to the zero. If program e halts, the zero realizes $\neg\neg\exists x\,\theta(e, x)$, as in the proof of Theorem 7.53, and the output's witness is a halting time, again bounded by a computable function of e. □

The same model treats every principle whose computational content is a bit.

Theorem 7.56 (Bit letters give neither MP nor LPO)**.** *Let Γ be a finite list of LLPO letters and WLPO letters, at any formulas. No derivation from Γ whose EM and LPO entries are false concludes MP_θ or LPO_θ; quantifier-free MP rules and induction on every formula are allowed. In particular LLPO letters at Σ_1 formulas give neither.*

Conditions and scope.

Base and premises

The calculus of Chapter 6; hypotheses that are LLPO or WLPO letters at arbitrary formulas. Proved in classical mathematics, in the model of Theorem 7.55.

Conclusion

However many yes-or-no answers of the LLPO and WLPO kind are supplied, at whatever formulas, they buy neither Markov's principle nor LPO at the halting matrix.

Proof. By Lemma 7.52(1) we may assume no added rule. Each letter is realized by a functional majorized by a computable one. For a WLPO letter at φ, at given values of its parameters and of e, return the tag 0 when no instance $\varphi(e, x)$ is realized, and 1 otherwise, with the zero realizer: in the first case the zero realizes $\forall x \neg\varphi$, and in the second $\forall x \neg\varphi$ has no realizer, so the zero realizes its negation. For an LLPO letter, return 0 when no $\varphi(e, x + x)$ is realized and 1 otherwise. In the second case some $\varphi(e, x + x)$ is realized; a realizer of the premise that φ has at most one witness turns realizers of $\varphi(e, x + x)$ and $\varphi(e, S(y + y))$ into a realizer of the false equation $x + x = S(y + y)$, so no $\varphi(e, S(y + y))$ is realized and the zero realizes the second side. Each functional returns a bit and zeros, so it is majorized by the functional returning 1 and zeros. The checks of the proof of Theorem 7.55 then bound the witness of MP_θ or of LPO_θ by a computable function of e when program e halts. □

LLPO gives neither WLPO nor LPO

Kleene's reading and the reading by terms of T treat all bits alike, so neither can separate LLPO from WLPO. Lifschitz realizability can (Definition 12.68): a realizer of a disjunction may leave two candidates open, which is exactly what LLPO needs and WLPO forbids.

Theorem 7.57 (LLPO letters give neither WLPO nor LPO)**.** *Let Γ be a finite list of LLPO letters at any formulas and MP letters at Σ_1 formulas. No derivation from Γ whose EM and LPO entries are false concludes WLPO_θ or LPO_θ; quantifier-free MP rules and induction on every formula are allowed. In particular $\mathrm{LLPO}_\lambda \to \mathrm{WLPO}_\theta$ has no derivation whose EM, LPO and MP entries are false.*

Conditions and scope.

Base and premises

The calculus of Chapter 6; hypotheses that are LLPO letters at arbitrary formulas and Markov letters at Σ_1 formulas. Proved in classical mathematics with Lifschitz realizability.

Conclusion

At the halting matrix, LLPO_λ does not reach WLPO_θ or LPO_θ, and adding Markov letters at Σ_1 formulas does not close the gap.

Proof. By Lemma 7.52 the derivation becomes one in the theory of Theorem 12.22, from the same letters, and Lemma 12.69 carries Lifschitz realizers of the letters to one of the conclusion.

The letters are realized. A Σ_1 Markov letter is realized by the search of Theorem 7.54, its answer returned as a singleton. For an LLPO letter at φ, take the given realizer of $\forall x(\varphi \vee \neg\varphi)$ and return the number whose two candidates are the answers "no even witness" and "no odd witness", each with the zero realizer, and whose program searches, dovetailing over x and over steps, for an x at which every number up to the bound of the decision realizer at x whose tag is not 0 has been struck, then strikes the even answer if x is even and the odd one if x is odd. Such an x is found exactly when $\varphi(e, x)$ is realized: the decision realizer always keeps a

candidate, every candidate is correct, and when $\varphi(e, x)$ is realized its negation is not, so every number with a nonzero tag is struck. By the premise there is at most one such x, so at most one candidate is struck, and every candidate left is correct.

WLPO and LPO at θ are not. Suppose $\{r\}(e)$, applied to the computable decision realizer $D(e)$, realizes $\forall x \neg\theta(e, x) \vee \neg\forall x \neg\theta(e, x)$ for every e; it is a pair (p, b) whose candidates $k \leq b$ all carry the correct tag. If program e halts, every candidate has the tag 1, so every $k \leq b$ with tag 0 is struck after finitely many steps; if it never halts, every $k \leq b$ with tag 1 is struck. Since there is a candidate, exactly one of these two finite waits ends. Running both decides halting, which no program does. The argument for LPO_θ is the same, with the tags of $\exists x\,\theta$ and $\forall x \neg\theta$. □

Two disjunctions that cost something

A disjunction of two instances is weaker than either. Is it free? Not here.

Theorem 7.58 (Two meets are not free)**.** *Neither* $\mathrm{WLPO}_\theta \vee \mathrm{MP}_\theta$ *nor* $\mathrm{LLPO}_\lambda \vee \mathrm{MP}_\theta$ *has a derivation whose EM and LPO entries are false, while each follows from each of its disjuncts with an empty ledger. Read in the two-sorted language,* LLPO_λ *is not a theorem of the theory of Theorem 12.22.*

Proof. Disjunction introduction gives each disjunction from each disjunct. By Lemma 7.52 a derivation with false EM and LPO entries gives a theorem of the theory of Theorem 12.22, which has a modified realizer, and a modified realizer of a disjunction carries a tag and a realizer of the side it names. MP_θ has no modified realizer (Theorem 7.53). WLPO_θ has none: its tag would decide halting, by the argument of Theorem 12.22. LLPO_λ has none: a realizer of $\lambda(e, x)$ makes it true (its only negation is of a formula decided by a program), so the premise that $\lambda(e, \cdot)$ has at most one witness is realized by the zero. A realizer of LLPO_λ, applied to e, to the computable realizer of $\forall x(\lambda(e, x) \vee \neg\lambda(e, x))$ and then to the zero, gives a tag, and that tag would be a total computable function of e separating A_0 from A_1. So neither disjunction has a modified realizer, and LLPO_λ is not a theorem of that theory. □

The disjunctive Markov principle at the halting matrix

$\mathrm{MP}_\theta^\vee$ says: if two programs cannot both run forever, then one of them cannot fail to halt. It asks for a choice between two programs, not for a halting time.

Theorem 7.59 (The price of the disjunctive Markov principle)**.**

1. $\mathrm{MP}_\theta^\vee$ *has no derivation whose EM and LPO entries are false, and read in the two-sorted language it is not a theorem of the theory of Theorem 12.22.*
2. *It has a derivation whose ledger is exactly* {EM}*, and one whose EM and MP entries are false; the second uses LPO rules and induction on formulas with quantifiers.*
3. *It has a derivation with an empty ledger from one MP letter, the letter at the* Σ_1 *formula* $\theta(e_1, x) \vee \theta(e_2, x)$*.*
4. *From* $\mathrm{MP}_\theta^\vee$ *and MP letters at* Σ_1 *formulas, no derivation whose EM and LPO entries are false concludes* WLPO_θ*,* LLPO_λ *or* LPO_θ*; with LLPO letters at any formulas added, none concludes* WLPO_θ *or* LPO_θ*.*

5. *From* $\mathrm{MP}^{\vee}_{\theta}$*, LLPO letters and WLPO letters at any formulas, no derivation whose EM and LPO entries are false concludes* MP_{θ} *or* LPO_{θ}*.*

Conditions and scope.

Base and premises

The calculus of Chapter 6; the halting matrix; the hypotheses named in each part. The non-derivations are proved in classical mathematics.

Conclusion

At the halting matrix, $\mathrm{MP}^{\vee}$ is not free, lies strictly below the Σ_1 Markov letter at $\theta(e_1, x) \vee \theta(e_2, x)$ (parts 3 and 5; that letter gives MP_{θ} when $e_1 = e_2$), is supplied by EM or by LPO, and with Markov letters buys none of the bit instances. Whether MP_{θ} gives it is open (Remark 7.63(4)).

Proof. (3) Assume $\neg(\neg\exists x\,\theta(e_1, x) \wedge \neg\exists x\,\theta(e_2, x))$. A witness of $\exists x(\theta(e_1, x) \vee \theta(e_2, x))$ would give one of the two existentials, so the negation of that existential gives both negations, against the assumption: we have $\neg\neg\exists x(\theta(e_1, x) \vee \theta(e_2, x))$. The MP letter at the Σ_1 formula $\theta(e_1, x) \vee \theta(e_2, x)$, with e_1, e_2 as its parameters, gives the existential; a case split on its witness gives $\exists x\,\theta(e_1, x)$ or $\exists x\,\theta(e_2, x)$, and each gives its double negation. No added rule and no induction is used.

(2) The EM rule at $\exists x\,\theta(e_1, x)$ gives the first side, or $\neg\exists x\,\theta(e_1, x)$; in that case the assumption gives $\neg\neg\exists x\,\theta(e_2, x)$. For the second derivation, replace the EM rule by its derivation from LPO rules (Theorem 6.10), which uses induction on formulas with quantifiers.

(1) As in Theorem 7.58, it suffices that $\mathrm{MP}^{\vee}_{\theta}$ has no modified realizer. A realizer, applied to e_1, e_2 and to the zero, gives a total computable tag $t(e_1, e_2)$. If at least one of the two programs halts, the premise is true and has no realizable negation, so the zero realizes it, and the tag names a side whose double negation is realized, that is, a program that halts. So t would be a total computable selector that picks a halting program from every pair of which one halts. None exists. By Kleene's recursion theorem for two programs [19, Chapter XII] there are programs a and b such that a computes $t(a, b)$ and halts if it is not 0, and loops otherwise, while b computes $t(a, b)$ and halts if it is 0, and loops otherwise. Exactly one of a, b halts, and $t(a, b)$ names the other.

(4) $\mathrm{MP}^{\vee}_{\theta}$ has a Kleene realizer: on e_1, e_2 and a realizer of the premise, run the two programs in parallel and return the side of the first that halts. A realizer of the premise exists only if one of them halts, so the search ends there. With Theorem 7.54 this gives the first claim. For the second, the same parallel search, returned as a singleton, is a Lifschitz realizer, and Theorem 7.57 applies. Both claims also follow from part 3, by cutting in its derivation for each hypothesis $\mathrm{MP}^{\vee}_{\theta}$.

(5) $\mathrm{MP}^{\vee}_{\theta}$ is realized in the model of Theorem 7.55 by a bit: at e_1, e_2 return the tag 0 when some $\theta(e_1, x)$ is realized and 1 otherwise, with the zero realizer; when the premise is realized and no $\theta(e_1, x)$ is, some $\theta(e_2, x)$ is, and the zero realizes the second side. This functional is majorized by the one returning 1, and Theorem 7.56 realizes the letters; the checks of Theorem 7.55 finish. □

Table 7.5 collects the non-derivations at the halting matrix. Each matches a separation of the schemas over HAS (Table 7.4), here proved one sentence at a time and against every induction tier of the calculus. Of the positive steps of Table 7.4, only those listed under the table are proved at the instances. The others are not settled here: Remark 7.63 lists $\mathrm{LPO}_\theta \to \mathrm{MP}_\theta$, $\mathrm{WLPO}_\theta \to \mathrm{MP}^\vee_\theta$ and $\mathrm{MP}_\theta \to \mathrm{MP}^\vee_\theta$, and neither $\mathrm{WLPO}_\theta \to \mathrm{LLPO}_\lambda$ nor $\mathrm{LLPO}_\lambda \to \mathrm{MP}^\vee_\theta$ is derived in this book.

From (with false EM and LPO entries)	Not derived	Model
nothing	MP_θ, $\mathrm{MP}^\vee_\theta$	modified realizability
Σ_1 MP letters, $\mathrm{MP}^\vee_\theta$	LPO_θ, WLPO_θ, LLPO_λ	Kleene realizability
WLPO_θ	LPO_θ, MP_θ	T and majorants
LLPO and WLPO letters, $\mathrm{MP}^\vee_\theta$	MP_θ, LPO_θ	T and majorants
LLPO letters, Σ_1 MP letters, $\mathrm{MP}^\vee_\theta$	WLPO_θ, LPO_θ	Lifschitz realizability
nothing	$\mathrm{WLPO}_\theta \vee \mathrm{MP}_\theta$, $\mathrm{LLPO}_\lambda \vee \mathrm{MP}_\theta$	modified realizability

Table 7.5. What does not follow at the halting matrix, in the calculus of Chapter 6 with quantifier-free MP rules and induction on every formula (Theorems 7.53–7.59). The positive steps proved at the instances: one EM rule gives MP_θ and $\mathrm{MP}^\vee_\theta$; one Σ_1 Markov letter gives $\mathrm{MP}^\vee_\theta$; LPO_θ gives WLPO_θ; WLPO_θ and MP_θ together give LPO_θ.

The successor step costs what equality costs

The empty ledger allows induction on quantifier-free formulas only. Proposition 6.14 showed that it does not decide equality. The induction that would decide it runs on y for the formula $x = y \vee \neg(x = y)$; its base $x = 0 \vee \neg(x = 0)$ is free, by a quantifier-free induction on x. What does its step cost?

Proposition 7.60 (The price of the successor step)**.** *Let* $\sigma(x)$ *be* $\forall y\big((x = y \vee \neg(x = y)) \to (x = S(y) \vee \neg(x = S(y)))\big)$.

1. *$\sigma(x)$ has no derivation whose EM, MP and induction entries are false, even with LPO rules permitted.*
2. *It has a derivation whose ledger is exactly* {EM} *and one whose ledger is exactly* {MP}*, neither containing an induction rule; and one with no added rule, which uses induction on a formula with quantifiers.*

In particular the posit-free derivation with quantifier-free induction that one might expect for this one-step formula does not exist.

Proof. (2) Each conclusion $x = S(y) \vee \neg(x = S(y))$ is an EM instance, and the remark after Theorem 6.10 derives it from one MP rule at a quantifier-free body, with no induction. With induction on every formula, (6.9) gives it with no added rule.

(1) Suppose such a derivation of $\sigma(x)$. Induction on y for the quantifier-free formula $x = y \vee \neg(x = y)$ takes the base $x = 0 \vee \neg(x = 0)$, which quantifier-free induction on x derives by the split into zero and successor, and the step $\sigma(x)$ with its universal quantifier eliminated; it

gives $\forall y(x = y \vee \neg(x = y))$, and universal introduction gives $\forall x\,\forall y(x = y \vee \neg(x = y))$. The added inductions are quantifier-free, so the ledger keeps false EM, MP and induction entries. Proposition 6.14 shows that no such derivation exists. □

Weak excluded middle for Σ^0_2 and Π^0_2 statements, and the shift

Back to schemas over HAS. Chapter 28 shows that WLPO gives neither Π^0_2 weak excluded middle nor its shift (Theorem 28.32). Where do these principles sit, and what of the corresponding statement one quantifier lower?

Definition 7.61 (Three more principles over HAS)**.** With $\langle N, m\rangle = (N + m)\cdot(N + m) + m$, $\mathrm{Row}(X, N)$ and $\mathrm{All}(X)$ as in Definition 28.31, put $\mathrm{Gap}(X) :\equiv \exists N\forall m\,\neg\langle N, m\rangle \in X$ (some row is empty) and

$$\Sigma^0_2\text{-WEM} :\equiv \forall X\big(\mathrm{Dec}(X) \to \neg\mathrm{Gap}(X) \vee \neg\neg\mathrm{Gap}(X)\big),$$
$$\mathrm{DNS}_{\mathrm{ar}} :\equiv \forall X\big(\mathrm{Dec}(X) \to \forall N\neg\neg\mathrm{Row}(X, N) \to \neg\neg\mathrm{All}(X)\big).$$

$\mathrm{DNS}_{\mathrm{ar}}$ is the double-negation shift for the rows of a decidable matrix: if no row can fail to have an entry, then it cannot fail that every row has one.

Theorem 7.62 (The order of the weak arithmetic principles)**.** *Over intuitionistic second-order arithmetic:*

1. *HAS derives* $\mathrm{WLPO} \leftrightarrow \Sigma^0_2\text{-WEM}$, $\Pi^0_2\text{-WEM} \to \mathrm{WLPO}$ *and* $\mathrm{MP} \to \mathrm{DNS}_{\mathrm{ar}}$.
2. *HAS derives none of* $\mathrm{WLPO} \to \mathrm{DNS}_{\mathrm{ar}}$, $\Pi^0_2\text{-WEM} \to \mathrm{DNS}_{\mathrm{ar}}$ *and* $\mathrm{WLPO} \wedge \mathrm{DNS}_{\mathrm{ar}} \to \mathrm{LPO}$.
3. *Consequently*

$$\mathrm{WLPO} \;<\; \Pi^0_2\text{-WEM} \;<\; \mathrm{WLPO} \wedge \mathrm{DNS}_{\mathrm{ar}} \;<\; \mathrm{LPO},$$

where each principle derives the one before it and none derives the one after it; $\mathrm{DNS}_{\mathrm{ar}}$ *is comparable with neither WLPO nor* Π^0_2*-WEM; MP derives* $\mathrm{DNS}_{\mathrm{ar}}$ *but not conversely, and is comparable with none of WLPO,* Π^0_2*-WEM and* $\mathrm{WLPO} \wedge \mathrm{DNS}_{\mathrm{ar}}$*; and* $\Pi^0_2\text{-WEM} \wedge \mathrm{DNS}_{\mathrm{ar}}$ *does not derive LPO.*

Conditions and scope.

Base and premises
HAS (Definition 7.14) and its systems of truth values (Definition 7.19). The non-derivations are proved in classical mathematics with the axiom of choice.

Conclusion
Σ^0_2 weak excluded middle is WLPO; Π^0_2 weak excluded middle lies strictly between WLPO and LPO; the shift for rows lies below MP and is independent of WLPO.

Proof. (1) *WLPO gives* Σ^0_2*-WEM.* Let X be decidable, let Y_N be the set of m with $\langle N, m\rangle \in X$, and let comprehension give the set B of the N with $\mathrm{Z}(Y_N)$. WLPO at each Y_N makes B decidable, and $\mathrm{Gap}(X)$ is $\mathrm{E}(B)$. WLPO at B gives $\mathrm{Z}(B)$, hence $\neg\mathrm{Gap}(X)$, or $\neg\mathrm{Z}(B)$, which is $\neg\neg\mathrm{Gap}(X)$.

Σ^0_2-WEM gives WLPO. Let X be decidable and let M be the set of n such that some $k < n$ lies in X; it is decidable by a bounded search. Then $\mathrm{Z}(X) \leftrightarrow \mathrm{Gap}(M)$. If X is empty, so is M, and row 0 is empty. If row N of M is empty and $k \in X$, then $\langle N, S(k)\rangle \in M$, because $k < S(k) \le \langle N, S(k)\rangle$. So $\neg\mathrm{Gap}(M) \vee \neg\neg\mathrm{Gap}(M)$ gives $\neg\mathrm{Z}(X) \vee \neg\neg\mathrm{Z}(X)$, and $\neg\neg\mathrm{Z}(X)$ gives $\mathrm{Z}(X)$.

Π^0_2-WEM gives WLPO. Use the complement M' of M, the set of n with no $k < n$ in X. Then $\mathrm{Z}(X) \leftrightarrow \mathrm{All}(M')$. If X is empty, $\langle N, 0\rangle \in M'$ for every N. If every row of M' has an entry and $k \in X$, row $S(k)$ has an entry $\langle S(k), m\rangle$, and $k < \langle S(k), m\rangle$: with $s = S(k) + m$ we have $S(k) \le s \le s \cdot s \le s \cdot s + m$. So the entry is not in M', a contradiction. Weak excluded middle for $\mathrm{All}(M')$ then decides $\mathrm{Z}(X)$ weakly, as before.

MP gives $\mathrm{DNS}_{\mathrm{ar}}$. Each Y_N is decidable and, by hypothesis, not empty up to double negation, so MP gives a member; hence $\mathrm{All}(X)$, and its double negation. This is the first step of the proof of Proposition 28.33.

(2) By Lemma 7.21 each claim needs one system of truth values.

WLPO without the shift. HAS derives $\mathrm{DNS}_{\mathrm{ar}} \to$ Shift, since the shift's conclusion $\neg\mathrm{All}(X) \vee \neg\neg\mathrm{All}(X)$ is a weakening of $\neg\neg\mathrm{All}(X)$. The bounding frame of Theorem 28.32 gives WLPO the value $\top$ and Shift another value, so $\mathrm{DNS}_{\mathrm{ar}}$ has a value other than $\top$ there.

Π^0_2-WEM without the shift: a new frame. Order the finite bit strings by extension, and call a set of strings *open* when it contains every extension of each of its members; the cone $[s]$ of a string s is the set of its extensions. An open set is *regular* when it equals the set of strings all of whose extensions have an extension in it. The regular open sets form a Boolean algebra R, with meet the intersection and complement a^{c} the set of strings with no extension in a; the join of a and b is the complement of $a^{\mathrm{c}} \cap b^{\mathrm{c}}$. The *ideals* of R, the sets of elements closed under finite joins and under passing to smaller elements, form a system of truth values I, ordered by inclusion: meet is intersection, the join of a family is the ideal it generates, and the distributive law (7.12) holds, because an element of $a \wedge \bigvee_j b_j$ lies below a finite join of elements of the b_j and its meet with each of them lies in a. We use three facts.

(i) *Weak excluded middle holds in I for every formula.* Let J be an ideal, and let u be the set of strings all of whose extensions have an extension in some member of J. Then u is regular, and an element of R disjoint from every member of J is disjoint from u. So the principal ideal of u^{c} lies in $\neg J$ and that of u in $\neg\neg J$; their join contains $u \vee u^{\mathrm{c}}$, the top element. Hence $\neg J \vee \neg\neg J$ is the top ideal for every J, and in particular Π^0_2-WEM has the value $\top$.

(ii) *A matrix with dense rows.* Number the bit strings as $s_0, s_1, \ldots$. Row N has the entries $c_{N,0}$, the regular set of strings incompatible with s_N, and $c_{N,k+1} = [s_N 0^k 1]$ for $k \ge 0$. Value $\langle N, m\rangle \in X$ by the principal ideal of $c_{N,m}$ and every other number by the zero ideal. A principal ideal and the principal ideal of the complement are disjoint and join to the top, so $\mathrm{Dec}(X)$ has the value $\top$. An element disjoint from every entry of row N lies in $[s_N]$ and misses every $[s_N 0^k 1]$, so it contains no cone and is empty; hence $\neg\mathrm{Row}(X, N)$ is the zero ideal and $\neg\neg\mathrm{Row}(X, N)$ is the top, for every N.

(iii) *The rows meet in zero.* The value of $\mathrm{All}(X)$ is the largest ideal below every row join. A nonzero element a in it contains a cone $[s_N]$ and lies below some finite join $c_{N,0} \vee \cdots \vee c_{N,M}$ of row N. But the cone of $s_N 0^M$ lies in a and misses each of these entries, hence their join. So $\mathrm{All}(X)$ has the zero value, $\neg\neg\mathrm{All}(X)$ is the zero ideal, and the instance of $\mathrm{DNS}_{\mathrm{ar}}$ at X has the

value $\top \Rightarrow \bot = \bot$.

WLPO with the shift, without LPO. Take the open sets of $\beta\mathbb{N}$ (Section 7.12). WLPO has the value $\top$ and LPO does not (Table 7.2). Every nonempty open set contains an isolated point $\{n\}$, an atom of the system of truth values. In such a system, $\bigwedge_N \neg\neg r_N \leq \neg\neg \bigwedge_N r_N$ for every countable family: otherwise an atom $\{n\}$ lies below every $\neg\neg r_N$ and below $\neg \bigwedge_N r_N$; an atom below $\neg\neg r_N$ meets r_N, hence lies in it, so it lies in every r_N, hence in their meet, which contradicts the last inequality. With r_N the value of $\mathrm{Row}(X, N)$, the value of $\mathrm{All}(X)$ is $\bigwedge_N r_N$, so the hypotheses of every instance of $\mathrm{DNS}_{\mathrm{ar}}$ lie below its conclusion, and $\mathrm{DNS}_{\mathrm{ar}}$ has the value $\top$.

(3) The derivations: (1); LPO $\to$ WLPO and LPO $\to$ MP (Proposition 7.35) with MP $\to \mathrm{DNS}_{\mathrm{ar}}$; and WLPO $\wedge\, \mathrm{DNS}_{\mathrm{ar}} \to \Pi^0_2$-WEM, from WLPO $\to$ (Shift $\to \Pi^0_2$-WEM) (Proposition 28.33) and $\mathrm{DNS}_{\mathrm{ar}} \to$ Shift. The steps are strict: the bounding frame validates WLPO and not Π^0_2-WEM (Theorem 28.32); the frame I validates Π^0_2-WEM and not $\mathrm{DNS}_{\mathrm{ar}}$; and $\beta\mathbb{N}$ validates WLPO $\wedge$ $\mathrm{DNS}_{\mathrm{ar}}$, and with it Π^0_2-WEM, and not LPO. The one-point compactification of an uncountable set validates MP, hence $\mathrm{DNS}_{\mathrm{ar}}$, and not WLPO, hence not Π^0_2-WEM; $\beta\mathbb{N}$ validates WLPO, Π^0_2-WEM and $\mathrm{DNS}_{\mathrm{ar}}$ and not MP. These, with (2), give every incomparability claimed. □

Remark 7.63 (Open questions at the halting matrix)**.** Four questions about the instances are not settled here. Each asks for a derivation with no added rule, that is, with false EM, LPO and MP entries, at any tier of induction.

1. *Decidability of the halting matrix.* Does $\forall e\, \forall x\, (\theta(e, x) \vee \neg\theta(e, x))$ have a derivation with no added rule? Its truth is decided by a program (Lemma E.11(iii)); the question is whether the calculus derives that decision.

2. *LPO and Markov's principle at θ.* Does $\mathrm{LPO}_\theta \to \mathrm{MP}_\theta$ have a derivation with no added rule? A positive answer to question 1 gives one, by a derivation with no added rule and no induction: under the decision of θ, the LPO instance either produces a witness or refutes $\neg\neg\exists x\, \theta$. With it, LPO_θ would be exactly $\mathrm{WLPO}_\theta \wedge \mathrm{MP}_\theta$.

3. *WLPO and the disjunctive Markov principle at θ.* Does $\mathrm{WLPO}_\theta \to \mathrm{MP}^\vee_\theta$ have a derivation with no added rule? A positive answer to question 1 gives one, with no added rule and no induction.

4. *Markov's principle and its disjunctive half at θ.* Does $\mathrm{MP}_\theta \to \mathrm{MP}^\vee_\theta$ have a derivation with no added rule? If both this and question 3 have positive answers, $\mathrm{MP}^\vee_\theta$ follows from each of WLPO_θ and MP_θ while it is not free (Theorem 7.59), so it lies in the meet of the two.

The calculus does derive $\mathrm{MP}^\vee_\theta$ from the single Σ_1 Markov letter at $\theta(e_1, x) \vee \theta(e_2, x)$ (Theorem 7.59(3)); question 4 asks for it from the letter at θ itself, which is a different sentence.

Questions to work through

1. In Lemma 7.4, identify each use of addition compatibility and each use of transitivity. Why would cancelling a common addition have been unjustified?

2. In the constructive classification, which supplied value bounds the search for the period? Which bounds the search for the index? At what point is the pairwise decision procedure used?

3. In (7.4), explain why an individual pair can be decided by finite search even though $\exists n\, \theta(n)$ may not yet have a witness.

4. Trace both branches of the EM reversal. Which branch supplies P, and which supplies $\neg P$? Why was no initial decision of P needed to define the congruence?

5. In the WLPO part of Theorem 7.9, show that $\sim_\theta$ is equality exactly when θ has no witness. In the forward direction, where are the decidability of equality and of order between two counts used, and why do those uses need no added principle?

6. Show that a decidable congruence has at most one first collision, and that the event of γ for $\sim_{i,p}$ sits at $2(i + p) + (i \bmod 2)$. Why does the sequence need to record only the parity of the smaller element?

7. Show directly that $\sim_{2,3}$ is not cancellative. Describe the congruence on $\mathbb{Z}_\delta$ whose restriction is $\sim_{0,3}$, using the integer reading of its records.

8. In the reversal for the period parity in Corollary 7.13, let α have its only event at $n = 3$. Which records does $\approx_\alpha$ relate to $[0, 0]$, and what is its period? Which alternative of LLPO does the corollary's assertion return? Why would the index parity of Theorem 7.9 be useless here?

9. In the open sets of $\beta\mathbb{N}$, take $A_n = \{n\}$. Compute the values of $\exists n\, \alpha(n) = 1$, of $\forall n\, \alpha(n) = 0$ and of its negation. Read off from (7.13) why Markov's principle and LPO fail at this sequence while WLPO holds.

10. Show that in a chain every sequence decided at the value $\top$ has only the values $\bot$ and $\top$. Why does this make LPO hold, and which value refutes excluded middle in the chain $0 < 1 < 2 < \cdots < \infty$?

11. In the convergent sequence $\widehat{\mathbb{N}}$, take $A_n = \{n\}$. Compute the values of $\mathrm{Z}(X)$, $\mathrm{E}(X)$, $\mathrm{Ev}(X)$ and $\mathrm{Od}(X)$, and say why MP, LLPO and $\mathrm{MP}^\vee$ fail at this sequence. Then take $B_n = \{n\}$ for even n and $B_n = \emptyset$ for odd n, and show that the instance of $\mathrm{W}(X)$ at B excludes ∞. Why does WMP hold at this sequence?

12. In the model X_2 of Proposition 7.46, take $A_n = \{n\}$. Show that ∞ lies in the value of $\mathrm{C}_i^3(X)$ for some $i < 3$. Can the construction tell you which i? Why is that no obstacle to the value of LLPO_3 being $\top$?

13. In the cone, take $A_0 = \{a\}$ and $A_n = \emptyset$ for $n \geq 1$. Compute the values of $\mathrm{Dec}(X)$, $\mathrm{E}(X)$ and $\mathrm{Z}(X)$, and check that the instance of LPO at this sequence is $\top$ while the instance of WEM is not. What does the focal point show about the difference between the two principles?

Chapter 8

Changing the representation

A number can be written as a numeral or as a list of identical marks. Changing between these descriptions should not change the answer to an arithmetic question. But that expectation needs a proof. The translations must preserve the operations, and quantified statements must range over corresponding objects in both descriptions.

We will construct the translations and prove what they preserve. There are three distinct objects to keep in view: the arithmetic carrier, the interpretation of a formula, and the finite derivation of that formula. A theorem about one does not automatically settle a question about the others. We therefore take them in turn. We change the carrier and follow every formula through the change, measure what survives when a translation runs in only one direction, and compare the derivations themselves with those of an independently stated proof system, Peano arithmetic, and with its intuitionistic counterpart, Heyting arithmetic.

8.1 An interpretation is fixed by its operations

Chapter 2 established a unique reading of the start-and-step orbit into any pointed step system. The terms of Chapter 5 have a larger signature: variables, zero, successor, addition and multiplication. The same structural method applies to this explicitly stated signature.

Definition 8.1 (A realization of the term signature)**.** A term realization consists of a carrier A, an assigned element v_n for every variable position, an element 0_A, and operations

$$S_A : A \longrightarrow A, \qquad +_A, \cdot_A : A \times A \longrightarrow A.$$

An interpretation preserving the signature sends each variable x_n to v_n, sends 0 to 0_A, and commutes with the three operations. No arithmetic law is assumed merely by giving these operations.

Theorem 8.2 (Unique interpretation of terms)**.** *For every term realization, exactly one interpretation preserves the specified signature. It is the structural evaluation*

$$\begin{aligned} F_A(x_n) &= v_n, & F_A(0) &= 0_A, \\ F_A(S(t)) &= S_A(F_A(t)), \\ F_A(t+u) &= F_A(t) +_A F_A(u), \\ F_A(t \cdot u) &= F_A(t) \cdot_A F_A(u). \end{aligned}$$

Proof. The displayed recursion gives an interpretation because every term is finite and each recursive call concerns an immediate constituent term. It preserves the signature by its defining clauses.

For uniqueness, let G be any signature-preserving interpretation and induct on the term. At a variable and at zero, preservation specifies the answer. At a successor, preservation gives $G(S(t)) = S_A(G(t))$, and the induction hypothesis replaces $G(t)$ by $F_A(t)$. At a sum or product, preservation expresses the result using the two interpreted constituents; the two induction hypotheses replace them by their F_A values. Thus $G(t) = F_A(t)$ in every constructor case. □

This theorem concerns expressions. It does not make every realization a model of arithmetic. To interpret the arithmetic rules soundly, the operations must satisfy those rules. The next construction supplies a concrete realization for which the necessary agreement is proved.

8.2 Arithmetic on lists of one mark

Let a list contain repetitions of a single mark $\bullet$. The empty list is written $[\,]$; placing a mark at the front of u is written $\bullet :: u$. There is no additional information in a list entry: every entry is the same mark.

Definition 8.3 (The list arithmetic)**.** Take finite lists of the single mark as the carrier. Define zero to be the empty list, successor to prepend one mark, and addition to be concatenation, written $u +_L v = u +\!\!+ v$. Define multiplication by recursion on the second list:

$$u \cdot_L [\,] = [\,],$$
$$u \cdot_L (\bullet :: v) = (u \cdot_L v) +_L u.$$

These operations act directly on lists. Concatenation joins the two lists; multiplication appends one copy of the first list for every mark in the second. We can therefore compare this construction with ordinary arithmetic after both descriptions have been specified.

Write $L(u)$ for the length of a list. Write $R(n)$ for the list of n repetitions of the mark.

Theorem 8.4 (The two translations are inverse)**.** *For every natural number n and every list u of the single mark,*

$$L(R(n)) = n, \qquad R(L(u)) = u.$$

Length preserves zero, successor, addition and multiplication:

$$L([\,]) = 0, \qquad L(\bullet :: u) = L(u) + 1, \tag{8.1}$$
$$L(u +_L v) = L(u) + L(v), \qquad L(u \cdot_L v) = L(u)L(v). \tag{8.2}$$

In particular, lists of this one mark are equal exactly when their lengths are equal.

Proof. Induction on n proves $L(R(n)) = n$: the empty repetition has length zero, and one further repetition increases length by one. Induction on u proves the other round trip. The empty list is immediate. A nonempty list has the form $\bullet :: v$, since there is only one mark. Replicating its length produces $\bullet :: R(L(v))$, which is $\bullet :: v$ by the induction hypothesis.

The start and successor equations are the definition of length. For concatenation, induct on the first list. The empty first list leaves v, while removing one mark from a nonempty first list reduces both sides by one and leaves the induction hypothesis.

For multiplication, induct on v. At the empty list, both sides are zero. At $\bullet :: v$, the definition and the addition identity give

$$\begin{aligned} L(u \cdot_L (\bullet :: v)) &= L(u \cdot_L v) + L(u) \\ &= L(u)L(v) + L(u) \\ &= L(u)(L(v) + 1). \end{aligned}$$

This is the required successor case.

Finally, equal lists have equal lengths. If their lengths are equal, their replications are equal, and the second round-trip identity gives equality of the original lists. □

Both carriers also satisfy the faithful-generation conditions of Chapter 2. A nonempty list is not empty, equality of two lists with a mark prepended gives equality of their tails, and list induction reaches every list. The corresponding facts for natural numbers were already used in Part I. Thus the translation preserves the starting point and step as well as the recursive arithmetic operations.

The one-mark condition is essential to this example. Its role is visible in the round-trip proof: knowing the length reconstructs every entry because every entry has already been specified to be the same mark.

8.3 Terms agree under corresponding assignments

A formula may contain variables, so carrier translations alone are insufficient. We must translate the assignments to those variables. Let ρ_N assign natural numbers and let ρ_L assign lists. Call the assignments *related* when

$$L(\rho_L(k)) = \rho_N(k) \qquad \text{for every position } k. \tag{8.3}$$

Evaluate a term on lists by the same recursive clauses as before, using the list operations just defined. Write this value as $E^L_{\rho_L}(t)$, and the natural-number value as $E^N_{\rho_N}(t)$. Figure 8.1 shows the two routes.

Lemma 8.5 (Evaluation commutes with the translation)**.** *For related assignments and every arithmetic term t,*

$$L(E^L_{\rho_L}(t)) = E^N_{\rho_N}(t).$$

Proof. Induct on t. The variable case is (8.3). The zero case is $L([\,]) = 0$. The successor case follows from the length identity for successor and the induction hypothesis. For a sum, length of concatenation is the sum of lengths; apply the two induction hypotheses. For a product, use the multiplication identity in (8.2) and the same two induction hypotheses. These are all the term constructors. □

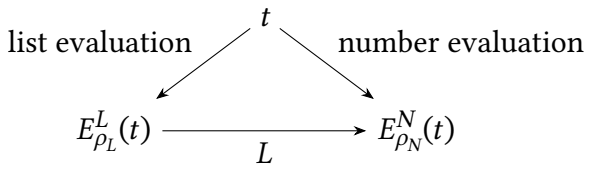


Figure 8.1. Related assignments make the two routes agree. The equation is proved for every term by following its construction.

8.4 Truth is preserved in both directions

Interpret formulas on lists directly. An equality says that its list values are equal. The propositional connectives keep their constructive readings. A universal quantifier ranges over every finite list of the mark, and an existential quantifier supplies one such list as a witness. Write this satisfaction relation as $\rho_L \vDash_L \varphi$.

Theorem 8.6 (Invariance of every arithmetic formula)**.** *If ρ_N and ρ_L are related, then for every formula of the language in Chapter 5,*

$$\rho_L \vDash_L \varphi \quad \Longleftrightarrow \quad \rho_N \vDash \varphi. \tag{8.4}$$

Proof. Induct on the formula, with the two assignments arbitrary. For an equality of terms, equal list values give equal lengths, and Lemma 8.5 gives equal number values. Conversely, equal number values give equal lengths, which give equal lists by Theorem 8.4. Falsehood is false on both sides.

For conjunction and disjunction, transport each component by its induction hypothesis. This preserves both the pair of proofs and the chosen side of a disjunction. For an implication, both directions of each component equivalence are needed. Given the implication in one interpretation, transport an antecedent from the other, apply the implication, and transport the consequent back. The reverse direction uses the same procedure with the roles interchanged.

For a universal formula, suppose first that its list interpretation holds. Choose any natural number n and use $R(n)$ for the new list variable. The extended assignments are related because $L(R(n)) = n$. The induction hypothesis transports the body to the number interpretation at n. Since n was arbitrary, the universal number interpretation holds.

Conversely, suppose the universal number interpretation holds and choose any list u. Use $L(u)$ for the new number variable. The extended assignments are related, and the induction hypothesis transports the body at $L(u)$ to the list body at u. This proves the list universal statement.

For an existential formula, a list witness u becomes the number witness $L(u)$; a number witness n becomes the list witness $R(n)$. In each direction the extended assignments are related, and the induction hypothesis transports the witness's property. This covers every formula constructor. □

The quantifier cases are why a one-way preservation result would have been insufficient. We needed every number to have a corresponding list and every list to have a corresponding

number, with the assigned values related. The round trips supply exactly that information. Section 8.5 measures what survives when only the first of them is available.

Corollary 8.7 (Validity is independent of these two presentations)**.** *A formula holds under every number assignment if and only if it holds under every list assignment. Every closed derivation with an empty ledger has a conclusion true in both presentations.*

Proof. For a given list assignment, take the lengths at every position to obtain a related number assignment. For a given number assignment, replicate each assigned number to obtain a related list assignment. Apply Theorem 8.6 in each direction. For the last assertion, soundness gives validity in the number interpretation, and the first assertion gives validity in the list interpretation. □

8.5 What survives a one-way translation

The list carrier came with translations in both directions. Many structures come with only one. Take a structure M for the arithmetic language, as in Definition 5.7: a carrier with an element 0_M and operations S_M, $+_M$ and $\cdot_M$, in which formulas are read by the clauses of Chapter 5 with quantifiers ranging over the whole carrier. Write $M, \sigma \vDash \varphi$ when an assignment σ of elements of M satisfies φ. Suppose that the six arithmetic rules of Chapter 5 hold in M as facts about its elements:

$$S_M(x) \neq 0_M, \qquad S_M(x) = S_M(y) \Rightarrow x = y,$$
$$x +_M 0_M = x, \quad x +_M S_M(y) = S_M(x +_M y), \quad x \cdot_M 0_M = 0_M, \quad x \cdot_M S_M(y) = x \cdot_M y +_M x.$$

Nothing is assumed about induction in M. Write $S^n(0)$ for the term with n successors, the *numeral* of n. Its value in M is $\nu(n) = S_M^n(0_M)$, and $\nu\rho$ denotes the assignment $k \mapsto \nu(\rho(k))$.

Theorem 8.8 (The numerals form a copy of the numbers)**.** *Under these six conditions, ν is injective and satisfies*

$$\nu(m + n) = \nu(m) +_M \nu(n), \qquad \nu(mn) = \nu(m) \cdot_M \nu(n).$$

Consequently, for every assignment ρ of natural numbers and every quantifier-free formula ψ:

1. *$M, \nu\rho \vDash \psi$ holds exactly when $\rho \vDash \psi$;*
2. *if $\rho \vDash \exists x_1 \cdots \exists x_k \psi$, then $M, \nu\rho \vDash \exists x_1 \cdots \exists x_k \psi$;*
3. *if $M, \nu\rho \vDash \forall x_1 \cdots \forall x_k \psi$, then $\rho \vDash \forall x_1 \cdots \forall x_k \psi$.*

Proof. The map ν is the reading of the orbit into the pointed step system $(M, 0_M, S_M)$ constructed in Chapter 2, written in natural-number coordinates. Its injectivity is the argument of Lemma 2.5. That argument used only separation and injectivity of the step, and its induction runs over the numbers, not over M. For addition, induct on n: $\nu(m + 0) = \nu(m) = \nu(m) +_M 0_M$, and

$$\nu(m + n + 1) = S_M(\nu(m + n)) = S_M(\nu(m) +_M \nu(n)) = \nu(m) +_M \nu(n + 1).$$

Multiplication follows the same pattern, using the addition identity at the successor step. By induction on terms, every term evaluates under $v\rho$ to the numeral of its value under ρ.

For (1), induct on the formula. An equation holds in M when two numerals are equal, which by injectivity happens exactly when the two values are equal. Falsehood and the connectives transport componentwise. For (2), induct on k: a witness n in $\mathbb{N}$ becomes the witness $v(n)$ in M, the extended assignments are again related by v, and (1) transports the quantifier-free body. For (3), instantiate each universal quantifier of M at numerals and apply (1). □

The two remaining directions fail, and one structure shows both. The structure W of Section 5.7 adds two points a and b to the natural numbers. It keeps the operations of $\mathbb{N}$ on numbers and sets, for every number n and every element x,

$$S(a) = a, \qquad S(b) = b, \qquad a + n = a, \qquad b + n = b, \qquad x + a = x + b = b;$$

a product is 0 when either factor is 0, and b when neither factor is 0 and one of them is a or b. Section 5.7 checks that the six conditions hold in W. For example, $x + S(a) = x + a = b = S(b) = S(x + a)$, and $a \cdot S(0) = b = 0 + a = a \cdot 0 + a$. The sentence $\exists x(S(x) = x)$ is true in W, witnessed by a, and false in $\mathbb{N}$. The sentence $\forall x \neg(S(x) = x)$ is true in $\mathbb{N}$ and false in W. So existential statements need not come down, and universal statements need not go up.

The same two sentences settle the question for every pattern of quantifiers. A *prefix* is a finite string $Q_1x_1 \cdots Q_kx_k$ in which each Q_j is $\forall$ or $\exists$ and each quantifier binds a new variable; placing it before a quantifier-free formula ψ gives a formula with that prefix. Say that a prefix *goes up* when, for every structure M obeying the six conditions, every quantifier-free ψ and every assignment ρ of natural numbers, the formula with that prefix that holds at ρ in $\mathbb{N}$ also holds at $v\rho$ in M. Say that it *comes down* when the converse holds.

Proposition 8.9 (No other prefix transfers)**.** *A prefix goes up exactly when all its quantifiers are existential, and comes down exactly when all its quantifiers are universal; only the empty prefix does both. If a prefix contains a universal quantifier, some quantifier-free ψ gives a formula with that prefix that is true in $\mathbb{N}$ and false in W under every assignment. If it contains an existential quantifier, some ψ gives one that is true in W and false in $\mathbb{N}$ under every assignment.*

Proof. Theorem 8.8 gives the positive directions, the empty prefix by its first assertion. Suppose the prefix contains a universal quantifier, and let x be its variable. Take ψ to be $\neg(S(x) = x)$. Every other quantifier of the prefix binds a variable that ψ does not mention, so it changes nothing, in $\mathbb{N}$ or in W, because both carriers have elements. The formula therefore holds exactly when $\forall x \neg(S(x) = x)$ does: in $\mathbb{N}$, and not in W, where $S(a) = a$. If the prefix contains an existential quantifier with variable x, take ψ to be $S(x) = x$. The formula then holds exactly when $\exists x(S(x) = x)$ does: in W, witnessed by a, and not in $\mathbb{N}$. □

The proposition is about all formulas with a given prefix, not about each formula separately. A single universal sentence can go up. The sentence $\forall x(x + 0 = x)$ holds in every structure obeying the six conditions, because it is one of them.

The failure has one source: the numerals miss a and b. In the list carrier the second round trip supplied every list as the expansion of its length; here nothing supplies a numeral for a. Supplying one is the work of the generation condition of Definition 2.4, which makes the

reading reach every element by Lemma 2.6. The theorem shows how far the first round trip reaches without it.

8.6 What happens to the proof and its ledger

Throughout this comparison, the formula and its derivation remain the same finite expressions. We changed their interpretation. A rule labelled EM in a derivation is still that rule, and a quantified induction formula still contains a quantifier. The resource scan of Theorem 6.5 reads those expressions; it does not read their interpreted number or list values. It therefore retains the same record for this change of interpretation.

For example, the reflexivity-and-disjunction proof of $(0 = 0) \vee \neg(0 = 0)$ remains the same empty-ledger derivation. Its formula holds whether zero is interpreted as a number or as the empty list. The alternative proof invoking EM still records that invocation. A change of carrier does not remove an occurrence from a proof tree.

This is the exact scope of the accounting claim. Here the syntax and rules are fixed, the carrier translation is explicit, and satisfaction is preserved and reflected for every formula. Translating the tree into a different proof language is a separate task, and each such translation needs its own correspondence theorem. Section 8.7 proves such a correspondence for the standard presentations of Peano arithmetic and of Heyting arithmetic. It does not cover other proof languages, or procedures for accepting trees in general.

8.7 A second proof language: Peano arithmetic

The standard first-order presentation of arithmetic is Peano arithmetic. It was formulated with no ledger in view, so it is a fair test of what the rules of Chapters 5 and 6 describe. We state it over the same terms and formulas, with the same lifting and substitution operations.

Definition 8.10 (Peano arithmetic)**.** The *axioms* of Peano arithmetic are all instances, for terms t and u, of

$$\neg(S(t) = 0), \qquad S(t) = S(u) \to t = u,$$
$$t + 0 = t, \qquad t + S(u) = S(t + u), \qquad t \cdot 0 = 0, \qquad t \cdot S(u) = t \cdot u + t.$$

A derivation of $\Gamma \vdash_{\mathrm{PA}} \varphi$ is a finite tree built from hypotheses in Γ, axioms, reflexivity, equality substitution, the introduction and elimination rules for $\wedge$, $\vee$, $\to$, $\forall$ and $\exists$ with the freshness conditions of Chapter 5, the rule for an impossible case, induction for every formula, and excluded middle, $\Gamma \vdash_{\mathrm{PA}} \varphi \vee \neg\varphi$ for every formula φ. No ledger appears in the definition.

The two systems differ in three visible places. Injectivity of the successor is an axiom here and a rule there. Excluded middle is an ordinary rule here and a recorded additional principle there. Induction is unrestricted in both, but only the rules of Chapter 6 record whether its formula contains a quantifier. Deleting excluded middle from the definition gives the usual presentation of Heyting arithmetic, the intuitionistic counterpart of Peano arithmetic.

Theorem 8.11 (The rules with excluded middle are Peano arithmetic)**.** *For every finite context* Γ *and every formula* φ*, the following are equivalent:*

1. $\Gamma \vdash_{\mathrm{PA}} \varphi$*;*
2. φ *has a derivation from* Γ *by the rules of Chapters 5 and 6 whose ledger has false LPO and MP entries.*

Moreover, the translation from (1) to (2) produces a derivation whose EM entry is true exactly when the Peano derivation uses excluded middle, and whose induction entry is true exactly when it uses induction on a formula containing a quantifier.

Proof. From (1) to (2), translate the Peano derivation node by node. Hypotheses, reflexivity, equality substitution, the connective and quantifier rules, the rule for an impossible case and induction have identical counterparts. Five of the axioms are rules of Chapter 5 with no premises. The injectivity axiom becomes a short derivation: assume $S(t) = S(u)$, apply the injectivity rule, and discharge the assumption. Excluded middle becomes an EM leaf. No LPO or MP leaf is ever produced. By Theorem 6.5, the ledger of the result is read from its EM leaves and its induction nodes, and these correspond exactly to the uses of excluded middle and the inductions of the original tree, with the same induction formulas.

From (2) to (1), translate the derivation node by node in the same way. The injectivity rule becomes the injectivity axiom followed by implication elimination, and an EM leaf becomes excluded middle. By Theorem 6.5, the ledger hypothesis excludes LPO and MP leaves, so no other case arises. Both translations preserve contexts, including the lifted contexts of the quantifier rules, because the two systems share their formulas and their lifting and substitution operations. □

Two consequences are worth recording. A Peano derivation that never uses excluded middle becomes a posit-free derivation. By Corollary 6.7, every environment satisfying its hypotheses then satisfies its conclusion, with none of EM, LPO or MP assumed in the metatheory. If, in addition, its inductions are on quantifier-free formulas only, the translation has the empty ledger.

Conversely, a posit-free derivation translates into a Peano derivation that never uses excluded middle. The posit-free derivations are therefore exactly the translations of derivations in Heyting arithmetic, the system obtained by deleting excluded middle from Definition 8.10, up to one point of form that the next corollary states exactly.

Corollary 8.12 (The posit-free derivations are Heyting arithmetic)**.** *For every finite context* Γ *and every formula* φ*, the following are equivalent:*

1. φ *has a derivation from* Γ *in Heyting arithmetic, that is, a Peano derivation that never uses excluded middle;*
2. φ *has a posit-free derivation from* Γ*.*

Moreover, the translation from (1) to (2) in Theorem 8.11 turns every derivation in Heyting arithmetic into a posit-free derivation, and every posit-free derivation has the judgment and the ledger of such a translation. The two trees differ only where the posit-free derivation applies the injectivity rule to a premise: the translation there derives $S(t) = S(u) \to t = u$, by the injectivity rule under a discharged assumption, and applies implication elimination.

Proof. Restrict both translations in the proof of Theorem 8.11 to trees without excluded middle. From (1) to (2), no EM leaf is produced, and no LPO or MP leaf ever is, so the result is posit-free. From (2) to (1), a posit-free derivation has no EM, LPO or MP leaf, by Theorem 6.5, so its translation never uses excluded middle. For the last assertion, translate a posit-free derivation d into Heyting arithmetic and back. Every node returns unchanged except an application of the injectivity rule, which returns as the injectivity axiom, rederived under a discharged assumption, followed by implication elimination. That short derivation has the same conclusion and the same context, and its only new leaf is a hypothesis, which adds nothing to the ledger. The induction nodes and their formulas are those of d. So the result has the judgment and the ledger of d. □

The theorem and its corollary compare proof systems over one language. They do not show that any of these systems is consistent, and they say nothing about languages with further types or symbols; Section 20.3 takes up that comparison for the typed language of Chapter 9, and Section 20.3.11 proves that Peano arithmetic and the typed language are consistent. What they settle are the names of the two fragments: allowing excluded middle and nothing else, the rules of Chapter 5 prove exactly the theorems of Peano arithmetic, and allowing no added principle, exactly the theorems of Heyting arithmetic. In both, the ledger shows where the quantified inductions occur, and in the first, where the classical step occurs.

8.8 A quotient represents what a reading retains

Chapter 7 studied readings that can identify different counts. Their quotients provide a second useful comparison of representations.

Let r preserve zero and addition from the orbit into an additive commutative monoid V. Its *range* consists of the values actually reached by r. Its kernel identifies x and y exactly when $r(x) = r(y)$. Write $[x]$ for the class of x under that kernel.

Theorem 8.13 (The kernel quotient and the range)**.** *The map*

$$[x] \longmapsto r(x)$$

is an addition-preserving bijection from the kernel quotient onto the range of r. If the kernel is the index–period congruence with parameters i, p, the corresponding index–period quotient represents that same range.

Proof. If x and y represent the same class, the definition of the kernel gives $r(x) = r(y)$. Hence the displayed map is well defined. It preserves zero and addition because r does. If two classes have the same image, their representatives have equal readings and therefore belong to the same kernel class; this proves injectivity. Every value in the range is $r(x)$ for a representative

x, so it is the image of $[x]$; this proves surjectivity. If the kernel is the specified index–period relation, replacing the relation by that equal description gives the corresponding quotient. □

The target of this bijection is the range, not automatically all of V. That distinction prevents unused target values from being counted as information carried by the reading. It also explains the role of the classification: the parameters describe precisely the retained identifications, while this theorem connects their quotient to the values the reader actually reaches.

8.9 The limits are part of the invariance theorem

The constructions in this chapter give positive reasons for independence from a notation. The arithmetic operations were defined in two ways, the translations were specified, and each connective and quantifier was checked under those translations. The resource record stayed attached to the same finite derivation.

The hypotheses exclude concrete failures. Consider a carrier with one element, using that element both as zero and as the successor of every element. It has operations for the term signature, so the theorem giving a unique interpretation of terms applies. But its zero is a successor. It therefore fails the faithful-generation conditions and does not supply the arithmetic interpretation just proved invariant. A mere assignment of operations is not enough.

The surrounding mathematics also remains present. Finite expressions, functions, equality, recursion and the interpretation of quantifiers are the tools used to state and prove the equivalence. We have proved a particular representation-independence result with those tools. The next part examines the different kinds of recursive construction themselves, including what becomes available when their rules grow stronger.

Questions to work through

1. In the list round-trip proof, identify the use of the assumption that the list has only one possible mark. Which later proof depends on the resulting injectivity of length?
2. Explain both directions of the universal-quantifier case in Theorem 8.6. Which translation supplies the new variable in each direction?
3. Why does the implication case require preservation and reflection of its components, rather than preservation alone?
4. In Theorem 8.13, locate the exact use of the kernel definition. Why is surjectivity onto the range immediate, while surjectivity onto the whole target was not claimed?
5. Verify in W that $x \cdot S(y) = x \cdot y + x$ when x is a number $n > 0$ and $y = a$, and when $x = y = a$. Why does $\forall x \neg(S(x) = x)$ fail in W while every numeral instance $\neg(S(S^n(0)) = S^n(0))$ holds there?
6. The one-element carrier at the end of this chapter has operations for the term signature. Which of the six conditions of Theorem 8.8 fails there, and which conclusion of the theorem fails with it?

7. Translate the Peano axiom $S(t) = S(u) \to t = u$ into a derivation by the rules of Chapter 5 and record its ledger. Then translate a use of the injectivity rule back into Peano arithmetic.

8. A Peano derivation of $\forall x(0 + x = x)$ uses induction on $0 + x = x$. What is the ledger of its translation? Compare it with the example in Chapter 6.

9. For the prefix $\exists y \forall x$, write the formula that Proposition 8.9 uses to show that it does not go up, and the one that shows it does not come down. Check each in $\mathbb{N}$ and in W. Why does $\forall x(x + 0 = x)$ not contradict the proposition?

Part III

Construction, iteration and the continuum

Chapter 9

Iteration, recursion and induction

Repeating an operation, using the position of a step, and proving a statement at every position are different abilities. On the finite tallies constructed in Chapter 1, we have all three. That can make their differences difficult to see. To separate them, we must ask what each rule would permit on a carrier whose structure has not already been settled.

This chapter does that with two pictures. In the first, two distinct roots lead to the same successor. An iterator works perfectly well there. In the second, two chains run alongside one another. Indexed recursion works on both, although only one begins at the designated start. These examples locate the work done by the stronger rules.

The conclusions concern a precisely specified language. We will give its rules, construct the examples, and prove that the examples respect every rule they are meant to interpret. That last step matters: an appealing picture alone cannot establish that a statement is underivable.

9.1 Three ways to proceed from a start

Write O for an orbit carrier, $0 : O$ for its designated start, and $S : O \to O$ for its successor operation. At this point, these are a type and two operations. We have not assumed that S is injective or that every element of O is reached from 0.

An *iterator* has a fixed target type B, a starting value $b : B$, and a step $g : B \to B$. Its equations are

$$\mathsf{I}(B, b, g, 0) \equiv b, \tag{9.1}$$

$$\mathsf{I}(B, b, g, Sx) \equiv g(\mathsf{I}(B, b, g, x)). \tag{9.2}$$

The symbol $\equiv$ means equality by the computation rules of the language. The step sees the previous result. It need not be told the position at which it is acting.

An *indexed recursor* still has a fixed target B, but its step receives both the position and the previous result:

$$g : \prod_{x : O} (B \to B).$$

Its equations are

$$\mathsf{R}(B, b, g, 0) \equiv b, \tag{9.3}$$

$$\mathsf{R}(B, b, g, Sx) \equiv g(x, \mathsf{R}(B, b, g, x)). \tag{9.4}$$

The notation $\prod_{x : O}$ denotes a dependent function type: a function takes an $x : O$ and returns a value of the type written after the product sign. Here that return type happens to be the same $B \to B$ for every x.

A *dependent eliminator* allows the target itself to depend on the position. Given a type expression $P(x)$, a value $b : P(0)$, and

$$g : \prod_{x:O}\big(P(x) \to P(Sx)\big),$$

it provides $\mathsf{E}(P, b, g, x) : P(x)$, with equations

$$\mathsf{E}(P, b, g, 0) \equiv b, \tag{9.5}$$

$$\mathsf{E}(P, b, g, Sx) \equiv g(x, \mathsf{E}(P, b, g, x)). \tag{9.6}$$

When $P(x)$ states a property, this is induction: supply a proof at the start and a method for carrying a proof through one step, then obtain a proof at each position. When $P(x)$ describes a kind of data, the same rule constructs that data.

A family asks a different question at each position

Here is the family that will later separate indexed recursion from dependent elimination. Use recursion to rebuild a position from the designated start:

$$c(x) = \mathsf{R}(O, 0, \lambda z : O.\, \lambda y : O.\, Sy, x).$$

Its equations say $c(0) \equiv 0$ and $c(Sx) \equiv S(c(x))$. Now ask for an identity proof between the rebuilt position and the supplied one:

$$P(x) = \mathrm{Id}_O(c(x), x).$$

The input x occurs in the question itself. At the start, the question is the identity of 0 with itself. At a successor, an answer to the previous question can be carried through S. A dependent eliminator would put those answers together into a proof of $P(x)$ at every x. The construction of the step proof is given explicitly in Lemma 9.5.

The two-chain picture will explain why the last step matters. Recursion can rebuild both chains into the first one, so rebuilding an element of the second chain need not recover that element. Having a recursive function c is therefore different from having the requested identity proof at every input. A dependent function of type $\prod_{x:O} P(x)$ must return a proof of the question belonging to the input it receives.

These displays already suggest two inclusions. Indexed recursion can ignore the position and act as an iterator. Dependent elimination can use a constant family $P(x) = B$ and act as indexed recursion. The interesting question is whether either inclusion can be reversed.

9.2 The language in which the question is exact

We use a small typed language. Its type expressions are generated by

$$O, \qquad \mathbf{0}, \qquad \prod_{x:A} B(x), \qquad \mathrm{Id}_A(t, u), \qquad D(t).$$

Here $\mathbf{0}$ is the empty type. An element of $\mathrm{Id}_A(t, u)$ is an identity proof between two terms of type A. The last form, $D(t)$, is optional; we will specify exactly when it is available.

A context is a finite list of typed variables. Each type in the list may use variables declared earlier. A type family $P(x)$ is a type expression in a context containing x; it is not an arbitrary family borrowed from the mathematics in which we discuss the language. There are no variables standing for types or type families, and no type whose elements are all types. In particular, there is no universe inside this language. There is also no dependent-pair constructor.

Terms consist of variables, 0, successors, functions, applications, identity proofs and their eliminator, empty-type elimination, and the three orbit operations just displayed. We consider three fragments: the iterator fragment permits I; the indexed fragment additionally permits R; the dependent fragment additionally permits E.

The typing rules are as follows. The empty context is valid, and a valid context may be extended by a type formed in it. The types O and $\mathbf{0}$ may be formed in every valid context. A product type requires its domain A in the context and its codomain $B(x)$ in the context extended by $x : A$. An identity type requires A and two terms of that type. The type $D(t)$, when enabled, requires $t : O$.

A declared variable has its declared type. The term 0 has type O, and S sends terms of type O to terms of type O. Function introduction and application have the familiar rules

$$\frac{\Gamma, x : A \vdash t : B(x)}{\Gamma \vdash \lambda x : A.\, t : \prod_{x:A} B(x)}, \qquad \frac{\Gamma \vdash f : \prod_{x:A} B(x) \quad \Gamma \vdash u : A}{\Gamma \vdash f(u) : B(u)},$$

with the displayed types required to be formed. The notation $\Gamma \vdash t : A$ says that the term t has type A under the assumptions in Γ.

Reflexivity supplies $\mathrm{refl}_x : \mathrm{Id}_A(x, x)$. The identity eliminator, called J, works as follows. Fix $x : A$ and a type family $Q(y, p)$ in the context $y : A,\ p : \mathrm{Id}_A(x, y)$. From $d : Q(x, \mathrm{refl}_x)$ it supplies

$$J(A, x, Q, d, y, p) : Q(y, p), \qquad J(A, x, Q, d, x, \mathrm{refl}_x) \equiv d.$$

Thus a construction on an identity proof is specified at reflexivity and transported along that proof. Empty-type elimination supplies a term of any formed type from a term of $\mathbf{0}$.

The typing rules for I, R and E require exactly the types of their arguments stated in the preceding section, and $x : O$. All earlier variables in the context remain available as parameters. Computation uses the six orbit equations, the J equation, and function evaluation

$$(\lambda x : A.\, t)(u) \equiv t[u/x].$$

An equation is available only in a fragment that permits its orbit operation.

We optionally add the *start discriminator*, with equations

$$D(0) \equiv \mathrm{Id}_O(0, 0), \qquad D(Sx) \equiv \mathbf{0}. \tag{9.7}$$

It gives a type-level distinction between the start and a successor. The superscripts $-$ and $+$ will mean that this discriminator is disabled or enabled, respectively.

For precision, raw conversion is the least equivalence relation containing the enabled computation equations and closed under every expression constructor and capture-avoiding

substitution. It is defined on raw expressions: an intermediate expression in a conversion chain need not be well typed. A *typed* equality requires both endpoints to have the stated type; a type equality requires both endpoint types to be formed. A term may be converted from A to B only through such a type equality. We do not turn an identity proof into a raw conversion, and we add neither function extensionality nor an η equation.

The word "raw" deserves an example. As an expression, the function application $(\lambda x : \mathbf{0}.\, x)(0)$ rewrites to 0. But its function expects an input from the empty type, while the supplied 0 has type O. The application cannot be typed. A soundness proof cannot therefore justify every intermediate raw rewrite merely by appealing to the typing of the original endpoints. Section 9.8 handles this issue in two steps: typed expressions satisfy a safety condition, and convertible safe endpoints reduce forward to a common expression without losing their interpreted value. That is why the later proof needs both safety and confluence.

Finally, a substitution from a context Γ to a context Δ assigns to each declared variable of Γ a term of its substituted type in Δ. Substitution acts on types, terms and typed equalities. Only the finitely many variables in the source context are required. These formation, introduction, elimination, conversion and substitution rules constitute the whole calculus used below.

Definition 9.1 (Admissibility)**.** A rule is admissible in a fragment if, whenever its premises have derivations in that fragment, its conclusion has a derivation there too. For indexed recursion and dependent elimination, we require a term with both displayed computation equations as typed equalities, uniformly over every finite parameter context and every permitted type expression.

This definition makes a negative result strong enough to be useful. It rules out every construction from the available rules, including one we have not thought to try. A single instance with derivable premises and an underivable conclusion refutes admissibility.

9.3 What indexed recursion immediately gives

The additional position argument in R has a direct use. Choose target O, base 0, and step $(x, y) \mapsto x$. Define

$$\operatorname{pred}(x) = \mathsf{R}(O, 0, \lambda z : O.\, \lambda y : O.\, z, x). \tag{9.8}$$

Then the computation rules give

$$\operatorname{pred}(0) \equiv 0, \qquad \operatorname{pred}(Sx) \equiv x.$$

The second equation is a left-inverse law for successor.

Theorem 9.2 (Indexed recursion supplies injectivity)**.** *In either indexed fragment, with or without D, there is a term of*

$$\prod_{x:O} \prod_{y:O} \big(\mathrm{Id}_O(Sx, Sy) \to \mathrm{Id}_O(x, y)\big).$$

Proof. The term in (9.8) is well typed because its step accepts two orbit elements and returns the first. Given $p : \mathrm{Id}_O(a, b)$, identity elimination applies pred to that path: use the family $Q(z, q) =$

$\mathrm{Id}_O(\mathrm{pred}(a), \mathrm{pred}(z))$ and the base $\mathrm{refl}_{\mathrm{pred}(a)}$. This constructs a proof of $\mathrm{Id}_O(\mathrm{pred}(a), \mathrm{pred}(b))$. Set $a = Sx$ and $b = Sy$. The two predecessor computations convert the resulting type to $\mathrm{Id}_O(x, y)$. Abstracting over x, y, p gives the required term. No discriminator is used. □

The start discriminator gives a different fact.

Theorem 9.3 (The discriminator separates the start)**.** *In every fragment with D enabled, there is a term of*

$$\prod_{x:O} \big(\mathrm{Id}_O(0, Sx) \to \mathbf{0}\big).$$

Proof. By (9.7), refl_0 is a term of $D(0)$. If $p : \mathrm{Id}_O(0, Sx)$, apply identity elimination with family $Q(y, q) = D(y)$ and that base term. The result has type $D(Sx)$, which converts to $\mathbf{0}$. Abstract over x and p. The construction uses neither recursion nor induction. □

The two arguments identify two different resources. The predecessor retrieves the input of a successor. The discriminator distinguishes the start from a successor. Neither argument silently assumes the other fact.

9.4 An iterator can merge two roots

Consider distinct elements $a, u, v_1, v_2, \ldots$, with designated start a and step

$$S(a) = v_1 = S(u), \qquad S(v_n) = v_{n+1} \quad (n \geq 1).$$

The subscripts here are the finite natural numbers of the ambient mathematics. They are being used to construct a model of the language; they do not add an induction rule to that language. Figure 9.1 draws the structure.

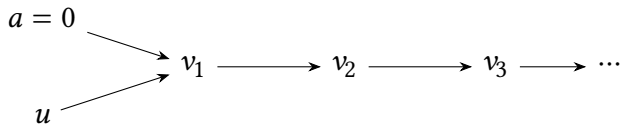


Figure 9.1. Two roots with one common tail. Iteration can give both roots the same starting value, so its equations do not detect the merger.

For any set B, any $b \in B$, and any function $g : B \to B$, define

$$I(a) = b, \qquad I(u) = b, \qquad I(v_n) = g^n(b). \tag{9.9}$$

Both roots satisfy $I(Sx) = g(I(x))$: each side is $g(b)$. At a tail element, the equation is $g^{n+1}(b) = g(g^n(b))$. Thus the iterator exists for every target and step, although successor is not injective.

The discriminator can be interpreted too. Give $D(a)$ and $D(u)$ the one-element set, and give each $D(v_n)$ the empty set. Then $D(a)$ interprets $\mathrm{Id}(a, a)$, and every $D(Sx)$ is empty. The extra root u is allowed to share the start's discriminator value; the discriminator equations never require the converse implication $D(x) \Rightarrow x = 0$.

Theorem 9.4 (The iterator boundary)**.** *In the fixed calculus of Section 9.2, the iterator fragment, with either choice of discriminator, has no closed term proving successor injectivity. Indexed recursion is not admissible in it.*

Proof. Interpret the orbit type by the merging structure. Interpret product types by sets of dependent functions, the empty type by the empty set, and an identity type by a singleton when its endpoints are equal and the empty set otherwise. The soundness proof in Section 9.8 verifies all the rules of the iterator fragment for this interpretation, in every finite parameter context and with either choice of D.

If an injectivity term existed, apply its interpreted function to a, u and the identity proof between $S(a)$ and $S(u)$. It would produce an element of the empty identity set between the distinct elements a and u. This is impossible.

For indexed recursion, use the closed premises $B = O$, $b = 0$ and $g(x, y) = x$, all expressible in the iterator fragment. An admissible recursor would give a function r with $r(Sx) = x$. In the model this would require both $r(v_1) = a$ and $r(v_1) = u$. Since $a \neq u$, no such function exists. Soundness therefore excludes a term with the required equations. □

The first example answers a common temptation: the ability to repeat an operation does not, by itself, make the underlying step reversible. The position argument in indexed recursion is doing real work.

9.5 Recursion can leave a whole chain unreached

Now let

$$O = \{0, 1\} \times \mathbb{N}_\delta, \qquad 0_O = (0, 0), \qquad S(i, n) = (i, Sn).$$

There are two disjoint successor chains. Only the first starts at the designated 0_O (Figure 9.2).

$$(0,0) \longrightarrow (0,1) \longrightarrow (0,2) \longrightarrow \cdots$$

$$(1,0) \longrightarrow (1,1) \longrightarrow (1,2) \longrightarrow \cdots$$

Figure 9.2. Indexed recursion can start with the same value on both chains. Its step can read the chain's index without making the second root reachable from the designated start.

For a fixed B, b, g, define the recursor on each chain by

$$R(i, 0) = b, \qquad R(i, Sn) = g((i, n), R(i, n)). \tag{9.10}$$

The initial equation holds at $(0, 0)$, and the successor equation holds everywhere by construction. There is no conflict in also starting the second chain at b. The equations specify no separate base value there. Ordinary iteration is obtained by choosing a step that ignores (i, n).

Set $D(0, 0)$ to the singleton and $D(i, n)$ to the empty set at every other point. The discriminator equations hold too. Thus even the indexed fragment with the discriminator leaves room for the second chain.

To expose what fails, we need a family expressible in the language. The informal property "belongs to the first chain" would not suffice unless the language could express it. Instead define the term

$$c(x) = \mathsf{R}(O, 0, \lambda z : O.\, \lambda y : O.\, Sy, x), \qquad P(x) = \mathrm{Id}_O(c(x), x). \tag{9.11}$$

The term c rebuilds the successor history from the designated start. It satisfies

$$c(0) \equiv 0, \qquad c(Sx) \equiv S(c(x)).$$

In the two-chain model, induction on the finite index gives

$$c(i, n) = (0, n). \tag{9.12}$$

At index zero the equation is the base clause of the recursor. If it holds at n, the successor clause applies S to $(0, n)$ and gives $(0, Sn)$. This proves the formula at every finite index.

Lemma 9.5 (A family with a base and a step)**.** *In either indexed fragment, $P(x)$ from* (9.11) *is a formed type family, with a closed term of $P(0)$ and a closed term of $\prod_{x:O}(P(x) \to P(Sx))$.*

Proof. The term $c(x)$ is an indexed recursor with a well-typed step, so its identity type with x is formed. At 0, its computation makes refl_0 a term of $P(0)$. For the step, take $p : \mathrm{Id}_O(c(x), x)$. Apply S to this path using J, with motive $\mathrm{Id}_O(S(c(x)), Sy)$ and reflexivity base. The result is a proof of $\mathrm{Id}_O(S(c(x)), Sx)$. Convert the first endpoint using $c(Sx) \equiv S(c(x))$. This is a proof of $P(Sx)$, and abstraction over x, p gives the required step. □

Theorem 9.6 (The indexed-recursion boundary)**.** *In either indexed fragment of the fixed calculus, there is no closed term of $\prod_{x:O} P(x)$ for this expressed family. Dependent elimination is not admissible. Nor is the rule asserting pointwise uniqueness of two maps with the same start and successor equations.*

Proof. The two-chain interpretation satisfies every rule of the indexed fragment by Section 9.8. At the second root, formula (9.12) gives

$$P(1, 0) = \mathrm{Id}_O((0, 0), (1, 0)).$$

Its interpretation is empty. A function inhabiting $\prod_{x:O} P(x)$ would have to return an element of this empty set at $(1, 0)$, so no such function, and hence no such term, exists.

Lemma 9.5 supplies the premises of dependent elimination inside the indexed fragment itself. If that rule were admissible, it would supply exactly the impossible term.

For uniqueness, compare c with $q(x) = x$. Both are terms $O \to O$; both satisfy $f(0) = 0$ and $f(Sx) = S(f(x))$, with identity proofs obtained from their computation equations. But at $(1, 0)$, c returns $(0, 0)$ and q returns $(1, 0)$. A term asserting their pointwise equality would again produce an element of an empty identity set there. All the proposed rule's premises have terms, while its conclusion does not. □

The failure occurs at a specific, expressible family. It does not rely on allowing an arbitrary external predicate as an induction motive. The two maps also show why merely solving recurrence equations does not establish uniqueness on a carrier whose generated part may be smaller than the whole.

9.6 Induction does not invent a discriminator

The preceding models separate the two weaker fragments. A different argument shows that, in this language, the start discriminator is still an additional resource even after dependent elimination is available.

Assign a truth value $v(A)$ to every raw type expression:

$$v(O) = \mathsf{true}, \qquad v(\mathbf{0}) = \mathsf{false},$$
$$v(\mathrm{Id}_A(t,u)) = \mathsf{true}, \qquad v(D(t)) = \mathsf{true},$$
$$v\left(\prod_{x:A} B\right) = v(A) \Rightarrow v(B).$$

Here $\Rightarrow$ is ordinary Boolean implication. The valuation deliberately ignores the terms inside a type. It is a tool for checking derivations, not a set interpretation in which identity means equality of denoted elements.

Lemma 9.7 (Substitution cannot change this truth value)**.** *Renaming variables or substituting terms in A preserves $v(A)$. When the discriminator is disabled, raw type conversion also preserves it.*

Proof. For substitution and renaming, induct on the type expression. The orbit and empty types do not change. Identity and discriminator types always have value true, whatever their contained terms. At a product, the two induction hypotheses preserve the antecedent and consequent of the Boolean implication.

For conversion, induct on its construction. Reflexivity, symmetry and transitivity preserve equality of truth values. Product congruence uses its two induction hypotheses. Identity and discriminator congruence preserve the constant value true. The substitution case uses the first part. The two special discriminator equations are unavailable when D is disabled. These are all the type-conversion cases; term conversions inside an identity or a discriminator cannot affect its value. □

Theorem 9.8 (Truth preservation without the discriminator)**.** *In any of the three fragments with D disabled, if $\Gamma \vdash t : A$ and every type declared in Γ has value true, then $v(A) = \mathsf{true}$. In particular every inhabited closed type has value true.*

Proof. Induct simultaneously on the finite derivations of terms, equalities and substitutions. A variable inherits the true value of its declared type; weakening does not change it by Lemma 9.7. Zero and successor have type O. Reflexivity has an identity type, whose value is true.

For a function, if the domain has value false, its product type has value true immediately. If the domain has value true, extend the true context by that domain and apply the induction hypothesis to the body. The codomain is true, so the product is true. For application, a true implication together with a true argument type gives a true codomain. Substitution for the argument preserves that value.

Identity elimination has a base term of $Q(x, \mathrm{refl}_x)$. The induction hypothesis makes this type true. Since substitution does not change the value of Q, the target $Q(y, p)$ is true as well. Empty-type elimination cannot have its premise in a true context: that premise would make $v(\mathbf{0})$ true.

For iteration and indexed recursion, the base term already makes the fixed target B true. For dependent elimination, the base term makes $P(0)$ true; substitution invariance then makes $P(x)$ true at every argument. These observations handle the orbit rules even when the strongest one is enabled.

Type equality preserves the valuation by Lemma 9.7; term equality inherits truth from either typed endpoint. Type conversion therefore preserves the term assertion. A derivable substitution sends a true target context to a true source context, since each source variable has a term of its substituted type in the target. Apply the induction hypothesis to the original judgment in that source context, then use substitution invariance. This handles substituted terms and equalities. Context and type formation add no assertion that a type is inhabited. All derivation rules have now been covered. □

Corollary 9.9 (A strict discriminator boundary)**.** *Without D, even the dependent fragment has no closed term of*

$$\prod_{x:O} \big(\mathrm{Id}_O(0, Sx) \to \mathbf{0}\big).$$

Proof. The inner implication has value $\mathsf{true} \Rightarrow \mathsf{false} = \mathsf{false}$. The outer product has the same false value because $v(O)$ is true. Theorem 9.8 excludes an inhabitant. □

The hypothesis disabling D is essential. When its equations are enabled, $D(S0) \equiv \mathbf{0}$ equates a type assigned true by this valuation with one assigned false. The valuation then ceases to preserve conversion, exactly where the positive construction in Theorem 9.3 becomes available.

9.7 The strict sequence of resources

Let $\mathscr{I}^-$, $\mathscr{R}^-$, and $\mathscr{E}^-$ denote the iterator, indexed and dependent fragments without D. Let $\mathscr{E}^+$ be the dependent fragment with D.

Theorem 9.10 (Three strict extensions)**.** *Every derivation in one of the following fragments is also a derivation, of the same judgment, in the next:*

$$\mathscr{I}^- \subsetneq \mathscr{R}^- \subsetneq \mathscr{E}^- \subsetneq \mathscr{E}^+.$$

The three strictness witnesses are, respectively, successor injectivity, the section $\prod_{x:O} P(x)$ *from* (9.11)*, and start separation.*

Proof. Each extension retains the earlier grammar and rules. Induct on a conversion derivation to retain its equations, and on a typing derivation to retain its judgment. At a rule that requires an enabled operation, the larger fragment still enables it. Every other rule and every context

is unchanged. Thus these are inclusions on actual finite derivations, not only comparisons of their interpretations.

Theorem 9.2 supplies the first witness in $\mathscr{R}^-$, and Theorem 9.4 excludes it from $\mathscr{I}^-$. Lemma 9.5 supplies the inputs for a dependent eliminator; applying that eliminator and abstracting over x gives the second witness in $\mathscr{E}^-$. Theorem 9.6 excludes it from $\mathscr{R}^-$. Finally, Theorem 9.3 supplies start separation in $\mathscr{E}^+$, and Corollary 9.9 excludes it from $\mathscr{E}^-$. □

The first two separations also hold with the discriminator enabled throughout: both countermodels interpret it. The third separation concerns adding the discriminator itself. How much arithmetic the top fragment $\mathscr{E}^+$ proves is a separate question. Chapter 20 answers it: $\mathscr{E}^+$ proves the negative translations of exactly the theorems of Peano arithmetic (Theorem 20.39).

There is no conflict with the arithmetic developed in Part I. There we constructed the finite tallies and used their structural elimination rule; generation from the start was built into the construction. Here we ask which conclusions follow from selected rules on a type O before its full structure has been established. The two countermodels identify precisely what is lost when the stronger rule is withheld. They are not alternative faithful representations of the already constructed tallies.

9.8 Why the countermodels cover every rule

The negative results used a soundness claim for the complete fixed calculus. This section proves that claim. A reader following the main structural argument can return to it after the chapter; its purpose is to close the gap between the two pictures and the statements about all derivations.

9.8.1 A total interpretation and its limits

Work in ordinary classical set theory for this external proof. Encode each orbit carrier injectively as a set, choosing the designated start different from the empty set. The two concrete encodings may be made from tagged ordered pairs of finite tally codes. Extend decoding to all sets by returning the designated start when a set is not an orbit code. Consequently S, I and R have total external interpretations, although their typing rules will permit only arguments in the appropriate domains. The unused indexed recursor in the merging model, and the unused dependent eliminator in both models, receive the default value $\varnothing$.

An environment ρ assigns a set to each variable. Define the interpretation of types recursively by

$$
\begin{aligned}
[\![O]\!]_\rho &= \text{the encoded orbit carrier},\\
[\![\mathbf{0}]\!]_\rho &= \varnothing,\\
\left[\!\!\left[\prod_{x:A} B\right]\!\!\right]_\rho &= \prod_{a \in [\![A]\!]_\rho} [\![B]\!]_{\rho[x:=a]},\\
[\![\mathrm{Id}_A(t,u)]\!]_\rho &= \begin{cases} \{\varnothing\}, & [\![t]\!]_\rho = [\![u]\!]_\rho,\\ \varnothing, & \text{otherwise.} \end{cases}
\end{aligned}
$$

The product consists of all graphs of functions with the stated domain and values in the corresponding fibers. Interpret $D(t)$ by the discriminator set at $[\![t]\!]_\rho$.

A variable is interpreted by its environment entry. Zero and successor use the model's start and step. A function is interpreted by the graph

$$[\![\lambda x \,:\, A.\, t]\!]_\rho = \{(a, [\![t]\!]_{\rho[x:=a]}) \,:\, a \in [\![A]\!]_\rho\}.$$

For application, look for an output paired with the given input in the interpreted graph: choose one if it exists, and use $\emptyset$ otherwise. On a function graph at an input in its domain the output is unique, so this choice has no effect. Reflexivity, empty-type elimination and the unavailable dependent eliminator are interpreted as $\emptyset$. The interpretation of $J(A, x, Q, d, y, p)$ is that of its base d. Finally I and R use the concrete operations (9.9) and (9.10), with interpreted base and step.

All raw expressions now have values. This does *not* make every raw computation equation preserve values. For example,

$$(\lambda x \,:\, \mathbf{0}.\, x)(0) \equiv 0$$

is a raw function-evaluation equation. The function's graph is empty, so the left side has default value $\emptyset$, whereas the right side is the designated nonempty start code. This application has no valid typing derivation. The soundness claim must use its typing hypotheses.

Two elementary facts about the interpretation will be used repeatedly. If $a \in A$, applying the graph of f on A to a returns $f(a)$. If a graph with domain A belongs to a dependent function set with domain B, then $A = B$: the first coordinates of the pairs in either description are exactly its domain. These facts follow directly from the definition of a function graph.

9.8.2 Substitution and safe expressions

For a substitution σ, let $\rho\sigma$ assign to each variable the value of its substituting term in ρ. Structural recursion on expressions gives

$$[\![A[\sigma]]\!]_\rho = [\![A]\!]_{\rho\sigma}, \qquad [\![t[\sigma]]\!]_\rho = [\![t]\!]_{\rho\sigma}. \tag{9.13}$$

For a variable the assertion is the definition of $\rho\sigma$. For a constant it is immediate. At a binder, extend both environments by the same new value and leave the bound variable fixed; the induction hypothesis then identifies the body values. At every remaining constructor, apply its defining operation to the equal constituent values. The J family has two binders and uses two such extensions. This proves both identities, together with their renaming versions.

We next define a property called *safety*, relative to a model and an environment. It records the domain conditions needed by evaluation, without claiming that the whole expression is typed. The definition follows the raw syntax.

For types, O and $\mathbf{0}$ are safe. A product is safe when its domain is safe and its codomain is safe at every value in that domain. An identity type is safe when its two endpoint terms are safe. A discriminator type is safe when its argument is safe.

For terms, variables, zero, reflexivity, empty-type elimination and the unavailable dependent eliminator are safe. A successor is safe when its argument is safe and has value in the orbit

carrier. A function is safe when its domain is safe and its body is safe at every value in that domain. An application is safe when both constituents are safe and there is a dependent function set containing the value of its function, whose domain contains the value of its argument. A J term is safe exactly when its base term is safe; the interpretation reads only that base.

An iterator is safe when its base, step and orbit argument are safe, the argument lies in the orbit carrier, and there is a set B containing the base value on which the step value is a function $B \to B$. For an indexed recursor, replace that last condition by a step function from the orbit carrier to $B \to B$. The type annotations ignored by these interpretations impose no additional safety condition.

Lemma 9.11 (Safety under substitution)**.** *Suppose e is safe in $\rho\sigma$, and every substituting term for a free variable of e is safe in ρ. Then $e[\sigma]$ is safe in ρ.*

Proof. Induct on e. A variable uses the corresponding substitution hypothesis. Constructors that are safe without conditions remain so. At a successor, application, iterator or recursor, the induction hypotheses give safety of the constituents, while (9.13) preserves all the required set memberships and function domains. At an identity or discriminator type, use the hypotheses for its term arguments. At a product or function, fix a value in the domain, extend the environment and lift the substitution beneath the binder. The new variable is safe by definition; all previous variables retain their hypotheses by renaming. Apply the body induction hypothesis in that extended environment. The J case uses its base hypothesis. This also shows why no condition is needed on substitution components outside the finite set of free variables. □

9.8.3 Joining conversions without changing the language

A conversion can pass through untyped intermediate expressions. We therefore cannot justify it by asserting that every step in its original zigzag preserves values. Instead we join the two endpoints by forward reductions that preserve values whenever they start safely.

Define an external parallel reduction relation $e \Rrightarrow e'$. It permits constituents to reduce in parallel under every constructor, including binders. It also permits a visible function application or enabled iterator/recursor equation to contract, with its constituents reduced in parallel. For example,

$$b \Rrightarrow b', \quad t \Rrightarrow t' \quad \Longrightarrow \quad (\lambda x : A.\, b)(t) \Rrightarrow b'[t'/x].$$

It includes the two enabled discriminator contractions. Finally, it has the external erasures

$$J(A, x, Q, d, y, p) \Rrightarrow d' \quad (d \Rrightarrow d'),$$

and sends reflexivity, empty-type elimination and raw dependent elimination to refl_0. These erasures reflect the chosen set interpretation. They are not new computation rules or new equalities in the object language.

Lemma 9.12 (Parallel substitution and joining)**.** *Parallel reduction is preserved by substitution. More generally, if $e \Rrightarrow e'$ and each component of σ reduces to the corresponding component of τ, then $e[\sigma] \Rrightarrow e'[\tau]$. The parallel relation has the diamond property, and its finite-step closure is confluent. In either weaker fragment, raw-convertible expressions have a common reduct in this external relation.*

Proof. Prove the substitution assertion by induction on the parallel derivation. Constructor cases use their constituent induction hypotheses; beneath a binder, lift both substitutions and retain the bound variable. In the function-contraction case, the substitution composition identity identifies the two results of substituting first or contracting first. The iterator and recursor contractions retain their form under substitution, as do the discriminator contractions. A J erasure reduces to the substituted base, and the other erasures still reduce to refl_0. This covers every parallel rule.

The diamond property follows by Takahashi's method of complete developments [85]. For each expression e, define its complete development e^* by developing the constituents and contracting the outer redex if its original outer form is a visible redex. Erase J to its developed base and the other erasable forms to refl_0. An operation disabled by the fragment retains its constructor. A redex newly exposed only after developing a constituent need not be contracted at this stage.

We claim that $e \Rrightarrow u$ implies $u \Rrightarrow e^*$. Induct on the displayed parallel step. For a constructor whose outer form is not a redex, reduce its constituents by the induction hypotheses. If an application originally has a lambda at its head, a congruence step leaves a lambda there, so contract it using the developed body and argument. If the first step already contracted it, parallel substitution gives the same developed result. For an iterator or recursor at zero, both possibilities lead to the developed base. At a successor, both lead to the developed step applied to the corresponding recursive expression; the recursor also receives the developed predecessor. With primes marking the results of the first step on the constituents, the congruence result $\mathsf{R}(B', b', g', Sx')$ and the contraction result $g'(x', \mathsf{R}(B', b', g', x'))$ both reduce in parallel to

$$\mathsf{R}(B, b, g, Sx)^* = g^*\big(x^*, \mathsf{R}(B^*, b^*, g^*, x^*)\big).$$

The discriminator cases lead to the same unit or empty type. A congruence step on J can be followed by erasure; an erasure step can be followed by development of its base. For the remaining erasable forms both paths end at refl_0. These exhaust the possible overlaps.

Thus any two parallel steps from e are joined at e^*. This is the diamond property. For finite reduction chains, use the diamond repeatedly to move a single step past a finite chain; induction on the other chain then fills a finite grid of diamonds and produces a common endpoint. This proves confluence of the finite-step closure.

Having a common reduct is consequently an equivalence relation: reflexivity and symmetry are immediate, and confluence joins the two reducts needed for transitivity. It is closed under constructors and substitution by the corresponding parallel rules. Each enabled computation equation has a common reduct: function, iterator, recursor and discriminator equations use their contractions, while the J equation uses erasure. The dependent computation equations are absent in the two weaker fragments. Induction on raw conversion now proves its inclusion in this common-reduct relation. □

Lemma 9.13 (Safe parallel reduction preserves values)**.** *In the merging model for the iterator fragment, and in the two-chain model for the indexed fragment, if e is safe and $e \Rrightarrow e'$, then e' is safe and $[\![e]\!] = [\![e']\!]$. Both choices of discriminator are allowed.*

Proof. The concrete iteration and recursion operations stay within their target set whenever their base and step do. For the merging model this follows by induction on the tail level in (9.9). For the two-chain model it follows by induction on the second coordinate in (9.10). The base and successor equations hold by those definitions. The discriminator equations hold by the definitions at the roots and tails.

Now induct on the parallel step, simultaneously for types and terms. At a product, the domain values agree by induction; at each point of that common domain, the fiber values agree too. Their sets of dependent functions are therefore equal. At a lambda, the same pointwise argument gives equality of graphs. Identity and discriminator types depend only on the relevant term values. Successor and application congruence preserve their domain witnesses and values using the constituent equalities. Iterator and recursor congruence similarly preserve their common target set and interpreted operations.

For a function contraction, safety of the application supplies a function-set domain containing the argument. The graph-domain fact identifies that domain with the lambda's declared domain. Graph application therefore equals evaluation of the body with that argument substituted. Use the induction hypotheses for the body and argument, then Lemma 9.11, to obtain both the value and safety of the contracted term. No off-domain default is used.

At an iterator or recursor zero contraction, use its base equation and the base induction hypothesis. At a successor contraction, safety supplies the predecessor in the orbit carrier. The target-closure fact proved above puts the recursive result in B. The step can therefore be applied, once for the iterator and twice for the indexed recursor; its concrete successor equation gives the claimed value. The induction hypotheses preserve all the needed memberships.

The discriminator contractions use its two equations. The interpretation and safety of J are precisely those of its base, so its congruence and erasure cases follow from the base induction hypothesis. Reflexivity, empty-type elimination and raw dependent elimination all have value $\varnothing$, as does refl_0; their erasures preserve safety and values directly. These are all the parallel rules. □

Corollary 9.14 (Coherence at safe endpoints)**.** *In either concrete interpretation, two raw-convertible expressions that are both safe in an environment have equal values there.*

Proof. By Lemma 9.12, the endpoints have a common external reduct. Apply Lemma 9.13 successively along each finite reduction chain. Each chain preserves safety and value. Both original values equal the value of the common reduct, hence equal one another. □

9.8.4 Soundness of the complete typing derivations

Theorem 9.15 (The two concrete models are sound)**.** *Every derivation in the iterator fragment has a valid interpretation in the merging model. Every derivation in the indexed fragment has a valid interpretation in the two-chain model. Both statements hold with or without the*

discriminator and in every finite parameter context. In particular, a typed term denotes a member of its type, and typed equalities denote equal values.

Proof. Prove a stronger assertion by simultaneous induction on the derivation. In an environment satisfying its context, every formed type is safe; every typed term is safe and belongs to its interpreted type; both endpoints of each typed equality are safe and have equal values. A derivable substitution takes a satisfying target environment to a satisfying source environment and has safe components at all declared source variables.

Context extension adds exactly the required membership. Variables use their context entries. Orbit and empty-type formation are safe directly. Product formation uses safety of the domain and the codomain in each extended environment; identity and discriminator formation use the safety of their term arguments. Zero belongs to the carrier, and the concrete successor preserves that carrier.

For function introduction, the body induction hypothesis gives a safe value in the correct fiber at every domain point. Its graph therefore belongs to the product type and is safe. For application, the two premises give a dependent function and an argument in its domain. Applying that graph yields a member of the required fiber, and (9.13) identifies the fiber with the substituted result type. The same premises give safety.

Reflexivity denotes $\emptyset$, which is an element of the singleton identity set at equal endpoints. For J, the identity premise says that the two endpoint values agree and the path value is $\emptyset$. Substituting these equal values into the motive makes the target fiber the same as the base fiber, which already contains the base value. Safety comes from that base. Empty-type elimination has no possible premise in a satisfying environment, since nothing belongs to $\emptyset$.

Iteration and indexed recursion use the proved closure of their concrete operations in the target set; their premises supply exactly the safety conditions stated above. Dependent elimination cannot be a rule of either fragment under consideration.

For type or term equality, the induction hypotheses supply safety of both endpoints. Corollary 9.14 then proves equality of their values, despite any untyped intermediates in the raw conversion. Converting a term's type preserves membership because the two interpreted types are equal.

For substitution, the premises type every component associated with a declared source variable. Their membership conclusions give a satisfying source environment, and their safety conclusions cover every variable that can occur in a source judgment. Formula (9.13) transports membership and equalities; the finite-variable form of Lemma 9.11 transports safety. This handles substitution of types, terms and both kinds of equality. It finishes the induction over all the original rules. □

The external set theory and its choices belong to this nonderivability proof. No universe, external family, semantic soundness assumption or additional equation has been inserted into the object calculus. The negative conclusions therefore apply to the exact rules stated at the beginning of the chapter.

Questions to work through

1. Check the iterator equation separately at a, at u, and at v_n in the merging model. Explain why setting $D(u)$ to the singleton does not violate either discriminator equation.

2. Carry out the two function-evaluation steps in $\mathrm{pred}(Sx)$ after the recursor equation has been applied. Identify the point at which the indexed step uses information unavailable to an ordinary iterator step.

3. Verify the base and successor equations for both c and $q(x) = x$ in the two-chain model. Locate the element at which their pointwise equality fails.

4. Why does dependent elimination preserve truth when the valuation ignores the motive's argument? Which discriminator equation invalidates this valuation?

Chapter 10

Finite construction and effective description

A finite tally can be written down. A rule for producing tallies can also be written down. These are different objects. The first records a result; the second may produce different results when given different inputs, and may fail to finish on some inputs.

There is a third distinction. Suppose that, for every requested accuracy, a description specifies a finite rational interval. Each interval is finite data. It does not follow that one finite rule produces all those intervals. The placement of the word "one" matters. This chapter makes that difference precise, while showing how finite arithmetic and finite programs fit into the same generated records.

10.1 An explicit budget for a tally

Write $L(x)$ for the number of steps in $x \in \mathbb{N}_\delta$, as in Chapter 1. The inverse reading takes an ordinary natural number k to the tally with k steps. The two readings are inverse by their construction. We use ordinary natural numbers to index budgets in the surrounding mathematics.

Definition 10.1 (A finite stage)**.** For a budget n, let

$$F_n = \{x \in \mathbb{N}_\delta \,:\, L(x) \le n\}.$$

If $n \le m$, the inclusion $j_{n,m} : F_n \longrightarrow F_m$ retains the same tally and regards it as satisfying the larger budget.

There are exactly $n+1$ elements in F_n: the tallies with lengths $0, 1, \dots, n$. This is an explicit correspondence. Send a tally to its length and send an index back to the tally with that many steps. The two composites are identities. In particular, a stage is finite because its elements have been enumerated, not because a finiteness assumption has been attached to it.

The inclusions satisfy

$$j_{n,n} = \mathrm{id}, \qquad j_{m,k} \circ j_{n,m} = j_{n,k} \quad (n \le m \le k).$$

Both equations say that enlarging a budget does not change a record. Every $x \in \mathbb{N}_\delta$ lies in its own finite stage $F_{L(x)}$. No search for a suitable stage is needed: the record supplies its budget.

This gives more than an increasing list of finite sets. It tells us exactly when calculations or readings defined separately on those sets determine a single reading of the generated orbit.

Definition 10.2 (Agreement across budgets)**.** Let A be a fixed target. A compatible family consists of maps $f_n : F_n \to A$ such that

$$f_m(j_{n,m}(x)) = f_n(x) \qquad (n \le m,\ x \in F_n). \tag{10.1}$$

Compatibility is a substantial condition. It says that the answer assigned to an already available record does not change merely because the reader is given a larger permitted budget.

Theorem 10.3 (Finite readings glue uniquely)**.** *For every compatible family (f_n), exactly one map $f : \mathbb{N}_\delta \to A$ restricts to f_n on every F_n. It is given explicitly by*

$$f(x) = f_{L(x)}(x).$$

Proof. The expression on the right is defined because $x \in F_{L(x)}$. If $x \in F_n$, then $L(x) \le n$, and compatibility gives

$$f_n(x) = f_n(j_{L(x),n}(x)) = f_{L(x)}(x) = f(x).$$

Thus f has every required restriction. Conversely, any map g with those restrictions satisfies $g(x) = f_{L(x)}(x)$ by applying its restriction at the particular budget $L(x)$. Hence $g(x) = f(x)$ for every x. □

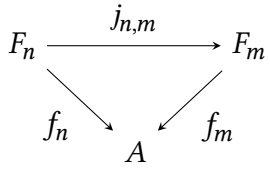


Figure 10.1. A larger budget retains the record. Agreement along the two routes is the hypothesis that makes the single global reading possible.

The proof uses one stage for each input. It does not require an apparatus that stores the whole orbit. Nor does it turn an arbitrary compatible family into an executable rule: the maps f_n were supplied as mathematical data. The theorem establishes precisely the unique assembly of those maps.

10.2 Arithmetic carries its own finite bounds

The length identities from Chapter 3 immediately give finite output budgets:

$$\begin{aligned} S &: F_n \longrightarrow F_{n+1}, \\ + &: F_n \times F_m \longrightarrow F_{n+m}, \\ \cdot &: F_n \times F_m \longrightarrow F_{nm}. \end{aligned} \tag{10.2}$$

For example, $L(x + y) = L(x) + L(y) \le n + m$. Similarly, $L(xy) = L(x)L(y) \le nm$. These are bounds for the size of the resulting tally. They do not count the intermediate work of a chosen calculating device.

If the input budgets are enlarged, the output tally is unchanged. For $n \le n'$ and $m \le m'$, adding after inclusion gives the same tally as adding first and then including into $F_{n'+m'}$. The multiplication statement is identical with the product budget. The proof is especially short: an inclusion retains its underlying tally, and the operations are the same operations on those tallies on both routes.

There is a useful discipline here. A claim of finite construction should identify what its budget measures. Equations (10.2) measure record length. An execution-time claim or a physical storage claim would require its own construction and bound; Chapter 15 gives one, for the solutions of polynomial differential equations whose computed trajectory passes a finite region certificate.

10.3 A record with a decoder

We can make the representation entirely explicit. Prepare a finite bit record $U(x)$ consisting of $L(x)$ repetitions of the bit 0. Define a budgeted decoder D_n as follows: count the length of the supplied bit record r. If that length is at most n, return the tally with that many steps, regarded as an element of F_n. Otherwise return failure.

The decoder uses the record's length. It does not inspect whether its bits are all zero. Preparing only zeros makes one canonical record for each tally; the decoder is allowed to accept other records of the same length too.

Proposition 10.4 (Exact finite recovery)**.** *For every $x \in F_n$,*

$$D_n(U(x)) = x, \qquad \text{length}(U(x)) \le n.$$

Every record longer than n is rejected.

Proof. The length of $U(x)$ is $L(x)$. The budget condition on x therefore selects the successful branch of D_n. Its answer is the tally reconstructed from $L(x)$, which is x by the length round trip. The same length identity proves the preparation bound. A longer input selects the failure branch by definition. □

This construction gives both directions of a representation, together with an exact recovery law. It is stronger than merely counting possible records. It is still a mathematical preparation and reading procedure. Identifying the physical actions that prepare, retain and read it is a further task, addressed in Part V.

10.4 A finite rule can describe an unbounded calculation

We now specify the programs used in this chapter. They are the usual finite expressions for partial recursive functions. A *partial* function may return a natural number on an input, or may remain undefined. Write $c(n)\downarrow v$ when program c returns v on input n. Returning a value is single-valued.

First fix an effective pairing of two natural numbers into one. One explicit choice is

$$\langle a, b\rangle = \begin{cases} b^2 + a, & a < b, \\ a^2 + a + b, & b \le a. \end{cases} \tag{10.3}$$

It lists pairs by successive square boundaries. Given k, set $s = \lfloor \sqrt{k} \rfloor$ and $t = k - s^2$. If $t < s$, recover (t, s); otherwise recover $(s, t - s)$.

To check the round trip, note that $0 \le t \le 2s$. In the first case, $t < s$ selects the first branch of (10.3), giving $s^2 + t = k$. In the second, $0 \le t - s \le s$ selects the second branch, again giving $s^2 + t = k$. Conversely, if $a < b$, the paired number lies between b^2 and $(b + 1)^2$, and its remainder is $a < b$. If $b \le a$, it lies between a^2 and $(a + 1)^2$, with remainder $a + b \ge a$. The decoder therefore returns the original pair in both cases.

Definition 10.5 (The program expressions)**.** There are four basic programs: the constant zero function, successor, and the two projections that recover the first and second members of a paired input. From programs f, g form the following further programs.

1. Pairing runs both on the same input and pairs their answers.
2. Composition runs g and then runs f on its answer.
3. Primitive recursion, on input $\langle a, n\rangle$, starts with $f(a)$ and performs n steps. At step y, with previous answer v, it runs $g(\langle a, \langle y, v\rangle\rangle)$.
4. Search, on input $\langle a, m\rangle$, runs f successively on $\langle a, m\rangle, \langle a, m + 1\rangle, \ldots$, returning the first tested position whose answer is zero.

Each program expression is a finite tree built from these eight constructors.

The partial behavior is part of the definition. Composition cannot proceed until its inner program returns. A recursive calculation cannot proceed past an undefined base or step. Search continues past a position only after that call returns a nonzero answer. If it waits forever at an earlier position, it does not skip ahead to a later zero.

Primitive recursion uses a supplied finite step count. Search need not have such a bound. Both have finite descriptions. Consequently, a finite program description is neither a certificate that execution terminates nor a bound on how long execution takes.

Proposition 10.6 (Expressions capture partial recursion)**.** *A partial function obtained from the four basic functions by pairing, composition, primitive recursion and unbounded search has a program expression. Conversely, every program expression defines such a partial recursive function.*

Proof. Induct on a construction of the partial recursive function. A basic function has the corresponding basic expression. At a pairing, composition or primitive-recursion step, apply that expression constructor to the programs given by the induction hypotheses. Their evaluation clauses give exactly the required function.

For a search beginning at zero, pair the input with zero and pass it to the search expression. This supplies its initial position. In the converse direction, induct on the finite program tree. The basic and three binary constructors are their corresponding closure rules. Search from a supplied position is ordinary search on the offset j, applying the test at $m + j$, then adding m to the successful offset. These are compositions and a search of partial recursive functions. Thus both directions include the stated search behavior. □

Here a total function is called *computable* when its value on every input is given by such a partial recursive function. This is the conventional mathematical definition used for the remainder of the chapter. No premise about the capabilities of every physical process enters that definition.

10.5 Storing the program preserves its execution

The expression tree can itself be assigned a natural number. To remove any ambiguity, here is a complete coding. Give the four basic programs the codes $0, 1, 2, 3$, in the order listed above. If the codes of f, g are u, v, set

$$\begin{aligned} e(\mathrm{pair}(f, g)) &= 4\langle u, v\rangle + 4, \\ e(\mathrm{comp}(f, g)) &= 4\langle u, v\rangle + 6, \\ e(\mathrm{rec}(f, g)) &= 4\langle u, v\rangle + 5, \\ e(\mathrm{search}(f)) &= 4u + 7. \end{aligned} \tag{10.4}$$

For a number at least four, subtract four and inspect its remainder modulo four. Remainders $0, 2, 1, 3$ select pairing, composition, recursion and search, respectively. Divide by four to recover the paired subcodes, or the single subcode for search. Decode the subcodes by the same procedure. Every subcode is smaller than the number being decoded, so the procedure terminates.

Lemma 10.7 (The program coding is reversible)**.** *Decoding $e(c)$ returns the identical finite expression c.*

Proof. Induct on the expression tree. The four basic cases follow from their distinct codes. Each composite code has the remainder selecting its own constructor. Integer division removes that remainder and the initial offset. For a binary constructor, the pairing round trip recovers the two subcodes; for search, it recovers the one subcode directly. Apply the induction hypotheses to those subcodes. The reconstructed constructor and children are exactly the original tree. □

Store c as the tally $P(c)$ with $e(c)$ steps. Recovery first reads its length and then applies the program decoder. This connects the program coding to the orbit constructed in Part I.

Theorem 10.8 (Faithful program storage)**.** *Recovery after storage returns c. In particular, storage is injective, and executing the recovered program has exactly the same partial function as executing c. Both returned answers and undefined behavior are preserved.*

Proof. The length of $P(c)$ is $e(c)$, so the recovery assertion is Lemma 10.7. If two stored tallies are equal, recovering both gives equality of their program expressions. Finally, execution after recovery is execution of the identical expression. Equality of these partial functions preserves undefined inputs as well as defined outputs; no termination assumption is needed. □

For a bit representation, prepare $U(P(c))$. Its length is exactly $e(c)$. The decoder with budget $e(c)$ first recovers $P(c)$ and then recovers c, by Proposition 10.4 and Theorem 10.8. Longer records are refused at the budget check. This unary representation is deliberately simple. Its size bound concerns this specified representation, with no claim that it is the shortest encoding of the program.

10.6 One program for every requested precision

We next apply program storage to rational intervals. The full construction of the continuum from these intervals is the subject of the next chapter. For now we need only the finite data and their consistency conditions.

Definition 10.9 (An interval protocol)**.** An interval protocol p assigns to each natural number n a closed rational interval

$$I_n(p) = [\ell_n(p), u_n(p)]$$

with $\ell_n(p) \le u_n(p)$, such that

$$I_{n+1}(p) \subseteq I_n(p), \qquad u_n(p) - \ell_n(p) \le \frac{1}{n+1}. \tag{10.5}$$

The index is a precision request. The interval at that index gives the specified rational enclosure, and later requests refine it. The definition requires a family with these properties. It does not yet require a method for producing its endpoints.

Fix the usual effective encoding of rational pairs by natural numbers, with an inverse decoder on valid codes. Rational numbers are finite data: they have reduced signed numerators and positive denominators, as developed in Chapter 4. Write $E_p(n)$ for the code of $(\ell_n(p), u_n(p))$. We use the injectivity of this coding when comparing answers.

Definition 10.10 (A uniform program name)**.** A program c names p if

$$\forall n \quad c(n)\!\downarrow E_p(n). \tag{10.6}$$

One finite expression must work at every precision. The nesting and width bounds remain properties of the specified protocol. A terminating program that prints arbitrary rational pairs does not acquire those properties merely by terminating.

Lemma 10.11 (A name specifies at most one protocol)**.** *If a program names both p and q, their interval families are equal.*

Proof. For each n, single-valued execution gives $E_p(n) = E_q(n)$. Injectivity of the pair encoding gives equality of both rational endpoints. Thus the intervals agree at every index, which is equality of the protocol families. □

Different programs may name the same family. The lemma asserts the other direction: a fixed program does not leave the answers undetermined.

The constant protocol $[q, q]$ has a concrete name for every rational q. If k is the code of (q, q), start with the zero program and compose with successor k times. Induction on k shows that this finite expression returns k on every input. The interval conditions hold because its width is zero and every interval is the same.

Theorem 10.12 (The criterion for an executable name)**.** *An interval protocol has a uniform program name if and only if its endpoint code function $n \mapsto E_p(n)$ is computable. The same criterion holds when the name must be stored as a distinction tally and recovered before execution.*

Proof. If c names p, its partial evaluation is the total function returning $E_p(n)$ at every n. Proposition 10.6 therefore gives computability. Conversely, that proposition supplies a finite expression for a computable endpoint function, and equality with that total function is precisely (10.6).

For the storage assertion, store a supplied name as $P(c)$ and use Theorem 10.8. In the reverse direction, recover the program from a purported stored name. Its execution is the required uniform name. Storage has changed the representation without changing the criterion. □

10.7 A finite prefix is a different quantifier

Suppose we ask only for the first N endpoint pairs. These form a finite table

$$T_{p,N} = (E_p(0), E_p(1), \dots, E_p(N-1)).$$

Given this table, a program can return its nth entry when $n < N$, and return zero otherwise. Finite list lookup is primitive recursive: it follows the list until the requested position or its end, with a default answer at the end. Its fixed entries are finite constants. Thus the expression construction of Proposition 10.6 gives a program for the lookup.

Theorem 10.13 (Every supplied finite prefix has a program)**.** *For every interval protocol p and every N, there is a program c such that*

$$\forall n < N \quad c(n)\!\downarrow E_p(n).$$

The program may depend on the entire supplied finite table.

Proof. Use the lookup program just constructed from $T_{p,N}$. Below N, its entry at n is $E_p(n)$ by the definition of the table. The default branch is irrelevant to the stated range of inputs. □

The premise is access to the table as data. The proof supplies no method for extracting it from an unknown physical quantity. Nor does the theorem supply a procedure that takes N and generates the required table.

For a fixed p, compare the two assertions:

$$\forall N\, \exists c\, \forall n < N \quad c(n)\!\downarrow E_p(n),$$
$$\exists c\, \forall n \quad c(n)\!\downarrow E_p(n).$$

The first permits a different finite table and program at each budget. The second requires one program for all budgets. Chapter 11 will prove that the first does not imply the second for general interval protocols in the classical real display. The proof will use a precise countability obstruction, not a failure to find a clever program.

This is the point at which finite construction needs uniformity. Every individual tally has a finite stage. Every finite program has a tally representation that preserves its execution. A whole family of rational intervals has such a program name exactly when its endpoint function is computable. Those are three established statements, each with its own object and its own quantifiers.

These program expressions have a second use. The terms of the typed language of Chapter 9 can be compiled into them and stored as numbers by (10.4). Identity proofs compile to the program that returns 0, identity elimination to the program of its base term, elimination from the empty type to a default program fixed by its target type, and E in the same way as R. A function is then a program code, applied through one fixed universal program that searches for a finite record of the computation and checks each candidate with one fixed test program. Chapter 20 uses this to compare the strongest typed fragment, $\mathscr{E}^+$, with Peano arithmetic: it proves the negative translations of exactly the theorems of Peano arithmetic. One direction of that comparison translates Peano derivations into typed terms. The other runs the typed terms as compiled programs. It needs derivations, by the arithmetic rules, that the test accepts a record of every computation and accepts only genuine records; that chapter shows how the record and its test are built so that the rules can supply them.

Questions to work through

1. In the proof of Theorem 10.3, identify the particular inclusion used for an input $x \in F_n$. Where is compatibility needed, and where is the own-budget construction sufficient?

2. Explain why increasing both input budgets in an addition changes neither its underlying inputs nor its result. What quantity is bounded by $n + m$, and what quantity has not been bounded?

3. Follow storage, bit preparation, budgeted decoding and program recovery for an arbitrary program c. Cite the round-trip identity used at each step. Why does this argument need no assertion that c terminates?

4. A program prints correct endpoint pairs for the first N requests and then prints zeros. Which conclusion of this chapter can it satisfy? Which additional obligation is required for it to be a uniform name?

Chapter 11

The continuum through rational enclosures

A rational interval tells us both an answer and how much remains unresolved. If a length lies between two rational bounds, narrowing those bounds improves the description without pretending that a finite record contains every digit. The question is how a coherent sequence of such records relates to a real number, and what is lost when only that number is retained.

We use the interval protocols of Definition 10.9. Their intervals are nested and have width at most $1/(n+1)$ at request n. The arithmetic of the endpoints is rational arithmetic. A protocol is a mathematical family of these intervals; a uniform executable name is the additional structure studied in Chapter 10.

This chapter compares protocols with the ordinary classical real line. That comparison uses its completeness and Archimedean property, together with classical reasoning. These are resources of the surrounding mathematics in which the comparison is proved. They are not new rules silently inserted into the object languages of Part II.

11.1 Why every later interval constrains every earlier one

Fix a protocol p, and abbreviate its interval by $I_n = [\ell_n, u_n]$. Regard its rational endpoints as real numbers when making the comparison. Nesting says

$$\ell_n \le \ell_{n+1}, \qquad u_{n+1} \le u_n.$$

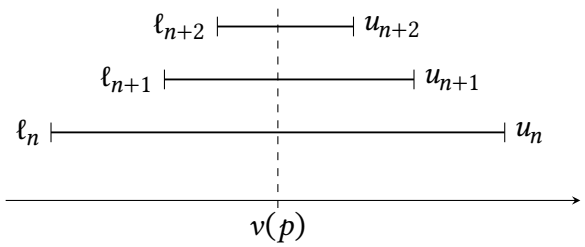


Figure 11.1. Nesting retains earlier constraints. The displayed value lies in every interval; the width condition makes that value unique. The drawing shows the containment pattern, not a prescribed sequence of widths.

Induction on the distance between indices therefore gives

$$a \le b \quad \Longrightarrow \quad \ell_a \le \ell_b \quad \text{and} \quad u_b \le u_a.$$

The indices in a lower and upper bound need not be the same. For any a, b, let $k = \max(a, b)$. Then

$$\ell_a \le \ell_k \le u_k \le u_b. \tag{11.1}$$

This simple observation is the reason a single value can satisfy every interval. Every upper endpoint bounds all the lower endpoints, even those arising at much later requests.

11.2 A unique value lies in every enclosure

The set of lower endpoints is nonempty and bounded above by u_0, using (11.1). Completeness of the classical real line therefore supplies its least upper bound. Define the *displayed value* by

$$v(p) = \sup\{\ell_n : n \in \mathbb{N}\}. \tag{11.2}$$

The word "displayed" identifies the comparison being made: one real number is retained from the more detailed interval family.

Proposition 11.1 (The value satisfies every finite enclosure)**.** *For every n,*

$$\ell_n \le v(p) \le u_n.$$

Proof. Every member of a set is at most its supremum, so $\ell_n \le v(p)$. Equation (11.1) says that u_n is an upper bound for the entire set of lower endpoints. The least-upper-bound property then gives $v(p) \le u_n$. □

Existence used nesting and completeness. Uniqueness uses the quantitative width bound. We first isolate the small estimate that will be used twice.

Lemma 11.2 (No positive quantity is below every reciprocal bound)**.** *If a real number a satisfies $0 \le a \le 1/(n+1)$ for every natural number n, then $a = 0$.*

Proof. If a were positive, the Archimedean property would give a natural number $n > 1/a$. Hence $n + 1 > 1/a$, and taking reciprocals of these positive numbers would give $1/(n+1) < a$, contradicting the supplied bound. Nonnegativity then leaves $a = 0$. □

Theorem 11.3 (Uniqueness of the displayed value)**.** *If y belongs to every interval of p, then $y = v(p)$.*

Proof. Both y and $v(p)$ lie between ℓ_n and u_n. Therefore

$$0 \le |y - v(p)| \le u_n - \ell_n \le \frac{1}{n+1}$$

for every n. Lemma 11.2 gives $|y - v(p)| = 0$, hence equality. □

This result provides a useful way to prove an identity of displayed values. Instead of calculating a supremum directly, show that the proposed answer lies in every interval. The uniqueness theorem then identifies it with the displayed value.

11.3 When two protocols describe the same point

Two different interval families may retain different finite information while describing the same limiting point. Exact equality of the families would be too strong for this comparison.

Definition 11.4 (Observational equality)**.** Protocols p and q are observationally equal, written $p \simeq q$, if their intervals overlap at every matching precision:

$$\forall n \quad \ell_n(p) \le u_n(q) \quad \text{and} \quad \ell_n(q) \le u_n(p). \tag{11.3}$$

Each pair of inequalities is a rational comparison. The definition requires all of them. It is a mathematical condition on two families, not a claim that a finite inspection decides the condition in every case.

Theorem 11.5 (Overlap at every precision is exactly equality of value)**.** *For all interval protocols p, q,*

$$p \simeq q \quad \Longleftrightarrow \quad v(p) = v(q).$$

Proof. Suppose first that the intervals overlap at request n. Since each value lies in its own interval, one direction of the difference satisfies

$$\begin{aligned} v(p) - v(q) &\le u_n(p) - \ell_n(q) \\ &\le \big(u_n(p) - \ell_n(p)\big) + \big(u_n(q) - \ell_n(q)\big). \end{aligned}$$

The second inequality uses $\ell_n(p) \le u_n(q)$. Interchanging p and q gives the same bound for the negative difference. Consequently,

$$|v(p) - v(q)| \le \frac{2}{n+1}.$$

Apply this bound at request $2m+1$. Its right side is $1/(m+1)$, so Lemma 11.2 implies equality of the values.

Conversely, if the values are equal, their common value belongs to both intervals at every request. Thus $\ell_n(p) \le v(p) = v(q) \le u_n(q)$, and the reverse cross inequality follows in the same way. These are precisely the overlap conditions. □

In particular, $\simeq$ is reflexive, symmetric and transitive: each property follows by translating to equality of values and back. Overlap of arbitrary single intervals is not what proves transitivity here. The theorem uses overlap at every precision, nesting and shrinking widths together.

There are therefore two distinct equalities in use. Equality of protocols retains every rational answer. Observational equality retains the displayed point. Chapter 10's uniqueness of a program name concerns the first equality, because the program returns the exact endpoint pair.

11.4 Rationals and arithmetic act on the intervals

A rational q determines the constant protocol $[q, q]$. It contains q at every request and has width zero, so its displayed value is q by Theorem 11.3. Two such constant protocols are observationally equal exactly when their rational values are equal. The rational arithmetic constructed earlier is thus preserved faithfully in this comparison.

For addition, the main issue is the precision budget. Adding intervals adds their widths. Reading both operands at request n would give only the bound $2/(n+1)$. Instead, read both at request $k = 2n+1$ and define

$$I_n(p+q) = [\ell_{2n+1}(p) + \ell_{2n+1}(q),\ u_{2n+1}(p) + u_{2n+1}(q)]. \tag{11.4}$$

Proposition 11.6 (Addition respects the protocol conditions)**.** *The intervals in* (11.4) *form an interval protocol, and*

$$v(p+q) = v(p) + v(q).$$

Proof. The lower endpoint does not exceed the upper endpoint because that is true for both operands. As n increases, the index $2n+1$ increases; monotonicity of both lower endpoint sequences and antitonicity of both upper sequences give nesting of the sum intervals.

The width of the sum interval is the sum of the two operand widths. Hence it is at most

$$\frac{1}{2n+2} + \frac{1}{2n+2} = \frac{1}{n+1}.$$

Finally, add the two lower bounds and the two upper bounds for the operand values. They show that $v(p) + v(q)$ lies in every sum interval. Apply Theorem 11.3 to identify its value. □

Negation reverses and negates the endpoints:

$$I_n(-p) = [-u_n(p), -\ell_n(p)]. \tag{11.5}$$

The new interval has the same width. The two nesting inequalities exchange roles under sign reversal, so it is again a protocol. Negating the value bounds shows that $-v(p)$ lies in every new interval, giving

$$v(-p) = -v(p).$$

Define $p - q = p + (-q)$. The two established value identities then give

$$v(p-q) = v(p) - v(q).$$

These constructions work on the rational enclosures themselves. Their agreement with real addition, negation and subtraction is proved afterward. The sum's reindexing is also visible: its finite answers need not be the same interval family as another construction with the same displayed sum. Observational equality is the comparison that retains the common value.

11.5 Every classical real has such a display

We have sent each protocol to a real number. The reverse coverage statement has an explicit interval formula. Fix a classical real r, and let

$$k_n = \lfloor r 2^n \rfloor, \qquad I_n^r = \left[\frac{k_n}{2^n}, \frac{k_n + 1}{2^n}\right]. \tag{11.6}$$

The endpoints are rational because k_n is an integer. The floor inequalities show that r lies in each interval.

Lemma 11.7 (The next dyadic floor stays in the previous interval)**.** *For every n,*

$$2k_n \le k_{n+1} \le 2k_n + 1.$$

Proof. The definition of floor gives $k_n \le r2^n < k_n + 1$. Multiply by two:

$$2k_n \le r2^{n+1} < 2k_n + 2.$$

Since both k_{n+1} and the bounds are integers, taking the floor gives the lower bound $2k_n$ and the upper bound $2k_n + 1$. □

Theorem 11.8 (Coverage of the classical real line)**.** *The family in* (11.6) *is an interval protocol with displayed value r. Consequently every classical real is the displayed value of an interval protocol.*

Proof. The lower dyadic bound in Lemma 11.7, divided by 2^{n+1}, says that the new lower endpoint is at least the old one. The upper bound, after adding one and dividing by 2^{n+1}, says that the new upper endpoint is at most the old one. This is nesting.

The interval width is 2^{-n}. We have $n + 1 \le 2^n$ by induction: it is equality at zero, and from $n + 1 \le 2^n$ and $1 \le 2^n$ we obtain $n + 2 \le 2^n + 2^n = 2^{n+1}$. Taking reciprocals proves $2^{-n} \le 1/(n + 1)$, the required width bound.

The floor inequalities put r in every interval, and uniqueness of the value gives $\nu(I^r) = r$. □

The formula takes r as a supplied classical real. Evaluating an exact floor at every request is not, by itself, a procedure for extracting these integers from an unknown quantity. The theorem establishes coverage by general interval families. A finite program name still requires the computability condition of Theorem 10.12.

Remark 11.9 (What the floor digits cost)**.** For $0 \le r < 1$, the integers k_n of (11.6) record the binary digits of r. Here $k_0 = 0$, and by Lemma 11.7 each digit $k_{n+1} - 2k_n$ is 0 or 1: it says which half of the previous interval contains r. The decision between the halves is classical. On the line $\mathbb{R}_\delta$ of Chapter 12, which assumes no classical completeness, the statement that every point of $[0, 1]$ has a binary expansion is equivalent to LLPO given countable choice, and binary expansions for all points of $[0, 1]$ already yield LLPO with no choice at all (Section 12.9). The obstacle is a point that may lie on a dyadic boundary: choosing its half requires a decision that no finite stage supplies. If a digit may also be -1, consecutive intervals overlap and the obstacle disappears. Every point of $[-1, 1]$ has a signed-digit expansion, each digit read off one stage of one sequence naming the point, with no omniscience principle and no choice principle.

Taken together, the results identify exactly what the display retains. It covers the classical real line; it identifies two protocols exactly when they overlap at every precision; and it preserves the rational embedding, addition, negation and subtraction constructed above. None of those assertions requires identifying all interval protocols with finite programs.

11.6 Why finite prefixes do not provide a uniform name

We can now prove the promised distinction between finite data at each precision and one program for every precision. The argument has two steps. There are only countably many program-named values. The classical real line has more than countably many points.

Lemma 11.10 (There are only countably many named values)**.** *The set of displayed values of uniformly program-named interval protocols is countable.*

Proof. A named protocol retains a finite program and the protocol it names. For a fixed program, Lemma 10.11 permits at most one such protocol. Retaining the program is therefore an injection of named protocols into the finite program expressions. Those expressions inject into the natural numbers by Lemma 10.7. Named protocols are consequently countable. Their displayed values form an image of a countable set, so that image is countable too. □

For completeness, the uncountability needed here can be seen directly from Cantor's argument. Binary sequences cannot be listed as $b_0, b_1, b_2, \ldots$: the sequence whose nth bit is the opposite of $b_n(n)$ differs from every sequence at its corresponding position.

These sequences inject into the classical real line. For a binary sequence b, with its bits regarded as 0 or 1, use the convergent sum

$$C(b) = \sum_{n=0}^{\infty} b(n)\, 3^{-n}. \tag{11.7}$$

Convergence follows by comparison with the geometric series. If two sequences first differ at position k, their contributions before k agree. The contribution at k differs by 3^{-k}, while all later contributions together can change the difference in the opposite direction by at most

$$\sum_{n>k} 3^{-n} = \frac{1}{2}\, 3^{-k}.$$

The first difference therefore cannot be canceled. The map is injective, and a countable real line would make the binary sequences countable as well, contradicting the diagonal argument. This is a classical cardinality argument; the infinite sums and the deduction from distinct sequences to a first differing position are used in that setting.

Theorem 11.11 (The finite-prefix and uniform-name boundary)**.** *In the classical comparison of this chapter, every interval protocol has a program for each supplied finite prefix, but some interval protocol has no single program naming all its intervals. The same failure holds if names are required to be stored in distinction tallies.*

Proof. The first assertion is Theorem 10.13. By Lemma 11.10 and the uncountability just proved, choose a real r outside the set of named values. By Theorem 11.8, it is the value of its dyadic protocol p. If p had a uniform name, its value r would belong to the excluded countable set, a contradiction.

If p had a stored program name, recovering it would give an ordinary uniform name by Theorem 10.8. The same contradiction therefore applies after storage. □

The result rules out a specific quantifier reversal. It does not say that a particular unnamed protocol has been obtained by a finite measurement, or that a physical process has performed a noncomputable operation. It concerns the difference between all families allowed by a mathematical definition and the families produced by finite programs.

There is a further question even for executable families. Producing successively narrower intervals is one task; deciding exact equality or choosing the sign of an arbitrary limiting quantity is another. The next chapter asks what additional logical principles those decisions require. It also takes apart the completeness used in this chapter. On a line built from the ratios of Chapter 4, with no classical real line assumed, limits, least upper bounds and suprema each receive their own exact price.

Questions to work through

1. Prove the cross-endpoint inequality (11.1) when $a \leq b$, then when $b \leq a$. Explain how the maximum-index proof combines the two cases.

2. In Theorem 11.5, identify where overlap is used and where the width bound is used. Why is the request $2m + 1$ convenient?

3. Check both nesting inequalities for the negated protocol. Then explain why proving enclosure of $-\nu(p)$ avoids a direct calculation with suprema.

4. Trace the contradiction in Theorem 11.11. Which result supplies a protocol for the selected real, and which result shows that a name for this protocol is impossible?

Chapter 12

Exact comparison has a cost

Suppose successive measurements enclose a quantity ever more closely, yet zero remains inside every enclosure received so far. More precision can eventually reveal that the quantity is positive or negative. What would justify a procedure that always decides whether it is exactly zero?

The difficulty is the word "always." A finite calculation can settle a strict inequality once the separation is larger than its error. An exact zero decision must also settle the case in which no separating witness ever appears. This chapter identifies the additional logical principle required by each of four precise forms of comparison: deciding exact zero, producing a witnessed sign, choosing one of two sides, and turning a proof that a quantity is not zero into a sign. Each result has two directions: the principle supplies the comparison, and the comparison supplies the principle. The chapter then asks the same question of completeness, the property of the line that the classical comparison of Chapter 11 took for granted.

We use a regular dyadic presentation to keep every finite comparison in integer arithmetic. This is a specified presentation with its own error condition. The classical completeness argument of Chapter 11 is not used in the proofs below. Nor do we silently transfer an operation between this presentation and the interval protocols of that chapter. Section 12.8 shows that nothing depends on the dyadic denominators: rounding matches these presentations, point for point, with regular sequences of the ratios of Chapter 4, and every price proved here holds there unchanged.

12.1 Finite stages and exact predicates

A dyadic rational is an integer divided by a power of two. At stage n, write the approximation as

$$x_n = \frac{q_n}{2^n}, \qquad q_n \in \mathbb{Z}.$$

The numerator is finite data. A sequence of arbitrary numerators need not describe one coherent quantity, so the presentation includes a regularity condition.

Definition 12.1 (Regular dyadic presentation)**.** A regular dyadic presentation consists of a sequence $(q_n)_{n\ge 0}$ of integers such that, for every m, n,

$$|q_m 2^n - q_n 2^m|(m+1)(n+1) \le 2^{m+n}(m+n+2). \tag{12.1}$$

In rational notation, the same condition reads

$$|x_m - x_n| \le \frac{1}{m+1} + \frac{1}{n+1}. \tag{12.2}$$

The rational notation abbreviates the integer condition: multiply by the positive denominator $2^{m+n}(m+1)(n+1)$ to recover (12.1).

We will write x for the whole presentation. This notation does not discard the numerators or their regularity. In particular, each decision made at a specified stage uses the numerator at that stage.

Definition 12.2 (Five comparisons with zero)**.** For a regular dyadic presentation x, define

$$\begin{aligned} Z(x) &: \quad \forall n,\ |q_n|(n+1) \le 2 \cdot 2^n, \\ P(x) &: \quad \exists n,\ 2 \cdot 2^n < q_n(n+1), \\ N(x) &: \quad \exists n,\ q_n(n+1) < -2 \cdot 2^n, \\ U(x) &: \quad \forall n,\ -2 \cdot 2^n \le q_n(n+1), \\ D(x) &: \quad \forall n,\ q_n(n+1) \le 2 \cdot 2^n. \end{aligned}$$

These are, respectively, the zero predicate, positive and negative witness predicates, and nonnegative and nonpositive predicates for this presentation.

In rational notation, $Z(x)$ says that every stage lies within $2/(n+1)$ of zero. The predicate $P(x)$ supplies a stage above this tolerance; $N(x)$ supplies one below its negative. The predicates $U(x)$ and $D(x)$ say that all stages respect the corresponding one-sided bounds. Their letters will keep the full quantified statements visible while we compare their logical strength.

This is the link with the enclosure picture of the preceding chapter. Finite data are tested against a stated error allowance. A positive or negative witness clears that allowance at a particular stage; the zero predicate keeps every stage within a shrinking band around zero. The nonnegative and nonpositive predicates prohibit witnesses on the other side. The regularity condition keeps the stages consistent enough that opposite witnesses cannot both occur, as the next lemma proves. These names refer to the displayed predicates of this presentation; using them does not invoke a map from an interval protocol or a completed classical real.

The factor two is part of these definitions. The theorems concern this regularity condition and these thresholds. They do not claim that an arbitrary choice of tolerances would give the same statements without a separate argument.

Every inequality at a fixed index is decidable by finite integer arithmetic. The quantifiers are where the questions change. To establish $P(x)$ it suffices to produce one index. To establish $Z(x)$ requires a reason that works at every index. Failing to find a witness in the stages already examined supplies neither such a reason nor an exact zero decision.

Lemma 12.3 (Opposite witnesses are incompatible)**.** *A regular dyadic presentation cannot satisfy both $P(x)$ and $N(x)$.*

Proof. Suppose m witnesses $P(x)$ and n witnesses $N(x)$. In rational notation,

$$x_m > \frac{2}{m+1}, \qquad x_n < -\frac{2}{n+1}.$$

Subtracting gives

$$x_m - x_n > \frac{2}{m+1} + \frac{2}{n+1}.$$

The right side is strictly greater than $1/(m+1)+1/(n+1)$, whereas regularity bounds $|x_m - x_n|$ by that smaller quantity. This is a contradiction. All denominators here are positive integers. Clearing them gives the same contradiction entirely in integer arithmetic. □

This lemma will do essential work in the proof of Theorem 12.8. Without regularity, a list could alternate between large positive and negative values, and that argument would fail.

12.2 Questions about a binary sequence

Let $\alpha : \mathbb{N} \to \{0, 1\}$ be a binary sequence. The value at each specified index is decidable. The principles WLPO, LPO and LLPO of Definition 7.8 in Chapter 7, together with the binary form $\mathrm{MP}_{\mathrm{bin}}$ of Markov's principle stated there, concern what can be concluded about all its values together.

WLPO decides whether the all-zero assertion holds. Its negative answer does not itself provide an index. LPO supplies an index in its second branch. LLPO has a different premise and conclusion: with at most one event, one parity class is entirely free of events. It need not tell us whether an event occurs in the other class. $\mathrm{MP}_{\mathrm{bin}}$ decides nothing: it turns a proof that not every value is zero into an index whose value is one.

The binary form of LPO makes the witness distinction of Chapter 6 concrete. These are principles about arbitrary sequences in the stated domain; there is no unmentioned restriction to sequences generated by a selected class of programs.

12.3 Encoding an event as a small jump

The reverse implications need a construction that turns information about a binary sequence into information about a regular dyadic presentation. It must be possible to construct each finite stage without first deciding whether the sequence ever contains a 1.

At stage n, inspect only

$$\alpha(0), \alpha(1), \dots, \alpha(n).$$

If they are all zero, report that no event has yet been found. Otherwise, report the least index $N \le n$ with $\alpha(N) = 1$. This is a bounded search. Once such a first index has been found, every later search reports the same index: all smaller entries have already been checked.

Choose a sign $\sigma(N) \in \{-1, 1\}$ for each possible first index. Define the jump numerators by

$$q_n = \begin{cases} 0, & \text{if no event has been found through stage } n, \\ \sigma(N)2^{n-N}, & \text{if the first event is at } N \le n. \end{cases} \tag{12.3}$$

Thus the rational stages are zero before the first event and $\sigma(N)2^{-N}$ from then on. The exponent $n - N$ in the numerator is nonnegative whenever that branch is used. No unbounded search is required to produce q_n.

Lemma 12.4 (The jump presentation is regular)**.** *For every binary sequence and every choice of signs of absolute value one,* (12.3) *defines a regular dyadic presentation.*

Proof. Compare stages m and n. If neither bounded search has found an event, both stages are zero. If both have found one, they report the same first index, so both rational stages equal $\sigma(N)2^{-N}$. In either case their difference is zero.

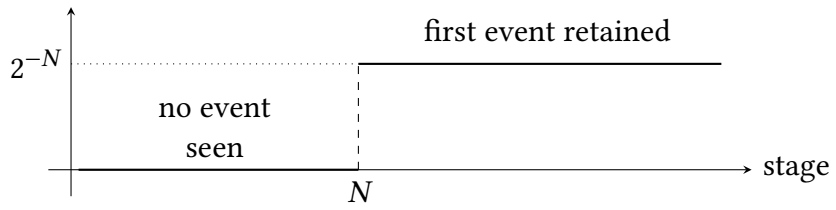


Figure 12.1. A positive jump presentation. The first event fixes the later rational stage. A later first event produces a smaller jump; the drawing shows the construction, with no claim that the vertical scale is uniform across different choices of N.

In the remaining case, interchange m, n if necessary so that the search at m has found no event and the search at n has found its first event at N. Then $m < N \le n$, and the absolute difference between the rational stages is 2^{-N}.

For every positive integer N, $N < 2^N$. This follows by induction: $1 < 2$, and $N+1 \le 2N < 2^{N+1}$ when $N \ge 1$ and $N < 2^N$. Since $m + 1 \le N$, it follows that

$$2^{-N} < \frac{1}{m+1} \le \frac{1}{m+1} + \frac{1}{n+1}.$$

This is the required regularity bound. The case $N = 0$ cannot occur in this branch, since it requires $m < N$. □

Let $j_{\sigma,\alpha}$ denote this presentation. Its decisive property is an exact relationship between its zero predicate and the original sequence.

Lemma 12.5 (A jump is zero exactly when no event occurs)**.** *For every allowed sign sequence,*

$$Z(j_{\sigma,\alpha}) \quad \Longleftrightarrow \quad \forall k,\ \alpha(k) = 0.$$

Proof. If every entry is zero, every jump numerator is zero, so every inequality in $Z(j_{\sigma,\alpha})$ holds.

Conversely, suppose $Z(j_{\sigma,\alpha})$. Fix an index k. Its binary value is decidable. If $\alpha(k) = 1$, bounded search through k supplies a first event $N \le k$. Choose the later stage

$$n_0 = 2^{N+1}.$$

This stage is at least N, and its rational absolute value is 2^{-N}. The zero predicate at n_0 would require

$$\frac{1}{2^N} \le \frac{2}{2^{N+1}+1}.$$

Multiplying by the positive denominators gives $2^{N+1} + 1 \le 2^{N+1}$, a contradiction. Therefore $\alpha(k) = 0$. Since k was arbitrary, every entry is zero. □

The displayed stage is a sufficient witness, not a claim about an optimal running time. Its purpose is to show exactly how a nonzero jump eventually exceeds the shrinking zero tolerance.

12.4 Exact zero decisions are equivalent to WLPO

Theorem 12.6 (The logical price of an exact zero decision)**.** *The following statements are equivalent:*

1. *Every regular dyadic presentation x satisfies $Z(x) \vee \neg Z(x)$.*
2. *WLPO holds for every binary sequence.*

Proof. Suppose first that the exact zero decision is available. Given α, construct its positive jump presentation, using $\sigma(N) = 1$ for every N. Apply the decision to it. By Lemma 12.5, the zero branch gives $\forall n, \alpha(n) = 0$. In the other branch, that all-zero assertion would imply the zero predicate, contradicting the decision. Thus it gives $\neg\forall n, \alpha(n) = 0$. This is WLPO.

Conversely, assume WLPO and let x be any regular dyadic presentation. Form a binary sequence of tolerance violations:

$$b(n) = \begin{cases} 1, & 2 \cdot 2^n < |q_n|(n+1), \\ 0, & |q_n|(n+1) \le 2 \cdot 2^n. \end{cases}$$

Each value is obtained by a decidable integer comparison. The assertion that every $b(n)$ is zero is exactly $Z(x)$. WLPO applied to b therefore gives $Z(x) \vee \neg Z(x)$. □

The second branch is still only a negative assertion. Asking instead for an exhibited positive or negative stage is a stronger request. The next theorem prices the full three-way answer; Theorem 12.11 prices the passage from the negative assertion to a sign on its own.

12.5 Sign trichotomy is equivalent to LPO

Theorem 12.7 (The logical price of a witnessed sign)**.** *The following statements are equivalent:*

1. *Every regular dyadic presentation x satisfies $P(x) \vee Z(x) \vee N(x)$.*
2. *LPO holds for every binary sequence.*

Proof. Assume the three-way comparison. Given α, construct its positive jump presentation. A zero answer gives that every entry of α is zero, by Lemma 12.5. A negative answer is impossible: every numerator of this presentation is nonnegative.

A positive answer supplies a stage n whose numerator exceeds the positive tolerance. The bounded search through n must therefore have found an event; otherwise the numerator would be zero. Its reported index N satisfies $\alpha(N) = 1$. These branches give LPO.

Conversely, assume LPO and use the violation sequence b from the preceding proof. If all entries vanish, $Z(x)$ holds. Otherwise LPO supplies an index n such that

$$2 \cdot 2^n < |q_n|(n+1).$$

The sign of the integer q_n is decidable. If $q_n \ge 0$, the inequality is a positive witness. If $q_n \le 0$, it becomes $q_n(n+1) < -2 \cdot 2^n$, a negative witness. Thus the required three-way comparison follows. □

Both reverse constructions use the same event encoding. What changes is the information returned by the comparison: a negative zero assertion in one case, an actual separating stage in the other.

12.6 One-sided order is equivalent to LLPO

The third comparison asks only for $U(x) \vee D(x)$: every presentation is nonnegative or nonpositive in the precise sense of Definition 12.2. It does not require a strict sign witness or a zero decision.

Theorem 12.8 (The logical price of order dichotomy)**.** *The following statements are equivalent:*

1. *Every regular dyadic presentation x satisfies $U(x) \vee D(x)$.*
2. *LLPO holds for every binary sequence with at most one event.*

Proof. Assume the order dichotomy. Let β have at most one entry equal to 1. Construct its jump presentation using the parity sign

$$\sigma(N) = \begin{cases} 1, & N \text{ even}, \\ -1, & N \text{ odd}. \end{cases}$$

If $U(j_{\sigma,\beta})$ holds, every odd entry of β must vanish. Indeed, an event at an odd index N would be its unique event, hence its first. At stage $n_0 = 2^{N+1}$ the rational value would be -2^{-N}. The nonnegative predicate would require

$$-\frac{2}{n_0 + 1} \leq -\frac{1}{2^N},$$

contradicting the same denominator calculation as in Lemma 12.5. The decidable binary value at each odd index is consequently zero. Similarly, $D(j_{\sigma,\beta})$ forces every even entry to vanish. These alternatives give LLPO.

For the converse, assume LLPO and fix a regular presentation x. First make a sequence of signed events by interleaving its finite comparisons:

$$\begin{aligned} e(2k) = 1 \quad &\Longleftrightarrow \quad 2 \cdot 2^k < q_k(k+1), \\ e(2k+1) = 1 \quad &\Longleftrightarrow \quad q_k(k+1) < -2 \cdot 2^k. \end{aligned}$$

Put zero in each remaining position. There may be many positive events, or many negative events. LLPO requires at most one event, so applying it directly to e would be invalid.

Instead retain only the first event. Define $f(j) = 1$ precisely when the bounded search through $e(0), \ldots, e(j)$ reports its first event at j. Put $f(j) = 0$ otherwise. Uniqueness of a first index proves that f has at most one event. LLPO can now be applied to f.

Suppose its even positions are all zero. We show $D(x)$. Fix k and decide the integer inequality defining $D(x)$ at k. If it holds, there is nothing more to prove. If it fails, k is a positive witness, so $e(2k) = 1$. Bounded search through $2k$ supplies a first event at an index j, with $f(j) = 1$.

If j is even, this contradicts the LLPO branch. If j is odd, it supplies a negative witness for x, contradicting the positive witness by Lemma 12.3. Hence the inequality must hold. Since k was arbitrary, $D(x)$ follows.

If instead all odd positions of f are zero, the symmetric argument gives $U(x)$. A hypothetical negative witness would force a first event that is either forbidden by that branch or is an incompatible positive witness. Thus $U(x) \vee D(x)$ follows in both cases. □

Two details make this proof work. Taking only the first event supplies the premise required by LLPO. Regularity then excludes a first event of the opposite sign. Neither step can be replaced by an appeal to the visual appearance of a converging sequence.

12.7 The difference between denying and finding

The form of Markov's principle used here is the binary form $\mathrm{MP}_{\mathrm{bin}}$ of Definition 7.8: for every binary sequence, $\neg(\forall n, \alpha(n) = 0)$ implies $\exists n, \alpha(n) = 1$. We keep to this form rather than silently changing between formulations of Markov's principle.

Proposition 12.9 (LPO separates into a decision and a witness principle)**.** *For the binary principles of Definition 7.8,*

$$\mathrm{LPO} \iff \mathrm{WLPO} \wedge \mathrm{MP}_{\mathrm{bin}}.$$

Moreover, LPO implies LLPO.

Proof. LPO gives either an all-zero assertion or an event. An event contradicts the all-zero assertion, so these alternatives give WLPO. If the all-zero assertion is already denied, LPO's all-zero branch is impossible, leaving an event. This gives $\mathrm{MP}_{\mathrm{bin}}$.

Conversely, WLPO decides the all-zero assertion. Keep its positive branch; apply $\mathrm{MP}_{\mathrm{bin}}$ in its negative branch to obtain an event. The resulting alternatives are exactly LPO.

Finally, suppose a sequence has at most one event. Apply LPO to its subsequence of even entries. If every even entry vanishes, one LLPO branch is already established. Otherwise there is an even event. An odd event would then be a different second event, contrary to the premise. Since each odd entry is binary, every odd entry is zero. This gives the other LLPO branch. □

WLPO's negative branch, applied to a presentation, gives only $\neg Z(x)$. Here is what that leaves undone. Let α be a binary sequence that is known not to be all zero, and let j be its positive jump presentation. Lemma 12.5 gives $\neg Z(j)$. Yet every stage of j before the first event is 0, so a stage above the positive tolerance can be exhibited only once an event of α has been found. For this presentation, turning $\neg Z(j)$ into $P(j)$ is the same task as turning $\neg(\forall n, \alpha(n) = 0)$ into an event. We now show that the passage from a proof that a presentation is not zero to an actual sign costs exactly $\mathrm{MP}_{\mathrm{bin}}$ in general. First we identify exactly what such a proof says.

Lemma 12.10 (Zero is the absence of both witnesses)**.** *For every regular dyadic presentation x,*

$$Z(x) \quad \Longleftrightarrow \quad \neg P(x) \wedge \neg N(x).$$

Consequently $\neg Z(x)$ holds exactly when $\neg\neg\big(P(x) \vee N(x)\big)$ holds.

Proof. Suppose $Z(x)$. A positive witness at stage n would give

$$2 \cdot 2^n < q_n(n+1) \leq |q_n|(n+1),$$

contrary to the zero bound $|q_n|(n+1) \leq 2 \cdot 2^n$ at the same stage. A negative witness fails in the same way, with $-q_n$ in place of q_n.

Conversely, suppose that neither witness exists, and fix n. The zero bound at n is a decidable integer comparison. If it failed, the sign of the integer q_n would turn the failure into a positive witness, when $q_n \geq 0$, or a negative witness, when $q_n \leq 0$, exactly as in the proof of Theorem 12.7. Both are excluded, so the bound holds at n. Since n was arbitrary, $Z(x)$ follows.

The equivalence says that $Z(x)$ is $\neg(P(x) \vee N(x))$. Negating both sides gives the second statement. □

A proof that x is not zero is therefore exactly a proof that a sign witness cannot be ruled out. Removing that double negation is the content of the next theorem.

Theorem 12.11 (The logical price of a sign for a nonzero quantity)**.** *The following statements are equivalent:*

1. *Every regular dyadic presentation x satisfies $\neg Z(x) \to P(x) \vee N(x)$.*
2. $\mathrm{MP}_{\mathrm{bin}}$ *holds for every binary sequence.*

Proof. Assume $\mathrm{MP}_{\mathrm{bin}}$, and let x satisfy $\neg Z(x)$. Form the violation sequence b from the proof of Theorem 12.6. The assertion that every $b(n)$ is zero is exactly $Z(x)$, so $\neg Z(x)$ says that b is not all zero. $\mathrm{MP}_{\mathrm{bin}}$ supplies an index n with $b(n) = 1$. The sign of q_n then gives a positive or a negative witness at n, as in the converse half of Theorem 12.7.

Conversely, assume the first statement, and let α be a binary sequence that is not all zero. Construct its positive jump presentation j, with $\sigma(N) = 1$ for every N. By Lemma 12.5, $Z(j)$ would make every entry of α zero, so $\neg Z(j)$. The first statement gives $P(j) \vee N(j)$. The negative branch is impossible because every numerator of j is nonnegative. A positive witness at stage n has a nonzero numerator, so the bounded search through n has found a first event $N \leq n$, and $\alpha(N) = 1$. This is the conclusion of $\mathrm{MP}_{\mathrm{bin}}$. □

The same price appears in a statement that mentions only one sign.

Corollary 12.12 (Stability of positivity)**.** *Every regular dyadic presentation satisfies* $\neg\neg P(x) \to P(x)$ *if and only if* $\mathrm{MP}_{\mathrm{bin}}$ *holds. The same is true of* $\neg\neg N(x) \to N(x)$*.*

Proof. The predicate $P(x)$ says that the binary sequence of positive tests, $p(n) = 1$ exactly when $2 \cdot 2^n < q_n(n+1)$, has an event. If $\neg\neg P(x)$, then p is not all zero, and $\mathrm{MP}_{\mathrm{bin}}$ supplies the event.

Conversely, let α be a binary sequence that is not all zero, and let j be its positive jump presentation. As in the preceding proof, $\neg Z(j)$, so Lemma 12.10 gives $\neg\neg(P(j) \vee N(j))$. Since $N(j)$ is impossible, this is $\neg\neg P(j)$. Stability gives $P(j)$, and the event of α is read off as before.

For N, replace every numerator q_n by $-q_n$. The regularity condition is unchanged, and the new presentation has a positive witness exactly where the old one had a negative witness. □

The four comparisons now fit together. For a single presentation, a witnessed sign is precisely an exact zero decision together with a sign for the nonzero case.

Proposition 12.13 (Trichotomy is a decision plus a sign)**.** *For every regular dyadic presentation* x*,*

$$P(x) \vee Z(x) \vee N(x) \quad \Longleftrightarrow \quad \big(Z(x) \vee \neg Z(x)\big) \wedge \big(\neg Z(x) \to P(x) \vee N(x)\big).$$

Proof. Suppose the three-way alternative. In the middle case $Z(x)$ holds, and the implication is satisfied because its premise is refuted. In either outer case, Lemma 12.10 gives $\neg Z(x)$, and the required sign is already present. Conversely, decide $Z(x)$. The positive answer is the middle alternative; the negative answer yields a sign through the implication. □

Quantify over all presentations and price the three parts with Theorems 12.7, 12.6 and 12.11. The result is again LPO $\Longleftrightarrow$ WLPO $\wedge$ $\mathrm{MP}_{\mathrm{bin}}$, this time obtained through the comparisons rather than from the binary sequences directly. Proposition 12.9 is the direct proof. Table 12.1 collects the four comparisons.

Comparison, for every regular dyadic presentation	Exact price	Proved in
$Z(x) \vee \neg Z(x)$	WLPO	Theorem 12.6
$P(x) \vee Z(x) \vee N(x)$	LPO	Theorem 12.7
$U(x) \vee D(x)$	LLPO	Theorem 12.8
$\neg Z(x) \to P(x) \vee N(x)$	$\mathrm{MP}_{\mathrm{bin}}$	Theorem 12.11

Table 12.1. Four comparisons with zero and their exact prices. Each row is an equivalence between the displayed statement, asserted for every presentation, and the named principle for every binary sequence in its domain.

The four equivalences therefore distinguish exact comparison tasks without assuming that they are interchangeable. They also show why approximation alone does not discharge the comparison obligation: the jump construction turns an arbitrary sequence question into that obligation.

An equivalence identifies the logical resource a task supplies and uses. Our claim is the explicit two-way reduction for the stated carrier, thresholds and binary principles. No classical completeness, choice principle, unbounded witness selection or physical measurement postulate is hidden in those reductions. That the four principles differ from one another is a separate, independence question, and Theorem 7.38 settles it: for each pair, a model of intuitionistic second-order arithmetic satisfies one of the two and refutes the other. In particular $\mathrm{MP}_{\mathrm{bin}}$ is comparable with neither WLPO nor LLPO.

12.8 One line built from ratios

The dyadic denominators kept every stage comparison in integer arithmetic. They are a convenience, not the source of the prices. To see this, build the continuum directly from the ratio records of Chapter 4.

Definition 12.14 (The line built from ratios)**.** A *regular sequence* is a sequence $s = (s_n)_{n\ge 0}$ of ratio records such that

$$|s_m - s_n| \le \frac{1}{m+1} + \frac{1}{n+1} \qquad \text{for all } m, n,$$

each inequality read in integer arithmetic after its positive denominators are cleared. Regular sequences s and t *name the same point*, written $s \approx t$, if for every k there is an N such that $|s_n - t_n| \le 1/(k+1)$ for every $n \ge N$. The *line* $\mathbb{R}_\delta$ is the set of regular sequences taken up to $\approx$. A ratio record q determines the point $\eta(q)$ of its constant sequence.

The relation $\approx$ is an equivalence relation; for transitivity, use the tolerance $1/(2k+2)$ for both given pairs and add the two bounds. Every regular dyadic presentation, read stage by stage as the ratio records $(q_n, 2^n)$, is a regular sequence: condition (12.2) is exactly the condition of the definition.

Naming the same point asks for eventual agreement within every tolerance, a condition with three quantifiers. For the point zero it collapses to a single bound at each stage: the zero predicate of this chapter.

Lemma 12.15 (Bishop's zero condition)**.** *A regular sequence s satisfies $s \approx \eta(0)$ if and only if $|s_n| \le 2/(n+1)$ for every n.*

Proof. If every stage satisfies the bound, take $N = 2k+1$ for the tolerance $1/(k+1)$: every $n \ge N$ has $|s_n| \le 2/(n+1) \le 1/(k+1)$.

Conversely, suppose $s \approx \eta(0)$, and fix m. The bound at m is a decidable comparison, so it suffices to refute its failure. If $|s_m| > 2/(m+1)$, regularity gives, for every n,

$$|s_n| \ge |s_m| - \frac{1}{m+1} - \frac{1}{n+1} > \frac{1}{m+1} - \frac{1}{n+1}.$$

For $n \ge 2m+2$ the right side is at least

$$\frac{1}{m+1} - \frac{1}{2m+3} = \frac{m+2}{(m+1)(2m+3)} \ge \frac{1}{2m+3}.$$

So every stage from $2m+2$ on lies outside the tolerance $1/(2m+3)$, contradicting $s \approx \eta(0)$ at that tolerance. □

A second estimate shows that a clearly positive stage is never lost.

Lemma 12.16 (A positive stage persists)**.** *If a regular sequence satisfies $s_N > 2/(N+1)$, then $s_m > 1/(2N+2)$ for every $m \geq 2N+1$.*

Proof. Since $m+1 \geq 2N+2$, regularity gives

$$s_m \geq s_N - \frac{1}{N+1} - \frac{1}{m+1} > \frac{1}{N+1} - \frac{1}{m+1} \geq \frac{1}{N+1} - \frac{1}{2N+2} = \frac{1}{2N+2}. \qquad \square$$

Call a point x *positive*, written $x > 0$, when a regular sequence naming it has a stage with $s_N > 2/(N+1)$. This belongs to the point, not to the sequence. To see it, let $t \approx s$, and take M such that $|s_m - t_m| \leq 1/(4N+4)$ for every $m \geq M$. For every $m \geq \max(M, 8N+7)$, Lemma 12.16 gives

$$t_m > \frac{1}{2N+2} - \frac{1}{4N+4} = \frac{1}{4N+4} \geq \frac{2}{m+1},$$

so t has a positive stage too. Define $x < 0$ in the same way, with a stage below $-2/(N+1)$. Write $x \geq 0$ for $\neg(x < 0)$ and $x \leq 0$ for $\neg(x > 0)$.

For a regular dyadic presentation, read as a regular sequence, these are the predicates of Definition 12.2. Its stages $q_n/2^n$ give $x > 0$ exactly when $P(x)$ and $x < 0$ exactly when $N(x)$: the inequalities are the same, with the denominator 2^n cleared. Each stage comparison is decidable, so $U(x)$, which says that no stage is a negative witness, is $\neg N(x)$, and $D(x)$ is $\neg P(x)$. In rational notation $Z(x)$ says that every stage lies within $2/(n+1)$ of zero, so by Lemma 12.15 it says that the point is $\eta(0)$.

The line contains nothing beyond the dyadic presentations.

Theorem 12.17 (One continuum, two presentations)**.** *For a regular sequence s, let r_n be the integer nearest to $2^n s_{2n+1}$, rounding halves upward. Then (r_n) is a regular dyadic presentation naming the same point as s. Two regular dyadic presentations name the same point exactly when $|q_n - q'_n|(n+1) \leq 2 \cdot 2^n$ for every n. Under this correspondence Z, P, N, U and D become $x = \eta(0)$, $x > 0$, $x < 0$, $x \geq 0$ and $x \leq 0$. No choice principle is used.*

Proof. The rounding error is at most one half, so $|r_n/2^n - s_{2n+1}| \leq 2^{-(n+1)}$, and $2n+2 \leq 2^{n+1}$ makes this at most $1/(2n+2)$. Using regularity of s at the stages $2m+1$ and $2n+1$ for the middle term,

$$\left|\frac{r_m}{2^m} - \frac{r_n}{2^n}\right| \leq \frac{1}{2m+2} + \left(\frac{1}{2m+2} + \frac{1}{2n+2}\right) + \frac{1}{2n+2} = \frac{1}{m+1} + \frac{1}{n+1}.$$

This is exactly (12.2), with nothing to spare. In the same way,

$$\left|\frac{r_n}{2^n} - s_n\right| \leq \frac{1}{2n+2} + \left(\frac{1}{2n+2} + \frac{1}{n+1}\right) = \frac{2}{n+1}.$$

A bound of $2/(n+1)$ at every stage gives agreement within $1/(k+1)$ from stage $2k+1$ on, as in the first half of Lemma 12.15. So the rounded presentation names the point of s.

The same remark proves one direction of the second assertion, whose bound reads $|x_n - x'_n| \leq 2/(n+1)$ in rational notation. Conversely, suppose that x and x' name the same point, and

fix n. For any k, choose a stage $p \geq k$ beyond the one from which the two presentations agree within $1/(k+1)$. Regularity of both presentations gives

$$|x_n - x'_n| \leq |x_n - x_p| + |x_p - x'_p| + |x'_p - x'_n| \leq \frac{2}{n+1} + \frac{2}{p+1} + \frac{1}{k+1} \leq \frac{2}{n+1} + \frac{3}{k+1}.$$

If $|x_n - x'_n|$ exceeded $2/(n+1)$, it would exceed it by a positive ratio (u, d) with $u \geq 1$; choosing $k + 1 > 3d$ makes $3/(k+1)$ smaller than that excess, a contradiction. The bound at n is a decidable comparison, so it holds. We call this argument the *Archimedean step*: a ratio that is at most $c + C/(k+1)$ for every k is at most c.

The predicates correspond as explained before the theorem. Every step is an explicit integer computation, a bounded estimate or the Archimedean step; no choice principle enters. □

Corollary 12.18 (The four prices on the line)**.** *For the points of $\mathbb{R}_\delta$, each of the following assertions, made for every point x, is equivalent to the principle named: $x = \eta(0) \vee x \neq \eta(0)$, to WLPO; $x > 0 \vee x = \eta(0) \vee x < 0$, to LPO; $x \geq 0 \vee x \leq 0$, to LLPO; $x \neq \eta(0) \to x > 0 \vee x < 0$, to $\mathrm{MP}_{\mathrm{bin}}$; and $\neg\neg(x > 0) \to x > 0$, to $\mathrm{MP}_{\mathrm{bin}}$.*

Proof. Through Theorem 12.17, each assertion for all points is equivalent to the corresponding assertion for all regular dyadic presentations. In one direction, every presentation names a point and the predicates correspond. In the other, an assertion about one given point is a proposition, so we may use one regular sequence naming that point and its rounded presentation; no selection among infinitely many points is needed. Theorems 12.6, 12.7, 12.8 and 12.11 and Corollary 12.12 then give the prices. □

The line also carries the arithmetic one expects, at no logical price. For regular sequences s and t, put

$$(s+t)_n = s_{2n+1} + t_{2n+1}, \qquad (-s)_n = -s_n, \qquad (st)_n = s_j t_j \quad \text{with } j = 2K(n+1) - 1,$$

where K is the larger of $\lfloor |s_0| \rfloor + 3$ and $\lfloor |t_0| \rfloor + 3$. Regularity keeps every stage within 2 of the first, since $|s_n - s_0| \leq 1/(n+1) + 1$, so $|s_n| \leq |s_0| + 2 < \lfloor |s_0| \rfloor + 3 \leq K$, and likewise $|t_n| < K$, for every n. Write $x < y$ for $y - x > 0$ and $x \leq y$ for $\neg(y < x)$.

Proposition 12.19 (The ordered field structure is free)**.** *These operations give regular sequences and respect $\approx$. They make $\mathbb{R}_\delta$ a commutative ring into which η embeds the ratios, preserving and reflecting their order; for records with positive denominators, $(u, d) < (v, e)$ means $ue < vd$. Moreover:*

1. *$<$ is irreflexive and transitive, and it is* cotransitive*: if $x < y$, then $x < z$ or $z < y$ for every z. Adding a point preserves $<$, and multiplying by a positive point preserves $<$;*
2. *$\leq$ is reflexive and transitive, and $x \leq y \leq x$ implies $x = y$;*
3. *x has a multiplicative inverse exactly when $x < 0$ or $x > 0$;*
4. *every point lies below $\eta(n)$ for some natural number n;*

5. *every two points have a least upper bound* $\max(x, y)$ *and a greatest lower bound* $\min(x, y)$*, and every* $x \geq 0$ *has exactly one square root* $y \geq 0$*.*

None of these statements uses an omniscience principle or a choice principle.

Proof. Throughout, s, t and u are regular sequences, and $K_s = \lfloor |s_0| \rfloor + 3$, so that the constant K of a product is the larger of K_s and K_t. Regularity gives $|s_n - s_0| \leq 1 + 1/(n+1) \leq 2$, so $|s_n| \leq K_s$ at every stage. Two tools, proved in Appendix F.4, carry the routine estimates. First, s and t name the same point exactly when $|s_n - t_n| \leq 2/(n+1)$ for every n, and a bound $|s_n - t_n| \leq C/(n+1)$ at every stage, for one fixed C, already gives $s \approx t$ (Lemma F.8). Second, say that s *follows* ratios $f(M)$ with constant C when

$$|s_n - f(M)| \leq C\Big(\frac{1}{n+1} + \frac{1}{M+1}\Big) \qquad \text{for all } n, M.$$

A regular sequence follows its own stages with constant 1; that is its regularity. If s follows f and t follows g, then $s + t$, $-s$ and st follow $f + g$, $-f$ and fg, with constants computed from those of s and t and a bound on f; and two sequences that follow ratios differing by at most $F/(M+1)$ name the same point (Lemma F.9). The second tool turns every identity of ratios into an identity of points.

The operations. The sum is regular: its stages at m and n read s and t at $2m+1$ and $2n+1$, so they differ by at most $2\big(1/(2m+2) + 1/(2n+2)\big) = 1/(m+1) + 1/(n+1)$. Negation changes no distance. For the product,

$$|s_i t_i - s_j t_j| \leq |s_i|\,|t_i - t_j| + |t_j|\,|s_i - s_j| \leq K\big(|s_i - s_j| + |t_i - t_j|\big),$$

and at $i+1 = 2K(m+1)$ and $j+1 = 2K(n+1)$ the right side is at most $K\big(2/(i+1) + 2/(j+1)\big) = 1/(m+1) + 1/(n+1)$. If $s \approx s'$ and $t \approx t'$, then $|s_M - s'_M|$ and $|t_M - t'_M|$ are at most $2/(M+1)$. The product st follows the ratios $s_M t_M$, the product $s't'$ follows $s'_M t'_M$, and these differ by at most $2(K_t + K_{s'})/(M+1)$; so $st \approx s't'$. Sums and negations respect $\approx$ by the same argument with simpler bounds. The operations are therefore operations on points.

A commutative ring. For each ring law the two sides follow the same ratios. For instance $(st)u$ follows $(s_M t_M)u_M$ and $s(tu)$ follows $s_M(t_M u_M)$, and these ratios are equal. In the same way, commutativity and associativity of sums and products, distributivity, the laws of $\eta(0)$ and $\eta(1)$, and $s + (-s) \approx \eta(0)$ reduce to the corresponding identities of ratios, applied to the stages.

The ratios. Sums, negatives and products of constant sequences are constant sequences, of the sum, negative and product, so η preserves the operations. Next, $\eta(q) = \eta(r)$ exactly when $|q - r| \leq 1/(k+1)$ for every k, that is when $q = r$. The point $\eta(r) - \eta(q)$ is named by the constant sequence $r - q$, which has a positive stage exactly when $r - q > 0$: if $r - q = \langle u; d\rangle$ with $u \geq 1$, any N with $N + 1 > 2d$ gives one, and if $r - q \leq 0$ no stage is positive. So $\eta(q) < \eta(r)$ exactly when $q < r$, which for records (u, d) and (v, e) with positive denominators says $ue < vd$.

Item 1. Positivity is closed under sums and products, and a positive stage of a sum gives a positive stage of a summand (Lemma F.10). The point $x - x$ is named by a sequence whose stages are all 0, so $x < x$ fails. If $x < y$ and $y < z$, then $z - x = (z - y) + (y - x)$ is positive. If $x < y$ and z is any point, then $y - x = (y - z) + (z - x)$, so $y - z$ or $z - x$ is positive: $z < y$ or $x < z$. Finally $(y + z) - (x + z) = y - x$, and $yz - xz = (y - x)z$ is a product of positive points when $x < y$ and $z > 0$.

Item 2. Reflexivity is irreflexivity of $<$. If $x \le y \le z$ and $z < x$, cotransitivity gives $z < y$ or $y < x$, and both are excluded; so $x \le z$. If $x \le y \le x$, then $x - y$ is neither positive nor negative, since $x < y$ says that $y - x = -(x - y)$ is positive. So no stage of a sequence naming $x - y$ lies above $2/(n+1)$ or below $-2/(n+1)$; each such comparison is decided, every stage lies within $2/(n+1)$ of 0, and Lemma 12.15 gives $x - y = \eta(0)$, that is $x = y$.

Item 3. If $x > 0$ or $x < 0$, the explicit inverse described below is a regular sequence whose product with x is $\eta(1)$ (Lemma F.11). Conversely, let s name x, let t name y, and suppose $xy = \eta(1)$. The product sequence names the positive point $\eta(1)$, so it has a positive stage N: with $j + 1 = 2K(N+1)$, $s_j t_j > 2/(N+1)$. Since $|t_j| \le K$,

$$|s_j| > \frac{2}{(N+1)K} = \frac{4}{j+1} > \frac{2}{j+1}.$$

The sign of the ratio s_j is decided, so $s_j > 2/(j+1)$ or $s_j < -2/(j+1)$, and $x > 0$ or $x < 0$.

Item 4. Every stage of s has $|s_m| \le K_s$, so the stages of $\eta(K_s + 1) - x$, which are $K_s + 1 - s_{2n+1}$, are at least 1, and the stage $n = 2$ exceeds $2/3$. Hence $x < \eta(K_s + 1)$.

Item 5. The maximum is computed stagewise, $\max(s,t)_n = \max(s_n, t_n)$, and $|\max(a,b) - \max(c,d)| \le \max(|a-c|, |b-d|)$ for ratios. So $\max(s,t)$ is regular, and by Lemma F.8 it respects $\approx$. It is an upper bound: if $\max(x,y) < x$, the difference $x - \max(x,y)$ has a positive stage N, which reads $s_{2N+1} - \max(s_{2N+1}, t_{2N+1}) > 2/(N+1)$, impossible. It is the least one: if $x \le z$ and $y \le z$ but $z < \max(x,y)$, a positive stage N of $\max(x,y) - z$ reads $\max(s_{2N+1}, t_{2N+1}) - u_{2N+1} > 2/(N+1)$; whichever of s_{2N+1}, t_{2N+1} is the larger, a decided comparison, gives a positive stage of $x - z$ or of $y - z$, which is excluded. The same argument gives the strict form: if $w < \max(x,y)$, then $w < x$ or $w < y$. The minimum is computed stagewise in the same way and is the greatest lower bound.

Two nonnegative square roots of x are equal. Let $y, z \ge 0$ with $yy = zz = x$, and suppose $y < z$. Then $0 < z$, because $0 \le y < z$, and cotransitivity applied to $0 < z$ gives $0 < z + y$ or $z + y < z$, the second meaning $y < 0$; so $z + y > 0$. The product $(z - y)(z + y)$ of positive points is positive, yet it equals $zz - yy = \eta(0)$. So $\neg(y < z)$, likewise $\neg(z < y)$, and item 2 gives $y = z$. A nonnegative square root exists by trisection (Lemma F.12). Fix a sequence s naming x and a natural number $K \ge 1$ with $x \le \eta(K^2)$. Starting from $[0, K]$, each step cuts the current interval $[a, b]$ of ratios into thirds at $c_1 < c_2$; since $a \ge 0$, the squares satisfy $c_2^2 - c_1^2 \ge (b-a)^2/3$. One stage of s, read with an error smaller than half this gap, decides by a comparison of ratios whether $\eta(c_1^2) < x$ or $x < \eta(c_2^2)$, and the step keeps $[c_1, b]$ or $[a, c_2]$ accordingly. The left ends form, after a fixed reindexing, a regular sequence; its point y is nonnegative, and yy and x lie in intervals of widths tending to 0, so they are equal. Each decision is computed from the one sequence s, so no choice is made, and by uniqueness the point y does not depend on which sequence naming x was used.

No principle. Every step above is a stage estimate, a comparison of two ratios, or the Archimedean step, and every sequence used is computed from given sequences. So none of these statements uses an omniscience principle or a choice principle. □

With two points the prices of Corollary 12.18 reappear, together with the field and lattice statements that name a side. Table 12.2 lists them.

Assertion, for all points x, y of $\mathbb{R}_\delta$	Exact price
$x < y \vee \neg(x < y)$	LPO
$x < y \vee x = y \vee y < x$	LPO
$x \le y \vee \neg(x \le y)$; also $x = y \vee x \neq y$	WLPO
$x \le y \vee y \le x$	LLPO
$\max(x, y) = x \vee \max(x, y) = y$	LLPO
$xy = \eta(0) \to x = \eta(0) \vee y = \eta(0)$	LLPO
$\neg(y \le x) \to x < y$; also $x \neq y \to x < y \vee y < x$	$\mathrm{MP}_{\mathrm{bin}}$
$x \neq \eta(0) \to x$ has a multiplicative inverse	$\mathrm{MP}_{\mathrm{bin}}$
$\neg\neg(x \le y) \to x \le y$	none

Table 12.2. Order and algebra on the line. Each priced row is equivalent to the named principle for binary sequences, in both directions, with no choice principle; the last row is proved outright.

Two rows deserve a word. Deciding $x < y$ is deciding $x \le y$ together with the stability $\neg(y \le x) \to x < y$, the same split as LPO = WLPO + $\mathrm{MP}_{\mathrm{bin}}$. And the classical field axiom, that a nonzero element is invertible, holds exactly under $\mathrm{MP}_{\mathrm{bin}}$. Without it the line is a field in the sense of item 3 of Proposition 12.19: what has an inverse is what is apart from zero.

Proof of the prices in Table 12.2. Every row compares two points through their difference, by the ring laws of Proposition 12.19. The statement $x < y$ says that $y - x > 0$; $y < x$ says that $x - y > 0$, that is $y - x < 0$, since negating a sequence negates its stages; $x \le y$ says $\neg(x - y > 0)$; $x = y$ says $x - y = \eta(0)$; and $z < \eta(0)$ says $z < 0$. An assertion about all pairs of points is therefore equivalent to one about all single points z: put $z = y - x$ in one direction, and $x = \eta(0)$ or $y = \eta(0)$ in the other. Two further facts are used. For a binary sequence α, let j_α also denote the point of its positive jump presentation. Then $j_\alpha > 0$ exactly when α has an event: a positive witness has a nonzero numerator, so the bounded search through its stage has found an event, as in the proof of Theorem 12.11; and an event at N gives, by bounded search through N, a first event $N' \le N$, which makes the stage $2^{N'+1}$ a positive witness, by the computation in the proof of Lemma 12.5. And for a point z, read through one regular dyadic presentation (Theorem 12.17), $z > 0$ exactly when the binary sequence p_z of positive tests, $p_z(n) = 1$ exactly when $2 \cdot 2^n < q_n(n+1)$, has an event.

Row 1. Assume LPO and let x, y be points. LPO applied to p_{y-x} gives an event, so $x < y$, or no event, so $\neg(x < y)$. Conversely, the assertion at $\eta(0)$ and j_α gives $j_\alpha > 0$, hence an event of α, or $\neg(j_\alpha > 0)$; in the second case no $\alpha(n)$ is 1, and each $\alpha(n)$, being binary, is 0. This is LPO.

Row 2. With $z = y - x$ the three alternatives are $z > 0$, $z = \eta(0)$ and $z < 0$, so the row is equivalent, in both directions, to the second assertion of Corollary 12.18, whose price is LPO.

Row 3. Assume WLPO and let $z = x - y$, so that $x \le y$ says $\neg(z > 0)$. WLPO applied to p_z gives that p_z is all zero, so $\neg(z > 0)$ and $x \le y$, or that it is not, so $\neg\neg(z > 0)$, which is $\neg(x \le y)$. Conversely, the assertion at j_α and $\eta(0)$ gives $\neg(j_\alpha > 0)$, so α is all zero, or $\neg\neg(j_\alpha > 0)$, so α is not all zero, since otherwise j_α would be $\eta(0)$. This is WLPO. For $x = y \vee x \neq y$, the reduction to $z = x - y$ gives the first assertion of Corollary 12.18, whose price is WLPO.

Row 4. With $z = x - y$, $x \le y$ says $\neg(z > 0)$, which is $z \le 0$, and $y \le x$ says $\neg(z < 0)$, which is $z \ge 0$. So the row is the third assertion of Corollary 12.18, whose price is LLPO.

Row 5. If $x \le y$, then y is an upper bound of x and y, so $\max(x,y) \le y \le \max(x,y)$ and $\max(x,y) = y$; if $y \le x$, then $\max(x,y) = x$. So row 4 gives row 5. Conversely, $\max(x,y) = x$ gives $y \le \max(x,y) = x$, and $\max(x,y) = y$ gives $x \le y$; so row 5 gives row 4, and both cost LLPO.

Row 6. Assume LLPO, and let $xy = \eta(0)$. Write $|x|$ for $\max(x,-x)$. By row 4, $|x| \le |y|$ or $|y| \le |x|$; suppose the first. If x were positive or negative, then $|x| > 0$, because $x \le |x|$ and $-x \le |x|$; so $|y| > 0$, because $u < v \le w$ gives $u < w$; so $y > 0$ or $-y > 0$ by the strict form of item 5 of Proposition 12.19. Then x and y would be invertible by item 3, and so would be $xy = \eta(0)$, which is neither positive nor negative. Hence x is neither positive nor negative, and $x = \eta(0)$ by item 2. The second case gives $y = \eta(0)$. Conversely, assume the domain law, and let β have at most one event. Let x and y be the points of the positive jump presentations of $k \mapsto \beta(2k)$ and $k \mapsto \beta(2k+1)$. No stage of either is negative. If xy were positive or negative, it would be invertible, so x and y would be invertible, hence positive, and each of the two sequences would have an event; these would be two different events of β. So xy is neither positive nor negative, $xy = \eta(0)$, and the domain law gives $x = \eta(0)$ or $y = \eta(0)$. By Lemma 12.5 and Theorem 12.17, every even entry of β is 0, or every odd entry is. This is LLPO.

Row 7. With $z = y - x$, $\neg(y \le x)$ is $\neg\neg(z > 0)$ and $x < y$ is $z > 0$, so the first form is the stability of positivity, the fifth assertion of Corollary 12.18. The second form, $x \ne y \to x < y \vee y < x$, reads $z \ne \eta(0) \to z > 0 \vee z < 0$, the fourth assertion. Both cost $\mathrm{MP}_{\mathrm{bin}}$.

Row 8. By item 3 of Proposition 12.19, x is invertible exactly when $x > 0$ or $x < 0$, so the row is the fourth assertion of Corollary 12.18, whose price is $\mathrm{MP}_{\mathrm{bin}}$.

Row 9. The statement $x \le y$ is the negation $\neg(y < x)$, and $\neg\neg\neg P$ implies $\neg P$ for every proposition P. So $\neg\neg(x \le y) \to x \le y$ holds outright.

In every row each point is read through one sequence naming it, as in the proof of Corollary 12.18, the jump presentations are explicit, and the principles are applied to explicit binary sequences. No choice principle is used in either direction. □

Remark 12.20 (Reasoning by cases costs nothing for stable conclusions)**.** Every priced row has a conclusion that names a side. A conclusion that names no side can be reached by cases for free. Any two points satisfy $\neg\neg(x \le y \vee y \le x)$: if neither held, we would have $\neg\neg(y < x)$ and $\neg\neg(x < y)$, hence $\neg\neg(x < x)$, contrary to irreflexivity. An inequality $x \le y$ is itself a negation, so it is stable under double negation, and an equation is stable because it follows from two such inequalities. Hence every equation or $\le$-inequality proved by cases on the order of finitely many points holds outright. Identities such as $\min(x,y)\max(x,y) = xy$ and $|x|\,|x| = x\,x$ are proved in this way. Allowing the same argument for an arbitrary conclusion costs exactly LLPO.

Proof of the last sentence of Remark 12.20. Allowing the argument for an arbitrary conclusion means the following principle: for every finite list of points and every proposition P, if P follows from the assumption that any two points of the list are comparable, that is $x \le y$ or $y \le x$, then P holds. The remark proves this principle for every P with $\neg\neg P \to P$. Any two points are not not comparable, and double negation passes through finite conjunctions, since $\neg\neg A$ and $\neg\neg B$ give $\neg\neg(A \wedge B)$; so the assumption is not not true, hence $\neg\neg P$, hence P. The two identities of the remark are instances. If $x \le y$, then $\min(x,y) = x$ and $\max(x,y) = y$, as

in row 5 of Table 12.2, so $\min(x,y)\max(x,y) = xy$, and symmetrically if $y \le x$. If $0 \le x$, then $-x \le 0 \le x$ and $|x| = \max(x,-x) = x$; if $x \le 0$, then $|x| = -x$; in both cases $|x|\,|x| = x\,x$.

For an arbitrary conclusion the principle is exactly LLPO. If LLPO holds, any two points are comparable, by row 4 of Table 12.2, so the assumption holds outright and P follows. Conversely, apply the principle to the list x, y and the conclusion $x \le y \vee y \le x$, which follows from the assumption at once. Every two points are then comparable, and row 4 gives LLPO. Neither direction uses a choice principle. □

12.9 What completeness costs

Chapter 11 used completeness of the classical line as one resource: a bounded set of lower endpoints had a supremum, and the displayed value was defined from it. On $\mathbb{R}_\delta$ that resource comes apart. Some forms of completeness cost nothing. One, that every located cut is the cut of a point, costs only countable choice, applied once to select an approximating pair of ratios for each precision. The other forms that need nothing beyond countable choice follow from that one with no further choice, and the limit of a Cauchy sequence needs only a choice of representatives. Each of the others costs exactly one principle: LPO, WLPO or LLPO, or, for arbitrary sets of points, weak excluded middle or excluded middle.

Countable choice decides nothing (Section 12.9.1). Sections 12.9.2–12.9.3 build limits from located cuts and show that the line is determined by its task; Sections 12.9.4–12.9.10 prove the rows of Table 12.3; Sections 12.9.11–12.9.14 show what replaces countable choice and what does not; Sections 12.9.15–12.9.16 give prices away from the line and the split of LPO.

12.9.1 Countable choice decides nothing

A principle of a different kind enters here.

Two readings of maps matter from here on. In this book a map is a relation with exactly one output for each input (Appendix C), so a relation with exactly one witness for each number already defines a sequence: the *principle of unique choice* is part of that reading. With excluded middle, unique choice gives countable choice for numbers, because the least witness is unique. The formal theory of Theorem 12.63 has variables for sequences instead and does not contain unique choice: in its model of computable sequences, the halting relation of that proof has exactly one witness for each program, and no computable sequence collects the witnesses. There, excluded middle and countable choice are separate principles. The theory of Theorem 12.22 contains countable choice for numbers, and with it unique choice, its case of a relation with exactly one witness for each number. Chapter 14 shows where unique choice enters linear algebra on the line, and Chapter 16 shows, inside the computable calculus of δ's own objects, that without it omniscience decides each question but cannot collect the answers into a sequence (Theorem 16.27).

Definition 12.21 (Countable choice)**.** *Countable choice*, AC_ω, states: if for every natural number n there is some x with $R(n, x)$, then there is a sequence f with $R(n, f(n))$ for every n.

Countable choice selects witnesses; it decides nothing. The next theorem makes this exact.

Theorem 12.22 (Countable choice decides nothing)**.** *Consider intuitionistic arithmetic with variables for sequences, induction for every formula, and countable choice both for numbers and for sequences. In that theory none of LPO, WLPO, LLPO and* $\mathrm{MP}_{\mathrm{bin}}$ *is provable, and none is refutable. More generally, if it proves* $\forall x\,(A(x) \vee \neg A(x))$ *for a property* A *built from equations and* $\bot$ *by conjunction, implication and universal quantification over numbers, then* A *is decided by a computer program, with any other free variables of* A *read as* 0 *and its free sequence variables as the zero sequence. With* $\mathrm{MP}_{\mathrm{bin}}$ *added, none of WLPO, LLPO and LPO becomes provable. The proof reasons in a classical metatheory.*

Conditions and scope.

Base and premises
Intuitionistic arithmetic with variables for sequences, induction for every formula, and countable choice for numbers and for sequences at every formula (hence unique choice); the metatheory is classical.

Conclusion
None of LPO, WLPO, LLPO and $\mathrm{MP}_{\mathrm{bin}}$ is provable or refutable; a provably decided property built from equations and $\bot$ by $\wedge$, $\to$ and $\forall x$ is decided by a program, with its other free variables read as in the theorem; with $\mathrm{MP}_{\mathrm{bin}}$ added, none of WLPO, LLPO and LPO becomes provable.

Proof. We fix the theory, read its formulas in two ways as tasks for programs, show that every theorem passes each reading, and show that none of the four principles does. The arguments about programs and models are made in ordinary classical mathematics: they describe what programs compute and which statements are true in a model, and they ask nothing of the theory under study.

The theory. Number terms are built from number variables, 0, the successor S, $+$, $\cdot$ and the application $\alpha(t)$ of a sequence variable α to a number term t. Formulas are built from equations $t = s$ and $\bot$ by $\wedge$, $\vee$ and $\to$, by the number quantifiers $\forall x$, $\exists x$ and by the sequence quantifiers $\forall \alpha$, $\exists \alpha$; $\neg A$ abbreviates $A \to \bot$. The rules are those of intuitionistic natural deduction for these connectives and quantifiers, in which $\forall x$ is instantiated and $\exists x$ introduced at any number term, and $\forall \alpha$ is instantiated and $\exists \alpha$ introduced at any sequence variable; reflexivity $t = t$ and the substitution of equals into every formula; the axioms $\neg(St = 0)$, $St = Ss \to t = s$, $t + 0 = t$, $t + Ss = S(t + s)$, $t \cdot 0 = 0$ and $t \cdot Ss = t \cdot s + t$; the induction rule, which infers $\forall x\, A(x)$ from $A(0)$ and $\forall x\,(A(x) \to A(Sx))$ for every formula A, sequence variables and sequence quantifiers included; and, for every formula A with any parameters, the two choice schemas

$$\forall x \exists y\, A(x, y) \to \exists \alpha \forall x\, A(x, \alpha(x)),$$
$$\forall x \exists \beta\, A(x, \beta) \to \exists \alpha \forall x \exists \beta \big(\forall y\, \beta(y) = \alpha(\langle x, y\rangle) \wedge A(x, \beta)\big),$$

where $\langle x, y\rangle = (x + y) \cdot (x + y) + y$. The first is countable choice for numbers. The second is countable choice for sequences: the pairing term takes different values at different pairs, so

the chosen sequences are the sections $y \mapsto \alpha(\langle x, y\rangle)$ of a single sequence. Write $\mathrm{Bin}(\alpha)$ for $\forall n\,(\alpha(n) = 0 \vee \alpha(n) = 1)$. With the two bits read as 0 and 1, the principles of Definition 7.8 are the sentences

$$\begin{aligned}
\mathrm{LPO} &: \ \forall\alpha\big(\mathrm{Bin}(\alpha) \to \forall n\,\alpha(n) = 0 \vee \exists n\,\alpha(n) = 1\big),\\
\mathrm{WLPO} &: \ \forall\alpha\big(\mathrm{Bin}(\alpha) \to \forall n\,\alpha(n) = 0 \vee \neg\forall n\,\alpha(n) = 0\big),\\
\mathrm{MP}_{\mathrm{bin}} &: \ \forall\alpha\big(\mathrm{Bin}(\alpha) \to \neg\forall n\,\alpha(n) = 0 \to \exists n\,\alpha(n) = 1\big),\\
\mathrm{LLPO} &: \ \forall\alpha\big(\mathrm{Bin}(\alpha) \to \forall m\,\forall n\,(\alpha(m) = 1 \to \alpha(n) = 1 \to m = n)\\
&\qquad\qquad \to \forall k\,\alpha(2k) = 0 \vee \forall k\,\alpha(2k+1) = 0\big).
\end{aligned}$$

None of the four is refutable. Read numbers as the natural numbers and sequences as all functions from the natural numbers to themselves. Every rule preserves truth in this reading; induction holds because the numbers are the natural numbers; the first choice schema holds because, when every x has some y, a function choosing one y for each x exists; and the second holds because sequences β_x chosen for the numbers x are assembled into one sequence by $\alpha(\langle x, y\rangle) = \beta_x(y)$, with $\alpha = 0$ off the values of the pairing term. These choices are made in the classical mathematics in which the model is described, which is exactly what a model of a theory with choice requires. In this model each of the four principles is true, by excluded middle applied to the given sequence. A refutation would be true there as well, so none of the four is refutable, in the theory or in the theory with $\mathrm{MP}_{\mathrm{bin}}$ added, which the model also satisfies. In particular both theories are consistent.

Programs. Fix a numbering of the programs that compute partial functions of one natural number, and write $\{e\}(a)$ for the output of program e on input a when that run halts. A program is *total* when it halts on every input. We use three standard facts of computability [19, Chapter XII]: a function obtained from given partial recursive functions by the operations of Definition 10.5 has a program (Proposition 10.6); from a program for a function of two arguments one computes, uniformly in the first argument, a program for the resulting function of the second; and functions defined by primitive recursion from computable ones are computable. Fix a computable bijection π from pairs of numbers onto numbers, and write $(r)_0$ and $(r)_1$ for the components of the pair numbered r. We arrange that $\pi(0, 0) = 0$ and that program 0 returns 0 on every input. A *valuation* assigns a number to every number variable and a total program to every sequence variable; the variable assigned i stands for the sequence $n \mapsto \{i\}(n)$. Under a valuation every term has a value, and every formula is true or false in the model whose sequences are the total computable ones.

Modified realizability. This is Kreisel's interpretation [7]. Types are built from $\mathbb{N}$ by $\sigma \to \rho$, $\sigma \times \rho$ and $\sigma + \rho$. A formula A has the type $\tau(A)$ given by

$$\begin{aligned}
&\tau(t = s) = \tau(\bot) = \mathbb{N}, && \tau(A \wedge B) = \tau(A) \times \tau(B), && \tau(A \vee B) = \tau(A) + \tau(B),\\
&\tau(A \to B) = \tau(A) \to \tau(B), && \tau(\forall x\,A) = \mathbb{N} \to \tau(A), && \tau(\exists x\,A) = \mathbb{N} \times \tau(A),\\
&\tau(\forall\alpha\,A) = (\mathbb{N} \to \mathbb{N}) \to \tau(A), && \tau(\exists\alpha\,A) = (\mathbb{N} \to \mathbb{N}) \times \tau(A).
\end{aligned}$$

The *hereditarily total operations* of each type form a set HRO_σ of numbers: $\mathrm{HRO}_{\mathbb{N}}$ contains every number; $\mathrm{HRO}_{\sigma\to\rho}$ contains the programs e such that $\{e\}(a)$ halts with a value in HRO_ρ for every a in HRO_σ; $\mathrm{HRO}_{\sigma\times\rho}$ contains the r with $(r)_0 \in \mathrm{HRO}_\sigma$ and $(r)_1 \in \mathrm{HRO}_\rho$; and $\mathrm{HRO}_{\sigma+\rho}$ contains the r with $(r)_0 = 0$ and $(r)_1 \in \mathrm{HRO}_\sigma$, or $(r)_0 \neq 0$ and $(r)_1 \in \mathrm{HRO}_\rho$. Thus $\mathrm{HRO}_{\mathbb{N}\to\mathbb{N}}$

is the set of total programs, and an induction on the type shows that 0 lies in every HRO_σ. Under a valuation, *r realizes* A, written $r \Vdash_{\mathrm{m}} A$, is defined by recursion on A:

- $r \Vdash_{\mathrm{m}} t = s$ when $t = s$ is true; no number realizes $\bot$;
- $r \Vdash_{\mathrm{m}} A \wedge B$ when $(r)_0 \Vdash_{\mathrm{m}} A$ and $(r)_1 \Vdash_{\mathrm{m}} B$;
- $r \Vdash_{\mathrm{m}} A \vee B$ when $(r)_0 = 0$ and $(r)_1 \Vdash_{\mathrm{m}} A$, or $(r)_0 \neq 0$ and $(r)_1 \Vdash_{\mathrm{m}} B$;
- $r \Vdash_{\mathrm{m}} A \to B$ when $r \in \mathrm{HRO}_{\tau(A)\to\tau(B)}$ and, for every $s \Vdash_{\mathrm{m}} A$, the value $\{r\}(s)$ realizes B;
- $r \Vdash_{\mathrm{m}} \forall x\, A(x)$ when, for every n, $\{r\}(n)$ halts with a value realizing $A(n)$;
- $r \Vdash_{\mathrm{m}} \exists x\, A(x)$ when $(r)_1 \Vdash_{\mathrm{m}} A((r)_0)$;
- $r \Vdash_{\mathrm{m}} \forall \alpha\, A(\alpha)$ when, for every total program i, $\{r\}(i)$ halts with a value realizing $A(i)$;
- $r \Vdash_{\mathrm{m}} \exists \alpha\, A(\alpha)$ when $(r)_0$ is total and $(r)_1 \Vdash_{\mathrm{m}} A((r)_0)$.

Here $A(n)$ is read with its variable assigned n, and $A(i)$ with its sequence variable assigned i. An induction on A shows that every realizer of A lies in $\mathrm{HRO}_{\tau(A)}$. The implication clause makes the reading *modified*: a realizer of $A \to B$ must be an operation of type $\tau(A) \to \tau(B)$, so it must answer on every input of that type, whether or not the input realizes A.

Kleene realizability. The relation $r \Vdash_{\mathrm{k}} A$ has the same clauses, except that $r \Vdash_{\mathrm{k}} A \to B$ asks only that $\{r\}(s)$ halt with a realizer of B for every $s \Vdash_{\mathrm{k}} A$. Nothing is asked about other inputs. This is the reading used in Chapter 6, extended to sequences.

Every theorem is realized. For a derivation of A from hypotheses $B_1, \dots, B_m$ we find one program c with two properties. Feed c the values of the free number variables, the total programs assigned to the free sequence variables, and numbers $s_1, \dots, s_m$ with $s_j \in \mathrm{HRO}_{\tau(B_j)}$. Then c halts with a value in $\mathrm{HRO}_{\tau(A)}$, and that value realizes A whenever each s_j realizes B_j. For Kleene realizability we ask only for the second property, on inputs whose s_j realize the B_j. The program is built by induction on the derivation; Appendix F.1 lists it rule by rule. Four steps carry the argument.

Implication introduction turns the program c for B, with the extra hypothesis A, into the program that receives the inputs and returns a program for $s \mapsto c(\dots, s)$. By the first property of c, this program halts with a value in $\mathrm{HRO}_{\tau(B)}$ on every s in $\mathrm{HRO}_{\tau(A)}$, so it lies in $\mathrm{HRO}_{\tau(A)\to\tau(B)}$, which is what the modified implication clause asks. The first property is carried through the induction for this step.

The induction rule is realized by primitive recursion. From realizers r_0 of $A(0)$ and r_1 of $\forall x(A(x) \to A(Sx))$ put $R(0) = r_0$ and $R(n+1) = \{\{r_1\}(n)\}(R(n))$. Each $\{r_1\}(n)$ lies in $\mathrm{HRO}_{\tau(A)\to\tau(A)}$, so by induction on n every $R(n)$ is defined and lies in $\mathrm{HRO}_{\tau(A)}$; when r_0 and r_1 are realizers, a second induction on n shows that $R(n)$ realizes $A(n)$.

The first choice schema is realized by the program E that sends e to $\pi(f_e, g_e)$, where f_e is a program for $n \mapsto (\{e\}(n))_0$ and g_e one for $n \mapsto (\{e\}(n))_1$, both computed from e. If e lies in $\mathrm{HRO}_{\mathbb{N}\to(\mathbb{N}\times\tau(A))}$, then f_e is total and g_e lies in $\mathrm{HRO}_{\mathbb{N}\to\tau(A)}$, which gives the first property. If e realizes $\forall x \exists y\, A(x,y)$, then for every n the value $r_n = \{e\}(n)$ has $(r_n)_1$ realizing $A(n, (r_n)_0)$, that

is $A(n, \{f_e\}(n))$. So g_e realizes $\forall x\, A(x, \alpha(x))$ with α assigned f_e, and $\pi(f_e, g_e)$ realizes the conclusion. The second schema is realized by $e \mapsto \pi(a_e, h_e)$, where a_e computes $\langle x, y\rangle \mapsto \{(\{e\}(x))_0\}(y)$, recovering x and y from the value of the pairing term by a bounded search, and returns 0 on numbers that are not such values; and h_e computes $x \mapsto \pi\big((\{e\}(x))_0, \pi(0, (\{e\}(x))_1)\big)$. If e lies in the type of the premise, every $(\{e\}(x))_0$ is total, so a_e is total and h_e lies in the type of the body; this is the first property. When e realizes $\forall x \exists \beta\, A(x, \beta)$, and with β assigned $(\{e\}(x))_0$ every equation $\beta(y) = \alpha(\langle x, y\rangle)$ is true, so program 0 realizes $\forall y\, \beta(y) = \alpha(\langle x, y\rangle)$ and $\{h_e\}(x)$ realizes the body of the conclusion. Neither program decides anything: each returns, rearranged, what the realizer of the premise supplies. This is the formal content of the sentence that countable choice selects. The same programs realize both schemas in Kleene's reading.

Finally, Kleene's reading also realizes $\mathrm{MP}_{\mathrm{bin}}$, by search. Let M send i to a program that sends b to a program that sends d to $\pi(\mu, 0)$, where μ is the first n with $\{i\}(n) = 1$. Suppose $b \Vdash_{\mathrm{k}} \mathrm{Bin}(\alpha)$ and $d \Vdash_{\mathrm{k}} \neg\forall n\, \alpha(n) = 0$, with α assigned i. Then α is binary, and it is not everywhere 0: otherwise program 0 would realize $\forall n\, \alpha(n) = 0$ and $\{d\}(0)$ would realize $\bot$. Hence some value is 1, the search stops at the first such n, and $\pi(\mu, 0)$ realizes $\exists n\, \alpha(n) = 1$.

Applied to derivations without hypotheses, under the valuation that assigns 0 to every number variable and program 0 to every sequence variable, this gives: every theorem of the theory has a modified realizer, and every theorem of the theory with $\mathrm{MP}_{\mathrm{bin}}$ added has a Kleene realizer.

Two problems that no program solves. Let $h(e, n) = 1$ if program e halts on input 0 within n steps, and $h(e, n) = 0$ otherwise. The function h is computable, because n steps of a run can be simulated, and program e halts on input 0 exactly when some $h(e, n)$ is 1. No total program decides, for every e, whether program e halts on input 0 [21]. Next let A_0 be the set of programs e with $\{e\}(0) = 0$ and A_1 the set with $\{e\}(0) = 1$. No total program d with values in $\{0, 1\}$ has $d(e) = 1$ for every $e \in A_0$ and $d(e) = 0$ for every $e \in A_1$. Indeed, by Kleene's recursion theorem [19, Chapter XII] there is a program c that returns 1 on every input if $d(c) = 1$, and 0 if $d(c) = 0$. If $d(c) = 1$, then $c \in A_1$, so $d(c) = 0$; if $d(c) = 0$, then $c \in A_0$, so $d(c) = 1$. Both cases are contradictory. The two sets are joined by one sequence for each program e. Put $\ell_e(2K) = 1$ when the run of e on input 0 halts after exactly K steps with output 0, put $\ell_e(2K + 1) = 1$ when it halts after exactly K steps with output 1, and put $\ell_e(n) = 0$ otherwise. Whether a run halts after exactly K steps is decided by running it for K steps, so $\ell_e(n)$ is computable from e and n. The sequence has at most one entry 1, since a run has only one halting time; it has an entry 1 at an even place exactly when $e \in A_0$, and at an odd place exactly when $e \in A_1$.

From a realizer to a program. Each of the four principles has the form $\forall \alpha(\mathrm{Bin}(\alpha) \to B(\alpha))$. Suppose r realizes such a principle, in either reading, and let $a(c, n)$ be a computable function with values in $\{0, 1\}$. A program i_c for $n \mapsto a(c, n)$ is computed from c, and it is total. The program b_c computing $n \mapsto \pi(a(c, n), 0)$ realizes $\mathrm{Bin}(\alpha)$ with α assigned i_c: its tag at n names the true side, and 0 realizes the true equation. Therefore

$$G(c) = \{\{r\}(i_c)\}(b_c)$$

halts for every c with a realizer of $B(\alpha)$, α assigned i_c, and G is computed by one program.

LPO is not realized. Take $a = h$, so that the sequence of c records the halting of program c. The tag $t(c) = (G(c))_0$ is a total computable function of c. If $t(c) = 0$, then $(G(c))_1$ realizes $\forall n\, \alpha(n) = 0$, so every $h(c, n)$ is 0 and program c does not halt. If $t(c) \neq 0$, then $(G(c))_1$ realizes

$\exists n\,\alpha(n) = 1$, so some $h(c,n)$ is 1 and program c halts. The tag decides halting, which no program does.

WLPO is not realized. With the same sequences, a tag $t(c) \neq 0$ now comes with a realizer of $\neg\forall n\,\alpha(n) = 0$. If every $h(c,n)$ were 0, program 0 would realize $\forall n\,\alpha(n) = 0$, and the realizer of the negation would turn it into a realizer of $\bot$. So not every $h(c,n)$ is 0, and program c halts. Again the tag decides halting. Both arguments use only the clauses that the two readings share.

$\mathrm{MP}_{\mathrm{bin}}$ *is not realized in the modified reading.* Again take $a = h$. Now $G(c)$ realizes $\neg\forall n\,\alpha(n) = 0 \to \exists n\,\alpha(n) = 1$, so it lies in $\mathrm{HRO}_{\sigma\to(\mathbb{N}\times\mathbb{N})}$ with $\sigma = \tau(\neg\forall n\,\alpha(n) = 0)$. Since 0 lies in HRO_σ, the run $\{G(c)\}(0)$ halts for every c, and $m(c) = (\{G(c)\}(0))_0$ is a total computable function of c. Suppose program c halts. Then $\forall n\,\alpha(n) = 0$ is false and has no realizer, so every element of HRO_σ realizes its negation, 0 included. Hence $\{G(c)\}(0)$ realizes $\exists n\,\alpha(n) = 1$, which says $h(c, m(c)) = 1$. So program c halts exactly when $h(c, m(c)) = 1$, a test that a program performs. This contradiction is where the two readings part. A Kleene realizer of the implication need not halt on the input 0 when program c never halts, because then the premise has no realizer; the search of the previous paragraph is such a realizer, and its m is not total.

LLPO is not realized. Take $a(c,n) = \ell_c(n)$. The formula $\forall m\,\forall n\,(\alpha(m) = 1 \to \alpha(n) = 1 \to m = n)$ is true for this sequence, and program 0 realizes it in both readings: on all arguments it returns 0, which lies in every HRO_σ and realizes the final equation, true whenever the premises are realized. So $z(c) = \{G(c)\}(0)$ halts for every c with a realizer of $\forall k\,\alpha(2k) = 0 \vee \forall k\,\alpha(2k+1) = 0$. Put $d(c) = 1$ if $(z(c))_0 \neq 0$ and $d(c) = 0$ otherwise. If $c \in A_0$, the sequence has an entry 1 at an even place, so the first side is false and unrealized; hence $(z(c))_0 \neq 0$ and $d(c) = 1$. If $c \in A_1$, the second side is false, so $(z(c))_0 = 0$ and $d(c) = 0$. The total program d separates A_0 from A_1, which no program does.

So none of the four principles has a modified realizer, while every theorem has one. None is provable, and none is refutable, as shown first.

Decided properties. Call a formula *negative* when it is built from equations and $\bot$ by $\wedge$, $\to$ and $\forall x$. For negative A, under every valuation, $r \Vdash_{\mathrm{m}} A$ holds exactly when $r \in \mathrm{HRO}_{\tau(A)}$ and A is true. This follows by induction on A. Equations and $\bot$ are immediate, and so are $\wedge$ and $\forall x$. If r realizes $A \to B$ and A is true, then 0 realizes A by the induction hypothesis, so $\{r\}(0)$ realizes B and B is true. Conversely, if $r \in \mathrm{HRO}_{\tau(A)\to\tau(B)}$ and $A \to B$ is true, a realizer s of A lies in $\mathrm{HRO}_{\tau(A)}$ and makes A true, so $\{r\}(s)$ lies in $\mathrm{HRO}_{\tau(B)}$ and B is true, and $\{r\}(s)$ realizes B. Now let $A(x)$ be negative, any other free variables read as in the valuation fixed above, and suppose the theory proves $\forall x(A(x) \vee \neg A(x))$. A realizer r gives, for every n, a value $z_n = \{r\}(n)$ realizing $A(n) \vee \neg A(n)$, and $d(n) = (z_n)_0$ is a total computable function. If $d(n) = 0$, then $(z_n)_1$ realizes $A(n)$, so $A(n)$ is true. If $d(n) \neq 0$, then $(z_n)_1$ realizes the negative formula $\neg A(n)$, so $A(n)$ is false. Thus $A(n)$ holds exactly when $d(n) = 0$, and the program computing d decides A.

With $\mathrm{MP}_{\mathrm{bin}}$ *added.* Every theorem of the extended theory has a Kleene realizer. The arguments for LPO, WLPO and LLPO above used only clauses that Kleene's reading shares with the modified one, and the realizer 0 of the true premise of LLPO is a Kleene realizer as well. So none of WLPO, LLPO and LPO has a Kleene realizer, and none is provable with $\mathrm{MP}_{\mathrm{bin}}$; the classical model shows that none is refutable there either. □

Below countable choice appears only as a stated hypothesis, and each result says where it is used. Every such use in Table 12.3 is countable choice for numbers or for sequences, so all choice used in that table lies in this theory, which proves none of the four principles.

12.9.2 Limits and located cuts: where countable choice enters

Distances are order statements. Say that x and y are *within* $1/(k+1)$ when $x \le y+\eta(1/(k+1))$ and $y \le x+\eta(1/(k+1))$. A sequence of points (x_n) *converges* to L if for every k there is an N such that x_m is within $1/(k+1)$ of L for all $m \ge N$. It is *Cauchy* if for every k there is an N such that x_m and x_n are within $1/(k+1)$ for all $m, n \ge N$. On sequences that name the points, "within" becomes a bound at every stage.

Lemma 12.23 (Two readings of "within")**.** *Let s and t be regular sequences naming the points x and y, and let $\varepsilon = 1/(k+1)$. Then x and y are within ε exactly when*

$$|s_n - t_n| \le \varepsilon + \frac{2}{n+1} \qquad \text{for every } n.$$

No choice principle is used.

Proof. By the definitions of sum and negation, the point $x-(y+\eta(\varepsilon))$ is named by the regular sequence $u_j = s_{2j+1} - t_{4j+3} - \varepsilon$. Regularity of s and t compares u_j with any stage n:

$$\left|u_j - (s_n - t_n - \varepsilon)\right| \le \frac{1}{2j+2} + \frac{1}{4j+4} + \frac{2}{n+1} = \frac{3}{4j+4} + \frac{2}{n+1}. \tag{12.4}$$

Suppose the stagewise bound holds, and suppose $y+\eta(\varepsilon) < x$. Positivity does not depend on the naming sequence, so $u_N > 2/(N+1)$ for some N. At the stage $n = 4N+3$, (12.4) with $j = N$ gives

$$s_n - t_n \ge u_N + \varepsilon - \frac{5}{4N+4} > \varepsilon + \frac{8}{4N+4} - \frac{5}{4N+4} > \varepsilon + \frac{2}{n+1},$$

contrary to the bound. So $x \le y+\eta(\varepsilon)$; exchanging the roles of s and t gives $y \le x+\eta(\varepsilon)$.

Conversely, suppose $x \le y+\eta(\varepsilon)$ and fix n. The comparison $s_n - t_n \le \varepsilon + 2/(n+1)$ is decidable, so it suffices to refute $s_n - t_n = \varepsilon + 2/(n+1) + \delta$ with a positive ratio δ. In that case (12.4) gives $u_j \ge \delta - 3/(4j+4)$ for every j, and for j with $4j+4 > 11/\delta$ this exceeds $8/(4j+4) = 2/(j+1)$. So u has a positive stage and $y+\eta(\varepsilon) < x$, contrary to the hypothesis. In the same way $y \le x+\eta(\varepsilon)$ gives $t_n - s_n \le \varepsilon + 2/(n+1)$. □

Corollary 12.24 (Equality from every tolerance)**.** *Two points are equal exactly when they are within $1/(k+1)$ of each other for every k. No choice principle is used.*

Proof. Let s and t name x and y. If the points are within every $1/(k+1)$, Lemma 12.23 gives $|s_n - t_n| \le 1/(k+1) + 2/(n+1)$ for all k and n, and the Archimedean step leaves $|s_n - t_n| \le 2/(n+1)$ at every stage. Then every $n \ge 2k+1$ has $|s_n - t_n| \le 1/(k+1)$, as in the first half of Lemma 12.15, so $s \approx t$ and $x = y$. Conversely, a point is within every $1/(k+1)$ of itself: $x+\eta(1/(k+1)) < x$ would give, after adding $-x$, $\eta(1/(k+1)) < \eta(0)$, and η reflects the order. □

A sequence of points is a sequence of classes. To calculate with it we want stages: for each k, a regular sequence naming a term x_N at an index N from which on the Cauchy condition holds at every tolerance $1/(j+1)$ with $j \le k$. Producing infinitely many such representatives at once is an application of AC_ω; the indices themselves are never selected. Once these representatives are supplied, the limit is explicit. A representative of every term is not enough by itself: the rational points c_s in the proof of Theorem 12.63 have evident representatives, and that sequence has no limit in the model used there.

Theorem 12.25 (The diagonal limit)**.** *Call regular sequences $s^{(0)}, s^{(1)}, \ldots$ a* Cauchy family *when their points are within $1/(k+1)$ of one another whenever both indices are at least k. The diagonal $d_n = s^{(2n+1)}_{2n+1}$ of a Cauchy family is a regular sequence, and the point of $s^{(m)}$ is within $1/(k+1)$ of the point of d whenever $k \le m$. So the points of the family converge to the point of d. No choice principle and no omniscience principle is used.*

Proof. By Lemma 12.23, the hypothesis says that $|s^{(i)}_p - s^{(j)}_p| \le 1/(c+1) + 2/(p+1)$ at every stage p whenever $i, j \ge c$.

Write $a = 2m+1$ and $b = 2n+1$. Compare $d_m = s^{(a)}_a$ with $d_n = s^{(b)}_b$ through an arbitrary stage p:

$$|d_m - d_n| \le |s^{(a)}_a - s^{(a)}_p| + |s^{(a)}_p - s^{(b)}_p| + |s^{(b)}_p - s^{(b)}_b|.$$

Regularity bounds the outer terms by $1/(a+1) + 1/(p+1)$ and $1/(b+1) + 1/(p+1)$. The hypothesis at $c = \min(a,b)$ bounds the middle term by $1/(c+1) + 2/(p+1)$, and $1/(c+1) \le 1/(a+1) + 1/(b+1)$. The total is

$$\frac{2}{a+1} + \frac{2}{b+1} + \frac{4}{p+1} = \frac{1}{m+1} + \frac{1}{n+1} + \frac{4}{p+1}.$$

Since p is arbitrary, the Archimedean step removes the last term, and this is the regularity of d.

For the second assertion, let $k \le m$, put $a = 2n+1$, and compare $s^{(m)}_n$ with $d_n = s^{(a)}_a$ through a stage p. Regularity of $s^{(m)}$ and of $s^{(a)}$ contributes $1/(n+1) + 1/(p+1)$ and $1/(p+1) + 1/(a+1)$. Neither m nor a is below $c = \min(k,a)$, so the hypothesis at c bounds $|s^{(m)}_p - s^{(a)}_p|$ by $1/(c+1) + 2/(p+1)$, and $1/(c+1) \le 1/(k+1) + 1/(a+1)$. Since $2/(a+1) = 1/(n+1)$, the total is $1/(k+1) + 2/(n+1) + 4/(p+1)$, and the Archimedean step leaves $|s^{(m)}_n - d_n| \le 1/(k+1) + 2/(n+1)$ at every stage n. By Lemma 12.23 the two points are within $1/(k+1)$, and the index $N = k$ witnesses convergence at the tolerance $1/(k+1)$. Every step is an integer estimate or the Archimedean step; nothing is chosen and nothing is decided. □

Theorem 12.26 (Cauchy sequences of points)**.** *A sequence of points has at most one limit; this uses no choice principle. Assuming AC_ω, every Cauchy sequence of points of $\mathbb{R}_\delta$ converges. Countable choice is used once, for regular sequences, to select representatives; no omniscience principle is used.*

Proof. Let L and L' be limits of (x_n), and fix k. Convergence at the tolerance $1/(2k+2)$ supplies indices for both; at an index m beyond the two, L is within $1/(2k+2)$ of x_m, which is within

$1/(2k+2)$ of L'. Adding the two order statements (if $a \le b$ then $a+c \le b+c$, since $b+c < a+c$ would give $b < a$ after adding $-c$, by Proposition 12.19(1); then use transitivity of $\le$ and $\eta(1/(2k+2)) + \eta(1/(2k+2)) = \eta(1/(k+1))$), L and L' are within $1/(k+1)$. This holds for every k, so $L = L'$ by Corollary 12.24. No choice is involved.

Now let (x_n) be a Cauchy sequence of points. Call an index N *good up to* k when the Cauchy condition holds from N on at every tolerance $1/(j+1)$ with $j \le k$. Every k has an index good up to k: by induction on k, the largest of an index good up to k and an index supplied by the Cauchy condition at $k+1$ is good up to $k+1$, since an index beyond a good one is good. For each k some regular sequence names x_N for some index N good up to k. Countable choice, applied once, supplies for each k such a regular sequence $s^{(k)}$; the index itself is never selected.

For $m, m' \ge k$ the indices behind $s^{(m)}$ and $s^{(m')}$ are both good at the tolerance $1/(k+1)$, and the condition at the smaller of them covers both, so the points of $s^{(m)}$ and $s^{(m')}$ are within $1/(k+1)$. These sequences therefore form a Cauchy family, so by Theorem 12.25 their diagonal names a point L, and the point of $s^{(k)}$ is within $1/(k+1)$ of L for every k. Given k, let N be an index behind $s^{(2k+1)}$. For $m \ge N$, the point x_m is within $1/(2k+2)$ of x_N, which is within $1/(2k+2)$ of L; adding the two order statements as in the first paragraph, x_m is within $1/(k+1)$ of L. So (x_n) converges to L. The indices are used only inside the proofs of these propositions. □

The proof uses countable choice once, to pick representatives; the limit itself is computed from the data. A second description of the continuum avoids even that step.

Definition 12.27 (Located cuts)**.** A *located cut* is a pair $C = (\mathscr{L}, \mathscr{U})$ of sets of ratios, both inhabited, with $\mathscr{L}$ closed downward and without a greatest element, $\mathscr{U}$ closed upward and without a least element, the two disjoint, and *located*: whenever $q < r$, either $q \in \mathscr{L}$ or $r \in \mathscr{U}$. The *cut of* a point x is $C_x = (\mathscr{L}_x, \mathscr{U}_x)$, where $\mathscr{L}_x = \{q : \eta(q) < x\}$ and $\mathscr{U}_x = \{r : x < \eta(r)\}$. For located cuts C and D and a positive ratio ε, write $C \preceq_\varepsilon D$ when every member of the lower set of C lies less than ε above every member of the upper set of D, that is, $q - r < \varepsilon$ whenever q is in the lower set of C and r in the upper set of D.

Lemma 12.28 (Laws of located cuts)**.** *Let $C = (\mathscr{L}, \mathscr{U})$, D and E be located cuts.*

1. *Every member of $\mathscr{L}$ lies below every member of $\mathscr{U}$.*
2. *A ratio r lies in $\mathscr{U}$ exactly when some $r' < r$ is not in $\mathscr{L}$. Hence two located cuts with the same lower set are equal.*
3. *If $C \preceq_\varepsilon D$ and $D \preceq_\delta E$, then $C \preceq_{\varepsilon+\delta} E$.*
4. *If $D \preceq_{1/(k+1)} E$ for every k, then every member of the lower set of D lies in the lower set of E.*

No choice principle is used.

Proof. 1. Let $q \in \mathscr{L}$ and $r \in \mathscr{U}$. If $r \le q$, take $q' \in \mathscr{L}$ with $q' > q$; then $r < q'$, so $r \in \mathscr{L}$, and r would lie in both sets. Comparisons of ratios are decidable, so $q < r$.

2. If $r \in \mathscr{U}$, take $r' \in \mathscr{U}$ with $r' < r$; it is not in $\mathscr{L}$, because the two sets are disjoint. Conversely, if $r' < r$ and $r' \notin \mathscr{L}$, locatedness at $r' < r$ gives $r \in \mathscr{U}$. So the upper set is determined by the lower set.

3. Let q be in the lower set of C and r in the upper set of E, and suppose $q - r \geq \varepsilon + \delta$. Take $q' > q$ in the lower set of C. Then $r + \delta < q' - \varepsilon$, and locatedness of D at this pair puts $r + \delta$ in the lower set of D or $q' - \varepsilon$ in its upper set. The first contradicts $D \preceq_\delta E$ at the pair $r + \delta$ and r, whose difference is δ; the second contradicts $C \preceq_\varepsilon D$ at the pair q' and $q' - \varepsilon$, whose difference is ε. Comparisons of ratios are decidable, so $q - r < \varepsilon + \delta$.

4. Let q be in the lower set of D, and take $q < q' < q''$ in it. Locate E at $q < q'$. If q' were in the upper set of E, then $q'' - q' < 1/(k+1)$ for every k, which fails once $k + 1 > 1/(q'' - q')$. So q is in the lower set of E. □

Call a sequence $C_0, C_1, \ldots$ of located cuts, with $C_n = (\mathscr{L}_n, \mathscr{U}_n)$, *Cauchy* when for every k there is an N such that $C_m \preceq_{1/(k+1)} C_n$ whenever $m, n \geq N$, and *regular* when $C_m \preceq_{1/(m+1)+1/(n+1)} C_n$ for all m and n. A located cut D is a *limit* of the sequence when for every k there is an N such that $C_n \preceq_{1/(k+1)} D$ and $D \preceq_{1/(k+1)} C_n$ for all $n \geq N$: the cuts approach D in the sense of the Cauchy condition, with D in place of either of the two cuts.

Theorem 12.29 (Cauchy sequences of located cuts)**.** *Every Cauchy sequence of located cuts has exactly one limit D. A ratio q lies in the lower set of D exactly when, for some k, $q + 1/(k+1) \in \mathscr{L}_m$ for every m from some index on; a ratio r lies in its upper set exactly when, for some k, $r - 1/(k+1) \in \mathscr{U}_m$ for every m from some index on. Every regular sequence of located cuts is Cauchy. No choice principle and no omniscience principle is used.*

Proof. Call an index N *good at* k when $C_m \preceq_{1/(k+1)} C_n$ for all $m, n \geq N$; the hypothesis says that every k has a good index. Let $\mathscr{L}^*$ consist of the ratios q for which there are k, an index N good at k, and $q' \in \mathscr{L}_N$ with $q \leq q' - 2/(k+1)$; let $\mathscr{U}^*$ consist of the ratios r for which there are k, an index N good at k, and $r' \in \mathscr{U}_N$ with $r \geq r' + 2/(k+1)$. We show that $D = (\mathscr{L}^*, \mathscr{U}^*)$ has the stated properties.

Inhabited. Let N be good at 0. A member a of $\mathscr{L}_N$ gives $a - 2 \in \mathscr{L}^*$, and a member b of $\mathscr{U}_N$ gives $b + 2 \in \mathscr{U}^*$.

Closed, and without end points. If $q \in \mathscr{L}^*$ and $p < q$, the same k, N and q' show $p \in \mathscr{L}^*$. If $q \in \mathscr{L}^*$ by k, N and q', take $q'' > q'$ in $\mathscr{L}_N$; then $q'' - 2/(k+1)$ lies in $\mathscr{L}^*$ by k, N and q'', and exceeds q. The upper set is treated in the same way.

Disjoint. Suppose $q \in \mathscr{L}^*$ by k, N, q' and $q \in \mathscr{U}^*$ by k', N', r'. Then $q' - r' \geq 2/(k+1) + 2/(k'+1)$. If $N \leq N'$, goodness of N at k, applied to the cuts C_N and $C_{N'}$, gives $q' - r' < 1/(k+1)$; if $N' \leq N$, goodness of N' at k' gives $q' - r' < 1/(k'+1)$. Both are impossible.

Located. Let $q < r$. Choose k with $4/(k+1) < r - q$ and an index N good at k. Since $q + 2/(k+1) < r - 2/(k+1)$, locatedness of C_N gives $q + 2/(k+1) \in \mathscr{L}_N$, so that $q \in \mathscr{L}^*$, or $r - 2/(k+1) \in \mathscr{U}_N$, so that $r \in \mathscr{U}^*$. The index N is used inside the proof of this disjunction and is never selected.

D is a limit. Given k, let N be good at k and let $n \geq N$. Take $q \in \mathscr{L}_n$ and $r \in \mathscr{U}^*$ by k', N', r', and suppose $q - r \geq 1/(k+1)$. Then $q - r' \geq 1/(k+1) + 2/(k'+1)$. If $N' \leq n$, goodness of N' at k', applied to C_n and $C_{N'}$, gives $q - r' < 1/(k'+1)$; if $n \leq N'$, then n and N' are at least N, and goodness of N gives $q - r' < 1/(k+1)$. Either way we have a contradiction, so $C_n \preceq_{1/(k+1)} D$. In the same way $D \preceq_{1/(k+1)} C_n$.

Uniqueness. Let D and E both be limits. Given k, take n beyond the indices that the limit property supplies for D and for E at the tolerance $1/(2k+2)$. Then $D \preceq_{1/(2k+2)} C_n$ and $C_n \preceq_{1/(2k+2)} E$, so $D \preceq_{1/(k+1)} E$ by Lemma 12.28(3). By Lemma 12.28(4) the lower set of D lies in that of E; by symmetry the two lower sets are equal, and $D = E$ by Lemma 12.28(2).

The lower set. Let D be the limit. If q lies in its lower set, take $q'' > q$ in it, choose k with $3/(k+1) < q'' - q$, and N with $D \preceq_{1/(k+1)} C_m$ for $m \geq N$. For $m \geq N$, locate C_m at $q + 1/(k+1) < q + 2/(k+1)$. If $q + 2/(k+1) \in \mathscr{U}_m$, then $q'' - (q + 2/(k+1)) < 1/(k+1)$, that is $q'' - q < 3/(k+1)$, which is false. So $q + 1/(k+1) \in \mathscr{L}_m$ for every $m \geq N$. Conversely, suppose $q + 1/(k+1) \in \mathscr{L}_m$ for every $m \geq N$. Locate D at $q < q + 1/(2k+2)$. If $q + 1/(2k+2)$ lay in the upper set of D, take N' with $C_m \preceq_{1/(2k+2)} D$ for $m \geq N'$; at $m = N + N'$ the pair $q + 1/(k+1) \in \mathscr{L}_m$ and $q + 1/(2k+2)$ would have difference exactly $1/(2k+2)$, which is not less than $1/(2k+2)$. So q lies in the lower set of D.

The upper set. If r lies in the upper set of D, take $r'' < r$ in it, choose k with $3/(k+1) < r - r''$, and N with $C_m \preceq_{1/(k+1)} D$ for $m \geq N$. For $m \geq N$, locate C_m at $r - 2/(k+1) < r - 1/(k+1)$. If $r - 2/(k+1) \in \mathscr{L}_m$, then $(r - 2/(k+1)) - r'' < 1/(k+1)$, that is $r - r'' < 3/(k+1)$, which is false. So $r - 1/(k+1) \in \mathscr{U}_m$ for every $m \geq N$. Conversely, suppose $r - 1/(k+1) \in \mathscr{U}_m$ for every $m \geq N$, and locate D at $r - 1/(2k+2) < r$. If $r - 1/(2k+2)$ lay in the lower set of D, take N' with $D \preceq_{1/(2k+2)} C_m$ for $m \geq N'$; at $m = N + N'$ the pair $r - 1/(2k+2)$ and $r - 1/(k+1) \in \mathscr{U}_m$ would have difference exactly $1/(2k+2)$. So r lies in the upper set of D.

Regular sequences. If $C_m \preceq_{1/(m+1)+1/(n+1)} C_n$ for all m and n, the index $2k+1$ is good at k, because $1/(m+1) + 1/(n+1) \leq 1/(k+1)$ when $m, n \geq 2k+1$.

Every step compares two indices or two ratios, which is decidable, or uses an index whose existence the hypothesis states, inside the proof of a proposition. Nothing is selected. □

The limit is read off from the sets alone: its lower and upper sets are described by the cuts of the sequence, with no choice principle at all. The properties of the limit that need an index from the Cauchy condition (disjointness, locatedness and the limit property) use that index only inside the proof of a proposition, so no index is ever selected. To compare cuts with points we need one more fact about the line: there is a ratio between any two points.

Lemma 12.30 (Ratios are dense)**.** *If $x < y$, there is a ratio q with $x < \eta(q) < y$. No choice principle is used.*

Proof. Let s and t name x and y. The point $y - x$ is named by $u_n = t_{2n+1} - s_{2n+1}$, and it is positive, so $u_N > 2/(N+1)$ for some N. By Lemma 12.16, $u_M > 1/(2N+2) = 4/(M+1)$ at $M = 8N+7$. Let q be the mean of s_{2M+1} and t_{2M+1}. Then $t_{2M+1} - q = q - s_{2M+1} = u_M/2 > 2/(M+1)$. These are the stages at M of the sequences $(t_{2j+1} - q)_j$ and $(q - s_{2j+1})_j$, which name $y - \eta(q)$ and $\eta(q) - x$, so both points are positive. □

Lemma 12.31 (The cut of a point)**.** *For every point x, the cut C_x is a located cut. Moreover $x \leq y$ exactly when $\mathscr{L}_x \subseteq \mathscr{L}_y$, so different points have different cuts. No choice principle is used.*

Proof. By Proposition 12.19(4), $x < \eta(n)$ and $-x < \eta(n')$ for some natural numbers n and n'; so $n \in \mathscr{U}_x$ and, adding $x - \eta(n')$ to the second inequality, $-n' \in \mathscr{L}_x$. If $p < q \in \mathscr{L}_x$, then

$\eta(p) < \eta(q) < x$, so $p \in \mathscr{L}_x$. By Lemma 12.30, $\eta(q) < x$ gives a ratio q' with $\eta(q) < \eta(q') < x$, and then $q < q' \in \mathscr{L}_x$, so $\mathscr{L}_x$ has no greatest element. The upper set is treated in the same way. A ratio in both sets would give $\eta(q) < x < \eta(q)$. If $q < r$, then $\eta(q) < \eta(r)$, and cotransitivity gives $\eta(q) < x$ or $x < \eta(r)$: the cut is located.

If $x \le y$ and $\eta(q) < x$, cotransitivity gives $\eta(q) < y$ or $y < x$, and the second is excluded. Conversely, if $\mathscr{L}_x \subseteq \mathscr{L}_y$ and $y < x$, Lemma 12.30 gives q with $y < \eta(q) < x$; then $q \in \mathscr{L}_x \subseteq \mathscr{L}_y$ and $\eta(q) < y < \eta(q)$, which is impossible. So $x \le y$. Equal cuts therefore give $x \le y \le x$, and $x = y$ by Proposition 12.19(2). □

A cut is the cut of a point exactly when it can be approximated by a sequence of pairs of ratios. Each single pair can be found without choice; assembling all of them into one sequence is exactly an instance of AC_ω at pairs of ratios, and nothing else is needed.

Lemma 12.32 (One approximating pair needs no choice)**.** *Let $(\mathscr{L}, \mathscr{U})$ be a located cut and n a natural number. There are $q \in \mathscr{L}$ and $r \in \mathscr{U}$ with $r - q < 1/(n+1)$. No choice principle is used.*

Proof. Take $a \in \mathscr{L}$ and $b \in \mathscr{U}$; then $a < b$ by Lemma 12.28(1). Choose a positive integer $G \ge 4(n+1)$ for which $K = (b-a)G$ is a natural number, for instance the product of the two denominators and $4(n+1)$, and put $p_j = a + j/G$. We prove by induction on j: either $p_j \in \mathscr{L}$, or there is an i with $p_i \in \mathscr{L}$ and $p_{i+2} \in \mathscr{U}$. For $j = 0$, $p_0 = a \in \mathscr{L}$. If $p_j \in \mathscr{L}$, locatedness at $p_{j+1} < p_{j+2}$ gives $p_{j+1} \in \mathscr{L}$, or $p_{j+2} \in \mathscr{U}$, and then $i = j$ serves; if the second alternative already holds for j, it holds for $j+1$. At $j = K$ we have $p_K = b \in \mathscr{U}$, so $p_K \notin \mathscr{L}$, and the second alternative holds: $q = p_i$ and $r = p_{i+2}$ satisfy $r - q = 2/G \le 1/(2n+2) < 1/(n+1)$. The induction proves a proposition at each step and selects nothing. □

Theorem 12.33 (Located cuts and points)**.** *A located cut $(\mathscr{L}, \mathscr{U})$ is the cut of a point exactly when it has an* approximation sequence*: a sequence of pairs $q_n \in \mathscr{L}$, $r_n \in \mathscr{U}$ with $r_n - q_n < 1/(n+1)$ for every n. Neither direction uses a choice principle. Assuming AC_ω for pairs of ratios, every located cut is the cut of exactly one point, and $x \mapsto C_x$ is a bijection from $\mathbb{R}_\delta$ onto the located cuts; countable choice is used once, at pairs of ratios, and no omniscience principle is used.*

Proof. Let x be named by s. For each n put $m = 4n+4$, $q_n = s_m - 2/(m+1)$ and $r_n = s_m + 2/(m+1)$. The point $x - \eta(q_n)$ is named by $(s_{2j+1} - q_n)_j$, and at $j = 3m+2$ regularity gives

$$s_{6m+5} - q_n \ge \frac{2}{m+1} - \frac{1}{m+1} - \frac{1}{6m+6} = \frac{5}{6m+6} > \frac{2}{j+1},$$

so $\eta(q_n) < x$; in the same way $x < \eta(r_n)$. And $r_n - q_n = 4/(4n+5) < 1/(n+1)$. Only the one sequence s naming the one point x is used.

Conversely, let (q_n, r_n) be an approximation sequence. Since $q_m < r_n$ (Lemma 12.28(1)) and $r_n - q_n < 1/(n+1)$, we have $q_m - q_n < 1/(n+1)$, and symmetrically $q_n - q_m < 1/(m+1)$; so (q_n) is a regular sequence. Let x be its point; we show $\mathscr{L} = \mathscr{L}_x$. If $\eta(q) < x$, the sequence $(q_{2j+1} - q)_j$ naming $x - \eta(q)$ has a positive stage, so $q < q_{2N+1} \in \mathscr{L}$ for some N, and $q \in \mathscr{L}$. If $q \in \mathscr{L}$, take $q' > q$ in $\mathscr{L}$ and put $\delta = q' - q$. Every k has $q' < r_k$, hence $q_k > q' - 1/(k+1)$, and so $q_{2j+1} - q > \delta - 1/(2j+2)$; for j with $2j + 2 > 5/\delta$ this exceeds $2/(j+1)$, so $\eta(q) < x$. The located cuts $(\mathscr{L}, \mathscr{U})$ and C_x have the same lower set, so they are equal by Lemma 12.28(2).

For the last sentence, Lemma 12.32 gives, for every n, some pair approximating the cut at precision n. Countable choice for pairs of ratios, applied once to this relation, gives an approximation sequence, and the first part gives the point. The point is unique by Lemma 12.31, which also makes $x \mapsto C_x$ injective. □

The cuts of the points of a Cauchy sequence of points form a Cauchy sequence of located cuts. This rests on reading "within" from the cuts.

Lemma 12.34 (Distances read from cuts)**.** *Let $\varepsilon = 1/(k+1)$. Points x and y are within ε exactly when $C_x \preceq_\varepsilon C_y$ and $C_y \preceq_\varepsilon C_x$. No choice principle is used.*

Proof. We show that $x \le y + \eta(\varepsilon)$ implies $C_x \preceq_\varepsilon C_y$, and that $C_x \preceq_\varepsilon C_y$ implies $x \le y + \eta(\varepsilon)$; exchanging x and y gives the lemma. Recall that $u \le v < w$ and $u < v \le w$ each give $u < w$, by cotransitivity.

Suppose $x \le y+\eta(\varepsilon)$, $\eta(q) < x$, $y < \eta(r)$ and $q-r \ge \varepsilon$. Then $y+\eta(\varepsilon) < \eta(r)+\eta(\varepsilon) = \eta(r+\varepsilon) \le \eta(q)$, so $x < \eta(q)$, contrary to $\eta(q) < x$. Comparisons of ratios are decidable, so $q - r < \varepsilon$.

Suppose $C_x \preceq_\varepsilon C_y$ and $y + \eta(\varepsilon) < x$. Lemma 12.30 gives q with $y + \eta(\varepsilon) < \eta(q) < x$; then $y < \eta(q - \varepsilon)$, so $q \in \mathscr{L}_x$ and $q - \varepsilon \in \mathscr{U}_y$, and their difference is ε, which is not less than ε. So $x \le y + \eta(\varepsilon)$. □

Corollary 12.35 (The limit of a Cauchy sequence as a cut)**.** *Let (x_n) be a Cauchy sequence of points. With no choice principle: the cuts of its terms form a Cauchy sequence of located cuts and so have exactly one limit D; a point L is a limit of (x_n) exactly when $C_L = D$. Hence (x_n) converges exactly when D is the cut of a point, and if every located cut is the cut of a point, every Cauchy sequence of points converges with no further choice. In particular, countable choice for pairs of ratios, applied once, makes every Cauchy sequence of points converge.*

Proof. By Lemma 12.34, the Cauchy condition for the points gives the Cauchy condition for their cuts, so Theorem 12.29 supplies the unique limit D. By the same lemma, "x_n is within $1/(k+1)$ of L for $n \ge N$" says that $C_{x_n} \preceq_{1/(k+1)} C_L$ and $C_L \preceq_{1/(k+1)} C_{x_n}$ for $n \ge N$. So L is a limit of (x_n) exactly when C_L is a limit of the cuts, that is, by uniqueness, exactly when $C_L = D$. The remaining assertions follow from Theorem 12.33. □

So the cuts of the points of any Cauchy sequence of points have a unique limit cut, with no choice; countable choice is used only to make that cut a point, and Theorem 12.63 shows that excluded middle, without unique choice, cannot replace it.

The same rung has an order form. A *set of points* is any property of points; it is *inhabited* if some point has it, and *bounded above* if some point lies at or above every member, an *upper bound*. A *least upper bound* of a set of points is an upper bound lying below every upper bound. Call a set of points *order-located* if, whenever $q < r$, either $\eta(r)$ is an upper bound of the set or some member exceeds $\eta(q)$.

Theorem 12.36 (Least upper bounds of order-located sets)**.** *Every inhabited, order-located set of points that is bounded above has a least upper bound exactly when every located cut is the cut of a point. Neither direction uses a choice principle. A set of points has at most one least upper bound, and if x is a least upper bound of an order-located set S, every ratio q with $\eta(q) < x$ has $\eta(q) < s$ for some $s \in S$.*

Proof. Two least upper bounds lie below each other, so they are equal by Proposition 12.19(2). Let x be a least upper bound of an order-located set S, and let $\eta(q) < x$. Lemma 12.30 gives a ratio q' with $\eta(q) < \eta(q') < x$. At $q < q'$, either $\eta(q')$ is an upper bound of S, and then $x \le \eta(q')$, which denies $\eta(q') < x$, or some member exceeds $\eta(q)$. Hence

$$\mathscr{L}_x = \bigcup_{s \in S} \mathscr{L}_s, \tag{12.5}$$

the inclusion $\supseteq$ holding because $s \le x$ for $s \in S$ (Lemma 12.31).

Suppose every located cut is the cut of a point, and let S be inhabited, order-located and bounded above by a point b. By Proposition 12.19(4) there is a ratio b' with $b < \eta(b')$, and $\eta(b')$ is also an upper bound. Let $\mathscr{L}^*$ be the set of ratios q with $\eta(q) < s$ for some $s \in S$, and $\mathscr{U}^*$ the set of ratios r above some ratio r' for which $\eta(r')$ is an upper bound of S. These sets form a located cut. A member of S has a ratio below it, and every $r > b'$ lies in $\mathscr{U}^*$. The lower set is closed downward and has no greatest element, because each $\mathscr{L}_s$ has these properties. The upper set is closed upward, and if $r' < r$ with $\eta(r')$ an upper bound, the mean of r' and r is a smaller member of $\mathscr{U}^*$. A ratio q in both sets would give $s \in S$ and $r' < q$ with $s \le \eta(r') < \eta(q) < s$. Finally, if $q < r$, order-locatedness at q and the mean m of q and r says either that $\eta(m)$ is an upper bound, so $r \in \mathscr{U}^*$, or that some member exceeds $\eta(q)$, so $q \in \mathscr{L}^*$. Let x be the point whose cut is $(\mathscr{L}^*, \mathscr{U}^*)$. For $s \in S$, $\mathscr{L}_s \subseteq \mathscr{L}_x$, so $s \le x$ by Lemma 12.31. If y is an upper bound and $q \in \mathscr{L}_x$, then $\eta(q) < s \le y$ for some $s \in S$, so $\mathscr{L}_x \subseteq \mathscr{L}_y$ and $x \le y$. Thus x is a least upper bound.

Conversely, suppose the least-upper-bound property, and let $(\mathscr{L}, \mathscr{U})$ be a located cut. The set $S = \{\eta(q) : q \in \mathscr{L}\}$ is inhabited, and $\eta(r)$ is an upper bound of it for every $r \in \mathscr{U}$. It is order-located: if $a < b$, then either $a \in \mathscr{L}$, and some $a' > a$ in $\mathscr{L}$ gives the member $\eta(a') > \eta(a)$, or $b \in \mathscr{U}$, and $\eta(b)$ is an upper bound. Let x be its least upper bound. By (12.5), $q \in \mathscr{L}_x$ exactly when $q < q'$ for some $q' \in \mathscr{L}$, that is, exactly when $q \in \mathscr{L}$, since $\mathscr{L}$ is closed downward and has no greatest element. So C_x and $(\mathscr{L}, \mathscr{U})$ have the same lower set, and they are equal by Lemma 12.28(2). No step selects anything. □

A finite set needs neither choice nor any decision. Call a set of points *finite and inhabited* when it is given by a list $x_0, x_1, \dots, x_n$ of points, $n \ge 0$, repetitions allowed.

Proposition 12.37 (Finite sets have least upper bounds)**.** *Let $x_0, x_1, \dots, x_n$ be points. Put $M_0 = x_0$ and $M_i = \max(x_i, M_{i-1})$ for $1 \le i \le n$. Then M_n is a least upper bound of $\{x_0, \dots, x_n\}$. Moreover, if $\eta(q) < M_n$ for a ratio q, then $\eta(q) < x_i$ for some $i \le n$. No choice principle and no omniscience principle is used.*

Proof. First, $a < \max(x, y)$ implies $a < x$ or $a < y$. Let c, s and t name a, x and y. The maximum is computed stagewise, so the point $\max(x, y) - a$ is named by $\max(s_{2j+1}, t_{2j+1}) - c_{2j+1}$, and

it has a stage N above $2/(N+1)$. Comparing the ratios s_{2N+1} and t_{2N+1}, the larger one, say s_{2N+1}, gives $s_{2N+1} - c_{2N+1} > 2/(N+1)$, a positive stage of $x - a$; so $a < x$.

Now induct on n. For $n = 0$, $x_0 \le x_0$, every upper bound of $\{x_0\}$ lies above x_0, and $\eta(q) < M_0$ is $\eta(q) < x_0$. Suppose the assertions hold for M_{n-1} and $\{x_0, \dots, x_{n-1}\}$. By Proposition 12.19(5), $x_n \le M_n$ and $M_{n-1} \le M_n$, so every x_i with $i < n$ satisfies $x_i \le M_{n-1} \le M_n$, and M_n is an upper bound. If y is an upper bound of $\{x_0, \dots, x_n\}$, it is one of $\{x_0, \dots, x_{n-1}\}$, so $M_{n-1} \le y$, and also $x_n \le y$; since $\max(x_n, M_{n-1})$ is the least upper bound of these two points, $M_n \le y$. Finally, $\eta(q) < M_n$ gives $\eta(q) < x_n$ or $\eta(q) < M_{n-1}$ by the first paragraph, and in the second case $\eta(q) < x_i$ for some $i < n$. □

With countable choice the two descriptions of the continuum coincide, and the price of identifying them is that single selection.

12.9.3 The line is determined by its task

The same selection settles a question asked of every stage of the number tower in Chapter 4: is the line the unique solution of its task? Its task is to receive exactly one map, preserving the arithmetic and the order, from every Archimedean ordered field: the structure that the next theorem defines by laws that Proposition 12.19 proves for the line.

Theorem 12.38 (The line is determined by its task)**.** *Call a commutative ring with a strict order $<$ an* Archimedean ordered field *when $0 < 1$, the order is irreflexive, transitive and cotransitive, two elements neither of which lies below the other are equal, adding an element preserves $<$, a product of positive elements is positive, every positive element has an inverse, and every element lies below some natural number. Then $\mathbb{R}_\delta$ is an Archimedean ordered field, and:*

1. *from every Archimedean ordered field there is at most one map into $\mathbb{R}_\delta$ that preserves 0, 1, addition, multiplication and $<$, and the only such map from $\mathbb{R}_\delta$ to itself is the identity;*
2. *every Archimedean ordered field has such a map into $\mathbb{R}_\delta$ exactly when every located cut is the cut of a point.*

Neither assertion uses countable choice. The forward direction of the second reads a relation with exactly one output for each input as a map, as Appendix C does; this reading is the principle of unique choice. With countable choice, therefore, every Archimedean ordered field maps into $\mathbb{R}_\delta$ in exactly one such way.

Conditions and scope.

Base and premises

The line $\mathbb{R}_\delta$ of this chapter and an Archimedean ordered field as defined in the statement. Neither part uses countable choice; the forward direction of part 2 uses unique choice, the book's reading of maps.

Conclusion

$\mathbb{R}_\delta$ is an Archimedean ordered field; from every Archimedean ordered field there is at most one structure-preserving map into $\mathbb{R}_\delta$, and the identity is the only one from $\mathbb{R}_\delta$ to

itself; every such field has one exactly when every located cut is the cut of a point. With countable choice every such field has exactly one.

Proof. Every map in the theorem is compared with one canonical passage: from an element of an Archimedean ordered field to its located cut. We place the ratios inside such a field, define the cut of an element, give the located cuts the arithmetic and the order read from their sets, and show that a map into the located cuts preserving the structure is unique. The two assertions then follow. The routine verifications are collected in Appendix F.2; the steps that carry the argument are here.

The line is an Archimedean ordered field. By Proposition 12.19, $\mathbb{R}_\delta$ is a commutative ring, and by its item 1 the order is irreflexive, transitive and cotransitive, and adding a point preserves it. Multiplying by a positive point y preserves $<$, so $0 < x$ gives $0 = 0 \cdot y < xy$: a product of positive points is positive. If neither of x, y lies below the other, then $x \le y$ and $y \le x$, and item 2 gives $x = y$. Positive points are invertible by item 3, every point lies below some $\eta(n)$ by item 4, and $\eta(0) < \eta(1)$ because η preserves the order of the ratios.

The ratios in an Archimedean ordered field. Let F be an Archimedean ordered field. Corollary 4.14 gives exactly one map from $\mathbb{Z}_\delta$ into F preserving 1, sums and products; write u_F for the image of the integer u. The axioms alone give three facts, proved in Lemma F.3: $u_F < v_F$ exactly when $u < v$; the inverse of a positive element is positive; and multiplying by a positive element preserves and reflects $<$. For a positive integer d the element d_F is therefore positive, so it has an inverse, which is unique because inverses in a commutative ring are. For the ratio $q = \langle u; d\rangle$ with $d > 0$ put $q_F = u_F d_F^{-1}$. Cross-equal records give the same element, the assignment $q \mapsto q_F$ preserves sums and products, and $q < r$ exactly when $q_F < r_F$. In $\mathbb{R}_\delta$ the image of q is $\eta(q)$, because η preserves sums and products and $\eta(\langle d; 1\rangle)\,\eta(\langle 1; d\rangle) = \eta(1)$.

The cut of an element. For x in F let $\kappa_F(x)$ be the pair of sets

$$\{q \,:\, q_F < x\}, \qquad \{r \,:\, x < r_F\}.$$

This is a located cut. Its lower set has no greatest element: if $q_F < x$, the positive element $x - q_F$ has an inverse e, the Archimedean condition gives a natural number N with $e < N_F$, so $1 < N_F(x - q_F)$, and the ratio $q + \langle 1; N\rangle$ still lies below x. Its lower and upper sets are located by cotransitivity: if $q < r$, then $q_F < r_F$, so $q_F < x$ or $x < r_F$. The remaining laws, and the facts that κ_F preserves and reflects $<$ and is injective, are proved in Lemma F.5. For $F = \mathbb{R}_\delta$ the pair $\kappa_F(x)$ is the cut C_x of the point x (Definition 12.27), since the image of q is $\eta(q)$; we write it $\kappa(x)$. The ratios themselves form an Archimedean ordered field, in which the order is decided and hence cotransitive; there $\kappa(q)$ is the pair $\{p \,:\, p < q\}, \{p \,:\, q < p\}$, the *cut of the ratio* q, which is also the cut of the point $\eta(q)$.

The arithmetic of located cuts. For located cuts $C = (\mathscr{L}_C, \mathscr{U}_C)$ and $D = (\mathscr{L}_D, \mathscr{U}_D)$ define $C + D$, $-C$ and CD by their lower and upper sets:

$$\begin{aligned}
C + D: &\quad \{q \,:\, q < a + b\}, \qquad \{r \,:\, a' + b' < r\},\\
-C: &\quad \{q \,:\, -q \in \mathscr{U}_C\}, \qquad \{r \,:\, -r \in \mathscr{L}_C\},\\
CD: &\quad \{q \,:\, q < ab,\ q < ab',\ q < a'b,\ q < a'b'\},\\
&\quad \{r \,:\, ab < r,\ ab' < r,\ a'b < r,\ a'b' < r\},
\end{aligned}$$

where in each set a ranges over $\mathscr{L}_C$, a' over $\mathscr{U}_C$, b over $\mathscr{L}_D$ and b' over $\mathscr{U}_D$, and the condition asks for some such ratios. Let $C < D$ mean that some ratio lies in $\mathscr{U}_C$ and in $\mathscr{L}_D$, and let zero and one be the cuts of the ratios 0 and 1. The product compares a ratio with all four corners of the brackets $[a, a']$ and $[b, b']$, so it never asks for the sign of a cut. Lemma F.6 proves that with these operations and this order the located cuts form an Archimedean ordered field, and that κ_F carries 0, 1, sums and products of F to those of the cuts. Two tools carry these verifications. Two located cuts are equal as soon as no lower member of either is an upper member of the other (Lemma F.1). And for every n a located cut has a lower member q and an upper member r with $r - q < 1/(n+1)$ (Lemma 12.32), between any given lower and upper member (Lemma F.2): starting from one lower and one upper member, the ratios of a fine grid between them are compared with the cut by locatedness, one after another, and an induction on the number of grid points proves that two grid points two steps apart lie on opposite sides. That proof establishes a disjunction step by step and selects nothing.

The cut of the ratio q is the image of q in this field: by the lemma for the field of ratios, cuts of ratios add and multiply as the ratios do. So the cut of an element C of the field of located cuts is C itself. Indeed, $\kappa(q) < C$ means that some ratio $p > q$ lies in $\mathscr{L}_C$, and this holds exactly when $q \in \mathscr{L}_C$, because $\mathscr{L}_C$ is closed downward and has no greatest element.

A map into the located cuts is unique. Let h send F into the located cuts preserving 0, 1, sums, products and $<$. We show $h(x) = \kappa_F(x)$ for every x. First, h sends q_F to the cut of q. For natural numbers this follows by induction, since $h(n_F + 1_F) = h(n_F) + h(1_F)$ and cuts of ratios add as ratios do. Since $h(y) + h(-y) = h(0)$ and negatives in a ring are unique, $h(-y) = -h(y)$, which covers the integers. For $q = \langle u; d\rangle$ we have $d_F\, q_F = u_F$, so $\kappa(d)\, h(q_F) = \kappa(u)$; multiplying by $\kappa(\langle 1; d\rangle)$ and using $\kappa(\langle 1; d\rangle)\kappa(d) = \kappa(1)$ gives $h(q_F) = \kappa(q)$. Now compare lower sets. If $q_F < x$, then $\kappa(q) = h(q_F) < h(x)$, so some ratio $p > q$ lies in the lower set of $h(x)$, and then so does q. Conversely, let q lie in the lower set of $h(x)$, and choose a larger ratio q' in that lower set. Cotransitivity in F, applied to $q_F < q'_F$, gives $q_F < x$ or $x < q'_F$. In the second case $h(x) < h(q'_F) = \kappa(q')$, so some ratio $s < q'$ lies in the upper set of $h(x)$; but $s < q'$ also puts s in the lower set of $h(x)$, and the two sets are disjoint. Hence $q_F < x$. The lower sets of $h(x)$ and $\kappa_F(x)$ agree, and a located cut is determined by its lower set, so $h(x) = \kappa_F(x)$.

The first assertion. Let f send F into $\mathbb{R}_\delta$ preserving 0, 1, sums, products and $<$. The composite $x \mapsto \kappa(f(x))$ preserves the same structure, because the cut of a point does, by Lemmas F.5 and F.6(b) for $F = \mathbb{R}_\delta$. By the uniqueness just proved, $\kappa(f(x)) = \kappa_F(x)$ for every x. A point is determined by its cut, so $f(x)$ is determined by x, and any two such maps agree. The identity of $\mathbb{R}_\delta$ is such a map, so it is the only one.

The second assertion. Suppose first that every located cut is the cut of a point, and let F be an Archimedean ordered field. For each x in F, the located cut $\kappa_F(x)$ is the cut of a point, and of only one, since different points have different cuts. Send x to the point $f(x)$ with that cut. This prescription specifies exactly one output for each input, which is what a map is (Appendix C); nothing is selected, since there is only one candidate. If maps are required instead to be operations given in advance, this sentence is the one place where the principle of unique choice is used: a relation with exactly one output for each input defines a map, here from the located cuts into $\mathbb{R}_\delta$. No other step needs it, because the first assertion and the converse below concern maps that are already given. The map f preserves the structure. For

sums,

$$\kappa\big(f(x+y)\big) = \kappa_F(x+y) = \kappa_F(x) + \kappa_F(y) = \kappa\big(f(x)\big) + \kappa\big(f(y)\big) = \kappa\big(f(x)+f(y)\big),$$

so $f(x+y) = f(x) + f(y)$ because points are determined by their cuts. Products, 0 and 1 are the same. If $x < y$, then $\kappa(f(x)) = \kappa_F(x) < \kappa_F(y) = \kappa(f(y))$, and the cut of a point reflects $<$, so $f(x) < f(y)$.

Conversely, suppose that every Archimedean ordered field has such a map into $\mathbb{R}_\delta$. Apply this to the field of located cuts, and call the map g. The composite $C \mapsto \kappa(g(C))$ sends the located cuts into themselves preserving 0, 1, sums, products and $<$. By the uniqueness of maps into the cuts, it is the cut map of that field, which is the identity. So $\kappa(g(C)) = C$: every located cut is the cut of the point $g(C)$.

Every step above is an explicit construction from given data, a comparison of ratios, an application of cotransitivity, or the approximation lemma, which proves a disjunction by induction. No choice principle is used. With countable choice, every located cut is the cut of a point (Theorem 12.33); by the second assertion every Archimedean ordered field then has such a map, and by the first it has exactly one. □

With countable choice the line is therefore determined by this task up to exactly one isomorphism. If an Archimedean ordered field T also receives exactly one such map from every Archimedean ordered field, the maps $\mathbb{R}_\delta \to T$ and $T \to \mathbb{R}_\delta$ compose to such maps of $\mathbb{R}_\delta$ and of T into themselves, and the first assertion, with the same uniqueness for T, makes both composites identities. No other isomorphism exists, since T receives only one such map from $\mathbb{R}_\delta$.

12.9.4 The table of completeness prices

Table 12.3 collects the prices of the familiar forms of completeness. Its terms have their usual meanings, read in the order of $\mathbb{R}_\delta$; recall that a least upper bound of a set of points is an upper bound lying below every upper bound. In the table, f is a function on the points of $\mathbb{R}_\delta$, and $[0,1]$ is the set of points x with $\eta(0) \le x \le \eta(1)$; the set $[-1,1]$ is defined in the same way. Such an f is *uniformly continuous on* $[0,1]$ when for every k there is a j such that points of $[0,1]$ within $1/(j+1)$ have values within $1/(k+1)$. A *root* of f is a point x with $f(x) = \eta(0)$. The function is *strictly increasing* on $[0,1]$ when $x < y$ in $[0,1]$ gives $f(x) < f(y)$, and it *attains a maximum* when some x in $[0,1]$ has $f(y) \le f(x)$ for every y in $[0,1]$. A *supremum* of f is an upper bound of its values on $[0,1]$ such that, for every k, some value $f(y)$ with y in $[0,1]$ lies within $1/(k+1)$ of it. Write $c_n = \sum_{i<n} e_i 2^{-(i+1)}$ for the nth partial sum of a digit sequence (e_i). A *binary expansion* of x has digits in $\{0,1\}$ and $\eta(c_n) \le x \le \eta(c_n + 2^{-n})$ for every n; a *signed-digit expansion* has digits in $\{-1,0,1\}$ and $\eta(c_n - 2^{-n}) \le x \le \eta(c_n + 2^{-n})$ for every n.

We now prove the rows of Table 12.3 priced LPO, WLPO, weak excluded middle and excluded middle, in both directions. Every passage from an assertion to its principle uses only the arithmetic and order of Proposition 12.19, and no choice principle. Every passage back says where countable choice enters. For exact roots and binary expansions, countable choice in the converse cannot be weakened to choice for the relations whose witnesses a search can find (Theorem 12.70). For the cut row it is replaced neither by excluded middle with such choice (Theorem 12.63) nor by unique choice (Theorem 12.66).

Assertion on $\mathbb{R}_\delta$	Price	Countable choice
The diagonal of a Cauchy family of regular sequences names the limit of their points (Theorem 12.25)	none	not used
Every Cauchy sequence of points converges (Theorem 12.26)	none	used once, to select representatives
Every Cauchy sequence of located cuts has a limit cut (Theorem 12.29)	none	not used
Every located cut is the cut of a point (Theorem 12.33)	none	used once, at pairs of ratios
Every bounded nondecreasing sequence converges (Theorem 12.45)	LPO	representatives; once at pairs of ratios
Every bounded nondecreasing sequence, given by regular sequences, is Cauchy (Theorem 12.46)	LPO	not used
Every bounded sequence has a convergent subsequence; every bounded sequence has a cluster point; every sequence has a monotone subsequence (Theorem 12.47)	LPO	representatives, decisions and indices; limits as in Theorem 12.45
Every sequence bounded above has a least upper bound (Theorem 12.48)	WLPO	representatives; once at pairs of ratios
… one that every smaller ratio is exceeded by some term (Theorem 12.49)	LPO	representatives; once at pairs of ratios
Every finite inhabited set has a least upper bound (Proposition 12.37)	none	not used
Every inhabited set of points bounded above has a least upper bound (Theorem 12.50)	WEM	once, at pairs of ratios
… one that every smaller ratio is exceeded by some member (Theorem 12.51)	EM	once, at pairs of ratios
A uniformly continuous f on $[0,1]$ with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$ has a root in $[0,1]$ (Theorem 12.54)	LLPO	once, one bit per dyadic ratio
such an f has, for each k, a point x of $[0,1]$ with $f(x)$ within $1/(k+1)$ of $\eta(0)$ (Theorem 12.55)	none	not used
such an f, strictly increasing, has a root in $[0,1]$ (Theorem 12.56)	none	used once, at pairs of ratios
A uniformly continuous f on $[0,1]$ attains a maximum (Theorem 12.60)	LLPO	bits for adjacent dyadic intervals; suprema as below
such an f has a supremum (Theorem 12.58)	none	used once, at pairs of ratios
Every point of $[0,1]$ has a binary expansion (Theorem 12.61)	LLPO	once, one bit per dyadic ratio
Every point of $[-1,1]$ has a signed-digit expansion (Proposition 12.62)	none	not used

Table 12.3. Completeness on the line built from ratios. "Price" is the principle equivalent to the assertion; "none" means that no omniscience principle is needed, so the assertion is proved outright, from countable choice where the last column says it is used. WEM is the weak excluded middle, $\neg P \vee \neg\neg P$ for every proposition P, and EM is excluded middle, $P \vee \neg P$ for every proposition P. In every priced row the passage from the assertion to the principle uses no choice principle, and the last column says where countable choice enters the passage back. Each row names the result that proves both directions.

For sequences and sets of points the terms are the following.

Definition 12.39 (Terms of the table)**.** Let (x_n) be a sequence of points of $\mathbb{R}_\delta$.

1. The sequence is *bounded above* if some point b satisfies $x_n \le b$ for every n, and *bounded* if, in addition, some point b' satisfies $b' \le x_n$ for every n. It is *nondecreasing* if $x_n \le x_{n+1}$ for every n, and then $x_m \le x_n$ whenever $m \le n$, by transitivity of $\le$; it is *nonincreasing* if $x_{n+1} \le x_n$ for every n; and it is *monotone* if it is nondecreasing or nonincreasing.

2. A *subsequence* is a sequence $(x_{\sigma(n)})$ with $\sigma(0) < \sigma(1) < \cdots$. By induction $\sigma(n) \ge n$ for every n.

3. A point L is a *cluster point* of (x_n) if for every k and every N some $m \ge N$ has x_m within $1/(k+1)$ of L.

4. The sequence is *given by regular sequences* when it is presented as a sequence $s^{(0)}, s^{(1)}, \ldots$ of regular sequences, its nth term being the point of $s^{(n)}$.

5. As before Theorem 12.36, a *set of points* is any property of points; it is *inhabited* if some point has it, and *bounded above* if some point b lies at or above every member. Upper bounds and least upper bounds of a sequence are those of the set of its terms.

6. A least upper bound L of a set S is *approached by its members* if every ratio q with $\eta(q) < L$ has a member s with $\eta(q) < s$. For a sequence we say *approached by its terms*.

The rows with the words "one that every smaller ratio is exceeded by some term" (or "member") ask for a least upper bound approached in the sense of item 6.

12.9.5 Order tools for the proofs of the table

Lemma 12.40 (Order tools)**.** *For points u, v, w, x, y, z and ratios q, r:*

(a) if $u \le v < w$ or $u < v \le w$, then $u < w$; and $u < v$ implies $u \le v$;

(b) if $x \le y$, then $x + z \le y + z$; consequently $x \le y$ exactly when $-y \le -x$;

(c) $\eta(q) + \eta(r) = \eta(q + r)$, and $\eta(q) < \eta(r)$ exactly when $q < r$, while $\eta(q) \le \eta(r)$ exactly when $q \le r$.

No omniscience principle and no choice principle is used.

Proof. (a) If $v < w$, cotransitivity gives $v < u$ or $u < w$, and $v < u$ denies $u \le v$. If $u < v$, cotransitivity gives $u < w$ or $w < v$, and $w < v$ denies $v \le w$. If $u < v$ and also $v < u$, transitivity gives $u < u$, contrary to irreflexivity; so $u < v$ gives $\neg(v < u)$, which is $u \le v$.

(b) If $y + z < x + z$, adding $-z$ gives $y < x$, which denies $x \le y$. Adding $-x - y$ to $x \le y$ gives $-y \le -x$, and adding $x + y$ reverses the step.

(c) This is the embedding of the ratios in Proposition 12.19, which preserves and reflects the order; the order of ratios is decidable, so $\neg(r < q)$ is $q \le r$. □

Two more facts about the order come from the located-cut material above. Between two points lies a ratio (Lemma 12.30): if $x < y$, then $x < \eta(q) < y$ for some ratio q. And $x \le y$ exactly when every ratio q with $\eta(q) < x$ also has $\eta(q) < y$ (Lemma 12.31). Distances are the order statements of "within", and four rules govern them.

Lemma 12.41 (Distances from order)**.** *Let a, c be ratios, y, z, w points and k a natural number.*

(a) If $\eta(a) \le y \le \eta(c)$ and $\eta(a) \le z \le \eta(c)$, where $c - a \le 1/(k+1)$, then y and z are within $1/(k+1)$.

(b) If $\eta(a) < y$ and $z < \eta(c)$, where $a - c \ge 1/(k+1)$, then y and z are not within $1/(k+1)$.

(c) If y, w and w, z are within $1/(2k+2)$, then y, z are within $1/(k+1)$.

(d) y and z are within $1/(k+1)$ exactly when $-y$ and $-z$ are.

No omniscience principle and no choice principle is used.

Proof. Write $\varepsilon = 1/(k+1)$.

(a) By Lemma 12.40(b),(c), $y \le \eta(c) = \eta(a) + \eta(c-a) \le z + \eta(c-a) \le z + \eta(\varepsilon)$, and in the same way $z \le y + \eta(\varepsilon)$.

(b) Suppose they were within $1/(k+1)$. Then $y \le z + \eta(\varepsilon)$. Adding $\eta(\varepsilon)$ preserves $<$, so $z + \eta(\varepsilon) < \eta(c) + \eta(\varepsilon)$, and $\eta(c) + \eta(\varepsilon) \le \eta(c) + \eta(a - c) = \eta(a)$. By Lemma 12.40(a), first $z + \eta(\varepsilon) < \eta(a)$ and then $y < \eta(a)$. Together with $\eta(a) < y$, transitivity gives $y < y$, contrary to irreflexivity.

(c) $y \le w + \eta(1/(2k+2)) \le z + \eta(1/(2k+2)) + \eta(1/(2k+2)) = z + \eta(\varepsilon)$, and symmetrically.

(d) Adding $-y - z$ to $y \le z + \eta(\varepsilon)$ gives $-z \le -y + \eta(\varepsilon)$, and adding $y + z$ reverses the step; Lemma 12.40(b) applies in both directions, and the same holds with y and z exchanged. □

In particular the points $\eta(0)$ and $\eta(1)$ are not both within $1/4$ of any point L: that would give $\eta(1) \le L + \eta(1/4) \le \eta(1/4) + \eta(1/4) = \eta(1/2)$, contrary to $\eta(1/2) < \eta(1)$.

12.9.6 Deciding a point against a ratio

An omniscience principle decides propositions about binary sequences. To use it on points, a question about points must be turned into a binary sequence, and for that the points need regular sequences naming them.

Lemma 12.42 (Deciding against a ratio)**.** *Let $s^{(0)}, s^{(1)}, \ldots$ be regular sequences naming the points $x_0, x_1, \ldots$, and let q be a ratio. Put $\beta_q(N) = 1$ if there are $n, j \le N$ with $s^{(n)}_{2j+1} - q > 2/(j+1)$, and $\beta_q(N) = 0$ otherwise. Then β_q is a binary sequence; it has an entry 1 exactly when $\eta(q) < x_n$ for some n, and it is all zero exactly when $x_n \le \eta(q)$ for every n. Consequently, with no choice principle:*

(a) LPO gives: $\eta(q) < x_n$ for some n, or $x_n \le \eta(q)$ for every n;

(b) WLPO gives: $x_n \le \eta(q)$ for every n, or not;

(c) $\mathrm{MP}_{\mathrm{bin}}$ *turns* $\neg(\forall n,\ x_n \le \eta(q))$ *into an* n *with* $\eta(q) < x_n$.

Proof. Each value $\beta_q(N)$ is a finite computation with ratios. The regular sequence $s^{(n)} + (-\eta(q))$ names the point $x_n - \eta(q)$, and its stage j is $s^{(n)}_{2j+1} - q$. Positivity belongs to the point, so $\eta(q) < x_n$ holds exactly when some stage j of this sequence exceeds $2/(j+1)$, and then $\beta_q(\max(n, j)) = 1$. Conversely, an entry $\beta_q(N) = 1$ exhibits such n and j. Since $x_n \le \eta(q)$ is $\neg(\eta(q) < x_n)$, the sequence β_q is all zero exactly when every $x_n \le \eta(q)$, and it is not all zero exactly when $\neg(\forall n,\ x_n \le \eta(q))$. Applying LPO, WLPO or $\mathrm{MP}_{\mathrm{bin}}$ to the one sequence β_q gives (a), (b) and (c). □

12.9.7 Two cuts that bound a set

The passages back build a candidate least upper bound as a located cut. There are two candidates, according to what can be decided.

Lemma 12.43 (The supremum cut)**.** *Let S be an inhabited set of points that is bounded above, and suppose that for every ratio q, either $\eta(q) < s$ for some member s or $s \le \eta(q)$ for every member s. Let $\mathscr{L}$ be the set of ratios q with $\eta(q) < s$ for some member s, and $\mathscr{U}$ the set of ratios r such that some ratio $r' < r$ has $s \le \eta(r')$ for every member s. Then $(\mathscr{L}, \mathscr{U})$ is a located cut, and every point whose cut is $(\mathscr{L}, \mathscr{U})$ is a least upper bound of S approached by its members. No choice principle is used.*

Proof. Let s_0 be a member and b an upper bound. By item 4 of Proposition 12.19 there are natural numbers n_0, n_1 with $-s_0 < \eta(n_0)$ and $b < \eta(n_1)$. The first gives $\eta(-n_0) < s_0$, so $-n_0 \in \mathscr{L}$. Every member satisfies $s \le b < \eta(n_1)$, hence $s < \eta(n_1)$ and $s \le \eta(n_1)$ by Lemma 12.40(a); so $n_1 + 1 \in \mathscr{U}$. If $q < q'$ and $\eta(q') < s$, then $\eta(q) < \eta(q') < s$, so $\mathscr{L}$ is closed downward; if $\eta(q) < s$, Lemma 12.30 gives a ratio q' with $\eta(q) < \eta(q') < s$, so $\mathscr{L}$ has no greatest element. $\mathscr{U}$ is closed upward by its definition, and if $r' < r$ witnesses $r \in \mathscr{U}$, the mean of r' and r is a smaller member of $\mathscr{U}$. If q lay in both sets, some member s and some ratio $r' < q$ would give $\eta(q) < s$ and $s \le \eta(r') < \eta(q)$, hence $\eta(q) < \eta(q)$ by Lemma 12.40(a) and transitivity. Finally let $q < r$. The hypothesis at q gives either a member with $\eta(q) < s$, and then $q \in \mathscr{L}$, or $s \le \eta(q)$ for every member, and then $r \in \mathscr{U}$ with $r' = q$. So the cut is located.

Let L have the cut $(\mathscr{L}, \mathscr{U})$, so that $\eta(q) < L$ exactly when some member exceeds $\eta(q)$. For a member s, every ratio q with $\eta(q) < s$ lies in $\mathscr{L}$, hence $\eta(q) < L$; by Lemma 12.31, $s \le L$. Let b be an upper bound and $\eta(q) < L$. Some member has $\eta(q) < s$, and cotransitivity gives $\eta(q) < b$ or $b < s$; the second denies $s \le b$. So $\eta(q) < b$, and Lemma 12.31 gives $L \le b$. Thus L is a least upper bound, and the description of $\mathscr{L}$ says that it is approached by the members. □

Lemma 12.44 (The least-upper-bound cut)**.** *Let S be an inhabited set of points that is bounded above, and suppose that for every ratio q, either $\eta(q)$ is an upper bound of S or it is not. Let $\mathscr{L}$ be the set of ratios q such that some ratio $q' > q$ has $\eta(q')$ not an upper bound of S, and $\mathscr{U}$ the set of ratios r such that some ratio $r' < r$ has $\eta(r')$ an upper bound of S. Then $(\mathscr{L}, \mathscr{U})$ is a located cut, and every point whose cut is $(\mathscr{L}, \mathscr{U})$ is a least upper bound of S. No choice principle is used.*

Proof. Let s be a member. As in Lemma 12.43, some ratio q has $\eta(q) < s$, and Lemma 12.30 gives q' with $\eta(q) < \eta(q') < s$. Then $\eta(q')$ is not an upper bound, so $q \in \mathscr{L}$. An upper bound

lies below some $\eta(n_1)$, which is then an upper bound, and $n_1 + 1 \in \mathcal{U}$. Both sets are closed in the right direction by their definitions, and neither has an end point: if $q' > q$ witnesses $q \in \mathcal{L}$, the mean of q and q' is a larger member of $\mathcal{L}$, and the mean of a witness $r' < r$ and r is a smaller member of $\mathcal{U}$. If q lay in both sets, there would be ratios $r' < q < q'$ with $\eta(r')$ an upper bound and $\eta(q')$ not; but every member satisfies $s \le \eta(r') \le \eta(q')$, so $\eta(q')$ is an upper bound. For $q < r$ let m be their mean. If $\eta(m)$ is an upper bound, then $r \in \mathcal{U}$; if it is not, then $q \in \mathcal{L}$. So the cut is located.

Let L have the cut $(\mathcal{L}, \mathcal{U})$. For a member s and a ratio q with $\eta(q) < s$, the first paragraph puts q in $\mathcal{L}$, so $\eta(q) < L$; by Lemma 12.31, $s \le L$. Let b be an upper bound and $\eta(q) < L$. Then $q \in \mathcal{L}$, so some $q' > q$ has $\eta(q')$ not an upper bound. Cotransitivity at $\eta(q) < \eta(q')$ gives $\eta(q) < b$ or $b < \eta(q')$, and the second would make $\eta(q')$ an upper bound, by Lemma 12.40(a). So $\eta(q) < b$, and Lemma 12.31 gives $L \le b$. □

12.9.8 Monotone sequences and sequential compactness cost LPO

Theorem 12.45 (Monotone convergence costs exactly LPO)**.** *Monotone convergence is equivalent to LPO, in the following sense.*

1. *If every bounded nondecreasing sequence of points of $\mathbb{R}_\delta$ converges, then LPO holds. The proof uses no choice principle.*

2. *Assume LPO and AC_ω. Then every nondecreasing sequence of points that is bounded above converges; in particular every bounded one does. Countable choice is used twice: to fix regular sequences naming the terms, and once at pairs of ratios, through Theorem 12.33.*

Proof. For the first statement, let α be a binary sequence. Put $t_n = \eta(1)$ if $\alpha(k) = 1$ for some $k \le n$, and $t_n = \eta(0)$ otherwise. Each term is fixed by a bounded search, so the sequence is given outright. It is nondecreasing, and it lies between $\eta(0)$ and $\eta(1)$, so it is bounded. By hypothesis it has a limit L; take N such that t_m is within $1/4$ of L for every $m \ge N$.

Search $\alpha(0), \dots, \alpha(N)$. If an event appears, the second alternative of LPO holds. Otherwise $t_N = \eta(0)$. Fix n. If $\alpha(n) = 1$, then $t_{N+n} = \eta(1)$, and $t_N = \eta(0)$ and $t_{N+n} = \eta(1)$ would both be within $1/4$ of L, which is impossible, as shown after Lemma 12.41. Since $\alpha(n)$ is binary, $\alpha(n) = 0$. This is the first alternative of LPO.

For the second, let (x_n) be nondecreasing and bounded above. Countable choice supplies regular sequences $s^{(n)}$ naming the points x_n. For every ratio q, LPO gives, through Lemma 12.42(a), either an n with $\eta(q) < x_n$ or $x_n \le \eta(q)$ for every n. So Lemma 12.43, applied to the set of terms, gives a located cut $(\mathcal{L}, \mathcal{U})$, and by Theorem 12.33, using countable choice at pairs of ratios, it is the cut of a point L. Thus $\eta(q) < L$ exactly when $q \in \mathcal{L}$, and $L < \eta(r)$ exactly when $r \in \mathcal{U}$.

Fix k. Lemma 12.32 gives $q \in \mathcal{L}$ and $r \in \mathcal{U}$ with $r - q < 1/(k+1)$. Since $q \in \mathcal{L}$, some N has $\eta(q) < x_N$; since $r \in \mathcal{U}$, some $r' < r$ has $x_n \le \eta(r')$ for every n. Let $m \ge N$. Then $x_N \le x_m$, so $\eta(q) < x_m$ by Lemma 12.40(a); and $x_m \le \eta(r') < \eta(r)$ gives $x_m < \eta(r)$. Hence $\eta(q) \le x_m \le \eta(r)$, and likewise $\eta(q) \le L \le \eta(r)$. By Lemma 12.41(a), x_m is within $1/(k+1)$ of L. So (x_n) converges to L. □

The proof of the first statement uses one limit point and one index N; it decides nothing about them beyond a bounded search.

When the terms come with regular sequences, the Cauchy condition costs the same principle, and then no choice is needed at all.

Theorem 12.46 (Monotone sequences are Cauchy exactly under LPO)**.** *The following are equivalent, and neither direction uses a choice principle:*

1. *every bounded nondecreasing sequence of points given by regular sequences is Cauchy;*
2. *LPO.*

Moreover, LPO makes every nondecreasing sequence given by regular sequences and bounded above Cauchy.

Proof. Assume the first statement, and let α be a binary sequence. Let $s^{(n)}$ be the constant sequence of the ratio 1 if $\alpha(k) = 1$ for some $k \le n$, and of the ratio 0 otherwise. Each $s^{(n)}$ is fixed by a bounded search, so the sequence of regular sequences is given outright. Its points are the terms t_n of the proof of Theorem 12.45: nondecreasing, and between $\eta(0)$ and $\eta(1)$. By hypothesis they form a Cauchy sequence; take N such that t_m and t_n are within $1/4$ whenever $m, n \ge N$. Search $\alpha(0), \dots, \alpha(N)$. If an event appears, the second alternative of LPO holds. Otherwise $t_N = \eta(0)$. Fix n. If $\alpha(n) = 1$, then $t_{N+n} = \eta(1)$, and $\eta(1)$ would be within $1/4$ of $\eta(0)$; then $\eta(1) \le \eta(0) + \eta(1/4) = \eta(1/4)$, contrary to $\eta(1/4) < \eta(1)$. Since $\alpha(n)$ is binary, $\alpha(n) = 0$. This is the first alternative of LPO.

Conversely, assume LPO, and let regular sequences $s^{(n)}$ be given whose points x_n are nondecreasing and bounded above. For every ratio q, LPO, through Lemma 12.42(a) applied to the given sequences, yields either an n with $\eta(q) < x_n$ or $x_n \le \eta(q)$ for every n. By Lemma 12.43 the sets $\mathscr{L}$ and $\mathscr{U}$ of that lemma form a located cut. Fix k, and take $q \in \mathscr{L}$ and $r \in \mathscr{U}$ with $r - q < 1/(k+1)$ (Lemma 12.32). Some N has $\eta(q) < x_N$, and some $r' < r$ has $x_n \le \eta(r')$ for every n. For $m \ge N$, monotonicity and Lemma 12.40(a) give $\eta(q) \le x_m \le \eta(r)$. So for $m, n \ge N$ the points x_m and x_n both lie between $\eta(q)$ and $\eta(r)$, and Lemma 12.41(a) puts them within $1/(k+1)$. The sequence is Cauchy.

No step selects anything: the regular sequences are given, each decision is one application of LPO to one binary sequence, the pair q, r is found inside the proof for one k, and the cut is never turned into a point. □

Theorem 12.47 (Sequential compactness costs exactly LPO)**.** *Consider the assertions*

(i) every bounded sequence of points of $\mathbb{R}_\delta$ has a convergent subsequence;

(ii) every bounded sequence of points has a cluster point;

(iii) every sequence of points has a monotone subsequence.

Each of them implies LPO, and the proof uses no choice principle; for (iii) the bounded sequences already suffice. Conversely, LPO and AC_ω imply all three. Countable choice is used in (iii) to fix regular sequences naming the terms and to collect three sequences of decisions and indices, and in (i) and (ii) through (iii) and Theorem 12.45.

Proof. (i) implies (ii). Let $(x_{\sigma(n)})$ converge to L, and let k and N be given. Take M such that $x_{\sigma(n)}$ is within $1/(k+1)$ of L for $n \geq M$. Then $m = \sigma(N+M) \geq N+M \geq N$, and x_m is within $1/(k+1)$ of L. So L is a cluster point. No choice is involved.

(ii) implies LPO. Let α be a binary sequence and (t_n) the test sequence of Theorem 12.45; it is bounded. Let L be a cluster point. With $k = 3$ and $N = 0$ there is an m with t_m within $1/4$ of L. Search $\alpha(0), \ldots, \alpha(m)$. If an event appears, the second alternative of LPO holds. Otherwise $t_m = \eta(0)$. Fix j. If $\alpha(j) = 1$, the cluster property with $k = 3$ and $N = j$ gives an $m' \geq j$ with $t_{m'}$ within $1/4$ of L, and $t_{m'} = \eta(1)$ because $j \leq m'$. Then $\eta(0)$ and $\eta(1)$ would both be within $1/4$ of L, which is excluded. Hence $\alpha(j) = 0$, the first alternative of LPO. With the first paragraph, (i) implies LPO as well.

(iii) implies LPO. Let α be a binary sequence. Put $z_n = \eta(1/(n+1))$ if $\alpha(k) = 1$ for some $k \leq n$, and $z_n = \eta(-1/(n+1))$ otherwise. Each term is fixed by a bounded search, and every term lies between $\eta(-1)$ and $\eta(1)$. Let $(z_{\sigma(n)})$ be a monotone subsequence.

If it is nondecreasing, every entry of α is 0. Indeed, if $\alpha(j) = 1$, an event occurs at or before $\sigma(j)$, since $\sigma(j) \geq j$, and hence at or before $\sigma(j+1)$. Then $z_{\sigma(j+1)} = \eta(1/(\sigma(j+1)+1)) < \eta(1/(\sigma(j)+1)) = z_{\sigma(j)}$, which denies $z_{\sigma(j)} \leq z_{\sigma(j+1)}$. Since each entry is binary, it is 0.

If it is nonincreasing, search $\alpha(0), \ldots, \alpha(\sigma(1))$. Should no event appear, none appears up to $\sigma(0) < \sigma(1)$ either, and $z_{\sigma(0)} = \eta(-1/(\sigma(0)+1)) < \eta(-1/(\sigma(1)+1)) = z_{\sigma(1)}$, which denies $z_{\sigma(1)} \leq z_{\sigma(0)}$. So the search finds an event, the second alternative of LPO.

LPO and AC_ω give (iii). Let (x_n) be a sequence of points, and use countable choice to fix regular sequences $s^{(n)}$ naming them. Call n a *peak* if $x_m \leq x_n$ for every $m > n$. For each n, $x_n < x_m$ holds exactly when some stage j of $s^{(m)} + (-s^{(n)})$, the ratio $s^{(m)}_{2j+1} - s^{(n)}_{2j+1}$, exceeds $2/(j+1)$. Let $\pi_n(N) = 1$ if some $i, j \leq N$ give such a stage for $m = n+i+1$, and $\pi_n(N) = 0$ otherwise. As in Lemma 12.42, π_n is all zero exactly when n is a peak, and has an entry 1 exactly when some $m > n$ has $x_n < x_m$; so LPO decides between the two. Countable choice collects these decisions into a binary sequence p, with $p(n) = 1$ when n is a peak and $p(n) = 0$ when a later term lies strictly above x_n. For each N, LPO applied to $p(N), p(N+1), \ldots$ decides whether a peak lies at or after N; countable choice collects these answers into a binary sequence c, with $c(N) = 0$ when some $n \geq N$ has $p(n) = 1$, and $c(N) = 1$ when $p(n) = 0$ for every $n \geq N$. Apply LPO to c.

If every $c(N)$ is 0, every index has a peak at or after it, and countable choice gives a sequence g with $g(N) \geq N$ and $p(g(N)) = 1$ for every N. Put $\sigma(0) = g(0)$ and $\sigma(i+1) = g(\sigma(i)+1)$. The indices increase, each is a peak, and so $x_{\sigma(i+1)} \leq x_{\sigma(i)}$: the subsequence is nonincreasing.

If $c(N) = 1$ for some N, every $n \geq N$ has a later term strictly above x_n, and countable choice gives a sequence g with $n < g(n)$ and $x_n < x_{g(n)}$ for every $n \geq N$. Put $\sigma(0) = N$ and $\sigma(i+1) = g(\sigma(i))$. The indices increase and stay at least N, and $x_{\sigma(i)} < x_{\sigma(i+1)}$, so $x_{\sigma(i)} \leq x_{\sigma(i+1)}$: the subsequence is nondecreasing.

LPO and AC_ω give (i), and hence (ii). Let (x_n) be bounded, with $b' \leq x_n \leq b$ for every n, and take a monotone subsequence $(y_n) = (x_{\sigma(n)})$ by (iii); it has the same bounds. If it is nondecreasing, Theorem 12.45 gives it a limit. If it is nonincreasing, $(-y_n)$ is nondecreasing and bounded above by $-b'$ (Lemma 12.40(b)), so it converges to some point M, and by Lemma 12.41(d) the subsequence (y_n) converges to $-M$. Then (ii) follows from the first paragraph of this proof. □

12.9.9 Least upper bounds cost WLPO, LPO, WEM or EM

Theorem 12.48 (Least upper bounds of sequences cost exactly WLPO)**.** *Least upper bounds of sequences are equivalent to WLPO, in the following sense.*

1. *If every bounded sequence of points of $\mathbb{R}_\delta$ has a least upper bound, then WLPO holds. The proof uses no choice principle.*
2. *Assume WLPO and AC_ω. Then every sequence of points that is bounded above has a least upper bound. Countable choice is used twice: to fix regular sequences naming the terms, and once at pairs of ratios, through Theorem 12.33.*

In particular "every sequence bounded above has a least upper bound" and "every bounded sequence has a least upper bound" are both equivalent to WLPO under AC_ω, and the passage to WLPO uses no choice.

Proof. For the first statement, let α be a binary sequence, and let $b_n = \eta(\alpha(n))$, reading each bit as the ratio 0 or 1. The sequence is bounded, below by $\eta(0)$ and above by $\eta(1)$; let L be a least upper bound. Since $\eta(0) < \eta(1)$, cotransitivity gives $\eta(0) < L$ or $L < \eta(1)$.

If $L < \eta(1)$, no entry of α is 1: an entry $\alpha(n) = 1$ would give $\eta(1) = b_n \leq L < \eta(1)$, hence $\eta(1) < \eta(1)$ by Lemma 12.40(a). Every entry is therefore 0.

If $\eta(0) < L$, then α is not all zero. Otherwise $\eta(0)$ would be an upper bound, and leastness would give $L \leq \eta(0)$, which denies $\eta(0) < L$.

These are the two alternatives of WLPO.

For the second, let (x_n) be bounded above, and use countable choice to fix regular sequences naming its terms. For every ratio q, WLPO decides, through Lemma 12.42(b), whether $\eta(q)$ is an upper bound of the terms. So Lemma 12.44, applied to the set of terms, gives a located cut, and every point with that cut is a least upper bound. By Theorem 12.33, using countable choice at pairs of ratios, the cut is the cut of a point. Finally, the assertion for sequences bounded above implies the assertion for bounded sequences, so statement 1 gives WLPO from either reading, and statement 2 gives both readings from WLPO. □

The proof of the first statement never decides whether L is $\eta(0)$ or $\eta(1)$; cotransitivity only places L away from one of them. If the least upper bound is also approached by the terms, the case $\eta(0) < L$ yields a term above $\eta(0)$, that is, an event, and the same argument gives LPO. The next theorem proves this and its converse.

Theorem 12.49 (Approached least upper bounds cost exactly LPO)**.** *Least upper bounds approached by the terms are equivalent to LPO, in the following sense.*

1. *If every sequence of points bounded above has a least upper bound approached by its terms, then LPO holds. The proof uses no choice principle.*
2. *Assume LPO and AC_ω. Then every sequence of points bounded above has a least upper bound approached by its terms. Countable choice is used to fix regular sequences naming the terms, and once at pairs of ratios, through Theorem 12.33.*

Proof. For the first statement, let α be a binary sequence and $b_n = \eta(\alpha(n))$, as in the proof of Theorem 12.48; the sequence is bounded above by $\eta(1)$. Let L be a least upper bound approached by the terms. Cotransitivity at $\eta(0) < \eta(1)$ gives $\eta(0) < L$ or $L < \eta(1)$. If $L < \eta(1)$, every entry of α is 0, exactly as in that proof. If $\eta(0) < L$, the ratio 0 is exceeded by some term: $\eta(0) < b_n$ for some n. If $\alpha(n)$ were 0, then $b_n = \eta(0)$ and $\eta(0) < \eta(0)$, which is impossible; so $\alpha(n) = 1$. These are the two alternatives of LPO.

For the second, let (x_n) be bounded above, and use countable choice to fix regular sequences naming its terms. LPO gives, through Lemma 12.42(a) and for every ratio q, an n with $\eta(q) < x_n$ or $x_n \le \eta(q)$ for every n. So the sets $\mathscr{L}$ and $\mathscr{U}$ of Lemma 12.43, for the set of terms, form a located cut. By Theorem 12.33, using countable choice at pairs of ratios, it is the cut of a point L, and by Lemma 12.43, L is a least upper bound approached by the terms. □

For arbitrary sets of points no binary sequence is available, and the decisions must be made about propositions. The prices become the weak and the full excluded middle.

Theorem 12.50 (Least upper bounds of sets cost exactly WEM)**.** *Least upper bounds of sets of points are equivalent to weak excluded middle, in the following sense.*

1. *If every inhabited set of points of $\mathbb{R}_\delta$ that is bounded above has a least upper bound, then $\neg P \vee \neg\neg P$ for every proposition P. The proof uses no choice principle.*

2. *Assume weak excluded middle and AC_ω. Then every inhabited set of points that is bounded above has a least upper bound. Countable choice is used once, at pairs of ratios, through Theorem 12.33.*

Proof. For the first statement, let P be a proposition and let T_P be the set of points s with $s = \eta(0)$, or with P and $s = \eta(1)$. It is inhabited by $\eta(0)$ and bounded above by $\eta(1)$; let L be a least upper bound. Since $\eta(0) < \eta(1/2)$, cotransitivity gives $\eta(0) < L$ or $L < \eta(1/2)$. If $L < \eta(1/2)$, then $\neg P$: from P we would get $\eta(1) \in T_P$, so $\eta(1) \le L < \eta(1/2) < \eta(1)$, and Lemma 12.40(a) with transitivity gives $\eta(1) < \eta(1)$. If $\eta(0) < L$, then $\neg\neg P$: from $\neg P$, the only member of T_P is $\eta(0)$, so $\eta(0)$ is an upper bound and leastness gives $L \le \eta(0)$, which denies $\eta(0) < L$.

For the second, let S be inhabited and bounded above. For a ratio q, the statement that $\eta(q)$ is an upper bound of S says that no member lies above $\eta(q)$. It is a negation, so its double negation implies it. Weak excluded middle therefore decides it: either $\eta(q)$ is an upper bound or it is not. Lemma 12.44 gives a located cut every point of which is a least upper bound of S, and Theorem 12.33, using countable choice at pairs of ratios, gives such a point. No regular sequences are needed: being an upper bound is a statement about the points themselves. □

Theorem 12.51 (Approached least upper bounds of sets cost exactly EM)**.** *Least upper bounds of sets approached by the members are equivalent to excluded middle, in the following sense.*

1. *If every inhabited set of points that is bounded above has a least upper bound approached by its members, then $P \vee \neg P$ for every proposition P. The proof uses no choice principle.*

2. *Assume excluded middle and* AC_ω*. Then every inhabited set of points that is bounded above has a least upper bound approached by its members. Countable choice is used once, at pairs of ratios, through Theorem 12.33.*

Proof. For the first statement, take the set T_P of the proof of Theorem 12.50 and a least upper bound L approached by its members. Cotransitivity at $\eta(0) < \eta(1/2)$ gives $\eta(0) < L$ or $L < \eta(1/2)$. If $L < \eta(1/2)$, then $\neg P$, as there. If $\eta(0) < L$, the ratio 0 is exceeded by some member s of T_P. The member $\eta(0)$ does not exceed $\eta(0)$, so s is a member of the second kind, and P holds.

For the second, let S be inhabited and bounded above. For every ratio q, excluded middle decides whether some member exceeds $\eta(q)$; if none does, every member s satisfies $s \leq \eta(q)$, since $s \leq \eta(q)$ means only that $\eta(q) < s$ fails. So Lemma 12.43 applies to S: its sets form a located cut, and every point with that cut is a least upper bound approached by the members. Theorem 12.33, using countable choice at pairs of ratios, gives such a point. □

Without countable choice the two assertions keep an exact price. Each of them makes every located cut the cut of a point, with no choice. Let $(\mathscr{L}, \mathscr{U})$ be a located cut. Every $q \in \mathscr{L}$ lies below every $r \in \mathscr{U}$, since $r \leq q$ would put r in $\mathscr{L}$. So the set of points $\eta(q)$ with $q \in \mathscr{L}$ is inhabited and bounded above by $\eta(r)$ for each $r \in \mathscr{U}$; let y be a least upper bound. If $q \in \mathscr{L}$, a larger $q' \in \mathscr{L}$ gives $\eta(q) < \eta(q') \leq y$. If $\eta(q) < y$, take a ratio q' with $\eta(q) < \eta(q') < y$; locatedness at $q < q'$ gives $q \in \mathscr{L}$ or $q' \in \mathscr{U}$, and $q' \in \mathscr{U}$ would make $\eta(q')$ an upper bound below y. So the lower set of y is $\mathscr{L}$, and in the same way its upper set is $\mathscr{U}$. Conversely, the proofs of the second statements use countable choice only through Theorem 12.33. Hence, with no choice principle, the assertion of Theorem 12.50 is equivalent to weak excluded middle together with the assertion that every located cut is the cut of a point, and the assertion of Theorem 12.51 is equivalent to excluded middle together with the same assertion. Theorem 12.63 shows that excluded middle, without unique choice, does not supply the second conjunct.

The passages back that produce a least upper bound or a limit share one pattern. The principle decides, for each ratio q, a question about the terms or members: whether some term exceeds $\eta(q)$, for LPO and excluded middle, or whether $\eta(q)$ is an upper bound, for WLPO and weak excluded middle. Those decisions make the candidate a located cut, by Lemma 12.43 or Lemma 12.44, and countable choice, applied once at pairs of ratios, turns the cut into a point; where the terms are given only as points, it also supplies regular sequences naming them. When the terms come with regular sequences and only the Cauchy condition is asked (Theorem 12.46), the cut is never turned into a point and no choice is used. The monotone subsequences of Theorem 12.47 follow a different pattern: the decisions concern peaks, and countable choice collects them into sequences. In the rows priced "none" that turn a cut into a point (the root of a strictly increasing f and the supremum), the cut is located with no omniscience principle and no choice: apart from decidable comparisons of ratios and finite searches, the only comparison used is the free one of cotransitivity. The root of a strictly increasing f, for example, is located by the signs of f at ratios, so only the passage from cut to point remains. The LLPO rows go back differently: they bisect. LLPO decides, at every dyadic ratio, a comparison that tells the bisection which half to keep, and countable choice tabulates these decisions in advance, one for every dyadic ratio, since a choice made along the path

would depend on the choices before it. The kept intervals then follow by recursion. For the maximum, the supremum of each dyadic interval comes from the supremum row, which uses countable choice at pairs of ratios.

12.9.10 Roots, maxima and expansions cost LLPO or nothing

The last seven rows of Table 12.3 concern functions on $[0,1]$ and digit expansions. Their proofs use a few facts about the order and about "within", collected first. None of them needs an omniscience principle or a choice principle.

Lemma 12.52 (Working with "within"). *Let x, y, z, u, v, w, L be points of $\mathbb{R}_\delta$ and k, m natural numbers.*

1. *If $x < z$, there are ratios $b < c$ with $x < \eta(b) < \eta(c) < z$.*
2. *If for every k some point $y \le z$ is within $1/(k+1)$ of L, then $L \le z$. If for every k some point $y \ge z$ is within $1/(k+1)$ of L, then $z \le L$.*
3. *If $x < z$, there is a K such that every point within $1/(K+1)$ of x lies below z; if $z < x$, there is a K such that every point within $1/(K+1)$ of x lies above z.*
4. *If a regular sequence s names x, then $\eta(s_n)$ and x are within $1/(n+1)$ for every n. For every x and m there are ratios $a < b$ with $\eta(a) < x < \eta(b)$ such that any two points between $\eta(a)$ and $\eta(b)$ are within $1/(m+1)$.*
5. *If x, x' are within $1/(k+1)$ and y, y' are within $1/(k+1)$, then so are $x+z$ and $x'+z$, $-x$ and $-x'$, $\max(x,y)$ and $\max(x',y')$, and $\min(x,y)$ and $\min(x',y')$. If x, y and y, z are within $1/(2k+2)$, then x, z are within $1/(k+1)$.*
6. *If $u < w$ and $v < w$, then $\max(u,v) < w$; if $w < u$ and $w < v$, then $w < \min(u,v)$. If $w < \max(u,v)$, then $w < u$ or $w < v$; if $\min(u,v) < w$, then $u < w$ or $v < w$.*

Proof. Two consequences of cotransitivity are used throughout: $u \le v < w$ gives $u < w$, and $u < v \le w$ gives $u < w$. In the first case cotransitivity applied to $v < w$ gives $v < u$ or $u < w$, and $v < u$ is excluded by $u \le v$; the second is the mirror image.

Item 1 is Lemma 12.30, used twice: first for a ratio b with $x < \eta(b) < z$, then for a ratio c with $\eta(b) < \eta(c) < z$; and $b < c$ because η reflects the order.

For the first assertion of item 2, suppose $z < L$. By item 1 there are ratios $r < q$ with $z < \eta(r) < \eta(q) < L$. Choose k with $1/(k+1) \le q - r$ and a point $y \le z$ within $1/(k+1)$ of L. Then

$$L \le y + \eta\big(1/(k+1)\big) \le z + \eta\big(1/(k+1)\big) < \eta\big(r + 1/(k+1)\big) \le \eta(q) < L,$$

which gives $L < L$, impossible. Hence $L \le z$. The second assertion is the mirror image.

For item 3, take ratios with $x < \eta(b) < \eta(c) < z$ by item 1, and K with $1/(K+1) \le c - b$. A point y within $1/(K+1)$ of x has $y \le x + \eta(1/(K+1)) < \eta(b + 1/(K+1)) \le \eta(c) < z$. The second assertion is the mirror image.

For item 4, regularity gives $|s_n - s_p| \le 1/(n+1) + 1/(p+1) \le 1/(n+1) + 2/(p+1)$ for every p; so, by Lemma 12.23 applied to the constant sequence of s_n and to s, $\eta(s_n)$ and x are within $1/(n+1)$. For the second assertion, which is an existence statement about one point, we may use one regular sequence s naming x. Take n with $n+1 \ge 4(m+1)$ and put $a = s_n - 2/(n+1)$ and $b = s_n + 2/(n+1)$. By the first assertion, $\eta(a) < \eta(s_n - 1/(n+1)) \le x \le \eta(s_n + 1/(n+1)) < \eta(b)$. Points u, v between $\eta(a)$ and $\eta(b)$ satisfy $u \le \eta(b) = \eta(a) + \eta(4/(n+1)) \le v + \eta(1/(m+1))$, and symmetrically.

Item 5 is read off the order form; write $e = 1/(k+1)$. From $x \le x' + \eta(e)$ and $y \le y' + \eta(e)$, both x and y lie below $\max(x', y') + \eta(e)$, so $\max(x, y)$ does; $\min(x, y)$ lies below $x' + \eta(e)$ and below $y' + \eta(e)$, hence below $\min(x' + \eta(e), y' + \eta(e))$, which is $\min(x', y') + \eta(e)$ because sums and minima are both computed stagewise. Translation and negation are immediate, and for the last assertion add the two order statements.

For item 6, let s, t, r name u, v, w; the maximum is named by the stagewise maximum. If $w < \max(u, v)$, a positive stage of $\max(u, v) - w$ is $\max(s_{2n+1}, t_{2n+1}) - r_{2n+1}$ for some n, which is the stage n of $u - w$ or of $v - w$; so $w < u$ or $w < v$. If $u < w$ and $v < w$, then by Lemma 12.16 the stages of $w - u$ and of $w - v$ exceed a fixed positive ratio δ from some index on. The stage n of $w - \max(u, v)$ is the smaller of the stages n of $w - u$ and $w - v$, so it exceeds $2/(n+1)$ once also $2/(n+1) < \delta$, and $\max(u, v) < w$. The two laws for the minimum are these applied to $-u$, $-v$ and $-w$, since $\min(u, v) = -\max(-u, -v)$. □

The test functions below move a given point into a fixed range without changing its sign.

Lemma 12.53 (A clamp keeps the sign)**.** *For a point a put $c(a) = \max\big(\eta(-\frac{1}{3}), \min(a, \eta(\frac{1}{3}))\big)$. Then $\eta(-\frac{1}{3}) \le c(a) \le \eta(\frac{1}{3})$; if $a > \eta(0)$ then $c(a) > \eta(0)$; and if $a < \eta(0)$ then $c(a) < \eta(0)$.*

Proof. The maximum lies above its first argument, and both arguments lie below $\eta(\frac{1}{3})$, so the maximum does. If $a > \eta(0)$, then also $\eta(\frac{1}{3}) > \eta(0)$, so $\min(a, \eta(\frac{1}{3})) > \eta(0)$ by Lemma 12.52, item 6, and $c(a)$, which lies above this minimum, is $> \eta(0)$. If $a < \eta(0)$, then $\min(a, \eta(\frac{1}{3})) \le a < \eta(0)$ and $\eta(-\frac{1}{3}) < \eta(0)$, so $c(a) < \eta(0)$ by item 6. □

Theorem 12.54 (The intermediate value theorem costs LLPO)**.** *Consider the assertion that every uniformly continuous f on $[0, 1]$ with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$ has a root in $[0, 1]$.*

1. *The assertion implies LLPO. The proof uses no choice principle.*
2. *Assuming AC_ω, LLPO implies the assertion. Countable choice is used once, to tabulate one bit for every dyadic ratio.*

Proof. Write $\frac{1}{3}$ and $\frac{2}{3}$ for $\eta(\frac{1}{3})$ and $\eta(\frac{2}{3})$, and let

$$g(x) = \max\big(\min(x - \tfrac{1}{3}, \eta(0)),\ x - \tfrac{2}{3}\big).$$

Then $g(\eta(0)) = -\frac{1}{3}$ and $g(\eta(1)) = \frac{1}{3}$. If $x \le \frac{2}{3}$, both arguments of the maximum are at most $\eta(0)$, so $g(x) \le \eta(0)$; if $x \ge \frac{1}{3}$, the first argument is $\eta(0)$, so $g(x) \ge \eta(0)$. By Lemma 12.52, item 5, points within $1/(k+1)$ have values of g within $1/(k+1)$.

For the first part, let a be a point, let $c = c(a)$ be its clamp (Lemma 12.53), and put $f(x) = g(x)+c$. Then $f(\eta(0)) = c-\frac{1}{3} \le \eta(0)$ and $f(\eta(1)) = \frac{1}{3}+c \ge \eta(0)$, and f is uniformly continuous on $[0,1]$: for each k take $j = k$. By the assertion f has a root x, so $c = -g(x)$. Since $\frac{1}{3} < \frac{2}{3}$, cotransitivity gives $\frac{1}{3} < x$ or $x < \frac{2}{3}$. In the first case $g(x) \ge \eta(0)$, so $c \le \eta(0)$; then $a > \eta(0)$ is impossible, because it would give $c > \eta(0)$, and so $a \le \eta(0)$. In the second case $g(x) \le \eta(0)$, so $c \ge \eta(0)$, and in the same way $a \ge \eta(0)$. Every point therefore satisfies $a \ge \eta(0)$ or $a \le \eta(0)$, and this is LLPO by Corollary 12.18. The function f is given outright, and the root was read with one use of cotransitivity, so no choice principle enters.

For the second part, assume LLPO and AC_ω, and let f satisfy the hypotheses. By Corollary 12.18 every point is $\ge \eta(0)$ or $\le \eta(0)$. So for every pair (n, j) of natural numbers there is a bit b such that $b = 1$ gives $f(\eta(j/2^n)) \le \eta(0)$ and $b = 0$ gives $f(\eta(j/2^n)) \ge \eta(0)$. Countable choice, applied to the pairs (n, j) listed as natural numbers, supplies one such bit $B(n, j)$ for every pair. Define natural numbers a_n by recursion, with no further choice: $a_0 = 0$, and $a_{n+1} = 2a_n + 1$ if $B(n+1, 2a_n+1) = 1$, otherwise $a_{n+1} = 2a_n$. The ratio $(2a_n+1)/2^{n+1}$ is the midpoint of $a_n/2^n$ and $(a_n+1)/2^n$, and induction on n gives

$$a_n + 1 \le 2^n, \qquad f\big(\eta(a_n/2^n)\big) \le \eta(0) \le f\big(\eta((a_n+1)/2^n)\big) \ :$$

if $B(n+1, 2a_n+1) = 1$, the value at the midpoint is at most $\eta(0)$ and the midpoint becomes the new left end, and otherwise it becomes the new right end. Put $l_n = a_n/2^n$ and $r_n = (a_n+1)/2^n$. The intervals are nested, so $l_m \le l_n < r_n \le r_m$ for $m \le n$, and $r_n - l_n = 2^{-n}$. Hence $|l_m - l_n| \le 2^{-\min(m,n)} \le 1/(\min(m,n)+1) \le 1/(m+1) + 1/(n+1)$: the ratios l_n form a regular sequence. Let L be its point. By Lemma 12.52, item 4, $\eta(l_n)$ is within $1/(n+1)$ of L. The points $\eta(r_{2n+1})$ and $\eta(l_{2n+1})$ are within $2^{-(2n+1)} \le 1/(2n+2)$, and $\eta(l_{2n+1})$ is within $1/(2n+2)$ of L, so $\eta(r_{2n+1})$ is within $1/(n+1)$ of L (item 5).

Every $\eta(l_n)$ lies in $[0,1]$, so item 2 puts L in $[0,1]$. Fix k and take j as in uniform continuity at k. The points $\eta(l_j)$ and L of $[0,1]$ are within $1/(j+1)$, so $f(\eta(l_j))$, which is at most $\eta(0)$, is within $1/(k+1)$ of $f(L)$. Since k was arbitrary, item 2 gives $f(L) \le \eta(0)$. In the same way $f(\eta(r_{2j+1})) \ge \eta(0)$ is within $1/(k+1)$ of $f(L)$, and so $f(L) \ge \eta(0)$. Hence $f(L) = \eta(0)$, by Proposition 12.19, item 2. □

Theorem 12.55 (An approximate root costs nothing)**.** *Let f be uniformly continuous on $[0,1]$ with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$. For every k some point x of $[0,1]$ has $f(x)$ within $1/(k+1)$ of $\eta(0)$. No omniscience principle and no choice principle is used.*

Proof. Fix k, and take j as in uniform continuity at the tolerance $1/(2k+2)$. Let $x_i = \eta(i/(j+1))$ for $0 \le i \le j+1$; neighbouring points x_{i-1} and x_i are within $1/(j+1)$. Since $-1/(k+1) < -1/(2k+2)$, cotransitivity gives, for each i,

$$(\mathrm{A}_i) \quad f(x_i) > \eta\big(-1/(k+1)\big) \qquad \text{or} \qquad (\mathrm{B}_i) \quad f(x_i) < \eta\big(-1/(2k+2)\big).$$

Going through $i = 0, 1, \dots, j+1$ in turn, induction on the number of steps gives either an $i \le j+1$ with (A_i) such that every earlier index satisfies (B), or (B_i) for every $i \le j+1$. The second alternative is impossible: (B_{j+1}) says $f(\eta(1)) < \eta(-1/(2k+2)) < \eta(0)$, against $\eta(0) \le f(\eta(1))$. So take the first i with (A_i), and put $x = x_i$. Then $f(x) \ge \eta(-1/(k+1))$. Moreover $f(x) \le \eta(0)$: for $i = 0$ this is the hypothesis $f(\eta(0)) \le \eta(0)$; for $i > 0$ the earlier index satisfies (B_{i-1}), and

$f(x_i)$, being within $1/(2k+2)$ of $f(x_{i-1})$, satisfies $f(x_i) \le f(x_{i-1}) + \eta(1/(2k+2)) < \eta(0)$. A point between $\eta(-1/(k+1))$ and $\eta(0)$ is within $1/(k+1)$ of $\eta(0)$. The number j was used for this one k, inside the proof, and every search was finite. □

Theorem 12.56 (The root of a strictly increasing function)**.** *Let f be uniformly continuous and strictly increasing on $[0,1]$, with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$. Let $\mathscr{L}$ be the set of ratios q such that $\eta(q) < \eta(0)$, or $\eta(q)$ lies in $[0,1]$ and $f(\eta(q)) < \eta(0)$; let $\mathscr{U}$ be the set of ratios r such that $\eta(1) < \eta(r)$, or $\eta(r)$ lies in $[0,1]$ and $f(\eta(r)) > \eta(0)$.*

1. *$(\mathscr{L}, \mathscr{U})$ is a located cut.*
2. *A point whose cut is $(\mathscr{L}, \mathscr{U})$ is a root of f in $[0,1]$, and f has at most one root in $[0,1]$.*
3. *Assuming countable choice at pairs of ratios, f has a root in $[0,1]$.*

The first two parts use no omniscience principle and no choice principle; the third uses countable choice once, to make the cut a point. So the assertion that every such f has a root in $[0,1]$ needs no omniscience principle, and it is proved from countable choice.

Proof. Comparisons between ratios are decided in integer arithmetic; besides cotransitivity they are the only case distinctions below.

The cut. The ratio -1 lies in $\mathscr{L}$ and 2 lies in $\mathscr{U}$. No ratio lies in both: each of the four combinations of clauses contradicts $\eta(0) < \eta(1)$, the definition of $[0,1]$, or $f(\eta(q)) < \eta(0) < f(\eta(q))$. Let $q < q'$ with $q' \in \mathscr{L}$. If $\eta(q') < \eta(0)$, then $\eta(q) < \eta(0)$. Otherwise $\eta(q')$ lies in $[0,1]$ with $f(\eta(q')) < \eta(0)$; then either $q < 0$, or $\eta(q)$ lies in $[0,1]$ and $f(\eta(q)) < f(\eta(q')) < \eta(0)$. In each case $q \in \mathscr{L}$, and the upper set is closed upward in the same way. To locate the cut, let $q < r$. If $q < 0$, then $q \in \mathscr{L}$; if $1 < r$, then $r \in \mathscr{U}$. Otherwise $\eta(q)$ and $\eta(r)$ lie in $[0,1]$, so $f(\eta(q)) < f(\eta(r))$, and cotransitivity at $\eta(0)$ gives $f(\eta(q)) < \eta(0)$, that is $q \in \mathscr{L}$, or $\eta(0) < f(\eta(r))$, that is $r \in \mathscr{U}$.

The lower set has no greatest member. If $\eta(q) < \eta(0)$, the ratio $q/2$ is a larger member. If $\eta(q)$ lies in $[0,1]$ and $f(\eta(q)) < \eta(0)$, take K as in Lemma 12.52, item 3, so that every point within $1/(K+1)$ of $f(\eta(q))$ is below $\eta(0)$; take j as in uniform continuity at K, and put $b = q + 1/(j+1)$. If $1 < b$, then $\eta(1)$ lies between $\eta(q)$ and $\eta(b)$, so it is within $1/(j+1)$ of $\eta(q)$, and $f(\eta(1)) < \eta(0)$, against the hypothesis. So $b \le 1$, $\eta(b)$ lies in $[0,1]$ within $1/(j+1)$ of $\eta(q)$, and $f(\eta(b)) < \eta(0)$: the ratio $b > q$ lies in $\mathscr{L}$. The mirror argument, with $f(\eta(0)) \le \eta(0)$, shows that $\mathscr{U}$ has no least member.

The root. Let x have the cut $(\mathscr{L}, \mathscr{U})$. If $x < \eta(0)$, a ratio r with $x < \eta(r) < \eta(0)$ would lie in $\mathscr{U}$, which contains no ratio below 0; so $\eta(0) \le x$, and in the same way $x \le \eta(1)$. Suppose $f(x) > \eta(0)$. Take K such that every point within $1/(K+1)$ of $f(x)$ is above $\eta(0)$, take j as in uniform continuity at K, and take ratios $a < b$ with $\eta(a) < x < \eta(b)$ such that any two points between $\eta(a)$ and $\eta(b)$ are within $1/(j+1)$ (Lemma 12.52, items 3 and 4). Then $a \in \mathscr{L}$. If $\eta(a) < \eta(0)$, then $\eta(0)$ lies between $\eta(a)$ and $\eta(b)$, so $f(\eta(0)) > \eta(0)$, against the hypothesis; otherwise $\eta(a)$ lies in $[0,1]$ and $f(\eta(a)) < \eta(0)$, against $f(\eta(a)) > \eta(0)$. So $f(x) \le \eta(0)$, and the mirror argument with $b \in \mathscr{U}$ gives $f(x) \ge \eta(0)$. Hence $f(x) = \eta(0)$. If x and y are roots in $[0,1]$ and $x < y$, then $\eta(0) = f(x) < f(y) = \eta(0)$; so neither lies below the other, and they are equal.

The point. With countable choice at pairs of ratios every located cut is the cut of a point (Theorem 12.33). Applied to $(\mathscr{L}, \mathscr{U})$, this gives a point, which by the second part is a root in $[0,1]$. □

A supremum is reached through an approximate maximum, which costs nothing.

Lemma 12.57 (An approximate maximum costs nothing)**.** *Let f be uniformly continuous on $[0,1]$. For every k some point x of $[0,1]$ satisfies $f(y) \le f(x) + \eta(1/(k+1))$ for every y in $[0,1]$. No omniscience principle and no choice principle is used.*

Proof. Fix k, take j as in uniform continuity at the tolerance $1/(2k+2)$, put $N = j+1$, and let $g_i = \eta(i/2^N)$ for $0 \le i \le 2^N$. Let M be the maximum of the finitely many values $f(g_0), \dots, f(g_{2^N})$. Since $M - \eta(1/(2k+2)) < M$, Lemma 12.52, item 6, applied along the finite maximum, gives an i with $M - \eta(1/(2k+2)) < f(g_i)$; so $M \le f(g_i) + \eta(1/(2k+2))$. Put $x = g_i$.

Let y be a point of $[0,1]$. Go through $i = 0, \dots, 2^N - 1$, using cotransitivity at $g_i < g_{i+1}$: $g_i < y$ or $y < g_{i+1}$. Either a first index has $y < g_{i+1}$, or $g_i < y$ for every $i < 2^N$. In every case there is a p with $p+1 \le 2^N$ such that y lies between g_p and $\eta((p+2)/2^N)$. These bounds are $2/2^N \le 1/(j+1)$ apart, so y and the grid point g_{p+1} are within $1/(j+1)$, and

$$f(y) \le f(g_{p+1}) + \eta\big(1/(2k+2)\big) \le M + \eta\big(1/(2k+2)\big) \le f(x) + \eta\big(1/(k+1)\big).$$

The searches are finite, and the one j is used inside the proof. □

Theorem 12.58 (A supremum needs only the passage from cut to point)**.** *Let f be uniformly continuous on $[0,1]$. Let $\mathscr{L}$ be the set of ratios q with $\eta(q) < f(y)$ for some y in $[0,1]$, and $\mathscr{U}$ the set of ratios r for which some ratio $r' < r$ has $f(y) \le \eta(r')$ for every y in $[0,1]$.*

1. *$(\mathscr{L}, \mathscr{U})$ is a located cut, and every point with this cut is a supremum of f. No omniscience principle and no choice principle is used.*
2. *Assuming countable choice at pairs of ratios, f has a supremum; the choice is used once, to make the cut a point.*

So the assertion that every uniformly continuous f on $[0,1]$ has a supremum needs no omniscience principle, and it is proved from countable choice.

Proof. A ratio below $f(\eta(0))$ lies in $\mathscr{L}$. By Lemma 12.57 at $k = 0$ some x has $f(y) \le f(x) + \eta(1)$ for all y in $[0,1]$; a ratio r' above $f(x) + \eta(1)$ (Proposition 12.19, item 4) gives $r' + 1 \in \mathscr{U}$. If $\eta(q) < f(y)$, every smaller ratio lies in $\mathscr{L}$, and a ratio q' with $\eta(q) < \eta(q') < f(y)$ is a larger member (Lemma 12.52, item 1). If $r' < r$ witnesses $r \in \mathscr{U}$, every larger ratio lies in $\mathscr{U}$ with the same witness, and every ratio strictly between r' and r is a smaller member. No ratio lies in both sets: $\eta(q) < f(y) \le \eta(r') < \eta(q)$ is impossible.

To locate the cut, let $q < r$, and choose k with $5/(k+1) < r - q$. Take x as in Lemma 12.57 at k. Since $q < q + 2/(k+1)$, cotransitivity gives $\eta(q) < f(x)$, so that $q \in \mathscr{L}$, or $f(x) < \eta(q + 2/(k+1))$. In the second case every y in $[0,1]$ has

$$f(y) \le f(x) + \eta\big(1/(k+1)\big) < \eta\big(q + 3/(k+1)\big) \le \eta\big(r - 2/(k+1)\big),$$

so $r' = r - 2/(k+1) < r$ witnesses $r \in \mathcal{U}$. The one k is used inside the proof of the disjunction.

Let s have the cut $(\mathcal{L}, \mathcal{U})$. If $f(y) > s$ for some y in $[0,1]$, a ratio q with $s < \eta(q) < f(y)$ would lie in $\mathcal{L}$ and in the upper set of the cut of s, which is $\mathcal{U}$; so s is an upper bound of the values. Given k, take x as in Lemma 12.57. Then $f(x) \le s$, and $s \le f(x) + \eta(1/(k+1))$: otherwise a ratio q with $f(x) + \eta(1/(k+1)) < \eta(q) < s$ lies in the lower set of the cut of s, which is $\mathcal{L}$, so $\eta(q) < f(y)$ for some y in $[0,1]$, while $f(y) \le f(x) + \eta(1/(k+1)) < \eta(q)$. So $f(x)$ is within $1/(k+1)$ of s, and s is a supremum. With countable choice at pairs of ratios every located cut is the cut of a point (Theorem 12.33), which gives the second part. □

The attained maximum is found by bisection on the suprema of dyadic intervals. For $j+1 \le 2^n$ write $I_{n,j}$ for the set of points between $\eta(j/2^n)$ and $\eta((j+1)/2^n)$; a *supremum of f on $I_{n,j}$* is an upper bound of the values of f on $I_{n,j}$ such that, for every k, some value at a point of $I_{n,j}$ lies within $1/(k+1)$ of it. The interval $I_{0,0}$ is $[0,1]$.

Lemma 12.59 (Suprema on dyadic intervals)**.** *Let f be uniformly continuous on $[0,1]$, let $j+1 \le 2^n$, and put*

$$\zeta(y) = \max\bigl(\eta(j/2^n),\ \min(\eta((j+1)/2^n),\ y)\bigr).$$

Then $f \circ \zeta$ is uniformly continuous on $[0,1]$, and a supremum of $f \circ \zeta$ is a supremum of f on $I_{n,j}$. Two suprema of f on $I_{n,j}$ are equal. Consequently, assuming countable choice at pairs of ratios, f has exactly one supremum on every $I_{n,j}$. No omniscience principle is used, and only the last sentence uses a choice principle.

Proof. Both arguments of the maximum lie below $\eta((j+1)/2^n)$, so $\zeta(y)$ lies in $I_{n,j}$ for every y; and if y lies in $I_{n,j}$, then $\min(\eta((j+1)/2^n), y) = y$ and $\zeta(y) = y$, by the lattice laws of Proposition 12.19, item 5. By Lemma 12.52, item 5, points within $1/(k+1)$ have images under ζ within $1/(k+1)$. Since $0 \le j/2^n$ and $(j+1)/2^n \le 1$, the interval $I_{n,j}$ lies in $[0,1]$, so $f \circ \zeta$ is uniformly continuous on $[0,1]$, with the same modulus as f: for each k, the number that uniform continuity of f supplies at k serves for $f \circ \zeta$. The values of $f \circ \zeta$ on $[0,1]$ are values of f on $I_{n,j}$, and every value $f(y)$ with y in $I_{n,j}$ is the value of $f \circ \zeta$ at the point y of $[0,1]$. So the upper bounds of the two sets of values are the same, and so are the values that approach a given point; a supremum of $f \circ \zeta$ is therefore a supremum of f on $I_{n,j}$.

If s and s' are suprema on $I_{n,j}$, then for every k some value, at most s', lies within $1/(k+1)$ of s; so $s \le s'$ by Lemma 12.52, item 2, and symmetrically $s' \le s$. The last sentence is Theorem 12.58 applied to $f \circ \zeta$. □

Theorem 12.60 (An attained maximum costs LLPO)**.** *Consider the assertion that every uniformly continuous f on $[0,1]$ attains a maximum.*

1. *The assertion implies LLPO. The proof uses no choice principle.*
2. *Assuming AC_ω, LLPO implies the assertion. Countable choice is used to tabulate one bit for every pair of adjacent dyadic intervals, and, through Lemma 12.59, at pairs of ratios for the suprema on dyadic intervals.*

Proof. For the first part, let a be a point with clamp $c = c(a)$ (Lemma 12.53), and let (with $\frac{1}{3}$ and $\frac{2}{3}$ again standing for $\eta(\frac{1}{3})$ and $\eta(\frac{2}{3})$)

$$h(x) = \max\big(\min(c,\ x - \tfrac{2}{3}),\ \min(-c,\ \tfrac{1}{3} - x)\big).$$

By Lemma 12.52, item 5, points within $1/(k+1)$ have values within $1/(k+1)$, so h is uniformly continuous on $[0,1]$. At $\eta(1)$ the first argument is $\min(c, \frac{1}{3}) = c$, so $h(\eta(1)) \geq c$; at $\eta(0)$ the second is $\min(-c, \frac{1}{3}) = -c$, so $h(\eta(0)) \geq -c$. If $c > \eta(0)$ and $x \leq \frac{2}{3}$, both arguments are at most $\eta(0)$, so $h(x) \leq \eta(0)$; the same holds if $c < \eta(0)$ and $x \geq \frac{1}{3}$. Let x be a maximum point of h. Cotransitivity gives $\frac{1}{3} < x$ or $x < \frac{2}{3}$. In the first case $a < \eta(0)$ is impossible: it would give $c < \eta(0)$ and then $h(\eta(0)) \geq -c > \eta(0) \geq h(x)$, against the maximum. So $a \geq \eta(0)$. In the second case $a > \eta(0)$ would give $h(\eta(1)) \geq c > \eta(0) \geq h(x)$, so $a \leq \eta(0)$. By Corollary 12.18 this is LLPO, and no choice entered.

For the second part, assume LLPO and AC_ω, and let f be uniformly continuous on $[0,1]$. Countable choice at pairs of ratios is a case of AC_ω, so by Lemma 12.59 f has exactly one supremum on each $I_{n,j}$. By Corollary 12.18, LLPO gives $v - u \geq \eta(0)$ or $v - u \leq \eta(0)$ for any two points u and v, with no choice principle; adding u (Lemma 12.40(b)), this is $u \leq v$ or $v \leq u$. So for every N and every q with $1 \leq q < 2^N$ there is a bit b with $s^- \leq s^+$ if $b = 1$ and $s^+ \leq s^-$ if $b = 0$, where s^- and s^+ are the suprema on $I_{N,q-1}$ and $I_{N,q}$; by uniqueness the bit does not depend on how the suprema are found, and at the other pairs any bit will do. Countable choice, applied to the pairs (N, q) listed as natural numbers, supplies such a bit $B(N, q)$ for all of them at once. A choice made only along the path of the bisection would depend on the earlier choices; listing every dyadic point in advance avoids that.

Define $a_0 = 0$, and $a_{n+1} = 2a_n + 1$ if $B(n+1, 2a_n + 1) = 1$, otherwise $a_{n+1} = 2a_n$. Let s be the supremum of f on $[0,1] = I_{0,0}$. We show by induction that s is the supremum on I_{n,a_n}. The two halves of I_{n,a_n} are $I^- = I_{n+1,2a_n}$ and $I^+ = I_{n+1,2a_n+1}$, meeting at $m = \eta((2a_n+1)/2^{n+1})$, with suprema s^- and s^+. Suppose $B(n+1, 2a_n+1) = 1$, so $s^- \leq s^+$. Every point y of I_{n,a_n} has $y \leq m$ or $m \leq y$, by LLPO as above, so it lies in I^- or in I^+, and $f(y) \leq s^+$ in both cases; values on I^+ approach s^+. So s^+ is a supremum on I_{n,a_n}, hence equal to s, and s is the supremum on $I^+ = I_{n+1,a_{n+1}}$. The case $B(n+1, 2a_n+1) = 0$ is symmetric.

As in the proof of Theorem 12.54, the ratios $a_n/2^n$ form a regular sequence; let L be its point, which lies in $[0,1]$. Fix k, take J as in uniform continuity at the tolerance $1/(2k+2)$, and let $n = 2J+2$. Some y in I_{n,a_n} has $f(y)$ within $1/(2k+2)$ of s. The width of I_{n,a_n} is $2^{-n} \leq 1/(2J+2)$, so y and $\eta(a_n/2^n)$ are within $1/(2J+2)$, and $\eta(a_n/2^n)$ is within $1/(n+1) \leq 1/(2J+2)$ of L; so y and L are within $1/(J+1)$, and $f(y)$ and $f(L)$ are within $1/(2k+2)$. Hence $f(L)$ is within $1/(k+1)$ of s; the number J was used for this one k, inside the proof. As k was arbitrary, $f(L) = s$ by Corollary 12.24, and $f(y) \leq s = f(L)$ for every y in $[0,1]$. Countable choice was used for the table of bits and, through Lemma 12.59, for the suprema of the dyadic intervals. □

Theorem 12.61 (Binary expansions cost LLPO)**.** *If every point of* $[0,1]$ *has a binary expansion, then LLPO holds, and the proof uses no choice principle. Conversely, assuming* AC_ω*, LLPO gives every point of* $[0,1]$ *a binary expansion; countable choice is used once, to tabulate one bit for every dyadic ratio.*

Proof. Suppose every point of $[0,1]$ has a binary expansion. By Corollary 12.18 it suffices to show that every point x satisfies $x \geq \eta(0)$ or $x \leq \eta(0)$; the passage from this dichotomy to LLPO uses no choice principle. Fix x, let $a = c(x)$ be its clamp (Lemma 12.53), and put $p = a + \eta(\frac{1}{2})$. Then $\eta(\frac{1}{6}) \leq p \leq \eta(\frac{5}{6})$, so p lies in $[0,1]$. Let (e_i) be a binary expansion of p. At $n = 1$ its bounds read $\eta(e_0/2) \leq p \leq \eta(e_0/2 + 1/2)$. If $e_0 = 0$, then $p \leq \eta(\frac{1}{2})$; subtracting $\eta(\frac{1}{2})$ gives $a \leq \eta(0)$, so $x > \eta(0)$ is impossible, and $x \leq \eta(0)$. If $e_0 = 1$, then $\eta(\frac{1}{2}) \leq p$, so $\eta(0) \leq a$, $x < \eta(0)$ is impossible, and $x \geq \eta(0)$. Only the first digit of one expansion of one point was used, inside the proof of a disjunction, so no choice principle enters.

Conversely, assume AC_ω and LLPO, and let x lie in $[0,1]$. By Corollary 12.18, applied to $x - \eta(j/2^n)$, and Lemma 12.40(b), LLPO gives $\eta(j/2^n) \leq x$ or $x \leq \eta(j/2^n)$ for every n and j, with no choice principle. So for every n and j there is a bit b such that $b = 1$ gives $\eta(j/2^n) \leq x$ and $b = 0$ gives $x \leq \eta(j/2^n)$. Countable choice, applied to the pairs (n, j) listed as natural numbers, supplies a table $B(n, j)$ of such bits, one for every dyadic ratio $j/2^n$. From here on nothing is selected. Put $a_0 = 0$, and let

$$e_n = B(n+1, 2a_n + 1), \qquad a_{n+1} = 2a_n + e_n,$$

so that e_n is the decision at the midpoint $(2a_n + 1)/2^{n+1}$ of $a_n/2^n$ and $(a_n + 1)/2^n$. By induction $a_n = \sum_{i<n} e_i 2^{n-1-i}$, that is, $a_n/2^n = c_n$. We show by induction that $\eta(a_n/2^n) \leq x \leq \eta((a_n+1)/2^n)$. At $n = 0$ this says that x lies in $[0,1]$. If $e_n = 1$, the midpoint lies at or below x; it is the new lower end $a_{n+1}/2^{n+1}$, and the new upper end $(2a_n + 2)/2^{n+1} = (a_n + 1)/2^n$ is the old one. If $e_n = 0$, then x lies at or below the midpoint, which is the new upper end $(2a_n + 1)/2^{n+1}$, and the new lower end $2a_n/2^{n+1}$ is the old one. So (e_n) is a binary expansion of x. □

Countable choice is used here to tabulate, in advance, a decision at every dyadic ratio; the bisection then reads its steps from the table. Choosing the decisions one step at a time along the bisection, where each midpoint depends on the decisions before it, would be a dependent choice instead. Allowing a digit -1 removes the decision altogether.

Proposition 12.62 (Signed digits cost nothing)**.** *Every point of* $[-1,1]$ *has a signed-digit expansion. The proof uses no omniscience principle and no choice principle: each digit is read off one stage of one regular sequence naming the point, together with the digits before it.*

Proof. Let $\eta(-1) \leq x \leq \eta(1)$. The assertion is an existence statement about the single point x, so we may use one regular sequence s naming it; nothing is selected. For each n let $r_n = s_{2m+1}$ with $m = 2^{n+3} - 1$. By regularity, $|s_i - r_n| \leq 1/(i+1) + 1/(2m+2) \leq 2^{-(n+3)} + 2/(i+1)$ for every i, so, by Lemma 12.23 with $1/(k+1) = 2^{-(n+3)}$, x and $\eta(r_n)$ are within $2^{-(n+3)}$. Write $h_n = 2^{-n}$, so that $2^{-(n+3)} = h_n/8$ and

$$\eta(r_n - h_n/8) \leq x \leq \eta(r_n + h_n/8).$$

Define integers t_n and digits d_n by $t_0 = 0$, $t_{n+1} = 2t_n + d_n$, where

$$d_n = \begin{cases} -1, & \text{if } r_n \le (t_n - \tfrac{1}{4})h_n, \\ 1, & \text{if not, and } r_n \ge (t_n + \tfrac{1}{4})h_n, \\ 0, & \text{otherwise.} \end{cases}$$

Each case is a comparison of two ratios, decided by integer arithmetic, so the recursion defines the digits outright. By induction $t_n = \sum_{i<n} d_i 2^{n-1-i}$, so $t_n h_n = c_n$. It remains to show

$$\eta\big((t_n - 1)h_n\big) \le x \le \eta\big((t_n + 1)h_n\big)$$

for every n, which is the signed-digit condition. At $n = 0$ it says that x lies in $[-1, 1]$. Suppose it holds at n, and write $t = t_n$, $h = h_n$. Since $h_{n+1} = h/2$, the condition at $n + 1$ reads $\eta\big((t + (d_n - 1)/2)h\big) \le x \le \eta\big((t + (d_n + 1)/2)h\big)$.

If $d_n = -1$, its lower bound $(t - 1)h$ is the hypothesis, and $x \le \eta(r_n + h/8) \le \eta((t - \tfrac{1}{8})h) \le \eta(th)$ is its upper bound.

If $d_n = 1$, its upper bound $(t + 1)h$ is the hypothesis, and $\eta(th) \le \eta((t + \tfrac{1}{8})h) \le \eta(r_n - h/8) \le x$ is its lower bound.

If $d_n = 0$, then $(t - \tfrac{1}{4})h < r_n < (t + \tfrac{1}{4})h$, so $\eta((t - \tfrac{1}{2})h) \le \eta(r_n - h/8) \le x \le \eta(r_n + h/8) \le \eta((t + \tfrac{1}{2})h)$, which are its two bounds.

So (d_n) is a signed-digit expansion of x. □

The three candidate intervals at each step overlap: the middle one, between $(t - \tfrac{1}{2})h$ and $(t + \tfrac{1}{2})h$, contains a neighbourhood of the point th where a binary digit would have to decide. So an approximation within $h/8$ always shows a safe digit, and no comparison of x itself with a dyadic ratio is ever decided.

12.9.11 Without unique choice, excluded middle does not replace countable choice

Theorem 12.63 (Without unique choice, excluded middle does not replace countable choice)**.** *Consider arithmetic with variables for sequences, induction for every formula, excluded middle for every formula, and countable choice for the relations whose witnesses a search can find: a decidable condition, or a decidable condition under one existential quantifier. Write each set of ratios, sequence of points and function on the points as a formula. Then none of the following is provable in that theory: every Cauchy sequence of points converges; every located cut is the cut of a point; every uniformly continuous, strictly increasing f on $[0, 1]$ with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$ has a root; every uniformly continuous f on $[0, 1]$ has a supremum. In the mathematics of this chapter, each of the four follows from countable choice at pairs of ratios. The theory does not contain the principle of unique choice; since excluded middle and unique choice together give countable choice for numbers, one of the two must be added. The proof reasons in a classical metatheory.*

Conditions and scope.

Base and premises

Arithmetic with variables for sequences, induction for every formula, excluded middle for every formula, and countable choice for numbers only at matrices that are decidable or a decidable condition under one existential quantifier; no unique choice. Sets of ratios, sequences of points and functions on the points are formulas. The metatheory is classical.

Conclusion

None of the four assertions is provable: Cauchy sequences of points converge; located cuts are cuts of points; a uniformly continuous, strictly increasing f on $[0,1]$ with $f(\eta(0)) \leq \eta(0) \leq f(\eta(1))$ has a root; a uniformly continuous f on $[0,1]$ has a supremum.

Proof. We describe the theory and a model of it in which the sequences are the computable ones. In that model there is a point that no computable sequence names, and each of the four assertions fails at data built from it. The model is described in ordinary classical mathematics.

The theory. Its language, logical rules and arithmetic axioms are those of the theory of Theorem 12.22, with induction for every formula. In place of the two choice schemas, its further axioms are excluded middle, $A \vee \neg A$ for every formula A, and countable choice for numbers, $\forall x \exists y\, A(x,y) \to \exists \alpha \forall x\, A(x, \alpha(x))$, for the formulas A that are quantifier-free, that is built from equations and $\bot$ by $\wedge$, $\vee$ and $\to$, or have the form $\exists z\, B$ with B quantifier-free. A quantifier-free formula is a condition decided by a finite computation from its number variables and finitely many values of its sequence variables, so these are the relations whose witnesses a search can find. Everything below applies to every theory with fewer axioms as well, for instance to the theory with no choice at all.

The four assertions as formulas. A sequence variable α codes the sequence of ratios whose stage n is $(\alpha(3n) - \alpha(3n+1))/(\alpha(3n+2)+1)$, and every sequence of ratios has a code. Regularity, naming the same point, the order relations $<$ and $\leq$ of points, being within $1/(k+1)$, and membership of a ratio in the lower or the upper set of the cut of a point are then formulas: each is the integer condition of the text with its denominators cleared, a ratio being three number variables. A set of ratios is a formula with a ratio variable. A sequence of points is a formula $P(n, \beta)$ such that for every n the β with $P(n,\beta)$ are exactly the regular sequences naming one point, and there is one. A function on the points is a formula $F(\beta, \beta')$ that relates every regular β to some regular β', to only one point, and to the same point for sequences naming the same point. These formulas may have parameters. Each of the four assertions thus becomes a schema with one instance for each formula of the relevant kind, closed by universal quantifiers over its parameters. To show that an assertion is not provable we exhibit one instance, with a single sequence parameter, that is not a theorem.

The model. Numbers range over the natural numbers and sequences over the total computable functions; truth is classical. Every rule of the theory preserves truth. Instantiating $\forall \alpha$ and introducing $\exists \alpha$ use sequence variables only, whose values lie in the model; induction holds over the natural numbers; and excluded middle holds because truth in the model is classical. If $\forall x \exists y\, A(x,y)$ is true with A quantifier-free and its sequence parameters computable, then $x \mapsto$ the least y with $A(x,y)$ is computable, since each candidate is tested by a finite computation; for $\exists z\, B$ the search runs over pairs (y,z) and keeps y. So the restricted choice schema is true as well, and every theorem of the theory is true in the model under every assignment of

computable functions to its sequence variables. Full countable choice is false there: with $\alpha(\langle x, n\rangle)$ the bit recording whether program x halts on input 0 within n steps, every x has a y with $y = 0 \wedge \forall n\, \alpha(\langle x, n\rangle) = 0$ or $y = 1 \wedge \exists n\, \alpha(\langle x, n\rangle) = 1$, but a computable choice of such y would decide halting. So the theory does not prove countable choice.

A point of the model is named by a computable regular sequence. Reasoning classically, every regular sequence s has a limit $\lim s$, a real number with $|s_n - \lim s| \le 1/(n+1)$ for every n. Two regular sequences name the same point exactly when their limits agree, and $<$, $\le$, being within $1/(k+1)$ and membership of a ratio q in the cut of a point are the familiar relations between the limits and q (Lemma F.7).

A point with no computable name. Let $b(e) = 1$ if program e halts on input 0 and $b(e) = 0$ otherwise, and let

$$c = \sum_{e \ge 0} b(e)\, 4^{-(e+1)}, \qquad x_n = \sum_{e < n} b(e)\, 4^{-(e+1)}.$$

The tail beyond n is at most $\sum_{e \ge n} 4^{-(e+1)} = 4^{-n}/3$, so $x_n \le c \le x_n + 4^{-n}/3$. No computable regular sequence s has limit c. Suppose one did. Fix e, put $n = 6 \cdot 4^{e+1} - 1$ and $D_e = 4^{e+1} x_{e+1} = \sum_{i \le e} b(i)\, 4^{e-i}$, an integer whose remainder on division by 4 is $b(e)$. From $|s_n - c| \le 1/(n+1)$ and $D_e \le 4^{e+1} c \le D_e + 1/3$ we get

$$D_e \le 4^{e+1} s_n + \frac{1}{6} \le D_e + \frac{2}{3},$$

so D_e is the integer part of the ratio $4^{e+1} s_n + 1/6$. A program would compute it from e, and its remainder modulo 4 would decide whether program e halts [21].

Approximations from below. Let $b_s(e) = 1$ if program e halts on input 0 within s steps and $b_s(e) = 0$ otherwise, and let $c_s = \sum_{e < s} b_s(e)\, 4^{-(e+1)}$. The ratio c_s is computed from s, and $c_s \le c$ because $b_s \le b$. The c_s approach c: given k, take e_0 with $4^{-e_0}/3 \le 1/(k+1)$; the finitely many programs $e < e_0$ that halt all do so within some number S of steps, and for $s \ge \max(S, e_0)$ we get $c_s \ge x_{e_0} \ge c - 1/(k+1)$. Let γ be a computable code of the sequence of ratios $s \mapsto c_s$; it is the parameter of the instances of the first two assertions below.

Located cuts. Let $\mathscr{L}(q)$ say that $q < c_s$ for some s, and $\mathscr{U}(r)$ that some ratio $r' < r$ satisfies $c_s \le r'$ for every s. In the model, with γ as above, $\mathscr{L}$ is the set of ratios below c and $\mathscr{U}$ the set of ratios above c. Indeed, a ratio $q < c$ lies below c_s once c_s is close enough to c, and $q < c_s$ gives $q < c$; a ratio $r' < r$ bounding every c_s satisfies $c \le r' < r$, and if $c < r$ any ratio strictly between c and r bounds every c_s. So in the model $(\mathscr{L}, \mathscr{U})$ is the cut of the real number c: both sets are inhabited, the lower set is closed downward with no greatest element, the upper set is closed upward with no least element, they are disjoint, and for $q < r$ either $q < c$ or $c < r$. The instance of "every located cut is the cut of a point" at these formulas would supply a computable regular β such that $q < \lim \beta$ exactly when $q < c$, for every ratio q; then $\lim \beta = c$, which is impossible. The instance is false in the model, so it is not a theorem.

Cauchy sequences of points. Let $P(n, \beta)$ say that β is regular and names the same point as the constant sequence c_n. In the model each section is the class of the rational point c_n, inhabited by a computable constant sequence, so P is a sequence of points. It is Cauchy: for each k some N has $|c_m - c_n| \le 1/(k+1)$ for all $m, n \ge N$, since the c_s approach c. A limit in the model would be a computable regular λ with c_n approaching $\lim \lambda$, so $\lim \lambda = c$, which is impossible. Here the representatives are listed by a program; the limit is what the model lacks.

Two functions on the points. We build an increasing map H of the real line onto itself with $H(0) = c$. For every s let

$$H(y) = x_s - 4^{-s} + \big(1 + \tfrac{b(s)}{3}\big)\big(y + 4^{-s}\big) \qquad \text{for } -4^{-s} \le y \le -4^{-(s+1)},$$
$$H(y) = x_{s+1} + 4^{-(s+1)} + \big(1 - \tfrac{b(s)}{3}\big)\big(y - 4^{-(s+1)}\big) \qquad \text{for } 4^{-(s+1)} \le y \le 4^{-s},$$

$H(y) = y$ for $|y| \ge 1$ and $H(0) = c$. Adjacent pieces agree at their common end, because $x_{s+1} = x_s + b(s)\,4^{-(s+1)}$; at $y = \mp 1$ the value is ∓ 1; and $x_s \mp 4^{-s}$ tends to c, so H is continuous at 0. Every slope lies between $2/3$ and $4/3$, so $\frac{2}{3}(y' - y) \le H(y') - H(y) \le \frac{4}{3}(y' - y)$ for $y < y'$, and H is a strictly increasing bijection of the real line. Its inverse G is strictly increasing, with $\frac{3}{4}(x' - x) \le G(x') - G(x) \le \frac{3}{2}(x' - x)$ for $x < x'$, $G(c) = 0$, $G(1) = 1$ and $G(0) \le G(c) = 0$, since $0 \le c \le 1/3$. Put $f(x) = H(-|G(x)|)$. Then f changes by at most twice the change of x, $f(x) \le H(0) = c$ for every x, and $f(x) = x$ when $x \le c$, because then $G(x) \le 0$.

Both functions send computable points to computable points. A computable point x differs from c, so $G(x) \ne 0$. Then $G(x)$ lies in a piece on which H is affine with rational coefficients, or $|G(x)| \ge 1$, and on that piece G is an affine map with rational coefficients, slope at most $3/2$. Applying it to the stages of a computable regular sequence naming x, read at stage $2n + 1$ in place of n, gives a computable regular sequence naming $G(x)$. The same holds for f.

To write the graphs as formulas, note that for a ratio $p \ne 0$ the value $H(p)$ is a ratio computed from the bits $b(0), \dots, b(s)$, where s is the index of the piece containing p. Computing it with b_t in place of b gives a ratio $H_t(p)$, computed from t and p, and equal to $H(p)$ for every large t. Let γ' be a computable code of $(t, p) \mapsto H_t(p)$; it is the parameter of the instances of the third and fourth assertions. Let $F_{15}(\beta, \beta')$ say that β and β' are regular and that for every ratio $p \ne 0$ and every ratio q equal to $H_t(p)$ for all large t, the ratio p lies below the point of β' exactly when q lies below the point of β, and above it exactly when q lies above. In the model, since H is increasing, $F_{15}(\beta, \beta')$ says that $\lim \beta' = G(\lim \beta)$. Let $F_{17}(\beta, \beta')$ say that β' is regular and that some δ, δ' satisfy $F_{15}(\beta, \delta)$, $F_{15}(\beta', \delta')$ and that δ and δ' name points u and $-|u|$ for some point u; in the model it says $\lim \beta' = f(\lim \beta)$. By the preceding paragraph both formulas are functions on the points of the model.

Strictly increasing functions. At F_{15} the premises of the third assertion hold in the model: G is uniformly continuous on $[0, 1]$ (for k, take j with $3/(2j + 2) \le 1/(k + 1)$), strictly increasing, and $G(0) \le 0 \le G(1)$. A root in the model would be a computable point x of $[0, 1]$ with $G(x) = 0$, that is $x = c$, which is impossible.

Suprema. At F_{17} the premise of the fourth assertion holds in the model: f is uniformly continuous. A supremum in the model would be a computable point w such that $f(x) \le \lim w$ for every computable x in $[0, 1]$, and such that for every k some computable x in $[0, 1]$ has $f(x)$ within $1/(k + 1)$ of $\lim w$. Every ratio x with $0 \le x < c$ has $f(x) = x$, so $\lim w \ge c$; and $\lim w \le f(x) + 1/(k + 1) \le c + 1/(k + 1)$ for every k, so $\lim w = c$, which is impossible.

In each case the failure comes from c. In the same model the four assertions hold at computable data: the cut of $1/2$ is the cut of the constant point $1/2$, the constant sequence of the point 0 converges, and the identity function has the root 0 and the supremum 1.

The last sentence. In the mathematics of this chapter, countable choice at pairs of ratios makes every located cut the cut of a point (Theorem 12.33). The other three assertions follow from that one with no further choice, because the limit, the root and the supremum are each

located as a cut first and only that cut is made a point (Corollary 12.35 and Theorems 12.56 and 12.58). □

In a classical theory, countable choice for numbers is the same as the principle of unique choice, that a relation with exactly one witness for each n defines a sequence: given witnesses, the least one is unique. So the theorem shows that one of the two must be added. Under the reading of maps of Appendix C, which contains unique choice, excluded middle therefore supplies countable choice for numbers, and with it the four assertions. Over an intuitionistic base that already contains unique choice, the second assertion still needs countable choice, and countable choice proves all four without excluded middle; we show both next.

12.9.12 A located cut that unique choice does not make a point

Let t be a classical real number, and let $\mathscr{L}_t$ and $\mathscr{U}_t$ be the sets of ratios below t and above t. For each t this is a located cut, and it moves continuously with t. A sequence of ratios that depends continuously on t cannot follow it. Each of its stages is a ratio, given by a triple of numbers, and a continuous function of t with values in a discrete set is constant on every interval of values of t. The ratio 0 lies in $\mathscr{L}_t$ for t slightly above 0 and not for t slightly below, while a sequence that is constant near $t = 0$ names one point on both sides. In the open sets of the real line this picture becomes a model of the system HAS of Chapter 7 in which unique choice holds and the assertion that every located cut is the cut of a point does not hold. That settles, for located cuts, the question left after Theorem 12.63. Two further results complete the account: in a theory without excluded middle, countable choice for numbers proves the four assertions of that theorem; and LLPO, with countable choice only for the relations whose witnesses a search can find, gives neither binary expansions, nor the directions of plane vectors, nor exact roots.

In the language of Definition 7.14, a triple $q = (a, b, d)$ of numbers stands for the ratio $(a - b)/(d + 1)$, and $\bar{q} = J(J(a, b), d)$ is its code. Comparisons, sums and differences of ratios are integer conditions with the denominators cleared, written with the order (7.9). A set S is read as a sequence of ratios: *stage n of S is q* means $J(n, \bar{q}) \in S$. By Lemma 7.17(h), distinct pairs have distinct codes.

Definition 12.64 (Located cuts and points in HAS)**.** Let S, L and U be sets.

1. A set S is a *point* when every stage of S is some triple, two triples at the same stage are equal as triples, and triples q at stage m and q' at stage n satisfy $|q - q'| \le 1/(m + 1) + 1/(n + 1)$.

2. A pair of sets (L, U) is a *located cut* when the sets of ratios $\{q : \bar{q} \in L\}$ and $\{q : \bar{q} \in U\}$ satisfy the conditions of Definition 12.27.

3. The *cut of the point S* consists of the set L_S of ratios q such that some n and some triple t at stage $2n + 1$ of S have $t - q > 2/(n + 1)$, and the set U_S of ratios q such that some n and t at stage $2n + 1$ have $q - t > 2/(n + 1)$.

4. LCP is the sentence: for all sets L and U, if (L, U) is a located cut, then some point S has $\bar{q} \in L \leftrightarrow q \in L_S$ and $\bar{q} \in U \leftrightarrow q \in U_S$ for every triple q.

5. Let $\varphi(n, q)$ be a formula with a number variable n, a triple q of number variables and any parameters. The instance UC_φ of *unique choice* says: if every n has a triple q with $\varphi(n, q)$, and any two triples q, q' with $\varphi(n, q)$ and $\varphi(n, q')$ are equal, then some set S, of which every stage is some triple and two triples at the same stage are equal, has stage n equal to q exactly when $\varphi(n, q)$, for all n and q.

Item 3 is the cut of Definition 12.27 for the point x that S names: the sequence $(s_{2n+1} - q)_n$ names $x - \eta(q)$, and $\eta(q) < x$ says that this sequence has a stage above $2/(n + 1)$.

Remark 12.65 (Unique choice is derivable in HAS)**.** Each instance UC_φ is derivable in HAS by comprehension. Comprehension, with the parameters of φ, gives a set S with $x \in S$ exactly when some n and some triple q have $x = J(n, \bar{q})$ and $\varphi(n, q)$. Stage n of S is then q exactly when $\varphi(n, q)$: from $J(n, \bar{q}) = J(n', \bar{q}')$, Lemma 7.17(h), applied three times, gives $n = n'$ and the three entries of q equal to those of q', and substitution of equals carries $\varphi(n', q')$ to $\varphi(n, q)$. So the two premises of UC_φ say that every stage of S is some triple and that two triples at the same stage are equal. Unique choice is therefore no axiom added to HAS, and part 1 of the next theorem also follows from the proof of Lemma 7.21, which gives every derivable formula the value $\top$ under every assignment.

Theorem 12.66 (A located cut need not be a point, although unique choice holds)**.** *Let* LCP *and the instances* UC_φ *be as in Definition 12.64.*

1. *In every system of truth values, every instance* UC_φ *of unique choice has the value* $\top$ *under every assignment.*
2. *In the open sets of the classical real line,* LCP *does not have the value* $\top$*.*
3. *In the two-valued system* $\{\bot, \top\}$*,* LCP *has the value* $\top$*.*

Consequently neither LCP *nor* $\neg$LCP *is derivable in HAS, alone or from any finite list of instances of unique choice: over HAS with unique choice, the assertion that every located cut is the cut of a point is independent. The proof reasons in a classical metatheory.*

Proof. Values are written $[\![\varphi]\!]$ as in Section 7.10, with the assignments dropped when they are fixed. A formula without set variables has the value $\top$ when it is true in $\mathbb{N}$ and $\bot$ when it is false, in every system of truth values: by induction on the formula, since meets, joins and implications of the elements $\bot$ and $\top$ are computed as in the two-valued system. This applies to every comparison of ratios.

Unique choice. Fix a system H, an assignment and a formula φ. Let $A : \mathbb{N} \to H$ send $J(n, \bar{q})$ to $[\![\varphi(n, q)]\!]$, for every number n and triple q, and every other number to $\bot$; this is well defined because distinct pairs and distinct triples have distinct codes. Assign A to S. Then "stage n of S is q" has the value $[\![\varphi(n, q)]\!]$. So the two conditions on S have the values a and b of the two premises of UC_φ, and the final equivalence has the value $\top$. The conclusion of UC_φ, a join over all assignments to S, therefore lies above $a \wedge b$, and UC_φ has the value $(a \wedge b) \Rightarrow [\![\text{conclusion}]\!] = \top$.

A located cut of the real line. Let the truth values be the open subsets of the classical real line $\mathbb{R}$. Assign to L the map sending $\bar{q}$ to the open set $\{y \in \mathbb{R} : q < y\}$, and to U the map sending

$\bar{q}$ to $\{y : y < q\}$, with $\emptyset$ at every number that is not the code of a triple. Each condition of a located cut then has the value $\mathbb{R}$. A universal quantifier over ratios has as its value the interior of an intersection, which is $\mathbb{R}$ when every member is $\mathbb{R}$, so it suffices to check each instance. Both sets are inhabited: the sets $\{y : q < y\}$, and also the sets $\{y : y < q\}$, cover $\mathbb{R}$. Downward closure at $q' < q$ asks for $\{y : q < y\} \Rightarrow \{y : q' < y\}$, which is $\mathbb{R}$ because the first set lies in the second; when $q' < q$ is false its value is $\emptyset$ and the implication is $\mathbb{R}$. The absence of a greatest element asks for $\{y : q < y\} \Rightarrow \bigcup_{q'>q}\{y : q' < y\}$, and the two sets are equal because the ratios are dense in $\mathbb{R}$. The upper set is treated in the same way. Disjointness asks for the negation of $\{y : q < y\} \cap \{y : y < q\}$; this intersection is empty, so its negation is $\mathbb{R}$. Locatedness at $q < r$ asks for $\{y : q < y\} \cup \{y : y < r\}$, which is $\mathbb{R}$.

No point has this cut. Suppose that LCP had the value $\mathbb{R}$. Its value lies below that of its instance at the assignment above, whose premise has the value $\mathbb{R}$. So the open sets $[\![S \text{ is a point}]\!] \cap [\![(L,U) \text{ is the cut of } S]\!]$, one for each assignment A to S, cover $\mathbb{R}$. The point 0 lies in one of them, and with it an interval $B = (-\varepsilon, \varepsilon)$, where $\varepsilon > 0$. Two facts about open sets are used: a point of $V \Rightarrow W$ that lies in V lies in W, and the interior of an intersection lies in each member.

First, every stage of S is constant on B. Fix n. Totality puts B inside the union of the open sets $A(J(n,\bar{q}))$ over all triples q. Single-valuedness puts B inside $A(J(n,\bar{q})) \Rightarrow \big(A(J(n,\bar{q}')) \Rightarrow \emptyset\big)$ for distinct triples q, q', so no point of B lies in both sets. Thus the sets $B \cap A(J(n,\bar{q}))$ are open, pairwise disjoint and cover B. An interval is connected, so one of them is B: some triple t_n has $B \subseteq A(J(n,\bar{t}_n))$.

Second, look at the ratio 0. The point $y = \varepsilon/2$ of B lies in $\{y : 0 < y\}$, the value of $\bar{0} \in L$, so it lies in the value of $0 \in L_S$. That value is the union of the sets $A(J(2n+1,\bar{t}))$ over the n and the triples t with $t > 2/(n+1)$, so y lies in one of them. Since $y \in B$, the triple is t_{2n+1}, and all of B lies in that set. Hence $-y$ lies in the value of $0 \in L_S$ as well, and by the equivalence read from right to left it lies in $\{y' : 0 < y'\}$, the value of $\bar{0} \in L$. That says $0 < -\varepsilon/2$, which is false. So LCP does not have the value $\mathbb{R}$.

The two-valued system. Here a set is a set of numbers and every formula has its ordinary meaning. Let (L,U) be a located cut, and let x be the supremum of its lower set of ratios, a real number. The lower set consists of the ratios below x: its members lie below x because it has no greatest member, and a ratio below x lies below a member. The upper set consists of the ratios above x: a member lies above some smaller member, which bounds the lower set, and if $x < r$, a ratio q with $x < q < r$ is not in the lower set, so locatedness puts r in the upper set. Let S consist of the codes $J(m,\bar{t}_m)$, where, with $z = \lfloor x(m+1) \rfloor$, $t_m = (\max(z,0), \max(-z,0), m)$ is a triple of the ratio $z/(m+1)$. Then $x - 1/(m+1) < t_m \leq x$, so S is a point. If $q < x$, take n with $2/(n+1) + 1/(2n+2) < x - q$; then $t_{2n+1} - q > x - q - 1/(2n+2) > 2/(n+1)$. Conversely, $t_{2n+1} - q > 2/(n+1)$ gives $q < t_{2n+1} \leq x$. So L_S is the lower set, and in the same way U_S is the upper set. Hence LCP is true: its value is $\top$.

Derivations. The proof of Lemma 7.21 shows that a derivation of $\Gamma \vdash \varphi$ gives $[\![\Gamma]\!]_{\rho,\sigma} \leq [\![\varphi]\!]_{\rho,\sigma}$ for every assignment. When Γ is a list of instances of unique choice, its value is $\top$ in every system, by the first part. A derivation of LCP from Γ would give LCP the value $\top$ in the open sets of $\mathbb{R}$, against the second part. A derivation of $\neg$LCP would give $\neg$LCP the value $\top$ in the two-valued system, where LCP has the value $\top$ and its negation the value $\bot$. $\square$

Models of truth values on the open sets of the real line are developed for analysis by Fourman

and Hyland [59]; see also [5]. Unique choice holds in every such model because a set of HAS is any assignment of truth values to numbers, so the set that an instance asks for is the relation itself, read through the codes. The theorem therefore answers, for the second assertion of Theorem 12.63, the question that theorem leaves: over an intuitionistic base with unique choice, countable choice cannot be dropped from the passage from a located cut to a point, where Theorem 12.33 uses it once. For the other three assertions of Theorem 12.63, that every Cauchy sequence of points converges, that every uniformly continuous, strictly increasing f on $[0,1]$ with $f(\eta(0)) \le \eta(0) \le f(\eta(1))$ has a root, and that every uniformly continuous f on $[0,1]$ has a supremum, the same question over HAS with unique choice is not settled here: the argument above shows only that the cut assertion is not derivable.

12.9.13 Countable choice proves the four assertions without excluded middle

Countable choice, on the other hand, proves all four assertions with no excluded middle at all.

Theorem 12.67 (Countable choice proves the four assertions intuitionistically)**.** *Let* $\mathrm{HA}_1 + \mathrm{AC}_{00}$ *be the theory with the language, logical rules, arithmetic axioms and induction of Theorem 12.22 whose only further axioms are the instances*

$$\forall x \exists y\, A(x,y) \to \exists \alpha \forall x\, A(x, \alpha(x))$$

of countable choice for numbers, for every formula A. Each of the four assertions of Theorem 12.63, written as a schema as in its proof, has every instance provable in $\mathrm{HA}_1 + \mathrm{AC}_{00}$*. Hence every instance is provable in the theory of Theorem 12.22, and in the theory of Theorem 12.63 with countable choice for numbers extended to every formula.*

Proof. We follow the proofs of the chapter and check that they are derivations in this theory. Two facts about the theory are used throughout. Every quantifier-free formula B satisfies $B \vee \neg B$: an equation between two terms is decided by induction, as in Lemma 7.17(a), and the connectives preserve decided formulas. The order of two terms, written with an existential quantifier as in (7.9), is decided by induction as in Lemma 7.17(b). So every comparison of ratios, and every condition on finitely many stages of given sequences, is decided, as the chapter's proofs assume. And a relation with exactly one witness for each number is a sequence, by countable choice for numbers.

Located cuts. Let $\mathscr{L}$ and $\mathscr{U}$ be formulas with a ratio variable, and assume that they form a located cut. By existential elimination take $a \in \mathscr{L}$ and $b \in \mathscr{U}$. For each n let G_n be the product of the denominators of a and b with $4(n+1)$, and put $p_{n,i} = a + i/G_n$. With $a = (a_1, a_2, a_3)$ and $g_n = G_n/(a_3+1)$, the triple $(a_1 g_n + i,\ a_2 g_n,\ G_n - 1)$ stands for $p_{n,i}$; its entries are terms in n, i and the entries of a and b. The induction in the proof of Lemma 12.32, made inside the theory for each n, proves

$$\forall n \exists i \left(p_{n,i} \in \mathscr{L} \wedge p_{n,i+2} \in \mathscr{U} \right),$$

and it selects nothing. One instance of countable choice for numbers, at a matrix that asks at the place $3n$ for the first entry of the triple of such a $p_{n,i}$ and at the places $3n+1$ and $3n+2$ for its other two entries, which do not depend on i, gives one sequence β whose stage n,

read as in the proof of Theorem 12.63, is a ratio q_n with $q_n \in \mathscr{L}$ and $q_n + 2/G_n \in \mathscr{U}$. Since $2/G_n < 1/(n+1)$, the pairs $(q_n, q_n + 2/G_n)$ form an approximation sequence, and the converse half of the proof of Theorem 12.33, from an approximation sequence to a point, an integer computation followed by Lemma 12.28, shows that β is regular and that its cut is $(\mathscr{L}, \mathscr{U})$.

The other three assertions. Corollary 12.35 and Theorems 12.56 and 12.58 reduce them to the cut assertion. Each writes down a pair of sets of ratios, the cut of the limit (by Theorem 12.29), of the root or of the supremum, proves that it is a located cut, and proves that a point with this cut is the limit, the root or the supremum. These proofs use no choice principle beyond collecting uniquely determined values into sequences, which the second fact supplies, and no excluded middle beyond decided comparisons of ratios, and the sets, sequences of points and functions they mention are formulas of the kinds fixed in the proof of Theorem 12.63. Together with the instance of the cut assertion at the formulas of that cut, they are derivations in $\mathrm{HA}_1 + \mathrm{AC}_{00}$.

The two other theories contain every axiom of $\mathrm{HA}_1 + \mathrm{AC}_{00}$, so they prove every instance as well. □

With Theorem 12.22, the four assertions are theorems of a theory that proves none of LPO, WLPO, LLPO and $\mathrm{MP}_{\mathrm{bin}}$. With Theorem 12.63, excluded middle with countable choice for the relations whose witnesses a search can find, *search-choice* for short, does not prove them. With Theorem 12.66, HAS does not prove the second, even with unique choice added. In these four forms, completeness of the line follows from countable choice, which decides nothing; excluded middle with search-choice does not replace it, and for the second form neither does unique choice.

12.9.14 LLPO does not replace countable choice

The rows of Table 12.3 for exact roots and binary expansions cost LLPO, and in Theorems 12.54 and 12.61 the passage back from LLPO uses countable choice once, to tabulate a bit for every dyadic ratio. That use cannot be dropped either. For exact roots the model of Theorem 12.63 shows it: excluded middle holds there, and so LLPO, while the function G of that proof is uniformly continuous with $G(\eta(0)) \le \eta(0) \le G(\eta(1))$ and has no root. For binary expansions that model is of no use. In it every point of $[0,1]$ has a binary expansion: a computable point that is not a dyadic ratio has digits computed by bisection, each comparison with a dyadic ratio being strict and found by a search, a dyadic ratio below 1 has an expansion ending in zeros, and 1 the expansion whose digits are all 1. The same holds for the directions defined below: $(1,0)$ at the origin, and otherwise the vector divided by its length, computed from programs for its coordinates. A realizability of another kind separates them, one whose realizers may leave a choice among finitely many candidates open and strike the wrong candidates later [57, 58].

A *direction* of a vector (a,b) of points is a pair of points (p,q) with $p^2 + q^2 = \eta(1)$, $aq = bp$ and $ap + bq \ge \eta(0)$. When $a^2 + b^2 > \eta(0)$ the direction is (a,b) divided by its length, and every unit vector is a direction of $(0,0)$; in each case $(a,b) = \ell(p,q)$ with $\ell = ap + bq$, the polar form of the vector.

We write three statements in the language of Theorem 12.22, with points given by codes of regular sequences and functions on the points by formulas, as in the proof of Theorem 12.63.

- *Binary expansions*: every point x of $[0,1]$ has a binary sequence β with $\eta(c_n) \le x \le \eta(c_n + 2^{-n})$ for every n, where $c_n = \sum_{i<n} \beta(i) 2^{-(i+1)}$. The theory has no symbol for 2^n, so the statement asks for the bounds $\neg(x < \eta(\nu(n)/(\pi(n)+1)))$ and $\neg(\eta((\nu(n)+1)/(\pi(n)+1)) < x)$ for all sequences π, ν with $\pi(0) = \nu(0) = 0$, $\pi(n+1) = 2\pi(n) + 1$ and $\nu(n+1) = 2\nu(n) + \beta(n)$; these force $\pi(n) = 2^n - 1$ and $\nu(n) = 2^n c_n$.

- *Directions*: for all points a, b in $[-1,1]$ there are points p, q whose stages satisfy

$$|p_k^2 + q_k^2 - 1| \le \tfrac{6}{k+1}, \qquad |a_k q_k - b_k p_k| \le \tfrac{6}{k+1}, \qquad a_k p_k + b_k q_k \ge -\tfrac{6}{k+1}$$

  for every k. A product of stages of two points of $[-1,1]$ lies within $3/(k+1)$ of the product of the points, so a direction satisfies these conditions; conversely, the conditions give $p^2 + q^2 = \eta(1)$, $aq = bp$ and $ap + bq \ge 0$, since the tolerances tend to 0.

- *Exact roots*: the third assertion of Theorem 12.63 without the hypothesis that f is strictly increasing. Its instance at the function G is the instance at the formula F_{15} of that proof.

Definition 12.68 (Lifschitz realizability)**.** We use the programs, the pairing π and the valuations of the proof of Theorem 12.22. The *candidates* of a number r are the numbers $k \le (r)_1$ at which the run of program $(r)_0$ on input k does not halt; a number $k \le (r)_1$ is *struck* when that run halts. Under a valuation, $r \Vdash_1 A$, read *r realizes A*, is defined by recursion on A. For equations, $\bot$, $\wedge$, $\to$, $\forall x$ and $\forall \alpha$ the clauses are those of Kleene realizability $\Vdash_{\mathrm{k}}$ in that proof. The other three clauses are:

- $r \Vdash_1 A \vee B$ when r has a candidate, and every candidate k has $(k)_0 = 0$ and $(k)_1 \Vdash_1 A$, or $(k)_0 \neq 0$ and $(k)_1 \Vdash_1 B$;

- $r \Vdash_1 \exists x\, A(x)$ when r has a candidate, and every candidate k has $(k)_1 \Vdash_1 A((k)_0)$;

- $r \Vdash_1 \exists \alpha\, A(\alpha)$ when r has a candidate, and every candidate k has $(k)_0$ total and $(k)_1 \Vdash_1 A((k)_0)$.

A realizer of a disjunction or of an existential statement names a nonempty finite set of answers, each of them correct, without saying which; its program may strike candidates over time, but never all of them. For every k_0 a program, computed from k_0, halts on every input except k_0; paired with the bound k_0 it has the single candidate k_0. We call this number the *singleton* of k_0.

Lemma 12.69 (Derivations have Lifschitz realizers)**.** *Consider a theory with the language, logical rules, arithmetic axioms and induction of Theorem 12.22, whose further axioms each have a realizer computed by one program from the values of the free variables of the axiom. For every derivation of A from hypotheses $B_1, \dots, B_m$ there is a program that, given the values of the free variables and realizers of $B_1, \dots, B_m$, halts with a realizer of A. In particular every sentence that the theory proves has a realizer. The proof reasons in a classical metatheory.*

Proof. The program is built by induction on the derivation, as for Kleene realizability in the proof of Theorem 12.22 and Appendix F.1. The clauses for equations, $\bot$, $\wedge$, $\to$, $\forall x$ and $\forall \alpha$ are Kleene's, so the programs given there for hypotheses, $\bot$, $\wedge$, $\to$, the universal quantifiers, equality, the arithmetic axioms and induction serve unchanged; an axiom returns its given realizer. Only the rules for $\vee$, $\exists x$ and $\exists \alpha$ change. An introduction returns the singleton of $\pi(0, w)$ or $\pi(1, w)$, of $\pi(v, w)$ with v the value of the term, or of $\pi(i_\beta, w)$ with i_β the program assigned to β, where w is the value of the premise program.

The eliminations rest on *mixing*: for every formula C there is a program M_C such that, whenever a number r has a candidate and a program f halts with a realizer of C at every candidate of r, the run $M_C(r, f)$ halts with a realizer of C. Given this, to infer C from $A \vee B$ and the two branches, run the first premise program to obtain r and return $M_C(r, f)$, where f at k runs the second premise program with $(k)_1$ as the realizer of A if $(k)_0 = 0$, and the third with $(k)_1$ as the realizer of B otherwise. For $\exists x\, A(x)$, the program f runs the branch with x given the value $(k)_0$ and with $(k)_1$ as the realizer of $A(x)$; for $\exists \alpha\, A(\alpha)$, with α assigned $(k)_0$, which is total when k is a candidate. At a candidate the answer of f realizes C; elsewhere f may run forever, and M_C never needs those answers.

Mixing is built by induction on C. For an equation return 0: some candidate carries a realizer, so the equation is true. No candidate carries a realizer of $\bot$, so for $\bot$ the hypothesis never holds. For $A \wedge B$ return $\pi(M_A(r, f_0), M_B(r, f_1))$, where f_j returns the component j of the answer of f. For $A \to B$ return a program that on input s returns $M_B(r, f_s)$, where f_s at k applies the answer of f at k to s; when s realizes A, every candidate gives a realizer of B. For $\forall x\, A$ and $\forall \alpha\, A$ apply the answers to the input n or i in the same way.

In the remaining cases C is $A \vee B$, $\exists x\, A$ or $\exists \alpha\, A$, and the answers v_k of f at candidates k have candidates of their own. Write $e = (r)_0$ and $b = (r)_1$. Search for a time t at which every $k \le b$ is *settled*: the run of e on k halts within t steps, or f halts on k within t steps with an answer v_k. Such a time exists, since a $k \le b$ that is not a candidate is struck at some time and a candidate is answered at some time. Let b' be the largest $(v_k)_1$ over the k answered by time t, and let e' be the program that on input j searches for a time $t' \ge t$ at which every $k \le b$ is struck within t' steps, or was answered by time t with $j > (v_k)_1$ or with the run of $(v_k)_0$ on j halting within t' steps. Return $\pi(e', b')$. The search of e' on j fails exactly when some k fails at every time t', since there are finitely many k and the condition on each persists once it holds. Such a k is never struck, so it is a candidate of r, and j is a candidate of v_k. So the candidates of $\pi(e', b')$ are exactly the candidates of the v_k with k a candidate of r; every candidate of r was answered by time t, so its v_k has bound at most b'. There is at least one, and each carries what the clause of C asks, because it does so as a candidate of a realizer v_k of C. So $\pi(e', b')$ realizes C. □

Theorem 12.70 (LLPO does not replace countable choice)**.** *Let* $\mathrm{T}_{\mathrm{LLPO}}$ *be the theory of Theorem 12.63 with excluded middle replaced by LLPO: the language, logical rules, arithmetic axioms and induction of Theorem 12.22, the sentence LLPO stated in its proof, and countable choice for numbers at the formulas that are quantifier-free or of the form* $\exists z\, B$ *with* B *quantifier-free. Let* $\mathrm{T}^{+}_{\mathrm{LLPO}}$ *have instead LLPO and countable choice for numbers at every formula.*

1. *In* $\mathrm{T}_{\mathrm{LLPO}}$*, binary expansions and directions are each neither provable nor refutable, and so is the instance of exact roots at the function G of Theorem 12.63, also when its hypothesis of uniform continuity is strengthened to the existence of a modulus sequence.*

2. *Binary expansions, directions and every instance of exact roots are provable in* $\mathrm{T}^{+}_{\mathrm{LLPO}}$.

The proof of the first part reasons in a classical metatheory.

Proof. Reading realizers. Call a formula *plain* when it is built from equations by $\wedge$, $\vee$ and $\exists x$, or from such formulas and their negations by $\wedge$ and $\forall x$. Under a valuation, a plain formula is true exactly when it has a realizer, and when it is true a realizer is computed from the values of its free variables by a fixed program. This is an induction on the formula: a true equation is realized by 0; a true disjunction or existential statement is realized by the singleton of a true side or of a witness, found by a search through all tuples of numbers, and a realizer has a candidate whose side or witness is correct; a negation $\neg A$ of a formula A of the first kind is realized by every number when A is false and by none when A is true. The same holds for quantifier-free formulas, in which a true implication $A \to B$ is realized by the program returning the computed realizer of B. Regularity of a sequence, membership of a point in $[0,1]$ or $[-1,1]$, being a binary sequence and the stage conditions of a direction are plain. In the formula for binary expansions, a realizer applied to programs for the two sequences π, ν that the recursion determines, and to the computed realizer of the recursion, gives a realizer of the plain bounds; so a realizer shows that the expansion is one.

The axioms of $\mathrm{T}_{\mathrm{LLPO}}$ *are realized.* LLPO is realized by a program that ignores the realizers of its two premises and, at a binary sequence α given by a total program, returns the number with bound $\pi(1,0)$ whose program, at input $\pi(0,0)$, searches the even places of α for an entry 1, at input $\pi(1,0)$ searches the odd places, and halts at once at every other input. Recall that $\pi(0,0) = 0 < \pi(1,0)$. The premises are realized, hence true; for the second, apply its realizer at m and n, and then to 0 and 0, when $\alpha(m) = \alpha(n) = 1$. So α has at most one entry 1, at most one search succeeds, and a candidate remains. A remaining $\pi(0,0)$ carries the tag 0 and the realizer 0 of $\forall k\, \alpha(2k) = 0$, which is true because α is binary and the search fails; a remaining $\pi(1,0)$ carries the tag 1 and the realizer 0 of the other side. An instance of countable choice at a matrix $\exists z\, B(x,y,z)$ with B quantifier-free is realized by the program that ignores the realizer s of the premise and returns the singleton of $\pi(i,w)$, where i computes $\alpha(x)$ as the y of the first pair (y,z) that satisfies B, and w computes realizers of the true formulas $\exists z\, B(x,\alpha(x),z)$. The premise is realized, hence true, so every search halts and i is total. A quantifier-free matrix is the case without z. The arithmetic axioms are realized as in Appendix F.1. By Lemma 12.69, every theorem of $\mathrm{T}_{\mathrm{LLPO}}$ has a realizer.

Binary expansions have no realizer. Suppose R realized them. For a total program c whose sequence is regular with its point in $[0,1]$, apply R to c, the result to the computed realizer $g(c)$ of regularity, and that result to 0, which realizes membership in $[0,1]$. The value r realizes the existence of a binary expansion, so every candidate k of r has a total program $(k)_0$ computing a binary expansion of the point. By Kleene's recursion theorem [19, Chapter XII] there is a program c computing the following sequence of ratios. It keeps a target $m/2^N$, first $1/2$, and its stage n is the target before step n. At step n it runs the three applications above, each for at most n steps. If they halt with a value r, it takes the least $k \le (r)_1$ that is not struck within n steps and whose program $(k)_0$ computes its first N entries within n steps, if there is one, and lets ν be the numerator $\sum_{i<N} \beta(i) 2^{N-1-i}$ of those entries $\beta(i)$. If $\nu = m$ it moves the target to $m/2^N - 4^{-(n+2)}$, and if $\nu + 1 = m$ to $m/2^N + 4^{-(n+2)}$. In every other case the target stays. A moved target is again of the form $m/2^N$, now with $N = 2n + 4$.

The moves from step n on total at most $\sum_{j\ge n} 4^{-(j+2)} = \frac{1}{3}4^{-(n+1)}$, so the stages form a regular sequence whose point x lies within $1/12$ of $1/2$, in $[0,1]$. So R produces r. Suppose a step n moved the target on account of a candidate k. Its expansion of x has $v/2^N \le x \le (v+1)/2^N$. If $v = m$, the target moved to $m/2^N - 4^{-(n+2)}$ and the later moves total at most $\frac{1}{3}4^{-(n+2)}$, so $x < m/2^N = v/2^N$; if $v+1 = m$, symmetrically $x > (v+1)/2^N$. Both are impossible, so the program never moves on account of a candidate. Now take a time T_1 beyond the steps the three applications need, at which every $k \le (r)_1$ that is not a candidate is struck. From step T_1 on, every k that the program examines is a candidate, so the target never moves again: it is a fixed $m/2^N$, and $x = m/2^N$. The number r has a candidate k_0, whose program is total and computes its first N entries within some T_2 steps. At step $\max(T_1, T_2)$ the program finds a candidate k, whose expansion has $v/2^N \le m/2^N \le (v+1)/2^N$, so $v = m$ or $v+1 = m$, and it moves the target, on account of a candidate. This contradiction shows that no realizer exists.

Directions have no realizer. Suppose R realized them. For total programs a, b whose sequences are regular with points in $[-1,1]$, six applications of R, to a, the computed realizer of its regularity, 0, then b, its regularity and 0, give a number r whose candidates k carry total programs $(k)_0$ computing regular sequences p such that some q makes (p,q) a direction of (a,b). By the recursion theorem there is a program c computing a pair of sequences a, b, with programs for them computed from c. Both are 0 until the first step s_0 at which the six applications have produced r within s_0 steps each and, with $M = (r)_1 + 2$ and $K = 4M^2$, every $k \le (r)_1$ is struck within s_0 steps or its program has computed stage K of its sequence p within s_0 steps. At step s_0 the program takes the least $j < M$, which exists as shown below, such that every $k \le (r)_1$ that is not struck within s_0 steps has its stage K at distance at least $1/(4M^2)$ from

$$x_j = \frac{M^2 - j^2}{M^2 + j^2}, \qquad \text{and puts} \qquad y_j = \frac{2jM}{M^2 + j^2},$$

and from step s_0 on the two sequences are the constant ratios εx_j and εy_j, with $\varepsilon = 2^{-(s_0+2)}$.

Such sequences are regular, since they change once, by at most $\varepsilon < 1/(s_0+1)$, and their points lie in $[-1,1]$. So R produces r, and the step s_0 exists: a non-candidate is struck at some time, and the program of a candidate is total. A suitable j exists. For $j < j' < M$,

$$x_j - x_{j'} = \frac{2M^2(j'^2 - j^2)}{(M^2 + j^2)(M^2 + j'^2)} \ge \frac{2M^2}{4M^4} = \frac{1}{2M^2},$$

so a ratio within less than $1/(4M^2)$ of one x_j is not that close to another; the at most $(r)_1 + 1 = M - 1$ unstruck stages exclude at most $M - 1$ of the M indices. Let J be the index taken. Then $(a,b) = \varepsilon(x_J, y_J)$ with $x_J^2 + y_J^2 = 1$ and $\varepsilon > 0$. Let k_0 be a candidate of r and (p,q) its direction. From $p^2 + q^2 = 1$, $x_J q = y_J p$ and $x_J p + y_J q \ge 0$,

$$(x_J p + y_J q)^2 = (x_J^2 + y_J^2)(p^2 + q^2) - (x_J q - y_J p)^2 = 1,$$

so $x_J p + y_J q = 1$ and $(p - x_J)^2 + (q - y_J)^2 = 2 - 2(x_J p + y_J q) = 0$: $p = x_J$. The candidate k_0 is not struck within s_0 steps, so its stage K lies at distance at least $1/(4M^2)$ from $x_J = p$, while a regular sequence has $|p_K - p| \le 1/(K+1) < 1/(4M^2)$. This contradiction shows that no realizer exists.

Exact roots, and the models. The model of Theorem 12.63, whose sequences are the computable ones, satisfies $\mathrm{T}_{\mathrm{LLPO}}$: excluded middle holds in it, and LLPO follows from excluded middle, since if not every even entry is 0 the binary sequence has its entry 1 at an even place and every odd entry is 0. The function G is uniformly continuous, with the computable modulus $k \mapsto 2k+1$ because G changes by at most twice the change of its argument, and $G(\eta(0)) \leq \eta(0) \leq G(\eta(1))$, but it has no root in the model. So the instance at F_{15}, with either form of uniform continuity, is not provable. In the classical model of the proof of Theorem 12.22, whose sequences are all functions, every axiom of $\mathrm{T}^{+}_{\mathrm{LLPO}}$ holds, LLPO by excluded middle in the metatheory. Every point of $[0,1]$ has a binary expansion there, with digits $\lfloor 2^{n+1}x \rfloor - 2\lfloor 2^{n}x \rfloor$, and all digits 1 at $x = 1$; every vector of $[-1,1]^2$ has a direction, (a,b) divided by its length or $(1,0)$ at the origin, named by the sequences $\lfloor u(k+1) \rfloor/(k+1)$ of the two coordinates u; and the classical intermediate value theorem gives every uniformly continuous function with a change of sign a root, named in the same way. So the three statements are true there, and none of their negations is provable, in $\mathrm{T}^{+}_{\mathrm{LLPO}}$ or in $\mathrm{T}_{\mathrm{LLPO}}$. With the realizers above this proves the first part.

The second part. Read in the language of the theory, the proofs of the second parts of Theorems 12.54 and 12.61, and the proof of directions below, use, besides intuitionistic reasoning with decided comparisons of ratios, LLPO and countable choice for numbers in three ways: to tabulate a bit for every dyadic ratio; to form, for a point x and a ratio u, the binary sequence of first detections from which LLPO gives $x \leq \eta(u)$ or $\eta(u) \leq x$ (Corollary 12.18); and to define sequences by recursion, such as $a_{n+1} = 2a_n + B(n+1, 2a_n + 1)$. The last two collect uniquely determined values, which this theory obtains only through countable choice. For a recursion, every n has a sequence that follows it up to n, by induction on n, where one instance of countable choice extends such a sequence by one step; two such sequences agree up to n, by induction; and one more instance collects their values at n into one sequence. So these proofs are derivations in $\mathrm{T}^{+}_{\mathrm{LLPO}}$.

For directions, let a, b lie in $[-1,1]$. LLPO gives every point u with $u \geq \eta(0)$ or $u \leq \eta(0)$ (Corollary 12.18). Applied to a, to b, and then to the difference of the two resulting coordinates, it lets us replace (a,b) by $(-a,b)$, $(a,-b)$ or (b,a) where needed until $0 \leq b \leq a \leq 1$. Each replacement carries a direction (p,q) of the new vector to a direction of the old one, by the same replacement: it keeps p^2+q^2 and $ap+bq$, and it keeps $aq-bp$ or changes its sign. So let $0 \leq b \leq a \leq 1$, and put $h(s) = as - b + bs^2/4$. Then $h(\eta(0)) = -b \leq \eta(0)$, $h(\eta(1)) = a - \frac{3}{4}b \geq \eta(0)$, and $|h(s) - h(s')| \leq \frac{3}{2}|s - s'|$ on $[0,1]$, so h is uniformly continuous. The exact intermediate value theorem gives s in $[0,1]$ with $h(s) = \eta(0)$. With $t = s/2$ put $p = (1-t^2)/(1+t^2)$ and $q = 2t/(1+t^2)$. Then $p^2+q^2 = \eta(1)$, $aq - bp = h(s)/(1+t^2) = \eta(0)$, and $ap+bq \geq \eta(0)$ because a, b, p and q are at least $\eta(0)$. □

The first parts of Theorems 12.54 and 12.61 derive LLPO from exact roots and from binary expansions with no choice. So in both rows of Table 12.3 LLPO is the price, and the way back from LLPO needs countable choice: LLPO with search-choice does not give the statement, and LLPO with countable choice does. Whether binary expansions or directions follow from excluded middle with search-choice is not settled here; the model of computable sequences, which settles the case of exact roots, satisfies both statements.

12.9.15 The same prices away from the line

The same prices appear for statements about binary sequences and colourings. Weak König's lemma [15] for decidable binary trees is equivalent to LLPO, and the infinite pigeonhole principle, for colourings with any finite number of colours, and Ramsey's theorem for pairs with two colours [16] are each equivalent to LPO. In each case the passage to the principle uses no choice, and countable choice is used only on the way back. The omniscience principles are the usual yardstick of constructive analysis [14, 4]; we prove the three equivalences here, because where choice enters is part of what the table records.

Definition 12.71 (Decidable binary trees)**.** A *binary string* is a finite sequence $s = (s_0, \dots, s_{n-1})$ of digits in $\{0, 1\}$; $n = |s|$ is its *length*, and sb is s followed by the digit b. For a binary sequence p write $\bar{p}(n) = (p(0), \dots, p(n-1))$. A *decidable binary tree* is a set T of binary strings, given by its characteristic function, a map from binary strings to $\{0, 1\}$, so that membership is decidable, which contains the empty string and contains s whenever it contains sb; so it contains every initial segment of each of its members. It is *infinite* when it has a member of every length. A *path* through T is a binary sequence p with $\bar{p}(n) \in T$ for every n. *Weak König's lemma* states that every infinite decidable binary tree has a path.

For such a tree the usual readings of "infinite" agree: a member of every length, members of arbitrarily large length, and infinitely many members. Whether T has a member of length n is decided by inspecting the 2^n strings of that length, and a tree with no member of length n has all its members among the $2^n - 1$ shorter strings.

Theorem 12.72 (Weak König's lemma is LLPO)**.** *If weak König's lemma holds, then LLPO holds; the proof uses no choice principle. Conversely, LLPO and* AC_ω *give weak König's lemma; countable choice is used once, to collect for every depth a finite table of extendable children.*

Proof. Assume weak König's lemma, and let α have at most one entry equal to 1. Let T_α consist of the strings s such that either every digit of s is 0 and $\alpha(m) = 0$ for every even $m < |s|$, or every digit of s is 1 and $\alpha(m) = 0$ for every odd $m < |s|$. Membership is decided by inspecting s and $\alpha(0), \dots, \alpha(|s| - 1)$. The empty string belongs to T_α, and deleting the last digit of a member preserves both conditions, so T_α is a decidable binary tree. It is infinite. Given n, inspect the even indices $m < n$. If $\alpha(m) = 0$ at each of them, the string of n zeros belongs to T_α. If $\alpha(m) = 1$ at one of them, that is the only entry 1 of α, so α vanishes at every odd index and the string of n ones belongs to T_α.

Let p be a path through T_α, and decide its first digit. Suppose $p(0) = 0$, and fix k. The member $\bar{p}(2k + 1)$ of T_α begins with 0, so it is not a string of ones. Hence it is a string of zeros, α vanishes at every even index below $2k + 1$, and in particular $\alpha(2k) = 0$. If instead $p(0) = 1$, the member $\bar{p}(2k + 2)$ is not a string of zeros, so $\alpha(2k + 1) = 0$, for every k. These are the two alternatives of LLPO. Every step is a bounded inspection of data already given; no choice principle is used.

For the converse, assume LLPO and AC_ω, and let T be an infinite decidable binary tree. Say that a string s *reaches depth* n when $se \in T$ for some string e of length n. This is decided by inspecting the 2^n strings e, and since T is closed under initial segments, a string that reaches depth n reaches every smaller depth. Call s *extendable* when it reaches every depth. The empty string is extendable, because T is infinite.

A child for each string, from LLPO. Fix s, and let $b_0(n)$ and $b_1(n)$ be the decidable statements that $s0$, respectively $s1$, reaches depth n. Define a binary sequence γ: $\gamma(2n) = 1$ when n is the first index at which b_0 or b_1 fails and b_0 fails there; $\gamma(2n+1) = 1$ when n is that first index, b_0 holds there and b_1 fails; γ is 0 at all other positions. Each value is a bounded search. The first index of a failure is unique, so γ has at most one entry equal to 1, and LLPO applies to it. If every even entry is 0, put $c_s = 0$; if every odd entry is 0, put $c_s = 1$. We show that sc_s is extendable whenever s is.

Let $c_s = 0$, assume that s is extendable, and fix n. The statement $b_0(n)$ is decidable; suppose it fails. Since s reaches depth $n+1$, one of $s0$ and $s1$ reaches depth n, so $b_1(n)$ holds. A bounded search finds the first index $j \leq n$ at which b_0 or b_1 fails. Since $\gamma(2j) = 0$, b_0 holds at j, so b_1 fails at j; but $b_1(n)$ holds and $j \leq n$, so $b_1(j)$ holds. This contradiction proves $b_0(n)$. The case $c_s = 1$ is symmetric: a failure of $b_1(n)$ forces $b_0(n)$, hence b_0 holds at the first failure $j \leq n$ and b_1 fails there, so $\gamma(2j+1) = 1$, contrary to the case. The child is named without deciding whether s is extendable, which is a universal statement.

Finitely many strings at once, with no choice principle. For each d there is a function f on the strings of length d such that $sf(s)$ is extendable for every extendable s of length d. By induction on d one proves, for every string t, that there is such a function on the strings te with $|e| = d$: for $d = 0$ take the child named for t, and for $d+1$ combine the functions given for $t0$ and for $t1$. The case of the empty string t is the claim.

Countable choice, once. AC_ω, applied to the relation "f is such a function for the strings of length d", gives a sequence $F_0, F_1, \ldots$ of these functions. Each is a finite table of digits, so the choice can be read as a choice of numbers.

The path. Let s_0 be the empty string, $s_{d+1} = s_d F_d(s_d)$, and $p(d) = F_d(s_d)$, so that $\bar{p}(d) = s_d$. By induction on d every s_d is extendable. An extendable string reaches depth 0, that is, it belongs to T. So every initial segment of p belongs to T, and p is a path. □

The step that names a child, taken alone, is equivalent to LLPO with no choice principle in either direction: the root of T_α is extendable, and its extendable child is the decision. Countable choice enters only to collect these decisions along all depths. Without countable choice weak König's lemma still has an exact price: it is equivalent to LLPO for a whole sequence of questions at once, with no choice principle in either direction (Proposition 13.9).

Definition 12.73 (Colourings and the infinite pigeonhole principle)**.** For a natural number k, a *colouring of* $\mathbb{N}$ *with* $k+1$ *colours* is a sequence c with values in $\{0, 1, \ldots, k\}$. A strictly increasing sequence σ of natural numbers is *homogeneous* for c when $c(\sigma(n))$ is the same colour for every n. The *infinite pigeonhole principle* states that, for every k, every colouring of $\mathbb{N}$ with $k+1$ colours has a homogeneous sequence.

A strictly increasing sequence of natural numbers satisfies $\sigma(n) \geq n$, by induction on n. Both converses below use one consequence of LPO and countable choice.

Lemma 12.74 (Splitting a double sequence)**.** *Assume LPO and* AC_ω*, and let* $f(N,m) \in \{0,1\}$ *for all* N, m*. Either every row has an entry* 1*, that is, for every* N *some* m *has* $f(N,m) = 1$*, or some row is entirely* 0*.*

Proof. For each N, LPO applied to the row $m \mapsto f(N,m)$ decides whether it has an entry 1. Countable choice collects these decisions into a binary sequence e, with $e(N) = 1$ exactly when row N has an entry 1. LPO applied to $1 - e$ gives either $e(N) = 1$ for every N, which is the first case, or an N with $e(N) = 0$, whose row is entirely 0. □

Theorem 12.75 (The infinite pigeonhole principle is LPO)**.** *If every colouring of* $\mathbb{N}$ *with two colours has a homogeneous sequence, then LPO holds; the proof uses no choice principle. Conversely, LPO and* AC_ω *give the infinite pigeonhole principle, and the passage from two colours to any finite number of colours uses no choice principle. So, under* AC_ω*, the infinite pigeonhole principle, for colourings with any finite number of colours, is equivalent to LPO, and the passage to LPO uses no choice. Countable choice is used in Lemma 12.74 and once more to pick later indices of colour* 1*.*

Proof. Assume the principle for two colours, and let α be a binary sequence. Colour n by $a(n) = 1$ if $\alpha(m) = 1$ for some $m \le n$, and by $a(n) = 0$ otherwise; each value is a bounded search. Let σ be homogeneous for a, with colour b. If $b = 1$, then $a(\sigma(0)) = 1$, and the search behind this value supplies an index $m \le \sigma(0)$ with $\alpha(m) = 1$: the second alternative of LPO. If $b = 0$, fix n. Since $\sigma(n) \ge n$ and $a(\sigma(n)) = 0$, the sequence α vanishes at every index up to $\sigma(n)$, and in particular $\alpha(n) = 0$: the first alternative. Nothing is selected. Two colours are the case $k = 1$ of the infinite pigeonhole principle, so the principle gives LPO with no choice.

Conversely, assume LPO and AC_ω, and let c be a colouring with two colours. Apply Lemma 12.74 to $f(N,m) = c(N+m)$, whose rows are the tails of c. If some tail, from N on, is entirely 0, then $\sigma(n) = N + n$ is homogeneous with colour 0. Otherwise every tail has an entry 1. Countable choice gives a sequence g with $g(n) > n$ and $c(g(n)) = 1$ for every n. Take an index a with $c(a) = 1$, and put $\sigma(0) = a$ and $\sigma(n+1) = g(\sigma(n))$. This sequence is strictly increasing and homogeneous with colour 1.

Finally, we show by induction on k that every colouring with values at most k has a homogeneous sequence, using only the principle for two colours. For $k = 0$ the colouring is constant, and $\sigma(n) = n$ serves. Let c have values at most $k+1$. Colour n by 1 if $c(n) = k+1$ and by 0 otherwise, and let σ be homogeneous for this colouring. If its colour is 1, then σ is homogeneous for c, with colour $k+1$. If its colour is 0, the colouring $n \mapsto c(\sigma(n))$ has values at most k; by the induction hypothesis it has a homogeneous sequence τ, with some colour i. Then $n \mapsto \sigma(\tau(n))$ is strictly increasing, since σ and τ are, and $c(\sigma(\tau(n))) = i$ for every n. This step selects nothing. □

Both uses of countable choice in the converse select uniquely determined data: each decision $e(N)$ in Lemma 12.74 is the only correct bit, and $g(n)$ can be taken to be the least index above n of colour 1. So the converse needs only the principle of unique choice, that a relation with exactly one witness for each n defines a sequence.

Definition 12.76 (Ramsey's theorem for pairs)**.** A *colouring of pairs with two colours* assigns a colour $c(x, y) \in \{0, 1\}$ to each pair of natural numbers $x < y$. A strictly increasing sequence σ is *homogeneous* for c when $c(\sigma(i), \sigma(j))$ is the same colour for all $i < j$. *Ramsey's theorem for pairs with two colours* states that every colouring of pairs with two colours has a homogeneous sequence.

Theorem 12.77 (Ramsey's theorem for pairs is LPO)**.** *Ramsey's theorem for pairs with two colours gives LPO; the proof uses no choice principle. Conversely, LPO and* AC_ω *give Ramsey's theorem for pairs with two colours; countable choice is used through Lemma 12.74, once to pick a successor for every node, and through Theorem 12.75.*

Proof. Assume Ramsey's theorem for pairs, and let a be a colouring of $\mathbb{N}$ with two colours. Colour each pair $x < y$ by $a(x)$, and let σ be homogeneous for this colouring of pairs, with colour b. For every n, $a(\sigma(n))$ is the colour of the pair $\sigma(n) < \sigma(n + 1)$, which is b; so σ is homogeneous for a. Theorem 12.75 turns this into LPO, and neither step uses a choice principle.

For the converse assume LPO and AC_ω, and fix a colouring c of pairs with two colours.

The tree of the colouring. The tree is that of Erdős and Rado [17]. By recursion on x, call x an *ancestor* of y, written $x \prec y$, when $x < y$ and $c(w, x) = c(w, y)$ for every ancestor w of x. Every ancestor of x is smaller than x, so each instance is decided by a finite computation. Four properties follow, each by induction on the smallest node involved.

(i) $0 \prec y$ for every $y > 0$, since 0 has no ancestor.

(ii) If $w \prec x$ and $x \prec y$, then $w \prec y$. Clearly $w < y$. For an ancestor v of w, induction gives $v \prec x$, and then $c(v, w) = c(v, x) = c(v, y)$.

(iii) If $w < x$, $w \prec y$ and $x \prec y$, then $w \prec x$. For an ancestor v of w, (ii) gives $v \prec y$, induction gives $v \prec x$, and so $c(v, w) = c(v, y) = c(v, x)$.

(iv) Let $z \prec x'$ and $b = c(z, x')$, and suppose that no $v < x'$ with $z \prec v$ has $c(z, v) = b$. Then $x' \prec y$ for every $y > x'$ with $z \prec y$ and $c(z, y) = b$. Indeed, let $w \prec x'$. If $w < z$, then $w \prec z$ by (iii), and $c(w, x') = c(w, z) = c(w, y)$. If $w = z$, both colours are b. If $z < w$, then $z \prec w$ by (iii) and $c(z, w) = c(z, x') = b$, which is impossible because $w < x'$.

Nodes with many descendants. Say that z has *many descendants* when for every N some $y \geq N$ has $z \prec y$. By (i) the root 0 has many descendants. Fix a node z, and apply Lemma 12.74 twice: to the double sequence that records whether $N + m$ is a descendant of z with $c(z, N + m) = 1$, and to the one with colour 0. Suppose that for one colour b every row has an entry 1, so that z has descendants y with $c(z, y) = b$ beyond every bound. Take one of them, y_0, and let x' be the least $v \leq y_0$ with $z \prec v$ and $c(z, v) = b$, found by a bounded search. Then $z \prec x'$, and by (iv) every later descendant y of z with $c(z, y) = b$ satisfies $x' \prec y$; there are such y beyond every bound, so x' has many descendants. If instead both applications give an all-zero row, at N_1 for colour 1 and at N_2 for colour 0, then z has no descendant $y \geq N_1 + N_2$, so z does not have many descendants; put $x' = 0$. In every case we have found x' such that, if z has many descendants, then $z \prec x'$ and x' has many descendants.

The path. Countable choice gives a sequence g such that, for every z with many descendants, $z \prec g(z)$ and $g(z)$ has many descendants. Put $x_0 = 0$ and $x_{k+1} = g(x_k)$. By induction on k, every x_k has many descendants and $x_k \prec x_{k+1}$, so in particular $x_k < x_{k+1}$. By (ii), $x_i \prec x_j$ whenever $i < j$. Hence $c(x_i, x_j) = c(x_i, x_{i+1})$ for $i < j$: for $j = i+1$ there is nothing to prove, and for $j > i+1$ it is the definition of $x_{i+1} \prec x_j$ at the ancestor x_i of x_{i+1}.

Homogeneity. Colour $\mathbb{N}$ by $d(i) = c(x_i, x_{i+1})$. By Theorem 12.75, LPO and AC_ω give a homogeneous sequence τ for d, with some colour b. Put $\sigma(n) = x_{\tau(n)}$. It is strictly increasing, and for $i < j$ we have $\tau(i) < \tau(j)$, so

$$c(\sigma(i), \sigma(j)) = c(x_{\tau(i)}, x_{\tau(i)+1}) = d(\tau(i)) = b. \qquad \square$$

What is particular here is the carrier: every row of Table 12.3 is proved on the line built from the ratios of Chapter 4, and every passage from a theorem to its principle, in the table and in the three theorems above, is made without choice.

12.9.16 Monotone convergence splits as LPO does

Table 12.3 repeats the split of LPO.

Proposition 12.78 (A nondecreasing limit is the least upper bound)**.** *If a nondecreasing sequence (x_n) converges to L, then L is a least upper bound of the sequence. No omniscience principle and no choice principle is used. (Limits are unique for every sequence, Theorem 12.26; here this also follows from the uniqueness of least upper bounds.)*

Proof. Suppose $L < x_n$ for some n. By Lemma 12.30, used twice, there are ratios $c < a$ with $L < \eta(c) < \eta(a) < x_n$; take k with $1/(k+1) \le a - c$. Some N has x_m within $1/(k+1)$ of L for every $m \ge N$. For $m = N + n$ we have $x_n \le x_m$, so $\eta(a) < x_m$ by Lemma 12.40(a), while $L < \eta(c)$; Lemma 12.41(b) says that x_m and L are not within $1/(k+1)$, a contradiction. So $x_n \le L$ for every n.

Let b be an upper bound, and suppose $b < L$. Choose ratios $c < a$ with $b < \eta(c) < \eta(a) < L$ and k with $1/(k+1) \le a - c$, and take N as before. Then $x_N \le b < \eta(c)$ gives $x_N < \eta(c)$, while $\eta(a) < L$, and Lemma 12.41(b) again contradicts the choice of N. So $L \le b$.

Two limits of the sequence are therefore two least upper bounds, each below the other, and they are equal by item 2 of Proposition 12.19. $\square$

Proposition 12.79 (Approaching a least upper bound costs exactly $\mathrm{MP}_{\mathrm{bin}}$)**.** *That least upper bounds of sequences are approached by the terms is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$*, in the following sense.*

1. *If every least upper bound of every sequence of points is approached by its terms, then* $\mathrm{MP}_{\mathrm{bin}}$ *holds. The proof uses no choice principle.*

2. *Assume* $\mathrm{MP}_{\mathrm{bin}}$ *and* AC_ω*. Then every least upper bound of a sequence of points is approached by its terms. Countable choice is used only to fix regular sequences naming the terms; for a sequence given by regular sequences no choice is used.*

Proof. For the first statement, let α be a binary sequence that is not all zero, and let $b_n = \eta(\alpha(n))$. The point $\eta(1)$ is an upper bound. It is the least: if an upper bound y had $y < \eta(1)$, every entry of α would be 0, as in the proof of Theorem 12.48, which is excluded; so $\eta(1) \leq y$ for every upper bound y. By hypothesis the ratio 0, which has $\eta(0) < \eta(1)$, is exceeded by some term b_n. Then $b_n \neq \eta(0)$, so $\alpha(n) = 1$.

For the second, let L be a least upper bound of (x_n) and $\eta(q) < L$. Then $\eta(q)$ is not an upper bound, since leastness would give $L \leq \eta(q)$. Fix regular sequences naming the terms, by countable choice unless they are given. Then $\mathrm{MP}_{\mathrm{bin}}$, through Lemma 12.42(c), gives a term with $\eta(q) < x_n$. □

So, with countable choice, monotone convergence is the least-upper-bound principle together with $\mathrm{MP}_{\mathrm{bin}}$, just as LPO is WLPO together with $\mathrm{MP}_{\mathrm{bin}}$: by Theorem 12.45 the first is LPO, by Theorem 12.48 the second is WLPO, and Proposition 12.9 joins them. Proposition 12.79 locates the difference: what monotone convergence adds to a least upper bound is that the terms approach it, and that costs exactly $\mathrm{MP}_{\mathrm{bin}}$.

12.10 A price table for the calculus on the line

Rolle's theorem promises a point where the derivative vanishes. Here is a function for which we can watch that point move. For a point a put

$$f_a(x) = \max(a, 0)\, x(1 - x)(2 - x) + \min(a, 0)\, x(1 - x)(1 + x).$$

Then $f_a(0) = f_a(1) = 0$. If $a > 0$ the derivative $a(2 - 6x + 3x^2)$ vanishes in $[0, 1]$ only at $1 - 1/\sqrt{3} \approx 0.42$; if $a < 0$ the derivative $a(1 - 3x^2)$ vanishes only at $1/\sqrt{3} \approx 0.58$; if $a = 0$ every point is critical. A procedure that returns a critical point c of every f_a would decide a side of every point: one free comparison places c above $9/20$, which excludes $a > 0$, or below $11/20$, which excludes $a < 0$. That is LLPO. Yet a point where the derivative is within a given tolerance of 0 is found by a finite search.

This section prices the calculus of the line; the rows are collected in Table 12.4. Throughout, a ratio q also denotes the point $\eta(q)$, and $|x| = \max(x, -x)$. Points x, y are within $1/(k + 1)$ exactly when $|x - y| \leq 1/(k + 1)$, since a maximum lies at or below a point exactly when both arguments do (Proposition 12.19, item 5). The rules $|u + v| \leq |u| + |v|$, $|uv| = |u|\,|v|$ and "$|u| \leq b$ exactly when $-b \leq u \leq b$" are equations or $\leq$-inequalities proved by cases on the order, so they hold with no principle (Remark 12.20).

Derivatives with a modulus

Definition 12.80 (Derivatives on the line)**.** Let f, g be functions on the points of $\mathbb{R}_\delta$.

1. g is a *uniform derivative* of f on $[0, 1]$ when for every k there is a j such that $|f(y) - f(x) - g(x)(y - x)| \leq |y - x|/(k + 1)$ for all x, y in $[0, 1]$ within $1/(j + 1)$.
2. f has *derivative* d at the point x when for every k there is a j such that $|f(y) - f(x) - d(y - x)| \leq |y - x|/(k + 1)$ for every y within $1/(j + 1)$ of x.

The first notion is Bishop's [38, 14]: one j serves every pair of points. Continuity of g follows.

Lemma 12.81 (A uniform derivative is uniformly continuous)**.** *A uniform derivative g of f on $[0,1]$ is uniformly continuous on $[0,1]$. No omniscience principle and no choice principle is used.*

Proof. For points x, y, w expanding the products gives

$$\big(g(y)-g(x)\big)(w-y) = \big[f(w)-f(x)-g(x)(w-x)\big] - \big[f(w)-f(y)-g(y)(w-y)\big] \\ - \big[f(y)-f(x)-g(x)(y-x)\big].$$

Fix k, let J be the number that the derivative condition supplies at $4k+3$, put $\varepsilon = 1/(4k+4)$ and $t = 1/(4J+4)$, and let x, y in $[0,1]$ be within t. Cotransitivity at $3/8 < 1/2$ gives $y < 1/2$ or $3/8 < y$; put $w = y + t$ in the first case and $w = y - t$ in the second. Since $t \le 1/4$, w lies in $[0,1]$, $|w-y| = t$ and $|w-x| \le 2t \le 1/(J+1)$, so each bracket is at most ε times the distance of its pair: $|g(y)-g(x)|\,t \le \varepsilon(2t+t+t)$. Multiplying by the ratio $1/t$ gives $|g(y)-g(x)| \le 4\varepsilon = 1/(k+1)$. □

Proposition 12.82 (Calculus without principles)**.** *Let g be a uniform derivative of f on $[0,1]$, and let x, y lie in $[0,1]$.*

1. *(The mean value inequality.) If $|g(z)| \le C$ for every z in $[0,1]$, then $|f(y)-f(x)| \le C|y-x|$.*
2. *If $g(z) \ge 0$ for every z in $[0,1]$ and $x \le y$, then $f(x) \le f(y)$. If $g(z) = 0$ for every z in $[0,1]$, then $f(x) = f(y)$.*
3. *If $g(z) > 0$ for every z in $[0,1]$ and $x < y$, then $f(x) < f(y)$.*

No omniscience principle and no choice principle is used.

Proof. For parts 1 and 2 fix k, take j from the derivative condition at k, and put $z_i = x + i(y-x)/(j+1)$ for $i \le j+1$; neighbours are within $1/(j+1)$. Adding the $j+1$ step estimates gives

$$\Big|f(y)-f(x)-\frac{y-x}{j+1}\sum_{i\le j} g(z_i)\Big| \le \frac{|y-x|}{k+1}.$$

In part 1 the sum has absolute value at most $C|y-x|$, so $|f(y)-f(x)| \le C|y-x| + 1/(k+1)$ for every k, and a point at most $1/(k+1)$ above b for every k lies at or below b (a ratio strictly between them is overtaken for large k, as in Lemma 12.52, item 2). In part 2 the sum and $y-x$ are nonnegative, so $f(x)-f(y) \le 1/(k+1)$ for every k; the second assertion is part 1 with $C = 0$. The estimate uses g only between x and y, so part 2 holds on every subinterval on which its hypothesis holds.

For part 3 let z be the midpoint of $x < y$, and take a ratio μ with $0 < \mu < g(z)$ (Lemma 12.52, item 1). By item 3 of that lemma and Lemma 12.81, some j makes $g > \mu$ at the points of $[0,1]$ within $1/(j+1)$ of z. Take a positive ratio $w \le 1/(j+1)$ below half of a ratio lying between 0 and $y-x$; the window $[z-w, z+w]$ lies in $[x,y]$. Part 2 on the window, applied to $f(u) - \mu u$ with derivative $g - \mu \ge 0$, gives $f(z+w) - f(z-w) \ge 2\mu w$, and part 2 on $[x, z-w]$ and $[z+w, y]$ gives $f(x) \le f(z-w)$ and $f(z+w) \le f(y)$. So $f(y) - f(x) \ge 2\mu w > 0$. □

The proof of part 3 seeks a positive lower bound of g only near one point. A bound on all of $[0,1]$ is a compactness statement: Julian and Richman [37] constructed, in Bishop's constructive mathematics with a consequence of Church's thesis, a uniformly continuous function on $[0,1]$ that is positive at every point and has infimum 0.

The fundamental theorem and integrals

Definition 12.83 (Riemann sums and integrals)**.** The *Riemann sum* of a function g on the points at depth m is $R_m(g) = \sum_{i<2^m} g\big(i/2^m\big)\,2^{-m}$. A point I is an *integral of g on* $[0,1]$ when the sums $R_m(g)$ converge to I.

Limits are unique (Theorem 12.26), so g has at most one integral on $[0,1]$. When an antiderivative is known, the integral is not found by a limit at all.

Theorem 12.84 (The fundamental theorem of calculus is free)**.** *Let g be a uniform derivative of f on $[0,1]$.*

1. *$f(1) - f(0)$ is an integral of g on $[0,1]$.*
2. *For every natural number n, $1/(n+1)$ is an integral of $x \mapsto x^n$ on $[0,1]$.*
3. *(Integration by parts.) If u', v' are uniform derivatives of u, v on $[0,1]$ and J is an integral of $u'v$ on $[0,1]$, then $u(1)v(1) - u(0)v(0) - J$ is an integral of uv' on $[0,1]$.*
4. *(Substitution.) If f is a uniform derivative of F and h' one of h on $[0,1]$, and h maps $[0,1]$ into $[0,1]$, then $F(h(1)) - F(h(0))$ is an integral of $(f \circ h)\,h'$ on $[0,1]$.*

No omniscience principle and no choice principle is used.

Proof. 1. Fix k, take j from the derivative condition at k, and let $m \geq j+1$ and $x_i = i/2^m$. Since $2^m > j+1$, neighbours are within $1/(j+1)$, and telescoping gives

$$\big(f(1) - f(0)\big) - R_m(g) = \sum_{i<2^m} \Big[f(x_{i+1}) - f(x_i) - g(x_i)\,2^{-m} \Big],$$

a sum of 2^m brackets each at most $2^{-m}/(k+1)$ in absolute value. So the sums reach the tolerance $1/(k+1)$ at the stated depth $j+1$.

2. The function $x^{n+1}/(n+1)$ has the uniform derivative x^n, since $y^{n+1} - x^{n+1} - (n+1)x^n(y-x) = (y-x)^2 \sum_{i\leq n} \sum_{l<i} x^{n-1-l}y^l$, a double sum of $n(n+1)/2$ terms in $[0,1]$. Part 1 applies.

3 and 4. A function uniformly continuous on $[0,1]$ is bounded there (Lemma 12.57 and Proposition 12.19, item 4); with Lemma 12.81 and part 1 of Proposition 12.82, the functions u, v, u', v', h', f are bounded and u, v, h have Lipschitz bounds. The product rule follows from the splitting $(uv)(y) - (uv)(x) - (u'v + uv')(x)(y-x) = u(y)\Delta_v + v(x)\Delta_u + (u(y) - u(x))v'(x)(y-x)$, where $\Delta_u = u(y) - u(x) - u'(x)(y-x)$ and Δ_v likewise, and the chain rule from $F(h(y)) - F(h(x)) - f(h(x))h'(x)(y-x) = \Delta_F + f(h(x))\Delta_h$, with Δ_F taken between the points $h(x), h(y)$ of $[0,1]$. So $u'v + uv'$ is a uniform derivative of uv and $(f \circ h)h'$ one of $F \circ h$. Part 1 applies to both; for parts, Riemann sums are additive and limits subtract. $\square$

Without an antiderivative an integral is a limit, and our proof takes it with countable choice. One dyadic grid serves every interval. For points a, b and a natural number A with $|a|, |b| \le A$, let $\kappa_i(x) = \max\big(i/2^n, \min((i+1)/2^n, x)\big)$ be the clamp to the cell $[i/2^n, (i+1)/2^n]$ and

$$S_n^A(f; a, b) = \sum_{i < 2A \cdot 2^n} f\big(i/2^n - A\big)\,\big(\kappa_i(b + A) - \kappa_i(a + A)\big).$$

The weight of a cell is the oriented length of its part of the shifted interval, so a and b need not be compared. A point I is an *integral of f from a to b* when the sums $S_n^A(f; a, b)$ converge to I for every natural number A with $|a|, |b| \le A$; for $a = 0$ and $b = 1$ they are the sums $R_n(f)$. A function is *uniformly continuous on the line* when for every k some j makes points within $1/(j+1)$ have values within $1/(k+1)$. It is *continuous at* a point b when for every k some j makes the points within $1/(j+1)$ of b have values within $1/(k+1)$ of $f(b)$.

Proposition 12.85 (Integrals over intervals)**.** *Let f be a function on the points of $\mathbb{R}_\delta$.*

1. *Assume AC_ω. If f is uniformly continuous on the line, it has an integral from a to b for all points a, b. Countable choice is used once, for the limit, at pairs of ratios (Corollary 12.35); no modulus is selected, and no omniscience principle is used.*
2. *If I_1, I_2, I_3 are integrals of f from a to b, from b to c and from a to c, then $I_1 + I_2 = I_3$.*
3. *If $F(x)$ is an integral of f from a to x for every point x, and f is continuous at b, then F has derivative $f(b)$ at b.*

Parts 2 and 3 use no omniscience principle and no choice principle.

Proof. 2. With A bounding $|a|, |b|, |c|$ the weights of each cell add, so $S_n^A(f; a, b) + S_n^A(f; b, c) = S_n^A(f; a, c)$ at every depth; limits add.

3. Given $\varepsilon = 1/(k+1)$, take j with $|f(y) - f(b)| \le \varepsilon$ for y within $1/(j+1)$ of b, let x be within $1/(2j+2)$ of b, and let $2^{-n} \le 1/(2j+2)$. By part 2, $S_n^A(f; a, x) - S_n^A(f; a, b) = S_n^A(f; b, x)$, whose weights add up to $x - b$ and have absolute values adding up to $|x - b|$; a cell of nonzero weight has its left end within $1/(j+1)$ of b. These are equations and inequalities proved by cases on the order of $b + A$, $x + A$ and the ends of the cells (Remark 12.20). So $|S_n^A(f; b, x) - f(b)(x - b)| \le \varepsilon |x - b|$, and in the limit $|F(x) - F(b) - f(b)(x - b)| \le \varepsilon |x - b|$.

1. From depth n to $n + e$ each cell splits into cells whose weights add up to its weight, and left ends move by at most 2^{-n}; so $|S_n^A - S_{n+e}^A| \le \varepsilon |b - a|$ when f varies by at most ε over 2^{-n}. Given k, uniform continuity at $\varepsilon = 1/((2A+1)(2k+2))$ supplies a j, and from $n = j + 1$ on the sums lie within $\varepsilon |b - a| \le 1/(2k+2)$ of S_n^A, hence within $1/(k+1)$ of one another. So the sums form a Cauchy sequence of points; nothing is selected, since the Cauchy condition asks only for some index at each k. Corollary 12.35, with one countable choice at pairs of ratios, gives the limit. The sums do not depend on A: added cells have weight zero, and shared cells are translates. □

The mean value theorems

Theorem 12.86 (An approximate mean value point costs nothing)**.** *Let g be a uniform derivative of f on $[0,1]$. For every k some point ξ of $[0,1]$ has $g(\xi)$ within $1/(k+1)$ of $f(1)-f(0)$. No omniscience principle and no choice principle is used.*

Proof. Write $c = f(1)-f(0)$ and $\varepsilon = 1/(2k+2)$. Let J be the larger of the numbers that uniform continuity of g (Lemma 12.81) and the derivative condition supply at $2k+1$, and let $x_i = i/(J+1)$ and $h = 1/(J+1)$. Then $g(x_i)$, $g(x_{i+1})$ are within ε and $|f(x_{i+1}) - f(x_i) - g(x_i)h| \le \varepsilon h$. For each $i \le J$, cotransitivity at $c-2\varepsilon < c-\varepsilon$ and at $c+\varepsilon < c+2\varepsilon$ gives one of: (L) $g(x_i) < c-\varepsilon$; (H) $g(x_i) > c+\varepsilon$; (M) $c-2\varepsilon < g(x_i) < c+2\varepsilon$. A finite search finds an i in case (M), which serves as ξ, or labels every i with (L) or (H). The second outcome is impossible. Neighbours labelled (L) and (H) would have values more than 2ε apart, so all labels agree. If all are (L), telescoping gives

$$c = \sum_{i\le J}\big(f(x_{i+1}) - f(x_i)\big) \le \sum_{i\le J}\big(g(x_i)+\varepsilon\big)h < \sum_{i\le J} c\,h = c,$$

and if all are (H) the reversed estimate gives $c > c$. □

Theorem 12.87 (Rolle's theorem and the mean value theorems cost LLPO)**.** *Consider the following assertions, each for all functions f, g on the points with uniform derivatives f', g' on $[0,1]$:*

(A) (Rolle.) If $f(0) = f(1)$, some c in $[0,1]$ has $f'(c) = 0$.

(B) (The mean value theorem.) Some c in $[0,1]$ has $f'(c) = f(1) - f(0)$.

(C) (Cauchy's form.) Some c in $[0,1]$ has $\big(g(1)-g(0)\big)f'(c) = \big(f(1)-f(0)\big)g'(c)$.

1. *Each of them implies LLPO. The proof uses no choice principle.*
2. *Assuming AC_ω, LLPO implies all three. Countable choice is used through Theorems 12.60 and 12.54.*
3. *The forms with $0 < c < 1$ imply these, and so they imply LLPO.*

Proof. Assertion (C) with $g(x) = x$ is (B), and (B) with $f(0) = f(1)$ is (A); so for the first statement it suffices that (A) implies LLPO. Let a be a point, $A = \max(a,0)$, $B = \min(a,0)$, $h_1(x) = x(1-x)(2-x)$ and $h_2(x) = x(1-x)(1+x)$. Since

$$h_1(y) - h_1(x) - (2 - 6x + 3x^2)(y-x) = (y-x)^2(y+2x-3),$$
$$h_2(y) - h_2(x) - (1 - 3x^2)(y-x) = -(y-x)^2(y+2x),$$

with both last factors in $[-3,3]$ on $[0,1]$, the function $f = Ah_1 + Bh_2$ has the uniform derivative $f' = Ah_1' + Bh_2'$, and $f(0) = f(1) = 0$. Let $f'(c) = 0$ with c in $[0,1]$. The factorizations

$$h_1'(c) + \tfrac{37}{400} = 3\big(c - \tfrac{9}{20}\big)\big(c - \tfrac{31}{20}\big), \qquad h_2'(c) - \tfrac{37}{400} = 3\big(\tfrac{11}{20} - c\big)\big(\tfrac{11}{20} + c\big)$$

give $h_1'(c) \le -37/400$ for $9/20 \le c \le 1$ and $h_2'(c) \ge 37/400$ for $0 \le c \le 11/20$. Cotransitivity at $9/20 < 11/20$ gives $9/20 < c$ or $c < 11/20$. In the first case $a > 0$ would give $A = a$, $B = 0$ and $f'(c) = ah_1'(c) < 0$; so $a \le 0$. In the second case $a < 0$ would give $f'(c) = ah_2'(c) < 0$; so $a \ge 0$. This is LLPO by Corollary 12.18, and c was compared once.

For the second statement let $f(0) = f(1)$. By Lemma 12.81 and Theorem 12.60, applied to f' and to $-f'$, there are c_1, c_2 in $[0,1]$ with $f'(c_1) \le f'(y) \le f'(c_2)$ for every y in $[0,1]$. Part 2 of Proposition 12.82, applied to $f(x) - f'(c_1)x$ and to $f'(c_2)x - f(x)$ at $0 \le 1$, gives $f'(c_1) \le 0 \le f'(c_2)$. Theorem 12.54, applied to $t \mapsto f'\big(c_1 + t(c_2 - c_1)\big)$, uniformly continuous since $|c_2 - c_1| \le 1$, gives a root t, and $c = c_1 + t(c_2 - c_1)$. This is (A). Applied to $f(x) - \big(f(1) - f(0)\big)x$ it gives (B), and applied to $\big(g(1) - g(0)\big)f - \big(f(1) - f(0)\big)g$ it gives (C). The third statement holds because $0 < c < 1$ gives $0 \le c \le 1$. □

For the open-interval forms no converse is proved: an extremal point of f' may sit at an end of the interval, and moving it inside is not a decision that LLPO supplies by this route.

The exponential function

Proposition 12.88 (The exponential function is free)**.** *For every point x the partial sums $E_N(x) = \sum_{m<N} x^m/m!$ converge. Their limit $\exp(x)$ defines a function on the points, and*

1. $\exp(0) = 1$ *and* $\exp(x + y) = \exp(x)\exp(y)$ *for all points x, y;*
2. $|\exp(h) - 1 - h| \le |h|^2$ *whenever* $|h| \le 1$*;*
3. $\exp$ *has derivative* $\exp(x)$ *at every point x, and it is continuous at every point; on* $[0,1]$*, and on every interval* $[a,b]$*, it is a uniform derivative of itself;*
4. $1 + x \le \exp(x)$ *for* $x \ge 0$*, so for every point M some point N has* $M \le \exp(x)$ *whenever* $x \ge N$*.*

No omniscience principle and no choice principle is used.

Proof. Take a natural number $V \ge 1$ with $|x| \le V$. For $m + 1 \ge 2V$ the term $V^{m+1}/(m+1)!$ is at most half of $V^m/m!$, and $V^N/N! \le V^{2V}2^{-(N-2V)} = (2V)^{2V}2^{-N}$; so $|E_{N+d}(x) - E_N(x)| \le 2(2V)^{2V}2^{-N}$ for $N \ge 2V$. The points $E_N(x)$ are computed from one regular sequence naming x by the ring operations, so at the indices given by this modulus they form a Cauchy family whose diagonal names the limit (Theorem 12.25). Another sequence naming x gives the same points and so the same limit: exp is a function on the points, and nothing is chosen.

1. By the binomial theorem $E_N(x)E_N(y) - E_N(x+y)$ is the sum of the terms $x^i y^j/(i!\,j!)$ with $i, j < N \le i + j$, at most a tail of the series at $|x| + |y|$ in absolute value; so the limits agree. 2. For $N \ge 2$, $|E_N(h) - 1 - h| \le |h|^2 \sum_{2 \le m < N} 1/m! \le |h|^2$, and the bound passes to the limit (Lemma 12.52, item 2). 3. By part 1, $\exp(y) - \exp(x) - \exp(x)(y - x) = \exp(x)\big(\exp(y - x) - 1 - (y - x)\big)$, which part 2 bounds by $A|y - x|^2 \le |y - x|/(k + 1)$ when $A \ge |\exp(x)|$ is a natural number and $|y - x| \le 1/((A + 1)(k + 1))$; a derivative at a point gives continuity there. On an interval $[a,b]$ take a natural number $V \ge 1$ with $|a| \le V$ and $|b| \le V$. For $|x| \le V$ every term satisfies $V^m/m! \le (2V)^{2V}2^{-m}$ (for $m \le 2V$ because $V^m \le V^{2V}$, and beyond by the halving above), so

$|E_N(x)| < 2(2V)^{2V}$ and $|\exp(x)| \le 2(2V)^{2V}$. The one number $A = 2(2V)^{2V}$ therefore serves every x of $[a,b]$ at once, and j with $j+1 \ge (A+1)(k+1)$ is a modulus for the whole interval: this is the condition of Definition 12.80(1) on $[a,b]$ in place of $[0,1]$. 4. For $x \ge 0$ the terms of $E_N(x) - 1 - x$ are nonnegative; and $x \ge |M|$ gives $M \le x < 1 + x \le \exp(x)$. □

Three more limits

Theorem 12.89 (Monotone limits, open-interval roots and images)**.** *Consider the assertions:*

(i) every nondecreasing sequence (x_n) of points tends to infinity (for every point M some N has $M \le x_n$ for every $n \ge N$) or converges;

(ii) for every f uniformly continuous on $[0,1]$ and every point v with $f(0) < v < f(1)$, some c with $0 < c < 1$ has $f(c) = v$;

(iii) for every f uniformly continuous on $[0,1]$ there are points m, M with $m \le f(y) \le M$ for every y in $[0,1]$, such that every v with $m \le v \le M$ is a value $f(x)$ at a point x of $[0,1]$.

Assertion (i) implies LPO, and (ii) and (iii) each imply LLPO, with no choice principle. Conversely, LPO and AC_ω give (i), and LLPO and AC_ω give (ii) and (iii).

Proof. (i) A sequence bounded by b does not tend to infinity, since $M = b + 1$ would give $b + 1 \le x_n \le b$; so (i) gives monotone convergence, hence LPO (Theorem 12.45). Conversely, fix regular sequences naming the terms, by countable choice. For each natural number M, Lemma 12.42(a) gives a term above M or the bound $x_n \le M$ for all n; countable choice collects these decisions into a binary sequence, and LPO applied to it gives either a term above every natural number, and then the sequence, being nondecreasing, tends to infinity, or a natural bound, and then Theorem 12.45 gives a limit.

(ii) For a point a let $f(x) = \min\bigl(x - \frac{1}{3}, 0\bigr) + \max\bigl(x - \frac{2}{3}, 0\bigr) + a\,x(1-x)$, uniformly continuous on $[0,1]$ with $f(0) = -\frac{1}{3} < 0 < \frac{1}{3} = f(1)$. Let $0 < c < 1$ with $f(c) = 0$, so that $c(1-c) > 0$. Cotransitivity gives $\frac{1}{3} < c$, where the first two terms add up to at least 0, so that $a > 0$ would make $f(c)$ positive, or $c < \frac{2}{3}$, where they add up to at most 0, so that $a < 0$ would make $f(c)$ negative. So $a \le 0$ or $a \ge 0$, which is LLPO. Conversely, Theorem 12.54 applied to $f - v$ gives c in $[0,1]$ with $f(c) = v$. Uniform continuity gives a j such that points of $[0,1]$ within $1/(j+1)$ of 0 have values below v, and cotransitivity at $1/(2j+2) < 1/(j+1)$ gives $c > 1/(2j+2)$, the other case being excluded. In the same way $c < 1$.

(iii) A point x with $f(x) = M$ is a maximum point, so Theorem 12.60 gives LLPO. Conversely, that theorem applied to f and $-f$ gives c_1, c_2 with $m = f(c_1) \le f(y) \le f(c_2) = M$, and Theorem 12.54 applied to $t \mapsto f\bigl(c_1 + t(c_2 - c_1)\bigr) - v$ gives the point for v. □

Finite cores of two limit theorems

A sequence is Cauchy when, for every k, its terms from some index on lie within $1/(k+1)$ of one another. Its no-counterexample form [55] asks only that for every k and every proposed spacing g some window $[N, N + g(N)]$ be stable, its terms within $1/(k+1)$ of one another. This is *metastability*; for monotone sequences it comes with a bound on where to look [56].

Theorem 12.90 (Metastability with an explicit rate)**.** *Let (x_n) be nondecreasing with $x_n \le x_0 + K$ for every n, K a natural number. Let k be a natural number and g a function from the natural numbers to themselves, and put $N_0 = 0$, $N_{i+1} = N_i + g(N_i)$. Then some $i \le K(2k+2)$ has x_m, x_n within $1/(k+1)$ for all m, n in $[N_i, N_i + g(N_i)]$. Consequently:*

1. *every bounded nondecreasing sequence is* metastable*: for every k and g some N has the terms with indices from N to $N + g(N)$ within $1/(k+1)$ of one another;*
2. *for every sequence bounded above and all k, g some N has $x_m \le \max(x_0, \dots, x_N) + 1/(k+1)$ for every $m \le N + g(N)$.*

None of this uses an omniscience principle or a choice principle. The assertion that every bounded nondecreasing metastable sequence is Cauchy implies LPO, with no choice principle, and follows from LPO and AC_ω.

Proof. Let $d_i = x_{N_{i+1}} - x_{N_i} \ge 0$. Cotransitivity gives $d_i < 1/(k+1)$ or $d_i > 1/(2k+2)$, and a search over $i \le K(2k+2)$ finds a first i of the first kind or shows that all are of the second. In the second case $x_{N_{K(2k+2)+1}} - x_0 = \sum_{i \le K(2k+2)} d_i > K$, against the bound. In the first, for m, n in $[N_i, N_{i+1}]$ monotonicity gives $x_{N_i} \le x_m, x_n \le x_{N_{i+1}} \le x_{N_i} + 1/(k+1)$. For item 1 an upper bound lies below $x_0 + K$ for a natural number K; for item 2 apply item 1 to the running maxima (Proposition 12.37). By item 1 the last assertion makes every bounded nondecreasing sequence Cauchy, which gives LPO by Theorem 12.46; conversely, LPO and AC_ω give a limit (Theorem 12.45). □

The bound $K(2k+2)$ does not depend on g. Sequential compactness (Theorem 12.47) has a finite core of the same kind.

Proposition 12.91 (A finite Bolzano–Weierstrass theorem)**.** *Let (x_n) be a sequence of points of $[0,1]$, and let k, M be natural numbers. Some indices $n_0 < n_1 < \cdots < n_M \le 3^{k+1}M$ have terms pairwise within $1/(k+1)$. No omniscience principle and no choice principle is used. If the terms are given by regular sequences, a program finds indices $n_0 < n_1 < \cdots < n_M \le (3^{k+2}+1)M$ with terms pairwise within $1/(k+1)$, by reading one stage of each of the first $(3^{k+2}+1)M + 1$ terms.*

Proof. Put $m = k+1$ and cover $[0,1]$ by the windows $W_q = [q/3^m, (q+2)/3^m]$, $q < 3^m$, of width $2/3^m \le 1/(k+1)$. A point y of $[0,1]$ is placed in a window by going through q with cotransitivity at $(q+1)/3^m < (q+2)/3^m$: the first q with $y < (q+2)/3^m$, or the last window, contains y. By induction each of the $3^m M + 1$ indices $0, \dots, 3^m M$ is assigned a window. There are 3^m windows, so one of them holds $M+1$ of these indices, and their terms are pairwise within $1/(k+1)$ by Lemma 12.41(a).

For the program put $D = 3^{k+2}$, and let r_n be stage $D-1$ of the regular sequence of x_n, so that $\eta(r_n)$ is within $1/D$ of x_n and $-1/D \le r_n \le 1 + 1/D$. Label n by the least integer $q \ge -1$ with $r_n \le (q+2)/D$, a comparison of ratios; then $q \le D-1$ and $q/D \le r_n \le (q+2)/D$. Among the $(D+1)M+1$ indices $n \le (D+1)M$ some $M+1$ share one of the $D+1$ labels. Their stages lie within $2/D$ of one another, so their terms lie within $4/D \le 1/(k+1)$, since $4(k+1) \le 3^{k+2}$. □

The infinite pigeonhole step, some window holds infinitely many terms, costs LPO (Theorem 12.75).

A unique root costs nothing

What LLPO pays for in Theorem 12.54 is a decision between roots. With a modulus of uniqueness, approximate roots converge.

Theorem 12.92 (Roots from a modulus of uniqueness)**.** *Let F be uniformly continuous on $[0,1]$, with a* modulus of uniqueness *μ: any two points of $[0,1]$ at which F is within $1/(\mu(k)+1)$ of 0 are within $1/(k+1)$. Let regular sequences $\xi_0, \xi_1, \ldots$ name points of $[0,1]$, with F within $1/(k+1)$ of 0 at the point of ξ_k. Put $\nu(0) = \mu(0)$ and $\nu(k+1) = \nu(k) + \mu(k+1) + 1$. Then the diagonal of the family $\xi_{\nu(0)}, \xi_{\nu(1)}, \ldots$ names a root of F in $[0,1]$. If moreover $F(0) \le 0 \le F(1)$, F has a modulus of uniform continuity given as a function, and regular sequences name the values of F at the points $j/3^n$, then the root is computed from these data. No omniscience principle and no choice principle is used.*

Proof. ν is nondecreasing with $\nu(k) \ge \mu(k)$ and $\nu(k) \ge k$. For $m, n \ge k$ the values of F at the points of $\xi_{\nu(m)}$ and $\xi_{\nu(n)}$ are within $1/(\mu(k)+1)$ of 0, so these points are within $1/(k+1)$: a Cauchy family. Theorem 12.25 gives its limit x, which lies in $[0,1]$ (Lemma 12.52, item 2). Given k, take j from uniform continuity at $2k+1$ and $m = \max(j, 2k+1)$: $F(x)$ is within $1/(2k+2)$ of F at the point of $\xi_{\nu(m)}$, which is within $1/(2k+2)$ of 0. So $F(x) = 0$ (Corollary 12.24).

In the presented case let ω be the given modulus, fix k, and take n with $3^{-n} \le 1/(\omega(2k+1)+1)$. Read stage $4k+3$ of the given sequence of each value $F(j/3^n)$, $j \le 3^n$. A stage above $-3/(4k+4)$ gives $F(j/3^n) > -1/(k+1)$, and a stage at or below it gives $F(j/3^n) \le -1/(2k+2)$. With these comparisons of ratios in place of cotransitivity, the search of Theorem 12.55 on the grid $j/3^n$ returns a grid ratio at which F is within $1/(k+1)$ of 0; its constant sequence serves as ξ_k. □

With Theorem 12.55 and one countable choice of regular sequences, every uniformly continuous F with $F(0) \le 0 \le F(1)$ and a modulus of uniqueness therefore has a root, at no omniscience price, as in the strictly increasing case of Theorem 12.56.

Banach's fixed point theorem

Theorem 12.93 (Banach's fixed point theorem needs no omniscience)**.** *Let T map $[0,1]$ into itself with $|Tx - Ty| \le r|x - y|$ for x, y in $[0,1]$, where $r = p/(q+1)$ for natural numbers $p \le q$. Write T^n for the nth iterate and $N(k) = (q+1)(k+1)$.*

1. *For x in $[0,1]$ and $m \le m'$, $|T^m x - T^{m'} x| \le r^m$; also $r^{N(k)} \le 1/(k+1)$, and T has at most one fixed point in $[0,1]$.*
2. *If a map $\hat{T}$ of regular sequences gives T, in that $\hat{T}s$ names T of the point of s, then for every regular s naming a point of $[0,1]$ the diagonal of the sequences $\hat{T}^{N(k)}s$ names a fixed point of T.*
3. *If regular sequences name the points $x, Tx, T^2x, \ldots$ of one orbit, x in $[0,1]$, then T has a fixed point in $[0,1]$.*

None of this uses an omniscience principle or a choice principle. Assuming AC_ω, every such T has a fixed point in $[0,1]$; countable choice is used once, for the representatives of one orbit.

Proof. 1. By induction $|T^n u - T^n v| \le r^n |u - v|$ on $[0,1]$; take $u = x$ and $v = T^{m'-m}x$, which are within 1. If $p = 0$, then $r^{N(k)} = 0$. If $p \ge 1$, then $q \ge 1$ and $r \le q/(q+1)$, and Bernoulli's inequality gives $(1+1/q)^{q+1} \ge 1+(q+1)/q \ge 2$, so $r^{q+1} \le 1/2$ and $r^{N(k)} \le 2^{-(k+1)} \le 1/(k+1)$. Fixed points a, b have $|a-b| \le r|a-b|$ with $r < 1$, so $a = b$.

2. By part 1 the points $T^{N(k)}x$ of the sequences $\hat{T}^{N(k)}s$ form a Cauchy family. Theorem 12.25 gives a point L of $[0,1]$ with $T^{N(k)}x$ within $1/(k+1)$ of L, and

$$|TL - L| \le |TL - T^{N(k)+1}x| + |T^{N(k)+1}x - T^{N(k)}x| + |T^{N(k)}x - L| \le \frac{3}{k+1}$$

for every k, so $TL = L$. 3. The same, with the given sequences. Countable choice, applied once, supplies them. □

The theorem is Banach's [54], read on $\mathbb{R}_\delta$. The choice in part 3 concerns one orbit only. Countable choice at pairs of ratios serves equally: by part 1 the orbit $x, Tx, T^2x, \ldots$ is a Cauchy sequence of points, Corollary 12.35 gives its limit with one such choice, and the estimate of part 2, taken at a large enough index, makes the limit a fixed point. Without unique choice, excluded middle with search-choice does not replace countable choice here, already for maps that send points within $1/(k+1)$ to values within $1/(4k+4)$, as every map with $r \le 1/4$ does.

Theorem 12.94 (Without unique choice, excluded middle does not give Banach's fixed point)**.** *In the theory of Theorem 12.63, with functions on the points written as formulas, the following is not provable: every function on the points that maps* $[0,1]$ *into itself and, for every* k*, sends points of* $[0,1]$ *within* $1/(k+1)$ *to values within* $1/(4k+4)$ *has a fixed point in* $[0,1]$*. The proof reasons in a classical metatheory.*

Proof. In the model of computable sequences, with H, G and the point c of the proof of Theorem 12.63, put $T(x) = H\big(G(x)/8\big)$. The slope bounds $4/3$ for H and $3/2$ for G give $|T(x) - T(y)| \le \frac{1}{4}|x - y|$. For x in $[0,1]$, $G(0) \le G(x)/8 \le 1/8$, because $G(0) \le 0$ and $G(x) \le G(1) = 1$; since H is increasing with $H(G(0)) = 0$ and $H(1) = 1$, $T(x)$ lies in $[0,1]$. A fixed point has $G(x)/8 = G(T(x)) = G(x)$, so $G(x) = 0$ and $x = c$. A computable point differs from c, so $G(x)/8$ is a computable point other than 0, where H is affine with rational coefficients, as in that proof; so T sends computable points to computable points. Its graph is the formula $F_T(\beta, \beta')$: some δ, δ' satisfy $F_{15}(\beta, \delta)$, $F_{15}(\beta', \delta')$, and the stages of δ' are those of δ divided by 8. At F_T the hypotheses hold in the model, and the conclusion would give a computable point x of $[0,1]$ with $x = c$. The instance is false in the model, so it is not a theorem. □

The table

In Table 12.4, f' is a uniform derivative of f on $[0,1]$; in the rows about values and roots f is uniformly continuous on $[0,1]$; "continuous" and "injective" have the meanings of Theorem G.9.

What the classical proofs move along an interval by finite partitions is free. In this table the points that compactness supplies, a critical point, a mean value point, a value taken in the open interval and a point of the image, cost LLPO; a root with a modulus of uniqueness and the fixed point of a contraction cost nothing beyond countable choice.

Assertion on $\mathbb{R}_\delta$	Price	Countable choice
A uniform derivative is uniformly continuous; the mean value inequality; $f' \geq 0$, $f' = 0$, $f' > 0$ make f nondecreasing, constant, strictly increasing (Lemma 12.81, Proposition 12.82)	none	not used
$f(1) - f(0)$ is the integral of f'; the integral of x^n; parts; substitution (Theorem 12.84)	none	not used
A function uniformly continuous on the line has integrals over intervals (Proposition 12.85)	none	used once, for the limit, at pairs of ratios
they add over adjacent intervals, and $F'(b) = f(b)$ where f is continuous (Proposition 12.85)	none	not used
exp exists, $\exp' = \exp$, and $\exp(x + y) = \exp(x)\exp(y)$ (Proposition 12.88)	none	not used
An approximate mean value point (Theorem 12.86)	none	not used
Rolle's theorem, the mean value theorem and Cauchy's form, the point in $[0, 1]$ (Theorem 12.87)	LLPO	in the converse
A value strictly between $f(0)$ and $f(1)$ is taken in the open interval; the values of f fill the interval from a minimum to a maximum (Theorem 12.89)	LLPO	in the converse
A root from a modulus of uniqueness and given approximate roots (Theorem 12.92)	none	not used
A nondecreasing sequence tends to infinity or converges (Theorem 12.89)	LPO	in the converse
A bounded nondecreasing sequence is metastable: a stable window starts at one of $N_0, \ldots, N_{K(2k+2)}$ (Theorem 12.90)	none	not used
... and Cauchy (Theorem 12.90)	LPO	in the converse
$M + 1$ of the first $3^{k+1}M + 1$ terms of a sequence in $[0, 1]$ lie pairwise within $1/(k + 1)$ (Proposition 12.91)	none	not used
A continuous injective map of the line is strictly monotone (Theorem G.9)	$\mathrm{MP}_{\mathrm{bin}}$	not used
A contraction of $[0, 1]$ given on regular sequences has a fixed point (Theorem 12.93)	none	not used
A contraction of $[0, 1]$ has a fixed point (Theorems 12.93 and 12.94)	none	used once, for one orbit (or once at pairs of ratios, Corollary 12.35); excluded middle with search-choice does not replace it (Theorem 12.94)

Table 12.4. The calculus on the line built from ratios, read as Table 12.3: "Price" is the principle equivalent to the assertion, "none" means that the assertion is proved outright, from countable choice where the last column says it is used, and in every priced row the passage to the principle uses no choice. The open-interval forms of the mean value theorems imply LLPO; no converse is proved for them.

12.11 Integrals over the whole line and space, and what completeness costs

How much of the classical calculus of π survives on $\mathbb{R}_\delta$ with no principle at all? The Gaussian integral, the value of $\int |y|e^{-|y|}\,d^3y$, $\zeta(2)$, $\zeta(4)$ and $\zeta(6)$ do, each with the line's own π. What the classical proofs take from completeness is a different matter, and the last part of this section prices it: countable choice of natural numbers gives Cauchy completeness, while completeness, even with unique choice, does not give that choice back over intuitionistic second-order arithmetic.

Two conventions run through the section. The *quarter period* q is the point with $0 < q \le 2$, $\cos q = 0$ and $\sin q = 1$, for the circle functions of Definition 21.6; by Theorem 21.8, $4q = 2\pi_\delta$, and we write π for $2q$. The bound $q \le 2$ matters: $5q$ also has cos $= 0$ and sin $= 1$, since $4q$ is a period of both functions. An *improper integral* is an iterated limit of midpoint sums. On the line, v is the integral of f when the integrals of f over $[-L, L]$, in the sense of Proposition 12.85, come within any $1/(k+1)$ of v for all large natural L. In space, write $\Sigma_s^N(F)$ for the midpoint sum of F over the cube $[-s, s]^3$ cut into $(N+1)^3$ equal cells; W is the *space value* of F when for every k there is S_0 such that for every $s \ge S_0$ some N_0 gives $|\Sigma_s^N(F) - W| \le 1/(k+1)$ for all $N \ge N_0$.

The Gaussian integral, and integrals over space

Theorem 12.95 (The Gaussian integral). *For every point $p > 0$ and every point c,*

$$\int_{-\infty}^{\infty} e^{-py^2+cy}\,dy = \gamma e^{c^2/(4p)}, \qquad \int_{-\infty}^{\infty} y e^{-py^2+cy}\,dy = \frac{c}{2p}\,\gamma e^{c^2/(4p)},$$

where $\gamma > 0$ is the point with $\gamma^2 p = \pi$, as improper integrals on $\mathbb{R}_\delta$; and the improper integral is homogeneous: if v is the integral of f, then Cv is the integral of Cf. No omniscience principle and no choice principle is used.

Proof. The Gaussian has no elementary antiderivative on the line, but in the plane $e^{-|y|^2}$ is the divergence of the field $y(1 - e^{-|y|^2})/(2|y|^2)$, because $\int_0^\rho te^{-t^2}\,dt = (1 - e^{-\rho^2})/2$. The proof follows that field. (i) The midpoint sum of $e^{-u^2-v^2}$ over a square grid is exactly the square of the midpoint sum of e^{-x^2} over the side. (ii) A discrete divergence theorem on the square grid turns the square sum into sums over the four edges; the field is regularized near the origin, within 2ε of the integrand. (iii) On an edge the field is, up to the factor $(1 - e^{-s^2-w^2})/2$, which tends to $\frac{1}{2}$, the flux density $s/(s^2 + w^2)$ of a point source in the plane, whose edge sum scales exactly to $\int_{-1}^{1} dw/(1+w^2) = \arccos(-1/\sqrt{2}) - \arccos(1/\sqrt{2}) = \pi/2$, with the line's arccosine. With the factor $\frac{1}{2}$ each edge gives $\pi/4$, and four edges give π, so the line sums tend to $\sqrt{\pi}$. (iv) Midpoint grids scale exactly, which gives the factor $p^{-1/2}$. For every c, complete the square and move the centre: the derivative in the centre of $\sum e^{-p(y-c')^2}$ is minus a sum of y-derivatives, so it changes only by boundary terms of size $e^{-p(L-|c'|)^2}$, and the range of integration never moves. (v) The integral of Proposition 12.85 uses dyadic left sums; on 2^m cells they differ from midpoint sums by $O(2^{-m})$ for a Lipschitz integrand. The first moment follows from $y\Phi = (c/2p)\Phi - (1/2p)\Phi'$, with $\Phi = e^{-py^2+cy}$, and the fundamental theorem for

left sums. Every step is an estimate with explicit constants, and each limit has an explicit modulus. □

Theorem 12.96 (Radial integration and an elementary space value)**.** *Let F be a function on $[0,\infty)$ with $|F(r)| \le B$ and $|F(r')-F(r)| \le \Lambda|r'-r|$, and let H satisfy $H(0) = 0$, have the derivative $r^2F(r)$ with a uniform quadratic remainder on every $[0,R]$, and tend to H_∞. Then $F(|y|)$ has the space value $4\pi H_\infty$. In particular $\int_{\mathbb{R}^3} |y|e^{-|y|}\,d^3y = 24\pi$ as a space value. No omniscience principle and no choice principle is used.*

Proof. The structure is that of Theorem 12.95 one dimension up. Away from the origin $F(|y|)$ is the divergence of the radial field $\Phi(y) = yH(|y|)/|y|^3$, since $H' = r^2F$. (i) On the line $1/|y|$ has no value at $y = 0$, so the field is regularized with $\sqrt{|y|^2+\varepsilon^2}$ in place of $|y|$; its divergence differs from $F(|y|)$ by at most a constant times ε everywhere, by the Lipschitz bound on F. (ii) A discrete divergence theorem on the cubic grid: the box midpoint sum of a divergence equals the sum of the face midpoint sums of the field up to an error of order $1/N$, whenever the field has uniform quadratic remainders along coordinate lines; the remainders come from sums, products, the chain rule, $1/x$, $\sqrt{x}$ and e^{-x}, each with an explicit constant. (iii) On a face $y_0 = s$ the field is $sH(\rho)/\rho^3$; since $H(\rho)$ tends to H_∞, the face sum is H_∞ times the face flux of the inverse-square field $y/|y|^3$, up to a deficit that the tail bound on H controls. Midpoint grids scale exactly, so that flux does not depend on s: it is the midpoint sum of $(1+u^2+v^2)^{-3/2}$ over $[-1,1]^2$. (iv) That sum tends to $2\pi/3$, the solid angle of a face of the cube: the inner integral has the antiderivative $v/(A\sqrt{A+v^2})$ with $A = 1+u^2$, so the inner integral over $[-1,1]$ is $2/(A\sqrt{A+1})$, and the outer one has the antiderivative $-2\arccos\bigl(u/\sqrt{2+2u^2}\bigr)$, so the outer integral is $2\bigl(\arccos(-\frac{1}{2}) - \arccos(\frac{1}{2})\bigr) = 2\pi/3$, with the line's arccosine. The six faces give 4π. For the particular value take $F(r) = re^{-r}$, so that $H(r) = 6 - e^{-r}(r^3+3r^2+6r+6)$ and $H_\infty = 6$: the space value is $24\pi = 48q$. □

The zeta values with the line's π

Theorem 12.97 (Three zeta values)**.** *On $\mathbb{R}_\delta$, with no omniscience principle and no choice principle,*

$$\sum_{n\ge1}\frac{1}{n^2} = \frac{\pi^2}{6}, \qquad \sum_{n\ge1}\frac{1}{n^4} = \frac{\pi^4}{90}, \qquad \sum_{n\ge1}\frac{1}{n^6} = \frac{\pi^6}{945},$$

as limits of their partial sums. The third statement is false if $\pi = 2q$ is read with q any point with $q > 0$, $\cos q = 0$ and $\sin q = 1$.

Proof. The route is Hofbauer's doubling [209]. For $c = \csc^2 x$, the cosecants $a = \csc^2(x/2)$ and $b = \csc^2(q - x/2)$ satisfy $a + b = 4c$ and $ab = 4c$, by the double-angle formulas. Summing over the 2^n angles $(2k+1)q/2^{n+1}$ gives exact values: $\sum \csc^2 = 2\cdot 4^n$, and for the fourth and sixth powers, from $a^2+b^2 = 16c^2 - 8c$ and $a^3+b^3 = 64c^3 - 48c^2$, closed forms such as $15\sum \csc^6 = 64\cdot 64^n + 40\cdot 16^n + 16\cdot 4^n$. On $[0,2]$ the bounds $1 \le x^2\csc^2 x \le 1 + 8x^2$, raised to the needed power, put $\sum_{k<2^n}(2k+1)^{-2j}$ within $c_j 2^{-n}$ of its limit, which is $q^2/2$, $q^4/6$ and $q^6/15$ for $j = 1,2,3$. For the partial sums $Z(M)$ of $\sum n^{-2j}$, the even terms give $Z(2M) = O(M) + Z(M)/4^j$, where $O(M)$ is the sum over the odd $n \le 2M$; so along $M = 2^n$ the defect from the claimed

value is the odd defect plus a fixed fraction of the previous defect, it contracts to 0, and the monotone partial sums converge. The sine bounds use the series of Definition 21.6 on $[0, 2]$, which is why $q \le 2$ enters.

The refutation. If the value held for every q with $q > 0$, $\cos q = 0$ and $\sin q = 1$, it would hold at the quarter period q and at $5q$; one sequence would converge to $(2q)^6/945$ and to $(10q)^6/945$, and limits are unique, so $(10q)^6 = (2q)^6$, against $q > 0$. The same argument, with the exponents 2 and 4, refutes the first two statements under that reading. □

Two explicit remainders and a step that needs choice

Proposition 12.98 (Remainders with explicit constants)**.** *For points $|u| \le 1$ and $|z| \le 1$,*

$$\left|1 - \cos u - \frac{u^2}{2} + \frac{u^4}{24}\right| \le \frac{|u|^5}{50}, \qquad \left|\cosh z - 1 - \frac{z^2}{2} - \frac{z^4}{24}\right| \le \frac{|z|^6}{600}.$$

If Φ has the derivative f on $[a, b]$ with a uniform quadratic remainder of constant C, the midpoint sum of f with $N + 1$ cells differs from $\Phi(b) - \Phi(a)$ by at most $C(b - a)^2/(2N + 2)$; in particular $\int_0^\infty t^3 e^{-t}\, dt = 6$ as an iterated limit of midpoint sums. No omniscience principle and no choice principle is used.

Proof. The two remainders bound the tails of the alternating and of the positive series beyond the stated terms, with the terms compared to a geometric series on $|u| \le 1$. For the midpoint sums, each cell of width $2w$ contributes $\Phi(e_{n+1}) - \Phi(e_n) - 2w f(m_n)$, which the remainder bounds by $2Cw^2$ on each half; summing over the cells gives the bound. For $\int_0^R t^3 e^{-t}\, dt$ take $\Phi(t) = -e^{-t}(t^3 + 3t^2 + 6t + 6)$; the value on $[0, R]$ is $6 - e^{-R}(R^3 + 3R^2 + 6R + 6)$, which tends to 6. □

The mean value step from pointwise derivatives is different.

Proposition 12.99 (Taylor's step from pointwise derivatives)**.** *Assume countable choice of natural numbers. Let f have the derivative $g(t)$ at every point t with $|t| \le \delta$, in the pointwise sense of Definition 12.80, let $M \ge 0$ and $|g(t) - g(0)| \le M|t|$ there. Then $|f(h) - f(0) - hg(0)| \le Mh^2$ for $|h| \le \delta$. With a uniform derivative in place of the pointwise one the statement holds with no choice.*

Proof. With a uniform derivative this is the mean value inequality of Proposition 12.82. With pointwise derivatives the classical proof bisects: each halving keeps a half on which the increment exceeds its share, and the nested dyadic intervals close down on a point at which the pointwise derivative contradicts that excess. Each halving is a comparison decided only up to a tolerance, so its outcome is asserted to exist, not computed; the halvings form a sequence of natural numbers, and one application of countable choice of natural numbers collects them. The goal is an inequality, a negative statement, but that does not remove the choice: it would need a double-negation shift over the whole sequence. □

Completeness and countable choice

Write $\mathrm{AC}_{\mathbb{N}}$ for countable choice of natural numbers: if every n has some m with $R(n,m)$, there is a sequence f with $R(n, f(n))$ for every n. *Cauchy completeness* of the line says that every Cauchy sequence of points converges.

Theorem 12.100 (Completeness from countable choice of natural numbers, and not back)**.**

1. $\mathrm{AC}_{\mathbb{N}}$ *implies Cauchy completeness of* $\mathbb{R}_\delta$*, with no omniscience principle.*
2. *Over intuitionistic second-order arithmetic with unique choice, Cauchy completeness does not give* $\mathrm{AC}_{\mathbb{N}}$*: no derivation in HAS from finitely many hypotheses, each an instance of Cauchy completeness or of unique choice or the universal closure of one, derives the sentence*

$$\forall X\Big(\forall n\,\exists m\,J(n,m)\in X \to \exists S\big(\forall n\,\exists m\,J(n,m)\in S \wedge \mathrm{Fun}(S) \\ \wedge\ \forall n\,\forall m\,(J(n,m)\in S \to J(n,m)\in X)\big)\Big),$$

where $\mathrm{Fun}(S)$ *says that* $J(n,m)\in S$ *and* $J(n,m')\in S$ *give* $m=m'$*. The same holds for countable choice of pairs of ratios.*

In part 2 a point *is a set S of codes $J(n,c)$, total and single-valued in n, with c the code of a ratio $S(n)$ and $|S(m)-S(n)|\le 1/(m+1)+1/(n+1)$; points S and T are* equivalent *when for every k some N has $|S(n)-T(n)|\le 1/(k+1)$ for all $n\ge N$; and* $\mathrm{Near}_k(S,T)$ *says $|S(n)-T(n)|\le 1/(k+1)+2/(n+1)$ for every n. A formula $P(n,S)$, with a number variable n, a set variable S and any parameters,* names a sequence of points, $\mathrm{Seq}(P)$*, when for every n some point satisfies $P(n,\cdot)$, every set that satisfies it is a point, any two such are equivalent, and every point equivalent to one that satisfies it satisfies it. The instance of Cauchy completeness at P is*

$$\mathrm{Seq}(P)\wedge\forall k\,\exists N\,\forall m,n\ge N\,\forall S,T\big(P(m,S)\wedge P(n,T)\to \mathrm{Near}_k(S,T)\big) \\ \to \exists L\Big(L \text{ is a point}\wedge\forall k\,\exists N\,\forall m\ge N\,\forall S\big(P(m,S)\to\mathrm{Near}_k(S,L)\big)\Big).$$

Conditions and scope.

Base and premises

Part 1: $\mathbb{R}_\delta$ with $\mathrm{AC}_{\mathbb{N}}$ as a hypothesis. Part 2: HAS (Definition 7.14), with a sequence of points named by a formula $P(n,S)$ as displayed above; proved in classical mathematics with the open sets of the real line as truth values.

Conclusion

Completeness costs at most countable choice of natural numbers, applied twice per sequence, to the stages and to the Cauchy indices, and over HAS it does not give that choice back. So "completeness is priced at exactly countable choice" is refuted in its reverse direction.

Proof. (1) Let (x_n) be Cauchy: for every k there is an N with x_m, x_n within $1/(k+1)$ for $m,n\ge N$. Each x_n is a class of regular sequences, and each regular sequence is determined by its stages, which are ratios coded by natural numbers. Apply $\mathrm{AC}_{\mathbb{N}}$ once to the relation between $\langle n,j\rangle$

and the code of a ratio within $1/(j+1)$ of x_n, which holds for some code because x_n is a point, and once to the relation between k and a Cauchy index N for $1/(k+1)$. The chosen ratios, read along the chosen indices, form a regular sequence (Theorem 12.25), and its point is the limit. No representative of a point is chosen beyond these stages, which is why choice into the sequences themselves is not needed.

(2) Take the open sets of $\mathbb{R}$ as truth values (the sheaf model of Fourman and Hyland [59]). Every instance of unique choice has the value $\top$, being part of comprehension. Every instance of Cauchy completeness has the value $\top$ too. On the open set O where its hypotheses hold, read the m-th point at each $z \in O$ as a classical real, the limit of the stages of any member of its class at z. A set that codes a sequence has constant stages on every connected open set on which it does so, so this real is constant on the connected component I of a given point of O; the constants form a classical Cauchy sequence, and a regular sequence of codes for its classical limit, with values $\emptyset$ or the whole space, witnesses the conclusion on all of I. By Lemma 7.21 it suffices that the displayed sentence does not have the value $\top$. Let X have, at the code $J(n,m)$, the open set of points z such that m codes a ratio r with $|r-z| < 1/(n+1)$. Every n has such an m near every z, so the premise has the value $\top$. Suppose an open neighbourhood of 0 lay in the value of the conclusion, with a set S. On a ball around 0 of radius ε, S is total and single-valued, so each n-th value is locally constant, hence constant on the ball, which is connected: one ratio for all z in the ball. For $1/(n+1) < \varepsilon/2$ no ratio lies within $1/(n+1)$ of both $\varepsilon/2$ and $-\varepsilon/2$. For pairs of ratios the argument is the same, with a pair in place of one code. □

Theorem 12.101 (Box integrals and completeness)**.** *Let q be the quarter period, and let F have a modulus of uniform continuity on the zone $[-\pi,\pi]^3$.*

1. *Cauchy completeness implies that F has a zone average: the midpoint averages over $(N+1)^3$ cells converge, and the corner-grid averages converge to the same point. Hence $\mathrm{AC}_{\mathbb{N}}$ implies it.*
2. *If the modulus is given as a function, the zone average exists given only completeness for sequences with the modulus $k \mapsto k$.*
3. *If moreover F has period 2π in each coordinate and v is its zone average, the averages over the box's own modes $2\pi m/(N+1)$ converge to v, with no principle.*

Proof. The midpoint averages form a Cauchy sequence with a modulus computed from the modulus of continuity, with no principle: two grids are compared through a common refinement, cell by cell. Completeness supplies the limit. The corner part compares the corner grid with the midpoint grid cell by cell in the same way. In (2), a modulus function lets the Cauchy indices be computed rather than chosen, so completeness at the fixed modulus $k \mapsto k$ suffices. In (3), the mode grid of a periodic function is a shifted midpoint grid, and the shift changes each average by at most the modulus of continuity at the cell width. □

The fan theorem decides nothing

Theorem 12.102 (The fan theorem adds no omniscience)**.** *Write the decidable fan theorem as a sentence* FAN *of HAS: every decidable binary tree, given as a set of node codes, that every path*

leaves is finite. Then FAN *has the value* $\top$ *in the open sets of every topological space. Consequently HAS derives none of* FAN $\wedge$ MP $\to$ LLPO, FAN $\wedge$ LLPO $\to$ WLPO *and* FAN $\wedge$ WLPO $\to$ MP.

Proof. Fix a space and an open set on which the tree T is decidable and barred. At a point x of it, the nodes whose membership value contains x form a classical decidable tree T_x, since each membership value is open and closed on the open set. Every classical path is a decidable set in the model, with values $\emptyset$ or the whole space, so the bar hypothesis makes it leave T_x. By Kőnig's lemma in classical mathematics [15], T_x has no node of some length N. Decidability at x puts x in the complement value of each of the finitely many nodes of length N; the intersection V of these finitely many open sets is an open neighbourhood of x, and on V, by the tree property, no node of length at least N lies in the tree. So x lies in the value of the conclusion, with the bound N. So the value of FAN is $\top$. The non-derivations follow from Table 7.2 and Theorem 7.42: the one-point compactification of an uncountable set validates MP and not LLPO, the remainder $\beta\mathbb{N} \setminus \mathbb{N}$ validates LLPO and not WLPO, and $\beta\mathbb{N}$ validates WLPO and not MP, and FAN holds in all three. □

Theorem 13.6 makes Heine–Borel for $[0, 1]$ cost exactly the decidable fan theorem. Its direction from the fan theorem to Heine–Borel uses no choice principle and handles only ratios, finite strings and the given sequences, so it goes through in HAS with sets for sequences and covers. So Heine–Borel for $[0, 1]$ buys none of the omniscience principles over HAS either.

Proposition 12.103 (Confined motion for all time)**.** *Let a polynomial vector field on* $\mathbb{R}^n_\delta$ *with ratio coefficients have a polynomial first integral* F *that is coercive:* $|x|^2 \le a + bF(x)$ *for natural numbers* a *and* b*. Then every state* y *with ratio coordinates lies on a motion* $X : \mathbb{R}_\delta \to \mathbb{R}^n_\delta$ *of the field with* $X(0) = y$*, uniformly differentiable on bounded intervals. No omniscience principle and no choice principle is used.*

Proof. The local flow. Fix a box radius R. On the box of radius $R + 1$ the field and its first derivatives are bounded by ratios M and Λ read off its coefficients. Choose k with $2^{-k}M \le 1$ and $2^{-k}\Lambda \le 1$. The field rescaled by 2^{-k} is then a law given by a formula that meets the hypotheses of the Picard–Lindelöf theorem on the box, and from every state y with $|y_i| \le R$ its Euler polygons on $[0, 1]$, computed on regular sequences naming y, stay in the box of radius $R + 1$ and converge uniformly with a modulus fixed by M and Λ. Their limit is a motion through y of the rescaled field, that is, a motion of the field on $[0, 2^{-k}]$. The polygons are computed on the sequences themselves and the limit has an explicit modulus, so the flow Φ is a function of the state formed with no choice, as in part 2 of Theorem 12.93.

All time. Along a motion the first integral is constant, so a motion from y stays in the ball $|x|^2 \le a + bF(y)$; take a natural number R with $R^2 \ge a + bF(y)$, which exists because y has ratio coordinates. The same k serves at every state of that ball. Apply Φ at y, then at the end point, and so on, on the intervals $[j2^{-k}, (j + 1)2^{-k}]$, and in the other time direction with the field reversed; the pieces agree at their end points. They glue to a motion on the whole line, defined on representatives: at a ratio time the piece that contains it is found by comparing ratios, and at a shared end point the two pieces give the same state; along the motion the speed is bounded by a ratio read off M, so the glued map is uniformly continuous on ratio times, and at every time it is the limit of its values along a regular sequence naming that time.

On each piece's interval this limit is the piece itself, which is continuous. A bounded interval meets finitely many pieces, so the motion is uniformly differentiable there. □

Remark 12.104 (Open questions about completeness and motion)**.** Three statements are not settled here.

1. *The converse of Theorem 12.101(1).* Does the existence of zone averages for every F with a modulus of uniform continuity on the zone imply completeness for sequences of points with the modulus $k \mapsto k$? The evident encoding of a sequence into one uniformly continuous function needs the function's values near an accumulation point, which are limits of points without representatives.

2. *Taylor's step with no choice.* Does Proposition 12.99 hold without countable choice of natural numbers? Its exact price is not known.

3. *Completeness from choice, inside HAS.* Does HAS derive every instance of Cauchy completeness, written as in Theorem 12.100(2), from the displayed sentence of countable choice of natural numbers? On $\mathbb{R}_\delta$ the implication holds (Theorem 12.100(1)); inside HAS it would follow from two steps: that sentence gives countable choice of natural numbers for every formula, and each instance of Cauchy completeness follows from one instance of that schema.

Monotonicity from a nonnegative derivative holds in the model of named objects of the computable calculus, where points and sequences have computable names (Section 28.7). That is a fact about the model, proved in classical mathematics; whether monotonicity is free on $\mathbb{R}_\delta$ is open (Section 28.7).

Questions to work through

1. Starting from (12.2), clear the positive denominators and recover (12.1). Identify where positivity is needed to preserve the direction of the inequality.

2. In the proof of Lemma 12.5, explain why the later stage 2^{N+1} is available once a finite search has found the first event N. Why is deciding whether any event exists unnecessary for constructing each jump stage?

3. Reconstruct both directions of Theorem 12.6, then identify the exact step that requires LPO instead of WLPO in Theorem 12.7. Use Proposition 12.13 and Theorem 12.11 to show that this step, taken alone, costs exactly $\mathrm{MP}_{\mathrm{bin}}$.

4. Explain why the interleaved sequence e in Theorem 12.8 need not satisfy LLPO's premise. Prove that retaining only its first event supplies that premise. Trace the two uses of parity in the converse proof.

5. Prove the converse half of Lemma 12.10 in integer arithmetic. Where is decidability of a single stage used, and why does the proof not need to decide $P(x)$?

6. Derive the statement of Corollary 12.12 for N from the statement for P. Check that replacing each q_n by $-q_n$ preserves (12.1).

7. In Theorem 12.45, show that the test sequence is nondecreasing and bounded. The proof uses one limit point and one index N. Explain why using them requires no choice principle.

8. The proof of Theorem 12.48 never decides whether L equals $\eta(0)$ or $\eta(1)$. Explain why cotransitivity is enough. Then, before reading the proof of Theorem 12.49, assume in addition that every ratio below L is exceeded by some term, and show that the same test sequence yields LPO.

Part IV

Compactness, algebra and computation

Part IV asks what three tools of analysis and algebra cost on the line built from ratios, namely compactness, linear algebra and the computation of solutions of differential equations, and what a calculus of δ's own objects can derive about them. Each price is an equivalence proved in ordinary constructive mathematics and stated with the choice it uses. Heine–Borel for compact sets presented with their nets, and positive lower bounds and infima of uniformly continuous maps that come with names of their values and a modulus, are exactly the decidable fan theorem of Chapter 13, with no choice principle; for bare uniformly continuous functions the passage back is proved with countable choice (Theorem 13.31). The linear-algebra prices of Chapter 14 use no choice from an assertion to its principle and at most unique choice in the other direction, which the book's reading of maps contains; the one exception is dominated division, priced LLPO, whose converse uses countable choice (Theorem 14.22).

Chapter 15 computes the solutions of polynomial differential equations whose computed trajectory passes a finite region certificate. The certificate is checked by computation; it is not an assumption about an exact solution. Chapter 16 builds a calculus of δ's own objects. In its computable form, whose terms are maps computed by programs, it shows that statements are not derivable by building a model: in that model LPO holds, while the decidable fan theorem and Heine–Borel for presented compact sets, written with equations, fail. The full calculus, whose terms may carry any map, derives LPO, $\mathrm{WLPO}_{\mathrm{seq}}$ and the scalar pair division of Chapter 14 outright, so its separations are proved for the computable calculus. The chapter reasons about its models in classical mathematics, as Chapters 7 and 12 do.

Chapter 13

What compactness costs

Suppose a sequence of open intervals with ratio endpoints covers $[0,1]$: every point of the interval lies in one of them. Classical analysis keeps finitely many of them. On $\mathbb{R}_\delta$ the cover is given, and so is the knowledge that each point is caught by some interval. What is missing is one number N such that the first N intervals already catch every point. What does it take to produce that number?

Here is a small case. Let $I_0 = (-1, \frac{1}{2})$, $I_1 = (\frac{1}{4}, \frac{3}{4})$ and $I_2 = (\frac{1}{2}, 2)$, and let the later intervals be anything. Halve $[0,1]$ repeatedly, and call a piece $[v/2^m, (v+1)/2^m]$ *caught* when, widened by its own length on each side, it lies strictly inside one of the intervals I_k with $k < m$. Whether a piece is caught is a comparison of ratios. At depth $m = 4$ every one of the sixteen pieces is caught: the widened piece $[(v-1)/16, (v+2)/16]$ lies inside I_0 for $v \leq 5$, inside I_1 for $6 \leq v \leq 9$, and inside I_2 for $v \geq 10$. So the halving stops, and the first three intervals cover $[0,1]$.

For a single point the halving always stops: once the pieces containing it are small enough, the widened piece fits inside an interval that contains the point. The difficulty is uniformity. Different points may need deeper pieces, and nothing in the cover says when to stop. The pieces not yet caught form a binary tree, every branch of it ends, and the question is whether the whole tree ends at one depth. Its answer is a fan theorem.

Heine–Borel for $[0,1]$ is equivalent to the decidable fan theorem FAN_Δ, with no choice principle in either direction (Theorem 13.6). FAN_Δ is not an omniscience principle: it follows from weak König's lemma, hence from LLPO with countable choice and from WLPO with unique choice, but excluded middle without unique choice, even with all the countable choice that a search can witness, does not prove it (Theorem 13.11). It is also the price of positive lower bounds and positive infima on every compact set that comes with its finite nets, for functions that come with names of their values and a modulus of uniform continuity. Uniform continuity of pointwise continuous functions implies FAN_Δ; for the converse we name further data that suffice. Bishop's compactness asserts the nets without handing them over, and asks completeness for Cauchy sequences of points rather than named families; every presented set is Bishop compact exactly when every Cauchy sequence of points converges, with no choice either way (Theorem 13.37). Table 13.1 collects the prices.

Two conventions hold throughout. A ratio compared with a point, or added to one, stands for its point: $a < x$ means $\eta(a) < x$, and $x \leq y + 1/(k+1)$ means $x \leq y + \eta(1/(k+1))$. Binary strings, $\bar{p}(n)$ and decidable binary trees are those of Definition 12.71. We write $\emptyset$ for the empty string, st for the string s followed by the string t, $s{\restriction}m$ for the initial segment of s of length $m \leq |s|$, and $s0^\omega$ for the binary sequence that begins with s and continues with zeros. The number $\nu(s)$ has s as its binary numeral, most significant digit first, and $w_0, w_1, \ldots$ lists all binary strings, by length and, within one length, by numeral.

13.1 Bars and the fan theorems

A set of binary strings can be read as a set of stopping points: a binary sequence stops when one of its initial segments lies in the set.

Definition 13.1 (Bars)**.** A set B of binary strings is a *bar* when every binary sequence p has an n with $\bar{p}(n) \in B$. It is a *uniform bar* when one bound serves all sequences: some N has the property that every binary sequence p has an $n \leq N$ with $\bar{p}(n) \in B$. A set of strings is *decidable* when it is given by a map D from binary strings to $\{0, 1\}$; we use D also for the set of strings s with $D(s) = 1$.

A bar asserts, for each sequence, that a stopping index exists. It supplies no map from sequences to indices. A uniform bar supplies one number for all sequences at once.

Definition 13.2 (Fan theorems)**.** The *decidable fan theorem*, FAN_Δ, states that every decidable bar is a uniform bar. The *c-set* of a decidable D consists of the strings s with $D(st) = 1$ for every string t; FAN_c states that whenever the c-set of a decidable D is a bar, it is a uniform bar. A Π^0_1 *set* of strings is given by a map E from pairs of a string and a number to $\{0, 1\}$: the string s belongs to it when $E(s, n) = 1$ for every n. The set is *monotone* when s belonging implies that sb belongs for each digit b, and E is a *monotone presentation* when $E(s, n) = 1$ implies $E(sb, n) = 1$ for all s, b and n. The Π^0_1 *fan theorem* $\mathrm{FAN}_{\Pi^0_1}$ states that every Π^0_1 bar is a uniform bar; the *monotone* Π^0_1 *fan theorem* states this for monotone Π^0_1 bars, and the Π^0_1 *fan theorem for monotone presentations* for Π^0_1 bars given by a monotone presentation. The *full fan theorem*, FAN, states that every bar, given by an arbitrary property of strings, is a uniform bar.

Brouwer introduced the fan theorem [29]. Its decidable and c-set forms, and the Π^0_1 forms between them and the full theorem, are the yardsticks of compactness in constructive reverse mathematics [32, 31, 34, 35].

Theorem 13.3 (The tree form of the decidable fan theorem)**.** *Say that a binary sequence p* leaves *a tree T when $\bar{p}(n) \notin T$ for some n. Then* FAN_Δ *holds if and only if every decidable binary tree that every binary sequence leaves has a length at which it has no member. Neither direction uses a choice principle or an omniscience principle.*

Proof. Assume FAN_Δ, and let T be a decidable binary tree that every binary sequence leaves. The strings outside T form a decidable bar, so some N has the property that every binary sequence has an initial segment of length at most N outside T. Whether T has a member of length N is decided by inspecting the 2^N strings of that length; suppose s is one. The sequence $s0^\omega$ has an initial segment $s{\restriction}n$ with $n \leq N$ outside T. But $s{\restriction}n$ is an initial segment of the member s, and T is closed under initial segments, so $s{\restriction}n \in T$. This contradiction shows that T has no member of length N.

Conversely, assume the tree form, and let D be a decidable bar. If $D(\emptyset) = 1$, the bound $N = 0$ serves. Otherwise say that s is *seen* when $D(s{\restriction}n) = 1$ for some $n \leq |s|$, a search through $|s| + 1$ initial segments, and let A consist of the strings that are not seen. A contains $\emptyset$, and it is closed under initial segments, since an extension of a seen string is seen; so A is a decidable binary tree. Every binary sequence p leaves A: the bar gives n with $D(\bar{p}(n)) = 1$, so $\bar{p}(n)$ is seen. By the tree form, A has no member of some length N. For a binary sequence p, the

string $\bar{p}(N)$ is therefore seen: some $n \le N$ has $D(\bar{p}(N)\restriction n) = D(\bar{p}(n)) = 1$. So N is a uniform bound. Every step is a finite inspection of data already given. □

The set of strings outside a bar need not be a tree, which is why the proof passes through the strings that are not yet seen. The forms of the fan theorem are ordered by strength.

Proposition 13.4 (From the full fan theorem down to the decidable one)**.** *Each of the following principles implies the next:* FAN*;* $\mathrm{FAN}_{\Pi^0_1}$*; the monotone* Π^0_1 *fan theorem; the* Π^0_1 *fan theorem for monotone presentations;* FAN_c*;* FAN_Δ*. No choice principle is used.*

Proof. A Π^0_1 set is a property of strings, so FAN gives $\mathrm{FAN}_{\Pi^0_1}$; and the monotone forms are special cases. If each approximant $E(\cdot, n)$ is closed under one-digit extension, so is the set of strings s with $E(s, n) = 1$ for every n; so a Π^0_1 set with a monotone presentation is monotone, which gives the third implication.

For the fourth, let D be decidable, and put $E(s, n) = D(w_n)$ if s is an initial segment of w_n and $E(s, n) = 1$ otherwise. Then $E(s, n) = 1$ for every n says exactly that $D(st) = 1$ for every string t, since every extension st is some w_n; so the Π^0_1 set of E is the c-set of D. The presentation is monotone: if sb is an initial segment of w_n, so is s, and then $E(s, n) = 1$ means $D(w_n) = 1$, which is $E(sb, n)$; otherwise $E(sb, n) = 1$. A c-set that is a bar is therefore a Π^0_1 bar with a monotone presentation, and its uniform bound is the one FAN_c asks for.

For the last, let D be a decidable bar, and let S be the set of seen strings of the proof of Theorem 13.3. It is decidable and closed under extension, so its c-set is S itself, and it is a bar because D is. FAN_c gives a uniform bound N for S, and a seen initial segment $\bar{p}(m)$ with $m \le N$ has an initial segment $\bar{p}(n)$, $n \le m$, with $D(\bar{p}(n)) = 1$. □

We prove no converse. Lubarsky and Diener separate the fan theorem from several of these weakenings by models of intuitionistic set theory [34], and Kawai shows that the monotone Π^0_1 fan theorem is equivalent to a strong continuity principle for complete separable metric spaces [35].

13.2 Heine–Borel on $[0, 1]$ is the decidable fan theorem

The covers of the opening are countable and have ratio endpoints. Every open set of the line is a union of such intervals, so nothing is lost for the question at hand, and every finite subcover lies inside an initial segment of the cover.

Definition 13.5 (Ratio covers and cover-compact sets)**.** A *ratio cover* is a sequence of pairs of ratios (a_k, b_k); its kth interval consists of the points x with $a_k < x < b_k$. It *covers* a set K of points when every member of K lies in one of its intervals. The set K is *cover-compact* when every ratio cover of K has an N such that every member of K lies in one of the first N intervals. *Heine–Borel for* $[0, 1]$ is the assertion that $[0, 1]$ is cover-compact.

Theorem 13.6 (Heine–Borel is the decidable fan theorem)**.** *The interval* $[0,1]$ *is cover-compact if and only if* FAN_Δ *holds. Neither direction uses a choice principle or an omniscience principle.*

Proof. By Theorem 13.3 it suffices to show that $[0,1]$ is cover-compact exactly when every decidable binary tree that every binary sequence leaves has a length with no member. We call this the *tree form.*

From the tree form to compactness. Let (a_k, b_k) be a ratio cover of $[0,1]$. A string s of length m stands for the dyadic piece $[\nu(s)/2^m, (\nu(s)+1)/2^m]$. Call s *caught* when some $k < m$ has

$$a_k < \frac{\nu(s)-1}{2^m} \quad \text{and} \quad \frac{\nu(s)+2}{2^m} < b_k,$$

so that the piece, widened by its length on each side, lies inside interval k; this is decided by comparing ratios. Let H consist of the strings none of whose initial segments, the string itself included, is caught. H contains $\emptyset$, which is not caught because there is no $k < 0$, and it is closed under initial segments; so H is a decidable binary tree.

Every binary sequence leaves H. Let p be one, and let x be the point named by the partial sums $c_n = \sum_{i<n} p(i)2^{-(i+1)}$; they form a regular sequence, since $0 \le c_n - c_m \le 2^{-m} \le 1/(m+1)$ for $m \le n$. With $\nu_n = \nu(\bar{p}(n)) = 2^n c_n$, every $m \ge n$ has $\nu_n/2^n \le c_m \le (\nu_n+1)/2^n$, and c_m is within $1/(m+1)$ of x (Lemma 12.52, item 4); so item 2 of that lemma gives $\nu_n/2^n \le x \le (\nu_n+1)/2^n$. In particular x lies in $[0,1]$, hence in some interval k_0 of the cover. By item 1 of the same lemma there are m_1 and m_2 with $a_{k_0} + 1/(m_1+1) < x < b_{k_0} - 1/(m_2+1)$. Put $n = k_0 + 2(m_1+m_2) + 3$ and $\nu = \nu_n$. Then $n > k_0$, and $2/2^n < 2/n$ is smaller than both $1/(m_1+1)$ and $1/(m_2+1)$. Hence

$$a_{k_0} < x - \frac{1}{m_1+1} \le \frac{\nu+1}{2^n} - \frac{1}{m_1+1} < \frac{\nu-1}{2^n}, \qquad \frac{\nu+2}{2^n} \le x + \frac{2}{2^n} < x + \frac{1}{m_2+1} < b_{k_0}.$$

So $\bar{p}(n)$ is caught by the interval $k_0 < n$, and $\bar{p}(n) \notin H$.

An empty level gives the subcover. By the tree form, H has no member of some length N. Let x lie in $[0,1]$. By Proposition 12.62, x has a signed-digit expansion, so some integer z satisfies $(z-1)/2^N \le x \le (z+1)/2^N$. Put $w = 0$ if $z < 0$, $w = 2^N - 1$ if $z \ge 2^N$, and $w = z$ otherwise. Using $0 \le x \le 1$ in the first two cases,

$$\frac{w-1}{2^N} \le x \le \frac{w+2}{2^N}.$$

The string t of length N with $\nu(t) = w$ is not in H, so some initial segment $t{\restriction}m$, $m \le N$, is caught by an interval $j < m$: with $u = \nu(t{\restriction}m)$, we have $a_j < (u-1)/2^m$ and $(u+2)/2^m < b_j$. Since $w = u2^{N-m} + r$ with $0 \le r < 2^{N-m}$,

$$\frac{u-1}{2^m} \le \frac{w-1}{2^N} \quad \text{and} \quad \frac{w+2}{2^N} \le \frac{(u+1)2^{N-m}+1}{2^N} \le \frac{u+2}{2^m}.$$

Hence $a_j < x < b_j$ with $j < N$. So every point of $[0,1]$ lies in one of the first N intervals.

From compactness to the tree form: the Cantor cells. Write $u_n = 1/(8 \cdot 4^n)$. For a string s of length n let $\kappa(s)$ be the number whose base-four digits are $3s_0, \dots, 3s_{n-1}$. The *Cantor point* of s is $c(s) = \kappa(s)/4^n$, and with $\nu = \kappa(s)$ its *Cantor interval* is $J(s) = [8\nu u_n, (8\nu+8)u_n]$. So $c(s)$ is

the left end of $J(s)$, and the intervals $J(s0) = [8v, 8v+2]u_n$ and $J(s1) = [8v+6, 8v+8]u_n$ lie inside $J(s)$, separated by the *gap* $(8v+2, 8v+6)u_n$. The open *cells* $C(s)$, intervals with ratio endpoints, are defined by recursion:

$$C(\emptyset) = (-1, 2), \qquad C(s0) = \big(\ell(s), (8v+5)u_n\big), \qquad C(s1) = \big((8v+3)u_n, r(s)\big),$$

where $\ell(s)$ and $r(s)$ are the ends of $C(s)$. The cells of the two children overlap in $(8v+3, 8v+5)u_n$, inside the gap. Figure 13.1 shows the first two levels. Three facts follow.

(a) *$J(s)$ lies inside $C(s)$, with room at both ends, and $C(sb) \subseteq C(s)$.* By induction on s: $J(\emptyset) = [0, 1]$ lies inside $(-1, 2)$. If $J(s)$ lies inside $C(s)$ with room, the new end $(8v+5)u_n$ of $C(s0)$ exceeds the right end $(8v+2)u_n$ of $J(s0)$ and lies below the right end $(8v+8)u_n$ of $J(s)$, hence inside $C(s)$; the left end of $C(s0)$ is $\ell(s)$, below the common left end of $J(s)$ and $J(s0)$. The string $s1$ is symmetric.

(b) *If $s \neq t$ have the same length, $C(s)$ and $J(t)$ are disjoint.* By induction on the common length; at length 0 there is nothing to prove. Write $s = s'b$ and $t = t'b'$. If $s' \neq t'$, the induction hypothesis applies, because $C(s) \subseteq C(s')$ and $J(t) \subseteq J(t')$. If $s' = t'$, with $v = \kappa(s')$ and $n = |s'|$, then $b \neq b'$: the cell $C(s'0)$ lies below $(8v+5)u_n$, while $J(s'1)$ lies above $(8v+6)u_n$; and $C(s'1)$ lies above $(8v+3)u_n$, while $J(s'0)$ lies below $(8v+2)u_n$. Consequently, *if $|t| \geq |s|$ and $c(t) \in C(s)$, then s is an initial segment of t*: the point $c(t)$ lies in $J(t) \subseteq J(t{\restriction}|s|)$, so $t{\restriction}|s| = s$.

(c) *Every point x of $[0, 1]$ has an address*, a binary sequence p with $x \in C(\bar{p}(n))$ for every n. This is an existence statement about one point, so we may use one regular sequence naming x; nothing is selected. For a string s of length n with $v = \kappa(s)$, let q be the stage $2^{2n+5} - 1$ of that sequence, which is within $2^{-(2n+5)}$ of x (Lemma 12.52, item 4), and give s the digit 0 if $q \leq (32v+16)/2^{2n+5}$ and the digit 1 otherwise. Let $p(n)$ be the digit of $\bar{p}(n)$. If $x \in C(s)$ and the digit is 0, then

$$x \leq q + 2^{-(2n+5)} \leq \frac{32v+17}{2^{2n+5}} < \frac{32v+20}{2^{2n+5}} = (8v+5)u_n,$$

so $x \in C(s0)$; if the digit is 1, then $x \geq q - 2^{-(2n+5)} > (32v+15)/2^{2n+5} > (8v+3)u_n$, so $x \in C(s1)$. Since $x \in C(\emptyset)$, induction gives $x \in C(\bar{p}(n))$ for every n.

From compactness to the tree form: the dead cover. Let T be a decidable binary tree that every binary sequence leaves. The *dead cover* has as its kth interval the empty interval $(0, 0)$ if $w_k \in T$, and the cell $C(w_k)$ if $w_k \notin T$. It covers $[0, 1]$: an address p of x leaves T at some n, and x lies in $C(\bar{p}(n))$, the interval whose index k has $w_k = \bar{p}(n)$. By compactness some N has every point of $[0, 1]$ in one of the first N intervals. Put $d = \sum_{k<N}(|w_k| + 1)$, so that $|w_k| < d$ for every $k < N$. Whether T has a member of length d is decided by inspection; suppose t is one. The point $c(t)$ lies in $[0, 1]$, so in some interval $k < N$. That interval is not empty, so $w_k \notin T$ and $c(t) \in C(w_k)$. As $|w_k| < |t|$, fact (b) makes w_k an initial segment of t, and T, closed under initial segments, contains w_k: a contradiction. So T has no member of length d.

Every step in both directions is a comparison of ratios, a finite inspection, or the use of one regular sequence naming one point. □

The tree H is the halving of the opening example, its members the pieces not yet caught. The second half runs backward: a tree becomes the cover by the cells of its dead strings, and a deep member supplies a Cantor point that no early dead cell contains.

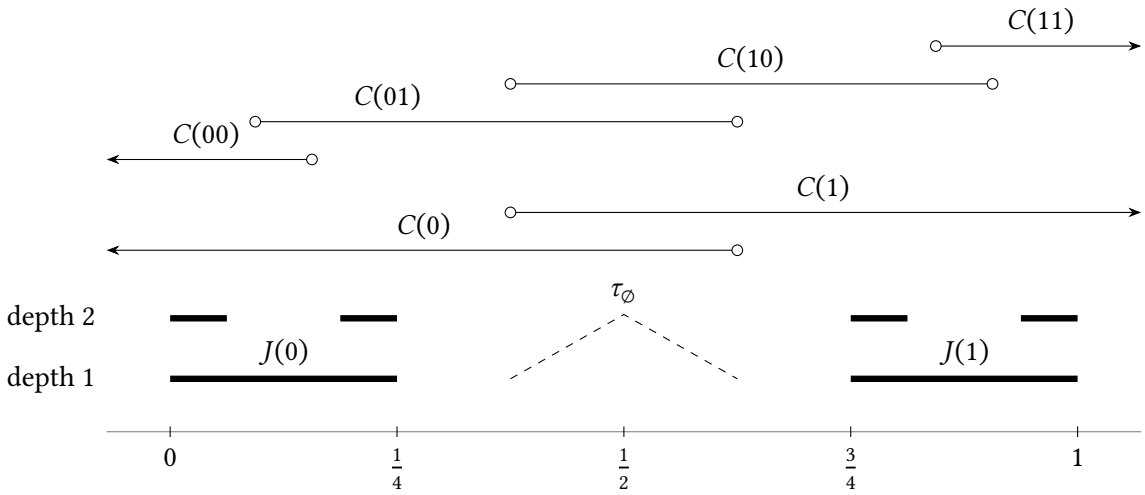


Figure 13.1. The first two levels of the Cantor intervals $J(s)$ (thick) and the open cells $C(s)$ (thin, open ends marked by circles; arrows continue to -1 and 2). The cells of two children overlap only inside the gap of their parent, where no Cantor interval of a later level reaches. The dashed tent $\tau_\varnothing$ of Section 13.4 rises over the overlap of $C(0)$ and $C(1)$.

13.3 Where the fan theorem sits

Weak König's lemma is the statement of Definition 12.71: every infinite decidable binary tree has a path. The fan theorem asks the opposite question about a tree that has no path. The link between them is a tree that is infinite for a reason that needs no decision.

Theorem 13.7 (Weak König's lemma gives the decidable fan theorem)**.** *Weak König's lemma implies* FAN_Δ. *No choice principle, no omniscience principle and no Markov principle is used.*

Proof. We prove the tree form. Let T be a decidable binary tree that every binary sequence leaves. The *padded tree* P consists of the strings s such that, for every $k \le |s|$, either $s{\restriction}k \in T$ or T has no member of length k. Both alternatives are decided by inspection. P contains $\varnothing$ and is closed under initial segments.

P is infinite, with no principle. Given n, decide whether T has a member of length n. If it has one, that member lies in P. If not, find the least $d \le n$ at which T has no member; $d \ge 1$ because $\varnothing \in T$. Take a member e of T of length $d-1$ and pad it with zeros to length n. Its initial segments of length less than d lie in T, and T has no member of any length $k \ge d$, because such a member would have an initial segment of length d in T. So the padded string lies in P.

Weak König's lemma gives a path p through P. Since p leaves T, some n has $\bar{p}(n) \notin T$, and $\bar{p}(n) \in P$ at $k = n$ then says that T has no member of length n. □

Applying weak König's lemma to T itself would need T to be infinite, which is the negation of what we want. The padded tree is infinite outright, and the empty level is read off the path, so no unbounded search for it is made. Ishihara gives a direct proof of the same implication [33].

Weak König's lemma has an exact price of its own, sharper than Theorem 12.72. It needs the omniscience principles for a whole sequence of questions at once.

Definition 13.8 (Omniscience for sequences of rows)**.** Let $A(i,n)$ be a double sequence of bits, with rows $A(i,\cdot)$. $\mathrm{WLPO}_{\mathrm{seq}}$ states that for every such A there is a binary sequence f with $f(i)=1$ exactly when $A(i,n)=0$ for every n. $\mathrm{LLPO}_{\mathrm{seq}}$ states that for every such A whose rows each have at most one entry 1 there is a binary sequence f such that $f(i)=1$ implies $A(i,2k)=0$ for every k, and $f(i)=0$ implies $A(i,2k+1)=0$ for every k.

Each single principle of Definition 7.8 is the case of one row. Conversely, AC_{ω} collects the answers for the rows into one sequence.

Proposition 13.9 (Weak König's lemma is LLPO for sequences)**.** *Weak König's lemma holds if and only if* $\mathrm{LLPO}_{\mathrm{seq}}$ *holds. No choice principle is used in either direction.*

Proof. Assume weak König's lemma, and let A have rows with at most one entry 1. Let S consist of the strings s such that, for every $i<|s|$, $s_i=1$ implies $A(i,m)=0$ for every even $m<|s|$, and $s_i=0$ implies $A(i,m)=0$ for every odd $m<|s|$. Membership is a finite inspection; S contains $\emptyset$ and is closed under initial segments, since a shorter string meets fewer conditions. S is infinite: for a length n, put $s_i=0$ if $A(i,m)=1$ for some even $m<n$, and $s_i=1$ otherwise. In the first case the single entry 1 of row i sits at an even place, so every odd entry is 0. Let f be a path through S. If $f(i)=1$ and k is given, the member $\bar{f}(n)$ with $n>\max(i,2k+1)$ gives $A(i,2k)=0$; if $f(i)=0$, it gives $A(i,2k+1)=0$. So f is the sequence that $\mathrm{LLPO}_{\mathrm{seq}}$ asks for.

Conversely, assume $\mathrm{LLPO}_{\mathrm{seq}}$, and let T be an infinite decidable binary tree. The proof of Theorem 12.72 attaches to each string s a binary sequence γ_s with at most one entry 1, built by bounded searches, and shows: if every even entry of γ_s is 0, then $s0$ is extendable whenever s is; if every odd entry is 0, then $s1$ is extendable whenever s is. Apply $\mathrm{LLPO}_{\mathrm{seq}}$ once, to the double sequence whose row i is γ_{w_i}. It gives f; put $c_s=0$ if $f(i)=1$ and $c_s=1$ if $f(i)=0$, where $s=w_i$. Then sc_s is extendable whenever s is. Let $s_0=\emptyset$ and $s_{d+1}=s_d c_{s_d}$. By induction every s_d is extendable, hence a member of T, and $p(d)=c_{s_d}$ defines a path. The children are named for all strings at once, so nothing is collected. □

Theorem 12.72 is the case in which countable choice turns LLPO into $\mathrm{LLPO}_{\mathrm{seq}}$; that is where its one use of AC_{ω} goes.

Corollary 13.10 (Principles that give the decidable fan theorem)**.**

1. $\mathrm{LLPO}_{\mathrm{seq}}$ *implies* FAN_{Δ}*, with no choice principle. In particular LLPO and* AC_{ω} *imply* FAN_{Δ}*; countable choice is used once, inside Theorem 12.72.*
2. $\mathrm{WLPO}_{\mathrm{seq}}$ *implies* $\mathrm{LLPO}_{\mathrm{seq}}$*, with no choice principle.*
3. *WLPO together with unique choice for relations from numbers to* $\{0,1\}$ *implies* $\mathrm{WLPO}_{\mathrm{seq}}$*, and hence* FAN_{Δ}*. Unique choice is used once, to collect the decisions of WLPO for the rows into one sequence.*

Proof. Item 1 combines Proposition 13.9 with Theorem 13.7, and Theorem 12.72 with Theorem 13.7.

For item 2, let the rows of A have at most one entry 1. Apply $\mathrm{WLPO}_{\mathrm{seq}}$ to the even parts, the double sequence $(i,k)\mapsto A(i,2k)$, to obtain f. If $f(i)=1$, every even entry of row i is 0. If

$f(i) = 0$, the even entries are not all 0, and every odd entry is 0: each is decidable, and an entry 1 at an odd place would be the only 1 of the row, making every even entry 0.

For item 3, WLPO decides for each row i whether it is identically 0. So the relation "$b = 1$ and row i is identically 0, or $b = 0$ and it is not" has exactly one b in $\{0, 1\}$ for each i, and unique choice turns it into a binary sequence. Items 2 and 1 finish. □

Under the reading of maps of Appendix C, which contains unique choice, WLPO therefore gives FAN_Δ with no countable choice. Without unique choice, excluded middle does not give the fan theorem. The witness is a tree of Kleene's, read in the model of computable sequences that Theorem 12.63 uses.

Theorem 13.11 (Without unique choice, excluded middle does not give the fan theorem)**.** *Consider the theory of Theorem 12.63: arithmetic with variables for sequences, induction for every formula, excluded middle for every formula, and countable choice for the relations whose witnesses a search can find. Write $v < w$ for $\exists z\,(v + Sz = w)$, $\langle n, v\rangle$ for the pairing term $(n + v)\cdot(n + v) + v$, and $\mathrm{Bin}(\alpha)$ as in the proof of Theorem 12.22. The decidable fan theorem, written as the sentence*

$$\forall\tau\Big[\Big(\tau\langle 0,0\rangle \neq 0 \wedge \forall n\,\forall v\,\forall b\,\big((b = 0 \vee b = 1) \to \tau\langle Sn, v + v + b\rangle \neq 0 \to \tau\langle n, v\rangle \neq 0\big)$$
$$\wedge\ \forall\alpha\,\big(\mathrm{Bin}(\alpha) \to \forall\gamma(\gamma(0) = 0 \wedge \forall n\,\gamma(Sn) = \gamma(n) + \gamma(n) + \alpha(n) \to \exists n\,\tau\langle n, \gamma(n)\rangle = 0)\big)\Big)$$
$$\to\ \exists N\,\forall\epsilon\,\big(\epsilon(0) = 1 \wedge \forall n\,\epsilon(Sn) = \epsilon(n) + \epsilon(n) \to \forall v(v < \epsilon(N) \to \tau\langle N, v\rangle = 0)\big)\Big],$$

is not provable in that theory, nor in any theory with fewer axioms. The proof reasons in a classical metatheory.

The sentence reads as follows. The sequence τ codes a set of strings: the string of length n with numeral v belongs when $\tau\langle n, v\rangle \neq 0$. The first two clauses say that the set contains $\emptyset$ and is closed under initial segments. In the third, $\gamma(n)$ is the numeral of $\bar{\alpha}(n)$, so the clause says that every binary sequence leaves the tree. In the conclusion, $\epsilon(N) = 2^N$, so some length has no member. This is the tree form of Theorem 13.3.

Proof. Use the model of the proof of Theorem 12.63: numbers range over the natural numbers, sequences over the total computable functions, and truth is classical. That proof shows that every theorem of the theory is true in the model. It suffices to find a computable τ at which the sentence is false.

Fix the enumeration of programs and the notation $\{e\}(a)$ of the proof of Theorem 12.22. *Kleene's tree* T_K consists of the strings l such that, for every $e < |l|$ for which program e halts on input 0 within $|l|$ steps with an output o, the digit l_e is 1 if $o = 0$ and 0 if $o \neq 0$ [51, 4]. Membership is computed by running $|l|$ programs for $|l|$ steps. T_K contains $\emptyset$, and it is closed under initial segments, because a program that halts within n steps halts within $n + 1$ steps with the same output, so a constraint visible at a length stays visible at every larger length.

T_K has a member of every length N: the string whose digit $e < N$ is 1 exactly when program e halts on input 0 within N steps with output 0 meets every constraint at length N.

Every computable binary sequence p leaves T_K. The function $(x, a) \mapsto p(x)$ is computable, so Kleene's recursion theorem [19] gives a program e with $\{e\}(a) = p(e)$ for every a. In

particular program e halts on input 0, within some k steps, with output $p(e)$. At every length $n > \max(e, k)$ the constraint on digit e asks for 1 if $p(e) = 0$ and for 0 if $p(e) = 1$, while $\bar{p}(n)$ has the digit $p(e)$ there. So $\bar{p}(n) \notin T_K$.

Let $\tau\langle n, v\rangle = 1$ when $v < 2^n$ and the string of length n with numeral v lies in T_K, and let τ be 0 at every other number; distinct pairs have distinct values $\langle n, v\rangle$. Then τ is computable, and the premises of the sentence hold in the model: the root and closure clauses hold because T_K is a tree and $v + v + b < 2^{n+1}$ forces $v < 2^n$, and the third clause holds because the binary sequences of the model are the computable ones, each of which leaves T_K. The conclusion fails, since T_K has a member of every length. So the sentence is false in the model and is not a theorem. □

The theorem concerns the fan sentence itself. With Theorem 13.6 it shows that the principle that prices Heine–Borel for $[0, 1]$ is not an omniscience principle. The model lacks a path that no program computes, much as the model of Theorem 12.63 lacks a point that no program names. In the computable calculus of δ's own objects, whose terms are maps computed by programs, the same computable reading applies to FAN_Δ, FAN_c, weak König's lemma for decidable trees and the presented forms of Heine–Borel, of the positive lower bounds and infima and of Heine–Cantor in this chapter, each written with equations between the calculus's terms: none of them is derivable from LPO, even with every instance of excluded middle added as a base rule (Chapter 16). The full calculus, whose terms may carry any map, derives each of them, because each is classically true.

13.4 Uniform continuity

The second classical consequence of compactness is that continuity on a compact set is uniform. On $\mathbb{R}_\delta$ this splits in two. Uniform continuity of pointwise continuous functions implies FAN_Δ; our proofs of the converse use data that pointwise continuity does not supply.

Definition 13.12 (Continuity on a set and on Cantor space)**.** Let f be a function on the points of $\mathbb{R}_\delta$ and K a set of points. The function is *pointwise continuous on K* when for every member x of K and every k there is a j such that every member y of K within $1/(j+1)$ of x has $f(y)$ within $1/(k+1)$ of $f(x)$. It is *uniformly continuous on K* when for every k one j serves all pairs: members x, y of K within $1/(j+1)$ have $f(x)$ and $f(y)$ within $1/(k+1)$. *Heine–Cantor on K* is the assertion that every function pointwise continuous on K is uniformly continuous on K.

A map a from binary sequences to natural numbers is *pointwise continuous* when every p has an n such that $a(q) = a(p)$ whenever $\bar{q}(n) = \bar{p}(n)$, and *uniformly continuous* when one n serves all p and q. A map F from binary sequences to points is *uniformly continuous* when for every k some n has $F(p)$ and $F(q)$ within $1/(k+1)$ whenever $\bar{p}(n) = \bar{q}(n)$.

For $K = [0, 1]$, uniform continuity is the notion of Table 12.3. In Bishop's mathematics continuity on a compact interval is defined as uniform continuity [14]; we name the pointwise notion separately because the two differ in price.

Theorem 13.13 (Heine–Cantor on $[0,1]$ gives the decidable fan theorem)**.** *If every function pointwise continuous on* $[0,1]$ *is uniformly continuous on* $[0,1]$*, then* FAN_Δ *holds. No choice principle is used beyond the unique choice that is part of the reading of maps, which defines the test function below.*

Proof. Let T be a decidable binary tree that every binary sequence leaves; by Theorem 13.3 it suffices to find a length with no member of T. We keep the notation of the proof of Theorem 13.6. For a string s of length n with $v = \kappa(s)$, let $g(s) = (8v+4)u_n$ be the centre of the gap of s, let $O(s) = \big((8v+3)u_n, (8v+5)u_n\big)$ be the overlap of the cells of its children, and let

$$\tau_s(x) = \max\big(0,\ 1 - |x - g(s)|/u_n\big)$$

be its *tent*. The tent has the value 1 at $g(s)$. At a point x not in $O(s)$ it has the value 0, because $|x - g(s)| < u_n$ would put x in $O(s)$, so $|x - g(s)| \geq u_n$. Points within $1/(j+1)$ have tent values within $8 \cdot 4^n/(j+1)$, by Lemma 12.52, item 5, and multiplication by the ratio $1/u_n$.

Two more facts about the overlaps. (d) *Overlaps of different strings are disjoint, and no Cantor point lies in an overlap.* By fact (b) and $J(s) \subseteq C(s)$, Cantor intervals of different strings of one length are disjoint, and $O(s)$ lies inside $J(s)$ but meets neither $J(s0)$ nor $J(s1)$. Let $s \neq s'$. If neither extends the other, they extend rb and rb' with $b \neq b'$ for some string r, and $O(s) \subseteq J(rb)$ is disjoint from $O(s') \subseteq J(rb')$. If s' extends s, it extends sb for a digit b, and $O(s') \subseteq J(sb)$ misses $O(s)$. A Cantor point does not change when zeros are appended to its string, so for $c(t)$ and $O(s)$ we may assume $|t| > |s|$; then $c(t)$ lies in $J(t{\restriction}(|s|+1))$, which is $J(s0)$, $J(s1)$, or disjoint from $J(s)$. (e) *If $u \notin T$ has length d, every tent τ_s of a member s of T with $|s| \geq d$ vanishes on $C(u)$.* Indeed $s{\restriction}d \neq u$, because T is closed under initial segments, so by fact (b) the cell $C(u)$ misses $J(s{\restriction}d)$, which contains $O(s)$.

The test function. For $N \geq 0$ let $H_N = \sum \tau_s$, summed over the members s of T of length less than N: a finite sum of tents. For a point x let $x' = \max(0, \min(x, 1))$, a point of $[0,1]$. From a regular sequence naming x', fact (c) computes an address p of x'; it leaves T, and we let d be the least n with $\bar{p}(n) \notin T$. Since x' lies in $C(\bar{p}(d))$, fact (e) gives $H_N(x') = H_d(x')$ for every $N \geq d$. Another regular sequence naming x' gives a depth d' with the same property, and $H_d(x') = H_{\max(d,d')}(x') = H_{d'}(x')$. So $f(x) = H_d(x')$ depends only on x. This prescription gives exactly one output for each point, so it defines a function on the points (Appendix C); if functions are required instead to be operations given in advance, this is the one place where unique choice is used, to make the least depth a function of the name. For every point y of $[0,1]$ we have $y' = y$, so $f(y) = H_N(y)$ for every N beyond the depth of some name of y.

Continuity on $[0,1]$. Let x lie in $[0,1]$, let d be its depth for one name, and $u = \bar{p}(d) \notin T$, so that $x \in C(u)$. By Lemma 12.52, item 3, some j_0 has every point within $1/(j_0+1)$ of x in $C(u)$. For such a point y of $[0,1]$, fact (e) gives $H_N(y) = H_d(y)$ for $N \geq d$, so $f(y) = H_d(y)$; likewise $f(x) = H_d(x)$. The function H_d is a sum of fewer than 2^d tents of strings shorter than d. Given k, choose $j \geq j_0$ with $2^d \cdot 8 \cdot 4^d/(j+1) \leq 1/(k+1)$. Then every y in $[0,1]$ within $1/(j+1)$ of x has $f(y)$ within $1/(k+1)$ of $f(x)$.

Values at Cantor points and gap centres. For every string t the point $c(t)$ lies in $[0,1]$ and in no overlap, so every tent vanishes there and $f(c(t)) = 0$. For every member s of T the point $g(s)$ lies in $O(s)$ and in no other overlap, so $H_N(g(s)) = \tau_s(g(s)) = 1$ for $N > |s|$, and $f(g(s)) = 1$.

Uniformity leaves no deep member. By hypothesis f is uniformly continuous on $[0,1]$, so at the tolerance $\frac{1}{2}$ some j has points of $[0,1]$ within $1/(j+1)$ with values within $\frac{1}{2}$. Whether T has a member of length $n = j+1$ is decided by inspection. If s were one, the points $c(s)$ and $g(s)$ of $[0,1]$ would be $4u_n = 2^{-(2n+1)} \leq 1/(j+1)$ apart, while their values 0 and 1 are not within $\frac{1}{2}$. So T has no member of length $j+1$. □

That Heine–Cantor on compact spaces implies the decidable fan theorem is known in Bishop-style analysis [36, 35]; the proof above works on $\mathbb{R}_\delta$ with one explicit test function. Our proof of the converse uses the local radii of continuity collected along one cover. A *control presentation* of f at the tolerance $1/(k+1)$ is a ratio cover (a_n, b_n) of $[0,1]$ together with a sequence of numbers r_n such that, for every n, members x of $[0,1]$ in interval n and members y of $[0,1]$ within $1/(r_n+1)$ of x have $f(x)$ and $f(y)$ within $1/(k+1)$.

Proposition 13.14 (Heine–Cantor from the fan theorem and control presentations)**.** *Assume that every function pointwise continuous on* $[0,1]$ *has a control presentation at every tolerance. Then Heine–Cantor on* $[0,1]$ *holds if and only if* FAN_Δ *holds. The passage from* FAN_Δ *uses no choice principle; the passage to it is Theorem 13.13.*

Proof. Assume FAN_Δ, let f be pointwise continuous on $[0,1]$, and fix k. Take a control presentation at $1/(k+1)$. By Theorem 13.6 the first N intervals of its cover cover $[0,1]$. Put $j = \max_{n<N} r_n$. If x and y in $[0,1]$ are within $1/(j+1)$, then x lies in some interval $n < N$ and y is within $1/(r_n+1)$ of x, so $f(x)$ and $f(y)$ are within $1/(k+1)$. □

On Cantor space the same question has a sharper answer.

Theorem 13.15 (Uniform continuity on Cantor space)**.**

1. FAN_c *implies that every pointwise continuous map from binary sequences to natural numbers is uniformly continuous. No choice principle is used.*
2. *If every pointwise continuous map from binary sequences to natural numbers is uniformly continuous, then* FAN_Δ *holds. Unique choice, part of the reading of maps, is used once, to make the first hit of a bar a map.*

Proof. For item 1, let a be pointwise continuous, and write $e(s) = a(s0^\omega)$. Let S be the decidable set of strings s with $e(s0) = e(s)$ and $e(s1) = e(s)$. The c-set of S is a bar: if $a(q) = a(p)$ whenever $\bar{q}(n) = \bar{p}(n)$, then for every extension s of $\bar{p}(n)$ and each digit b the sequences $s0^\omega$ and $sb0^\omega$ agree with p in their first n digits, so $e(sb) = a(p) = e(s)$, and every extension of $\bar{p}(n)$ lies in S. FAN_c gives N such that every p has an $n \leq N$ with $\bar{p}(n)$ in the c-set. If u is in the c-set, then $e(ut) = e(u)$ for every string t, by induction on t. Moreover $a(p) = e(\bar{p}(n))$ whenever $\bar{p}(n)$ is in the c-set: continuity at p gives an m, and with $K = \max(n,m)$ the sequence $\bar{p}(K)0^\omega$ agrees with p in m digits, so $a(p) = e(\bar{p}(K)) = e(\bar{p}(n))$. Now let $\bar{p}(N) = \bar{q}(N)$, and take $n \leq N$ with $\bar{p}(n)$ in the c-set. Then $\bar{q}(n) = \bar{p}(n)$, and $a(q) = e(\bar{q}(n)) = e(\bar{p}(n)) = a(p)$.

For item 2, let D be a decidable bar. For each p there is exactly one least n with $D(\bar{p}(n)) = 1$, the *first hit* $h(p)$; the bar gives a hit and the least one is unique. The relation "n is the first hit of p" has exactly one output for each p, so it is a map by the reading of maps; this is the use of unique choice. The map h is pointwise continuous: a sequence q that agrees with p in

its first $h(p)$ digits has the same first hit. If h is uniformly continuous with bound N, it takes at every p the value $h(\bar{p}(N)0^\omega)$, so it is bounded by the largest of the 2^N values $h(s0^\omega)$ with $|s| = N$, and that largest value is a uniform bound for D. □

Our converse of item 1 uses one more hypothesis.

Definition 13.16 (Last exits)**.** Let D be decidable and p a binary sequence. A number n is a *last exit* of p when $D(\bar{p}(m)) = 1$ for every $m \geq n$, and for every $j < n$ some $m \geq j$ has $D(\bar{p}(m)) = 0$. *Last-exit selection* states that for every decidable D whose c-set is a bar there is a map L from binary sequences to natural numbers such that $L(p)$ is a last exit of p for every p.

Lemma 13.17 (Last exits are unique)**.** *If the c-set of a decidable D is a bar, every binary sequence has exactly one last exit. No choice principle is used.*

Proof. The bar gives N with $\bar{p}(N)$ in the c-set, so $D(\bar{p}(m)) = 1$ for every $m \geq N$. A search below N finds the last $m < N$ with $D(\bar{p}(m)) = 0$; then $m + 1$ is a last exit, and if there is no such m, 0 is one. If $n < n'$ were both last exits, the second clause for n' at $j = n$ would give some $m \geq n$ with $D(\bar{p}(m)) = 0$, against the first clause for n. □

Theorem 13.18 (FAN_c is uniform continuity with last-exit selection)**.** FAN_c *holds if and only if every pointwise continuous map from binary sequences to natural numbers is uniformly continuous and last-exit selection holds. In particular, under last-exit selection* FAN_c *is equivalent to the uniform continuity of these maps. No choice principle is used.*

Proof. Assume FAN_c. Theorem 13.15 gives uniform continuity. For selection, let the c-set of D be a bar, with uniform bound N. Every p has $\bar{p}(n)$ in the c-set for some $n \leq N$, so the search of Lemma 13.17 can always stop at N. Let $L(p)$ be one more than the last $m < N$ with $D(\bar{p}(m)) = 0$, or 0 if there is none. This is a map, computed by a finite search, and $L(p)$ is a last exit.

Conversely, let the c-set of D be a bar, and let L select last exits. L is pointwise continuous: take N with $\bar{p}(N)$ in the c-set; a sequence q with $\bar{q}(N) = \bar{p}(N)$ has $\bar{q}(N)$ in the c-set too, and the searches below N that find the last exits of p and of q read the same digits, so $L(q) = L(p)$ by uniqueness. By uniform continuity L is bounded, as in the proof of Theorem 13.15, by some B. Then $\bar{p}(B)$ lies in the c-set for every p: for a string t, the sequence $r = (\bar{p}(B)t)0^\omega$ has its last exit at most $B \leq |\bar{p}(B)t|$, so $D(\bar{p}(B)t) = D(\bar{r}(|\bar{p}(B)t|)) = 1$. □

Theorem 13.18 keeps the selection as a hypothesis and uses no unique choice of its own. Last-exit selection is an instance of unique choice. The principle of unique choice for relations from binary sequences to numbers states: if a relation R has, for every binary sequence p, exactly one number n with $R(p, n)$, then some map L from binary sequences to numbers has $R(p, L(p))$ for every p.

Corollary 13.19 (FAN_c is uniform continuity on Cantor space)**.** *Assume unique choice for relations from binary sequences to numbers. Then* FAN_c *holds if and only if every pointwise continuous map from binary sequences to natural numbers is uniformly continuous. The passage from* FAN_c *uses no choice principle; the converse uses unique choice once, to select last exits.*

Proof. The passage from FAN_c is item 1 of Theorem 13.15. For the converse, let the c-set of a decidable D be a bar, and let $R(p, n)$ say that n is a last exit of p. By Lemma 13.17 this relation has exactly one output for each binary sequence, so unique choice, applied to it, gives a map L with $L(p)$ a last exit of p for every p. This is last-exit selection for D, and Theorem 13.18 gives FAN_c. □

Under the reading of maps of Appendix C, which contains unique choice, FAN_c is therefore exactly the uniform continuity of pointwise continuous maps from binary sequences to numbers. This is Berger's equivalence. The c-set form is his, and in Bishop's constructive mathematics he proves FAN_c equivalent to that uniform continuity [31, Section 6, Proposition 8]; his earlier paper matches FAN_Δ with uniform continuity for maps that come with a pointwise continuous modulus [30].

Theorem 13.20 (Positive lower bounds on Cantor space)**.**

1. *Suppose that every uniformly continuous map F from binary sequences to points with $F(p) > 0$ for every p has a k with $F(p) \geq 1/(k+1)$ for every p. Then* FAN_Δ *holds. Unique choice, part of the reading of maps, is used once, to make the first hit of a bar a map.*

2. *Assume* FAN_Δ*. Let F map binary sequences to points, with $F(p) > 0$ for every p, and suppose that F comes with* endpoint names*, a map R from strings to regular sequences such that $R(s)$ names $F(s0^\omega)$, and with a* modulus*, a map μ from numbers to numbers such that $F(p)$ and $F(q)$ are within $1/(k+1)$ whenever $\bar{p}(\mu(k)) = \bar{q}(\mu(k))$. Then some k has $F(p) \geq 1/(k+1)$ for every p. No choice principle is used.*

Proof. For item 1, let D be a decidable bar with first hit h, a map as in Theorem 13.15, and put $F(p) = 1/(h(p)+1)$. It is positive. It is uniformly continuous: if $\bar{p}(k) = \bar{q}(k)$, either $h(p) \leq k$ and then $h(q) = h(p)$, or both first hits exceed k and both values lie in $(0, 1/(k+2)]$. A bound $F \geq 1/(k+1)$ gives $h(p) \leq k$ for every p.

For item 2, let a string s pass when some $k \leq |s|$ has $\mu(k) \leq |s|$ and stage $2|s|+1$ of $R(s)$ at least $3/(k+1)$; this is decidable. The passing strings form a bar. Given p, take g with $1/(g+1) < F(p)$ (Lemma 12.52, item 1), put $k = 8g+7$, $n = \max(k, \mu(k))$ and $s = \bar{p}(n)$. The sequences $s0^\omega$ and p agree in $n \geq \mu(k)$ digits, so $F(s0^\omega)$ is within $1/(k+1)$ of $F(p)$, and stage $2n+1$ of $R(s)$ is within $1/(2n+2) \leq 1/(k+1)$ of $F(s0^\omega)$ (Lemma 12.52, item 4). So that stage is at least $1/(g+1) - 2/(k+1) = 3/(4g+4) > 3/(k+1)$, and s passes. FAN_Δ gives N such that every p has a passing $\bar{p}(n)$ with $n \leq N$, with its $k \leq n$. Then $F(p)$ is within $1/(k+1)$ of $F(\bar{p}(n)0^\omega)$, which is at least $3/(k+1) - 1/(2n+2)$, so $F(p) \geq 1/(k+1) \geq 1/(N+1)$. □

Lemma 13.21 (Finitely many cylinders suffice)**.** *Assume* FAN_Δ*. If* $w_0', w_1', \ldots$ *is a sequence of strings such that every binary sequence extends some* w_i'*, then some* N *has every binary sequence extending some* w_i' *with* $i < N$*. No choice principle is used.*

Proof. Let s pass when one of $w_0', \ldots, w_{|s|}'$ is an initial segment of s. This is decidable, and it is a bar: if p extends w_i', then $\bar{p}(i + |w_i'|)$ passes. A uniform bound N' gives, for every p, some $n \le N'$ and $i \le n$ with w_i' an initial segment of $\bar{p}(n)$; take $N = N' + 1$. □

13.5 Compact sets presented with their nets

The proofs for $[0, 1]$ used two features of the interval: names of points at every precision, and the fact that a limit of named members is again a member. A presented compact set carries both as data.

Definition 13.22 (Presented compact sets)**.** A set K of points is *presented* when it is given with a *presentation*, which consists of the following data and properties.

1. *Nets.* For every n, a finite list of regular sequences, each naming a member of K, such that every member of K is within 2^{-n} of a point named in the list.
2. *Completeness for named sequences.* Whenever regular sequences $s^{(0)}, s^{(1)}, \ldots$ name members of K and form a Cauchy family in the sense of Theorem 12.25, the point named by their diagonal belongs to K.

The lists are handed over with the set; they are not merely asserted to exist. A presented set is decidably empty or inhabited: if the list at $n = 0$ is empty, no member exists, since each would lie within 1 of an entry. The empty set is presented by empty lists. The interval $[0, 1]$ is presented: at n take the names of the ratios $i/2^{n+1}$ for $0 \le i \le 2^{n+1}$. As in the proof of Lemma 12.57, cotransitivity at consecutive grid points places each y of $[0, 1]$ between two grid points $2/2^{n+1}$ apart, so y is within $1/2^{n+1}$ of the grid point between them. A Cauchy family naming points of $[0, 1]$ names a limit of those points (Theorem 12.25), and that limit lies in $[0, 1]$ by Lemma 12.52, item 2.

One decidable test of closeness is used repeatedly. Say that regular sequences s and t *pass the test at* g when $|s_j - t_j| \le 3/(4g + 4)$ at the stage $j = 8g + 7$. If their points are within $1/(2g + 2)$, they pass, by Lemma 12.23, since $1/(2g + 2) + 2/(j + 1) = 3/(4g + 4)$. If they pass, their points are within $1/(g + 1)$: regularity gives, at every stage n,

$$|s_n - t_n| \le |s_n - s_j| + |s_j - t_j| + |t_j - t_n| \le \frac{2}{n+1} + \frac{2}{j+1} + \frac{3}{4g+4} = \frac{1}{g+1} + \frac{2}{n+1},$$

and Lemma 12.23 applies.

Proposition 13.23 (Presented sets are closed)**.** *Let K be presented, and let x be a point such that for every k some member of K is within $1/(k+1)$ of x. Then x belongs to K. No choice principle is used.*

Proof. The conclusion is a proposition about one point, so we may use one regular sequence r naming x. For each n, let β_n be the first entry of the net at precision $2^{-(n+3)}$ that passes the test against r at $g = 2^{n+1} - 1$, and the first entry if none passes. Some entry passes: a member y of K lies within $2^{-(n+3)}$ of x, an entry lies within $2^{-(n+3)}$ of y, and that entry is within $2^{-(n+2)} = 1/(2g+2)$ of x. So each β_n names a member of K within $2^{-(n+1)}$ of x. For $m, n \geq k$ the points of β_m and β_n are within $2^{-(m+1)} + 2^{-(n+1)} \leq 1/(k+1)$, so the β_n form a Cauchy family. By completeness for named sequences the point of its diagonal lies in K. By Theorem 12.25 that point is a limit of the points of the β_n, which converge to x; limits are unique (Theorem 12.26), so it is x. □

A presented set with an inhabitant can be parametrized by Cantor space, uniformly and onto, by a map computed from its names. We build the map from *named nets* of a set K: a regular sequence b naming a member of K, and for every k a finite list L_k of regular sequences naming members of K, such that every member of K is within $1/(k+1)$ of a point named in L_k. A presented set with a named member has named nets: the member and the nets of its presentation, since $2^{-k} \leq 1/(k+1)$.

The construction. Fix a coding $\langle m, i \rangle$ of pairs of numbers by numbers, one-to-one, with a decoding that recovers m and i and assigns a pair to every number. Write $g(n) = 2^n - 1$, $B(n) = 4g(n+1) + 3$ and $M(n) = 4B(n) + 3$. For a binary sequence p, the *candidate* $\chi_n(p)$ at stage n is the entry of $L_{M(n)}$ at the largest index i, below the length of that list, with $p(\langle n, i \rangle) = 1$; it is the entry at index 0 if there is no such i, and b if the list is empty. The *branch* is $\rho_0 = \chi_0(p)$, and $\rho_{n+1} = \chi_{n+1}(p)$ if ρ_n and $\chi_{n+1}(p)$ pass the test at $g(n+1)$, while $\rho_{n+1} = \rho_n$ otherwise. Each ρ_n names a member of K, and consecutive terms have points within $1/(g(n+1)+1) = 2^{-(n+1)}$; adding these bounds, the points of ρ_n and ρ_{n+d} are within 2^{-n}, so the branch is a Cauchy family. The *Cantor map* sends p to the point $\varphi(p)$ named by the diagonal of the branch. Let $R(n)$ be the least number above every code $\langle m, i \rangle$ with $m \leq n$ and i below the length of $L_{M(m)}$, and put $\pi(k) = R(2k+1)$.

Lemma 13.24 (The Cantor map of named nets)**.** *Let K have named nets, with Cantor map φ and the numbers $\pi(k)$ as above.*

1. *If K is presented, $\varphi(p)$ belongs to K for every binary sequence p.*
2. *Every member of K is $\varphi(p)$ for some p.*
3. *If $\bar{p}(\pi(k)) = \bar{q}(\pi(k))$, then $\varphi(p)$ and $\varphi(q)$ are within $1/(k+1)$.*
4. *For every string s a regular sequence naming $\varphi(s0^\omega)$ is computed from s.*
5. *If K is presented, then for every n there is a j such that any two members x, y of K within $1/(j+1)$ are $\varphi(p)$ and $\varphi(q)$ for some p, q with $\bar{p}(n) = \bar{q}(n)$.*

No choice principle and no omniscience principle is used.

Proof. Item 1 is completeness for named sequences, applied to the branch. Item 4 holds because the diagonal of the branch of $s0^\omega$ is computed from s.

For item 2, let x be a member and r one regular sequence naming it. The *target sequence* σ_r has the digit 1 at a number whose decoding is (n, i) exactly when $L_{M(n)}$ has an entry at index i and that entry passes the test against r at $2B(n) + 1$. Some entry of $L_{M(n)}$ is within $1/(M(n) + 1) = 1/(4B(n) + 4)$ of x and passes; so the candidate $\chi_n(\sigma_r)$, the entry at the largest passing index, passes, and its point is within $1/(2B(n) + 2)$ of x. By induction the branch of σ_r accepts every candidate: if $\rho_n = \chi_n(\sigma_r)$, then ρ_n and $\chi_{n+1}(\sigma_r)$ are within $2/(2B(n) + 2) = 1/(4g(n + 1) + 4)$, which is at most $1/(2g(n + 1) + 2)$, so they pass the test at $g(n + 1)$. The points of the branch therefore converge to x, since $B(n) \geq n$. By Theorem 12.25 and uniqueness of limits, $\varphi(\sigma_r) = x$.

For item 3, the branch up to stage n is computed from the candidates at stages $m \leq n$, which read only the digits at the codes below $R(n)$. If $\bar{p}(R(n)) = \bar{q}(R(n))$, the branches agree up to stage n. By Theorem 12.25, $\varphi(p)$ and $\varphi(q)$ are each within $1/(n + 1)$ of the point of that common term ρ_n. With $n = 2k + 1$ they are within $1/(k + 1)$.

For item 5, let c be the largest first component among the decodings of the numbers below n, or 0 if $n = 0$, and put $j = 2B(c) + 1$. Let x, y be members within $1/(j + 1)$, named by r and t. The *hybrid* sequence σ takes the digit of σ_r at numbers whose decoding has first component at most c, and the digit of σ_t elsewhere. Then $\bar{\sigma}(n) = \bar{\sigma}_r(n)$ and $\varphi(\sigma_r) = x$. The candidates of σ are those of σ_r up to stage c and those of σ_t afterwards, so its branch agrees with that of σ_r up to stage c, where its point is within $1/(2B(c) + 2)$ of x, hence within $1/(B(c) + 1) = 1/(4g(c + 1) + 4)$ of y. The candidate at stage $c + 1$ is within $1/(2B(c + 1) + 2) \leq 1/(8g(c + 1) + 8)$ of y. The two are within $3/(8g(c + 1) + 8) < 1/(2g(c + 1) + 2)$, so the candidate is accepted, and from there on the branch accepts every candidate of σ_t, as in item 2. So $\varphi(\sigma) = y$, and $p = \sigma_r$, $q = \sigma$ serve. □

The map is far from one-to-one: digits at codes $\langle m, i\rangle$ with i beyond the length of $L_{M(m)}$ are never read.

13.6 Heine–Borel, positive bounds and infima for presented sets

With the Cantor map, the theorems for $[0, 1]$ extend to every presented set. For Heine–Borel, and for maps presented with names of their values and a modulus, every choice premise disappears, because the data that a choice would select are part of the presentations of the set and of the map. For bare uniformly continuous maps the passage from the fan theorem back to the bounds is proved with countable choice, or with the pullback data that it selects (Theorem 13.31).

Theorem 13.25 (Heine–Borel for presented sets)**.** *Every presented set is cover-compact if and only if* FAN_Δ *holds. Neither direction uses a choice principle or an omniscience principle.*

Proof. If every presented set is cover-compact, so is the presented $[0,1]$, and Theorem 13.6 gives FAN_Δ.

Conversely, assume FAN_Δ, let K be presented, and let (a_i, b_i) be a ratio cover of K. If K is empty, $N = 0$ serves. Otherwise the first entry of the net at $n = 0$ names a member, and Lemma 13.24 gives the Cantor map φ, onto K, with the numbers $\pi(k)$. For a string s and a number k let $e_k(s)$ be stage k of the computed name of $\varphi(s0^\omega)$; it is a ratio within $1/(k+1)$ of $\varphi(s0^\omega)$.

Enumerate all triples (r, s, i) of a number, a string and an index, and *flag* a triple when $\pi(2r+1) \le |s|$ and, with $e = e_{2r+1}(s)$,

$$a_i < e - \frac{1}{r+1} \quad \text{and} \quad e + \frac{1}{r+1} < b_i,$$

a comparison of ratios. If (r, s, i) is flagged and p extends s, then $\varphi(p)$ is within $1/(2r+2)$ of $\varphi(s0^\omega)$, which is within $1/(2r+2)$ of e; so $\varphi(p)$ lies in interval i. Every p extends the string of some flagged triple: $\varphi(p)$ lies in some interval i, and by Lemma 12.52, item 3, some k' has every point within $1/(k'+1)$ of $\varphi(p)$ in that interval; with $r = 2k'+1$ and $s = \bar{p}(\pi(2r+1))$, the ratios $e \pm 1/(r+1)$ are within $2/(r+1) = 1/(k'+1)$ of $\varphi(p)$, so (r, s, i) is flagged. Some triple is flagged, since K is inhabited, and we fix one, (r^*, s^*, i^*). List the triples in order, replacing each unflagged one by (r^*, s^*, i^*); this gives strings w_n' with labels $\lambda(n)$ such that every binary sequence extends some w_n', and every sequence extending w_n' is sent into interval $\lambda(n)$. By Lemma 13.21 some N' has every binary sequence extending some w_n' with $n < N'$. Let N be one more than the largest label $\lambda(n)$ with $n < N'$. Every member of K is $\varphi(p)$ for some p, and p extends some w_n' with $n < N'$, so the member lies in interval $\lambda(n) < N$. □

A positive lower bound turns into a finite subcover through one function.

Lemma 13.26 (The cover series)**.** *Let $c = (a_i, b_i)$ be a ratio cover. Put*

$$\theta_i(x) = \max\bigl(0, \min(1, x - a_i, b_i - x)\bigr), \qquad \omega_i = \frac{1}{i+1} - \frac{1}{i+2}, \qquad H_n = \sum_{i<n} \omega_i \theta_i.$$

1. *The points $H_0(x), H_1(x), \ldots$ have names computed from a name of x, and they form a Cauchy family; the point $f_c(x)$ named by its diagonal does not depend on the name of x. So f_c is a function on the points, and a name of $f_c(x)$ is computed from a name of x. Moreover f_c has a* modulus*: a map μ from numbers to numbers such that any two points within $1/(\mu(k)+1)$ have values within $1/(k+1)$.*

2. *If x lies in an interval of c, then $f_c(x) > 0$.*

3. *If $f_c(x) \ge 1/(k+1)$ for every member x of a set K, then for some N every member of K lies in one of the first N intervals of c.*

No choice principle is used.

Proof. Each θ_i is built from translation, negation, max and min, so by Lemma 12.52, item 5, points within $1/(j+1)$ have values within $1/(j+1)$; and $0 \le \theta_i \le 1$. Since $\sum_{n \le i < m} \omega_i = 1/(n+1) - 1/(m+1)$, the points $H_m(x)$ and $H_n(x)$ are within $1/(n+1)$ for $m \ge n$, and weighting the bounds for the θ_i shows that $H_n(x)$ and $H_n(y)$ are within $1/(j+1)$ whenever x and y are. Names of the $H_n(x)$ are computed from a name of x by the operations on regular sequences, so they form a Cauchy family, and its diagonal names a limit of the points $H_n(x)$ (Theorem 12.25). Two names of x give limits of the same points, which coincide (Theorem 12.26). The value $f_c(x)$ is within $1/(n+1)$ of $H_n(x)$; with $n = 4k+3$ and x, y within $1/(2k+2)$, the values are within $1/(4k+4) + 1/(2k+2) + 1/(4k+4) = 1/(k+1)$. So $\mu(k) = 2k+1$ is a modulus.

For item 2, if $a_i < x < b_i$, then $\min(1, x - a_i, b_i - x) > 0$ and $\theta_i(x) > 0$, by Lemma 12.52, item 6. So $H_{i+1}(x) \ge \omega_i \theta_i(x) > 0$, and $f_c(x) \ge H_{i+1}(x)$ by item 2 of that lemma, since $H_m(x) \ge H_{i+1}(x)$ for $m \ge i+1$.

For item 3, put $N = 2k+1$. Then $H_N(x) \ge f_c(x) - 1/(N+1) \ge 1/(2k+2) > 0$. A positive finite sum has a positive term: if $0 < u + v$, cotransitivity at the point u gives $0 < u$ or $u < u + v$, that is $v > 0$. So some $i < N$ has $\theta_i(x) > 0$; then the min is positive, so $x - a_i > 0$ and $b_i - x > 0$, and x lies in interval i. □

The cover series is an example of a map that comes with its data.

Definition 13.27 (Presented uniform maps)**.** Let K be a set of points. A *presented uniform map* on K is a function f on the points together with a *transformer*, a map T sending each regular sequence r that names a member of K to a regular sequence naming f of that member, and a *modulus*, a map μ from numbers to numbers such that members x, y of K within $1/(\mu(k)+1)$ have $f(x)$ and $f(y)$ within $1/(k+1)$.

A uniformly continuous function asserts a j for each k; a presented uniform map hands over the function μ and the names. By Lemma 13.26, the cover series f_c, with its names and its modulus, is a presented uniform map on every set.

Theorem 13.28 (Positive lower bounds for presented maps)**.** FAN_Δ *holds if and only if, for every presented set K and every presented uniform map f on K with $f(x) > 0$ at every member x, some k has $f(x) \ge 1/(k+1)$ for every member x of K. Neither direction uses a choice principle.*

Proof. Assume the bound principle, let K be presented, and let c be a ratio cover of K. The cover series f_c is a presented uniform map, positive at every member of K (Lemma 13.26, item 2). A bound $f_c \ge 1/(k+1)$ on K puts every member in one of the first N intervals for some N (item 3). So every presented set is cover-compact, and Theorem 13.25 gives FAN_Δ.

Conversely, assume FAN_Δ, and let f, with transformer T and modulus μ_f, be positive on the presented set K. If K is empty, any k serves. Otherwise take the Cantor map φ of Lemma 13.24 and put $F(p) = f(\varphi(p))$. It is positive. Its endpoint names are T applied to the computed names of the points $\varphi(s0^\omega)$, which are members of K, and $k \mapsto \pi(\mu_f(k))$ is a modulus. Theorem 13.20, item 2, gives k with $F \ge 1/(k+1)$, and since φ maps onto K, $f \ge 1/(k+1)$ on K. □

For a presented uniform map the infimum itself is constructed.

Theorem 13.29 (The infimum of a presented map)**.** *Let K be presented, let f be a presented uniform map on K with transformer T and modulus μ, and let b name a member of K. Let S_n be the stagewise minimum of $T(b)$ and of $T(r)$ for the entries r of the net at precision $2^{-\mu(n)}$. Then $S_0, S_1, \ldots$ is a Cauchy family, and the point m^* named by its diagonal is an* infimum *of f on K: $m^* \le f(x)$ for every member x, and every point a with $a \le f(x)$ for every member x satisfies $a \le m^*$. No choice principle and no omniscience principle is used.*

Proof. The stagewise minimum of finitely many regular sequences is regular and names the minimum of their points, as in the proof of Lemma 12.52, item 6. So S_n names the minimum of finitely many values of f at members of K, and every lower bound a of f on K satisfies $a \le S_n$. For a member x, some entry r of the net at precision $2^{-\mu(n)} \le 1/(\mu(n)+1)$ names a point within that distance of x, so f there is within $1/(n+1)$ of $f(x)$, and $S_n \le f(x) + 1/(n+1)$. Hence $S_n - 1/(n+1)$ is a lower bound of f on K, and $S_n - 1/(n+1) \le S_m$ for all m and n. For $m, n \ge k$ this makes S_m and S_n within $1/(k+1)$: the family is Cauchy. Its diagonal point m^* is within $1/(n+1)$ of S_n for every n (Theorem 12.25). At $n = 2k+1$, the point $S_n - 1/(n+1)$ lies below $f(x)$ and within $1/(k+1)$ of m^*, so $m^* \le f(x)$ by Lemma 12.52, item 2. A lower bound a lies below every S_k, each within $1/(k+1)$ of m^*, so $a \le m^*$ by the same item. □

The theorem gives no point at which the infimum is attained, and none can be had for free: the negative of the test function in the proof of Theorem 12.60 is a presented uniform map on $[0,1]$, so attained minima of presented maps would give LLPO. For $[0,1]$ and a uniformly continuous function, Theorem 12.58 uses countable choice at pairs of ratios to turn the cut of the supremum into a point; a presented map needs none.

Corollary 13.30 (Positive infima for presented maps)**.** FAN_Δ *holds if and only if every presented uniform map that is positive at every member of an inhabited presented set K has an infimum $m > 0$ on K. Neither direction uses a choice principle.*

Proof. Assume FAN_Δ. The statement to prove is a proposition, so one name of one member serves as b. Theorem 13.28 gives $f \ge 1/(k+1)$ on K, and the infimum m^* of Theorem 13.29, a greatest lower bound, satisfies $m^* \ge 1/(k+1) > 0$. Conversely, a positive infimum m exceeds some ratio $1/(k+1)$ (Lemma 12.52, item 1), which is then a lower bound; empty sets, decided by the net at $n = 0$, need nothing. So the bound principle of Theorem 13.28 holds, and with it FAN_Δ. □

For $[0,1]$ and uniformly continuous functions, the equivalence of positive infima with the decidable fan theorem FAN_Δ is due to Julian and Richman [37, Corollary 2.6], in Bishop's constructive mathematics. The presented forms extend it to every presented set and separate what the set supplies, its names and nets, from what the theorems ask of the map, its names and its modulus.

Bare maps. Without the data of the map, one direction still uses no choice. Write (LB) for the assertion that for every presented set K and every function f uniformly continuous on K with $f(x) > 0$ at every member, some k has $f \ge 1/(k+1)$ on K, and (INF) for the assertion that every such f on an inhabited presented K has an infimum $m > 0$ on K. The *pullback data* principle states: for every presented set K, all named nets of K with Cantor map φ, and every f uniformly continuous on K, the map $p \mapsto f(\varphi(p))$ has endpoint names and a modulus, as in Theorem 13.20.

Theorem 13.31 (Bare uniformly continuous maps)**.**

1. *Each of (LB) and (INF) implies* FAN_Δ*. No choice principle is used.*
2. *Assume the pullback data principle. Then (LB), (INF) and* FAN_Δ *are equivalent, with no further choice.*
3. *Assuming* AC_ω*, the pullback data principle holds, so (LB), (INF) and* FAN_Δ *are equivalent. Countable choice is used for regular sequences, to select endpoint names, and for numbers, to select a modulus.*
4. *Assume that every Cauchy sequence of points converges. Then every function uniformly continuous on an inhabited presented set has an infimum on it, and (INF) is equivalent to (LB).*

Proof. For item 1, the cover series of a cover of a presented set is uniformly continuous and positive there, and a lower bound gives a finite subcover (Lemma 13.26); so (LB) makes every presented set cover-compact, and Theorem 13.25 applies. As in the proof of Corollary 13.30, (INF) implies (LB).

For item 2, assume FAN_Δ and the pullback data. For (LB), let K be presented. If K is empty, which the net at $n = 0$ decides, every k serves. Otherwise take a named member and the Cantor map, and apply Theorem 13.20, item 2, to $F = f \circ \varphi$ with its endpoint names R and modulus μ; since φ maps onto K, the bound holds on K. For (INF), let S_n be the stagewise minimum of the $R(s)$ over the strings s of length $\mu(n)$. Every p has $F(p)$ within $1/(n+1)$ of $F(\bar{p}(\mu(n))0^\omega)$, so the argument of Theorem 13.29 shows that the diagonal of the S_n names an infimum of F, which is an infimum of f on K because φ maps into K and onto it. By (LB) it is at least some $1/(k+1)$. Item 1 gives the remaining implications.

For item 3, the strings are enumerated, and for each string s some regular sequence names $F(s0^\omega)$; countable choice gives R. For each k, uniform continuity of f on K gives some j, and then $\pi(j)$ works for F at k; countable choice gives μ. Item 2 finishes.

For item 4, let K be inhabited, with a member x_0, and let z_n be the minimum of $f(x_0)$ and of f at the points of the net at precision 2^{-n}: a sequence of points, defined from the data with no choice. Given k, uniform continuity gives one j, used inside this proof only; for $n \geq j$ every member x is within $1/(j+1)$ of a net point, so $z_n \leq f(x) + 1/(k+1)$, and for $m, n \geq j$ the minima z_m and z_n are within $1/(k+1)$. So (z_n) is Cauchy and has a limit L. As in the proof of Theorem 13.29, $L \leq f(x)$ for every member x, and every lower bound lies below every z_n, hence below L. Then a positive lower bound $1/(k+1)$ gives $L \geq 1/(k+1) > 0$, and a positive infimum gives a positive lower bound as in item 1. □

By Theorem 12.26, the hypothesis of item 4 follows from AC_ω, used once to select representatives.

13.7 Heine–Cantor for presented sets

Heine–Cantor for presented sets lies between FAN_Δ and the full fan theorem. Under two bridges between continuity on a presented set and continuity on Cantor space, it is exactly FAN_c.

Theorem 13.32 (Heine–Cantor for presented sets lies between FAN_Δ and FAN)**.**

1. *If every function pointwise continuous on a presented set is uniformly continuous on it, then* FAN_Δ *holds. No choice principle is used beyond the unique choice that defines the test function of Theorem 13.13.*

2. *The full fan theorem implies Heine–Cantor for presented sets and the assertion (LB) of Theorem 13.31. No choice principle is used.*

Proof. For item 1, the hypothesis applies to the presented $[0,1]$, and Theorem 13.13 finishes.

For item 2, let f be pointwise continuous on a presented set K, and fix k. If K is empty there is nothing to prove; otherwise take named nets and the Cantor map φ. Let B be the set of strings s for which there are a member z of K and a number $j \le |s|$ such that $\varphi(q)$ is within $1/(2j+2)$ of z for every q extending s, and every member y of K within $1/(j+1)$ of z has $f(y)$ within $1/(2k+2)$ of $f(z)$. This is a property of strings, not a decidable set. It is a bar: for a sequence p, let $z = \varphi(p)$, let j be given by continuity of f at z at the tolerance $1/(2k+2)$, and let $n = \max(j, \pi(2j+1))$; every q extending $\bar{p}(n)$ agrees with p in $\pi(2j+1)$ digits, so $\varphi(q)$ is within $1/(2j+2)$ of z, and $\bar{p}(n) \in B$. The full fan theorem gives N such that every p has an $n \le N$ with $\bar{p}(n) \in B$.

Let x and y be members within $1/(2N+2)$. By Lemma 13.24, $x = \varphi(p)$ for some p; take $n \le N$ with $\bar{p}(n) \in B$, and its z and $j \le n \le N$. Then x is within $1/(2j+2)$ of z, and y, within $1/(2N+2) \le 1/(2j+2)$ of x, is within $1/(j+1)$ of z. So $f(x)$ and $f(y)$ are both within $1/(2k+2)$ of $f(z)$, hence within $1/(k+1)$ of each other, and $2N+1$ serves at k.

For (LB), let f be uniformly continuous and positive on a presented K. If K is empty, which the net at $n = 0$ decides, every k serves. Otherwise take named nets and φ as before. Let B be the set of strings s for which some $g \le |s|$ has $f(\varphi(q)) \ge 1/(g+1)$ for every q extending s. It is a bar: given p, take g_0 with $1/(g_0+1) < f(\varphi(p))$, put $g = 2g_0 + 1$, let j be given by uniform continuity at the tolerance $1/(g+1)$, and let $n = \max(g, \pi(j))$. For q extending $\bar{p}(n)$,

$$f(\varphi(q)) \ge f(\varphi(p)) - \frac{1}{g+1} > \frac{1}{g_0+1} - \frac{1}{2g_0+2} = \frac{1}{g+1}.$$

The full fan theorem gives N, and every member $x = \varphi(p)$ has $f(x) \ge 1/(N+1)$. □

Definition 13.33 (Two bridges)**.** The *pullback approximation principle* states: for every presented set K, all named nets of K with Cantor map φ, every function f pointwise continuous on K, and every k, there is a pointwise continuous map a from binary sequences to numbers such that $a(p) = a(q)$ implies that $f(\varphi(p))$ and $f(\varphi(q))$ are within $1/(k+1)$.

A *Cantor presentation* of a set K is a map ψ from binary sequences into K and onto K, together with a map π' from numbers to numbers such that $\psi(p)$ and $\psi(q)$ are within $1/(k+1)$ whenever

$\bar{p}(\pi'(k)) = \bar{q}(\pi'(k))$. It *reflects prefixes* when for every n some j has: $\psi(p)$ and $\psi(q)$ within $1/(j+1)$ force $\bar{p}(n) = \bar{q}(n)$. It *realizes last exits* when for every decidable D whose c-set is a bar some function g on the points has, for every p, $g(\psi(p)) = n$ for a last exit n of p. The *last-exit realization principle* states that some presented set has a Cantor presentation that reflects prefixes and realizes last exits.

Theorem 13.34 (Heine–Cantor and FAN_c through the bridges)**.**

1. *Assume* FAN_c*. Then Heine–Cantor for presented sets holds if and only if the pullback approximation principle holds.*
2. *Under the last-exit realization principle, Heine–Cantor for presented sets implies* FAN_c*.*
3. *Under both principles, Heine–Cantor for presented sets is equivalent to* FAN_c*.*

No choice principle is used.

Proof. For item 1, assume Heine–Cantor for presented sets, and let K, its named nets with Cantor map φ, f and k be given. Some j has members within $1/(j+1)$ with values within $1/(k+1)$. Let $a(p)$ be the index of $\bar{p}(\pi(j))$ in the list $w_0, w_1, \ldots$. It is pointwise continuous, and $a(p) = a(q)$ puts $\varphi(p)$ and $\varphi(q)$ within $1/(j+1)$, so their values are within $1/(k+1)$. Conversely, assume the pullback approximation, let f be pointwise continuous on a presented K, and fix k. If K is empty there is nothing to prove; otherwise take named nets, the Cantor map φ, and an approximation a at k. By Theorem 13.15, a is uniformly continuous, with some bound n; so $\bar{p}(n) = \bar{q}(n)$ gives $a(p) = a(q)$ and values $f(\varphi(p))$, $f(\varphi(q))$ within $1/(k+1)$. By Lemma 13.24, item 5, some j has any two members within $1/(j+1)$ of the form $\varphi(p)$, $\varphi(q)$ with $\bar{p}(n) = \bar{q}(n)$. So j serves at k.

For item 2, let K, ψ and π' be as in the realization principle, assume Heine–Cantor for presented sets, let the c-set of a decidable D be a bar, and let g realize its last exits. The function g is pointwise continuous on K. A member x is $\psi(p)$ for some p; take N with $\bar{p}(N)$ in the c-set, and j from prefix reflection at N. A member y within $1/(j+1)$ of x is $\psi(q)$ for some q with $\bar{q}(N) = \bar{p}(N)$, and, as in the proof of Theorem 13.18, p and q have the same last exit; so $g(y) = g(x)$. By Heine–Cantor, g is uniformly continuous on K; at the tolerance $\frac{1}{2}$ some j has members within $1/(j+1)$ with values within $\frac{1}{2}$. If $\bar{p}(\pi'(j)) = \bar{q}(\pi'(j))$, then $\psi(p)$ and $\psi(q)$ are within $1/(j+1)$, so the last exits of p and q, numbers whose points are within $\frac{1}{2}$, are equal. So the last exit of p is at most the largest last exit B of the sequences $s0^\omega$ with $|s| = \pi'(j)$, and as in the proof of Theorem 13.18, $\bar{p}(B)$ lies in the c-set for every p. Item 3 combines the first two. □

Proposition 13.35 (What the realization bridge requires)**.**

1. *The last-exit realization principle implies last-exit selection. No choice principle is used.*
2. *If a Cantor presentation ψ sends two binary sequences p and q with $\bar{p}(n) \neq \bar{q}(n)$ for some n to the same point, it neither reflects prefixes nor realizes last exits.*

Proof. For item 1, let g realize last exits on ψ. A point equal to a natural number n determines n: every regular sequence s naming it has $|s_7 - n| \le 1/8$ (Lemma 12.52, item 4), so n is the integer nearest to s_7 for every name. So $L(p)$, the number with $g(\psi(p)) = L(p)$, is computed from that point, and L selects last exits.

For item 2, if ψ reflected prefixes, the j given at n would force $\bar{p}(n) = \bar{q}(n)$, since $\psi(p) = \psi(q)$. Suppose ψ realizes last exits, and let D reject exactly the string $s = \bar{p}(n)$. Its c-set is a bar, since every string of length $n + 1$ lies in it. The last exit of p is $n + 1$, because $D(\bar{p}(m)) = 0$ only at $m = n$, and the last exit of q is 0, because $\bar{q}(n) \neq s$. A realizing g would take the values $n + 1$ and 0 at the same point. □

So a bridging presentation must be one-to-one, and the Cantor map of Lemma 13.24 is not. We do not decide here whether FAN_c gives the pullback approximation, whether a bridging presentation exists, or whether FAN_Δ alone gives Heine–Cantor on $[0, 1]$. So this chapter locates the price of Heine–Cantor for presented sets only between FAN_Δ and FAN.

13.8 Bishop compactness

Bishop's compact sets have finite nets and are complete [38, 14]. Read on $\mathbb{R}_\delta$ without names, the nets exist at each precision without being handed over, and completeness concerns sequences of points rather than named families.

Definition 13.36 (Bishop compactness)**.** A set K of points has *internal nets* when for every k there is a finite list of members of K such that every member of K is within $1/(k + 1)$ of one of them. It is *complete* when every Cauchy sequence of members of K converges to a member of K. It is *Bishop compact* when it has internal nets and is complete.

Theorem 13.37 (Forgetting the presentation costs Cauchy completeness)**.** *Every presented set is Bishop compact if and only if every Cauchy sequence of points of $\mathbb{R}_\delta$ converges. Neither direction uses a choice principle. By Theorem 12.26, every presented set is therefore Bishop compact under AC_ω, with countable choice used once, to select representatives.*

Proof. Assume that Cauchy sequences converge, and let K be presented. The net at precision $2^{-k} \le 1/(k + 1)$ is a list of members, so K has internal nets. A Cauchy sequence of members converges to some L; every tolerance has a member within it of L, so L belongs to K by Proposition 13.23.

Conversely, assume that every presented set is Bishop compact. Then the presented $[0, 1]$ is complete. Let (x_n) be a Cauchy sequence of points, take N such that x_m and x_n are within 1 for $m, n \ge N$, put $a = x_N$, and let $y_n = (x_{\max(n,N)} - a + 1)/2$. Each y_n lies in $[0, 1]$, and (y_n) is Cauchy, so it converges to some L in $[0, 1]$. Then (x_n) converges to $2L + a - 1$. The index N is used inside this proof only. □

The proof shows that $[0, 1]$ is Bishop compact exactly when every Cauchy sequence of points converges.

Assertion	Price	Choice
$[0,1]$ is cover-compact (Theorem 13.6)	FAN_Δ	not used
Every presented set is cover-compact (Theorem 13.25)	FAN_Δ	not used
A presented map positive on a presented set has a positive lower bound, and on an inhabited one a positive infimum (Theorem 13.28, Corollary 13.30)	FAN_Δ	not used
… for a bare map (Theorem 13.31)	FAN_Δ	in the converse
Every Bishop compact set is cover-compact (Theorem 13.39)	FAN_Δ	AC_ω, both directions
Pointwise continuous maps from binary sequences to numbers are uniformly continuous, and last exits can be selected (Theorem 13.18)	FAN_c	not used
Pointwise continuous maps from binary sequences to numbers are uniformly continuous (Theorems 13.15 and 13.18, Lemma 13.17, Corollary 13.19)	FAN_c	unique choice, in the converse
Functions pointwise continuous on $[0,1]$ are uniformly continuous there (Theorem 13.13)	implies FAN_Δ	unique choice
… on every presented set (Theorem 13.32)	between FAN_Δ and FAN	unique choice
Positive uniformly continuous maps from binary sequences to points have positive lower bounds (Theorem 13.20)	implies FAN_Δ	unique choice
A presented map on an inhabited presented set has an infimum (Theorem 13.29)	none	not used
A bare map on an inhabited presented set has an infimum (Theorem 13.31)	none	used once, for representatives
Every presented set is Bishop compact (Theorem 13.37)	none	used once, for representatives
Every two-point set is complete (Proposition 13.38)	LLPO	in the converse
Uniformly continuous images of Bishop compact sets are Bishop compact (Proposition 13.38)	implies LLPO	not used
Weak König's lemma for decidable binary trees (Proposition 13.9)	$\mathrm{LLPO}_{\mathrm{seq}}$	not used
FAN_Δ (Corollary 13.10, Theorem 13.11)	below $\mathrm{LLPO}_{\mathrm{seq}}$; not from excluded middle without unique choice	from WLPO, with unique choice

Table 13.1. Compactness on the line built from ratios and on Cantor space. A presented map is a presented uniform map, and a bare map a function uniformly continuous on the set. "Price" is the principle equivalent to the assertion; "implies" and "between" record the proved implications, "below P" means implied by P, and "none" means that no omniscience or fan principle is needed. The last column says which choice principle is used and where ("in the converse": on the way from the principle back to the assertion); unique choice, part of the reading of maps, makes a uniquely determined value, such as the first hit of a bar or the last exit of a sequence, a map. The theorems named state the conditional converses.

Proposition 13.38 (Images of Bishop compact sets)**.**

1. *If K has internal nets and f is uniformly continuous on K, the image $\{f(x) : x \in K\}$ has internal nets. No choice principle is used.*

2. *If every two-point set $\{x, y\}$ is complete, LLPO holds. Hence, if the image of every Bishop compact set under every function uniformly continuous on it is Bishop compact, LLPO holds. No choice principle is used.*

3. *Assuming LLPO and AC_ω, every two-point set is complete; countable choice is used once, through Theorem 12.26.*

Proof. For item 1, given k, uniform continuity gives j; the images of an internal net at $1/(j+1)$ are members of the image, and every $f(x)$ is within $1/(k+1)$ of one of them.

For item 2, let x and y be named by s and t. Put $z_n = y$ if $t_{2n+1} \geq s_{2n+1}$ and $z_n = x$ otherwise, a sequence of members of $\{x, y\}$ defined by comparing ratios. The point $\max(x, y)$ is named by the stagewise maximum, whose stage $2n+1$ is the stage $2n+1$ of the name of z_n; by Lemma 12.23, two points whose names agree at a stage $2n+1$ are within $1/(n+1)$. So (z_n) converges to $\max(x, y)$. If $\{x, y\}$ is complete, the limit is x or y, and limits are unique, so $\max(x, y) = x$ or $\max(x, y) = y$: that is, $y \leq x$ or $x \leq y$. With $y = 0$ this is the dichotomy of Corollary 12.18, which is LLPO. For the second assertion, $\{0, 1\}$ is Bishop compact with no choice (a Cauchy sequence in it is eventually constant), and $z \mapsto x + z(y - x)$ is uniformly continuous, since $|y - x|$ lies below some integer, with image $\{x, y\}$.

For item 3, a Cauchy sequence in $\{x, y\}$ converges to some L, by Theorem 12.26. LLPO, through Corollary 12.18, gives $|L-x| \leq |L-y|$ or $|L-y| \leq |L-x|$. In the first case, for every k some term within $1/(k+1)$ of L is x or y, and either way $|L - x| \leq 1/(k+1)$; so $L = x$ by Corollary 12.24. The second case gives $L = y$. □

Under countable choice the two notions of compactness meet.

Theorem 13.39 (Bishop compactness under countable choice)**.**

1. *If every function uniformly continuous on a Bishop compact set and positive at each of its members has a lower bound $1/(k+1)$ there, then every Bishop compact set is cover-compact. No choice principle is used.*

Assume AC_ω. Then:

2. *a set has a presentation if and only if it is Bishop compact;*

3. *every Bishop compact set is cover-compact if and only if FAN_Δ holds;*

4. *the lower-bound principle of item 1 is equivalent to FAN_Δ;*

5. *the principle that every function uniformly continuous and positive on an inhabited Bishop compact set has an infimum $m > 0$ there is equivalent to FAN_Δ.*

Proof. Item 1 is Lemma 13.26: the cover series of a cover of K is uniformly continuous and positive on K, and a lower bound gives a finite subcover.

For item 2, a presented set is Bishop compact by Theorem 13.37. Conversely, let K be Bishop compact. The internal net at $k = 0$ decides whether K is empty; an empty set is presented by empty lists. Otherwise, for each n there is a finite list of members within 2^{-n} of every member, and each member has a name, so there is a list of names with the same property; AC_ω selects such a list for every n. If regular sequences naming members form a Cauchy family, their points form a Cauchy sequence of members, which converges to a member; the diagonal names a limit of the same sequence (Theorem 12.25), so it names that member. Items 3 to 5 use countable choice only through item 2 and Theorems 13.37 and 13.31.

For item 3, if FAN_Δ holds, a Bishop compact set is presented by item 2, hence cover-compact by Theorem 13.25. Conversely, $[0, 1]$ is presented, hence Bishop compact by item 2; if it is cover-compact, Theorem 13.6 gives FAN_Δ.

For item 4, the principle makes Bishop compact sets cover-compact (item 1), and item 3 gives FAN_Δ; for item 5, a positive infimum gives a positive lower bound, and empty sets need nothing. The converses are Theorem 13.31, item 3, applied to the presentation of item 2. □

The choice-free theorems on $\mathbb{R}_\delta$ belong to presented sets; Bishop's notion reaches the same prices through countable choice (Table 13.1).

Questions to work through

1. In the proof of Theorem 13.3, give a decidable bar whose complement is not a tree. Why does a length with no member of A bound the bar D itself?

2. In the proof of Theorem 13.7, where would an unbounded search for an empty level of T need Markov's principle, and how does the path through the padded tree avoid it?

3. List the members of the tree H in the proof of Theorem 13.6 for the cover of the opening example, and say which interval catches each of the sixteen pieces of length 4.

4. Explain why a search finds the first hit of a decidable bar, while the last exit of a c-set is found by a search only once FAN_c supplies a bound (Theorems 13.15 and 13.18).

5. Show that $-h$, for the test function h of Theorem 12.60, is a presented uniform map on $[0, 1]$, and that attained infima of presented maps give LLPO.

Chapter 14

What linear algebra costs

Suppose an instrument reports a few vectors, each coordinate a point of the line $\mathbb{R}_\delta$ that we can know to any precision we ask for. Gaussian elimination would reduce the list to a basis of the subspace it spans: find an entry that is not zero, divide by it, clear its column, and repeat. Every step asks whether an entry is zero. Over a field with decidable equality, a discrete field in the sense of constructive algebra [39], that question is answered by computation. Chapter 12 priced it on the line: deciding $x = \eta(0) \vee x \neq \eta(0)$ for every point is WLPO (Corollary 12.18). Does linear algebra on the line cost exactly its zero tests, or does elimination need more?

Take the smallest case. Two points a and d are two vectors (a) and (d) of the one-dimensional space $\mathbb{R}^1_\delta$. A basis of their span drawn from them is one of three things: no vector, when both are zero; the vector (a) alone, when a is not zero and d is a multiple $c\,a$; or (d) alone, in the mirror case. Both together are never independent, because $d \cdot (a) + (-a) \cdot (d) = 0$. So selecting a basis from two measured numbers means knowing that both vanish, or naming one that is not zero and divides the other. With $a = d = x$, the selection decides $x = \eta(0) \vee x \neq \eta(0)$.

This chapter shows that the smallest case carries the whole subject. Every finite family of vectors in every $\mathbb{R}^n_\delta$ has a basis drawn from it exactly when every pair of points can be divided in this sense (Theorem 14.13), and the same holds for rank and nullity (Theorem 14.14). Each of these is equivalent to WLPO (Theorem 14.16), when independence means that only the zero combination vanishes (Definition 14.1). The passage from linear algebra to WLPO uses no choice. The passage back uses the principle of unique choice to collect a table of WLPO decisions into a sequence; in this book unique choice is part of the reading of maps (Section 12.9), so under that reading the price of linear algebra on the line is exactly WLPO.

The proof that WLPO suffices divides one point by another, and division has prices of its own. Dividing b by a when $|b| \leq |a|$, with a allowed to be zero, costs exactly LLPO. Dividing by a point that is merely known not to be zero, again with $|b| \leq |a|$, costs exactly a principle that follows from LLPO and from $\mathrm{MP}_{\mathrm{bin}}$: the principle $\mathrm{MP}^\vee$, which turns $x \neq \eta(0)$ into $x \geq \eta(0) \vee x \leq \eta(0)$. In both cases the passage to the principle uses no choice; the passage back uses countable choice for LLPO and unique choice for $\mathrm{MP}^\vee$. A table at the end of the chapter collects the prices, and the last section states what remains open when unique choice is not available.

Some conventions hold throughout. We write 0 and 1 for the points $\eta(0)$ and $\eta(1)$, and a ratio q for the point $\eta(q)$ when it is used as a point. As in Chapter 12, $x \neq y$ means $\neg(x = y)$; a point is *apart from* 0 when it is positive or negative, which by item 3 of Proposition 12.19 is exactly when it has a multiplicative inverse; and $|x| = \max(x, -x)$. A vector of $\mathbb{R}^n_\delta$ is an n-tuple of points, with sums and multiples taken coordinatewise. When we say that a proof uses *no choice*, we mean that it uses neither countable choice nor unique choice: every sequence and every table it uses is computed from the data by an explicit rule, and a search that is proved to terminate counts as an explicit rule. Where a proof does use unique choice or countable choice, we say so at that step.

14.1 Selections, bases and three kinds of division

A basis drawn from a finite family is recorded by saying which members are kept. This is a finite string of bits, so it can be named without any decision about the points.

Definition 14.1 (Selections and bases)**.** Let $v_0, \dots, v_{k-1}$ be vectors of $\mathbb{R}^n_\delta$. A *selection* is a string $b \in \{0,1\}^k$; it *selects* v_i when $b_i = 1$. For points $c_0, \dots, c_{k-1}$, the combination of the selected vectors is $\sum_{b_i=1} c_i v_i$.

1. The selection is *independent* when only the zero combination vanishes: whenever $\sum_{b_i=1} c_i v_i = 0$, every selected i has $c_i = 0$.
2. It *spans* the family when every v_i is a combination of the selected vectors.
3. It is a *basis selection* when it is independent and spans the family.

The family is *independent* when the selection of all its members is independent.

Independence here is the negative form, "only the zero combination vanishes." It does not ask that a combination with some coefficient apart from 0 have a coordinate apart from 0. Constructive analysis calls that stronger form linear independence, and studies a metric form of independence beside it [41]; we use neither. The prices of this chapter depend on this choice. With the stronger form, a basis selection of the single vector (x) decides $x = 0$ or x apart from 0: if nothing is selected, spanning gives $x = 0$; if (x) is selected, the combination $1 \cdot (x)$ has a coefficient apart from 0, so x is apart from 0. Deciding this for every point is LPO, by the second row of Table 12.2. A basis selection of a family is a basis, drawn from the family, of the subspace the family spans; so the assertion that every finite family of vectors in every $\mathbb{R}^n_\delta$ has a basis selection says that every finitely generated subspace of $\mathbb{R}^n_\delta$ has a basis drawn from its generators.

An $n \times m$ matrix A has entries A_{ij} for $i < n$ and $j < m$; its columns $\mathrm{col}_j = (A_{0j}, \dots, A_{n-1,j})$ are vectors of $\mathbb{R}^n_\delta$, and it sends $x \in \mathbb{R}^m_\delta$ to Ax, with $(Ax)_i = \sum_j A_{ij} x_j$. Write $e_0, \dots, e_{m-1}$ for the unit vectors of $\mathbb{R}^m_\delta$.

Definition 14.2 (Rank and nullity)**.** *Rank–nullity data* for an $n \times m$ matrix A consist of a selection $b \in \{0,1\}^m$ of its columns and vectors $K_0, \dots, K_{m-1}$ of $\mathbb{R}^m_\delta$ such that:

(i) b is a basis selection of the columns of A;

(ii) $AK_j = 0$ for every j with $b_j = 0$;

(iii) the selection of the K_j with $b_j = 0$ is independent;

(iv) every x with $Ax = 0$ is a combination of the K_j with $b_j = 0$.

Rank–nullity is the assertion that every matrix over $\mathbb{R}_\delta$ has rank–nullity data.

The selected columns are then a basis of the column space and the unselected K_j a basis of the kernel. The two selections are complementary, so the dimension of the column space and the dimension of the kernel add up to m. The vectors K_j with $b_j = 1$ play no role.

Definition 14.3 (Three kinds of division)**.** We consider three statements about dividing one point by another.

(a) *Scalar pair division*: for all points a and b, either $a = 0$ and $b = 0$; or $a \neq 0$ and $ac = b$ for some point c; or $b \neq 0$ and $bc = a$ for some point c.

(b) *Dominated division*: for all points a and b with $|b| \leq |a|$ there is a point c with $|c| \leq 1$ and $ac = b$.

(c) *Dominated division by nonzero points*: for all points a and b with $a \neq 0$ and $|b| \leq |a|$ there is a point c with $ac = b$.

In (a) and (c) the divisor is only known not to be zero; it need not be apart from 0, and no bound on the quotient is asked for. In (b) the divisor may be zero; then $b = 0$ and any c with $|c| \leq 1$ works, but the statement does not tell us whether a is zero. The bound $|b| \leq |a|$ in (c) matters: dividing 1 by every nonzero point is the inversion of the eighth row of Table 12.2, which costs $\mathrm{MP}_{\mathrm{bin}}$.

We use some facts about absolute values; each follows from Proposition 12.19 and Lemma 12.40. The maximum $|x|$ is an upper bound of x and $-x$. It is not negative: if $|x| < 0$, then $x < 0$ and $-x < 0$, and $0 < x < 0$ is impossible. If $x \geq 0$, then $-x \leq 0 \leq x$, so x is an upper bound of x and $-x$ and $|x| = x$; likewise $x \leq 0$ gives $|x| = -x$. Conversely $|x| = x$ gives $x \geq 0$, and $|x| = -x$ gives $x \leq 0$. If $|x| = 0$, then $x \leq 0$ and $-x \leq 0$, so $x = 0$ by item 2; hence $x \neq 0$ gives $|x| \neq 0$. If x is apart from 0, then $|x| > 0$; conversely $|x| > 0$ makes x positive or $-x$ positive, by the strict form of item 5. The equations $|xy| = |x|\,|y|$, $|-x| = |x|$ and $\big||x|\big| = |x|$ follow by cases on the signs; an equation proved by cases on the order of finitely many points needs no principle (Remark 12.20). Finally, if $0 \leq c$ and $x \leq y$, then $cx \leq cy$: the product of two nonnegative points is nonnegative, as in fact (iii) of the proof of Lemma F.12, and $cy - cx = c(y - x)$.

14.2 Cancelling without inverting

A point that is only known not to be zero cannot be inverted without $\mathrm{MP}_{\mathrm{bin}}$. It can still be cancelled. The lemmas of this section do the arithmetic of the rest of the chapter, and none of them uses an omniscience principle or a choice principle.

Lemma 14.4 (Cancelling a nonzero factor)**.** *If $a \neq 0$ and $ax = ay$, then $x = y$.*

Proof. Suppose $x - y$ were apart from 0. It would have an inverse u, and then $a = a(x - y)u = (ax - ay)u = 0$, contrary to $a \neq 0$. So $x - y$ is neither positive nor negative, and Lemma F.8(c) gives $x - y = 0$, that is $x = y$. □

Lemma 14.5 (Cancelling a nonnegative nonzero factor)**.** *If $0 \leq a$, $a \neq 0$ and $ax \leq ay$, then $x \leq y$.*

Proof. Suppose $y < x$. Then $d = x - y$ is positive and has an inverse u. The point u is not negative, since $u < 0$ would make $du = 1$ negative. From $ax \leq ay$ we get $ad \leq 0$, and multiplying by $u \geq 0$ gives $a = adu \leq 0$. With $0 \leq a$ this gives $a = 0$ by item 2 of Proposition 12.19, contrary to $a \neq 0$. So $\neg(y < x)$, which is $x \leq y$. □

Lemma 14.6 (Quotients of dominated numerators). *If $a \neq 0$, $|b| \leq |a|$ and $ac = b$, then $|c| \leq 1$.*

Proof. We have $|a|\,|c| = |ac| = |b| \leq |a| = |a| \cdot 1$, with $|a| \geq 0$ and $|a| \neq 0$. Lemma 14.5 gives $|c| \leq 1$. □

Lemma 14.7 (Reversing a dominated quotient). *If $b \neq 0$, $|b| \leq |a|$ and $bc = a$, then $ad = b$ for some point d.*

Proof. Here $|b| \cdot 1 = |b| \leq |a| = |b|\,|c|$, so $1 \leq |c|$ by Lemma 14.5. Cotransitivity of $0 < 1$ at $|c|$ gives $0 < |c|$ or $|c| < 1$, and the second is excluded by $1 \leq |c|$. So $|c| > 0$, and c is apart from 0. With d its inverse (item 3 of Proposition 12.19), $ad = bcd = b$. □

Only the quotient c is inverted, never the nonzero point b. The next lemma extracts a sign from a quotient by an absolute value, again with no bound on the quotient and no decision whether x is zero.

Lemma 14.8 (A weak sign from an absolute quotient). *If $|x|\,c = x$ for some point c, then $x \geq 0$ or $x \leq 0$.*

Proof. Cotransitivity of $-1 < 1$ at c gives $-1 < c$ or $c < 1$. In the first case $x \geq 0$, that is $\neg(x < 0)$: if $x < 0$, then $|x| = -x$ is positive and $(-x)c = x = (-x)(-1)$, so cancelling the invertible point $-x$ gives $c = -1$, contrary to $-1 < c$. In the second case $x \leq 0$ by the mirror argument: if $x > 0$, then $xc = x = x \cdot 1$ gives $c = 1$, contrary to $c < 1$. □

Lemma 14.9 (Absolute quotients of nonzero points). *If $x \neq 0$ and $|x|\,c = x$, then c is apart from 0.*

Proof. Taking absolute values, $|x|\,|c| = \big||x|c\big| = |x| = |x| \cdot 1$. Since $|x| \neq 0$, Lemma 14.4 gives $|c| = 1$, so $|c| > 0$ and c is apart from 0. □

Lemma 14.10 (The affine quotient). *If $0 \leq a \leq b$, $b \neq 0$ and $bc = a$, then $0 \leq c \leq 1$. Any two such points c are equal.*

Proof. Here $b \geq 0$. From $b \cdot 0 = 0 \leq a = bc$ and $bc = a \leq b = b \cdot 1$, Lemma 14.5 gives $0 \leq c$ and $c \leq 1$. If also $bc' = a$, then $bc = bc'$ and Lemma 14.4 gives $c = c'$. □

14.3 A basis is a sequence of divisions

Gaussian elimination needs a pivot in each column it treats. We isolate exactly what a pivot provides: an entry that is not zero, together with the quotients of the other entries by it.

Lemma 14.11 (Pivots give bases)**.** *An independent selection never selects a zero vector. Moreover, suppose every finite list $a_0, \dots, a_m$ of points is either all zero or has a* pivot*: an index p with $a_p \neq 0$ and points $q_0, \dots, q_m$ with $a_p q_i = a_i$ for every i. Then every finite family of vectors in every $\mathbb{R}_\delta^n$ has a basis selection. No omniscience principle and no choice is used.*

Proof. For the first statement, suppose an independent selection b selects $v_p = 0$. The combination with $c_p = 1$ and every other coefficient 0 is $v_p = 0$, so independence gives $1 = 0$, which is false. So $b_p = 0$.

For the second we argue by induction on the number n of coordinates, for all families at once. When $n = 0$, select nothing: every vector is the empty tuple, which is the empty combination, and the empty selection is independent because it selects nothing. Let $n \geq 0$, and let $v_0, \dots, v_{k-1}$ be vectors of $\mathbb{R}_\delta^{n+1}$. If $k = 0$ there is nothing to select. Otherwise write a_i for the first coordinate of v_i and $\bar{v}_i \in \mathbb{R}_\delta^n$ for the remaining ones, and apply the hypothesis to $a_0, \dots, a_{k-1}$.

If every a_i is 0, take a basis selection of $\bar{v}_0, \dots, \bar{v}_{k-1}$, given by the induction hypothesis. It is a basis selection of the v_i, because first coordinates contribute 0 to every combination.

Otherwise let p be a pivot, with $a_p q_i = a_i$. Then $q_p = 1$, by Lemma 14.4 applied to $a_p q_p = a_p = a_p \cdot 1$. Put

$$w_i = \bar{v}_i - q_i \bar{v}_p \in \mathbb{R}_\delta^n,$$

so that $w_p = 0$. The induction hypothesis gives a basis selection b of $w_0, \dots, w_{k-1}$, and $b_p = 0$ by the first statement. Let B be b with p added. We show that B is a basis selection of the v_i.

Independence. Suppose $\sum_B c_i v_i = 0$, and put $S = c_p + \sum_b c_i q_i$. The first coordinate reads $c_p a_p + \sum_b c_i q_i a_p = S a_p = 0 = a_p \cdot 0$, so $S = 0$ by Lemma 14.4. The other coordinates read $\sum_B c_i \bar{v}_i = 0$, and therefore

$$\sum_b c_i w_i = \sum_B c_i \bar{v}_i - S \bar{v}_p = 0.$$

Independence of b gives $c_i = 0$ for every i selected by b, and then $c_p = S = 0$.

Spanning. Fix i, and take coefficients c with $w_i = \sum_b c_l w_l$. Put $S = \sum_b c_l q_l$, and use the coefficients c_l for l selected by b and $q_i - S$ at p. The first coordinate of the combination is $(q_i - S)a_p + \sum_b c_l q_l a_p = q_i a_p = a_i$. The other coordinates are

$$(q_i - S)\bar{v}_p + \sum_b c_l \big(w_l + q_l \bar{v}_p\big) = w_i + q_i \bar{v}_p = \bar{v}_i.$$

So the combination is v_i. Only the cancellation of Lemma 14.4 was used; the pivot a_p is never inverted. Every case distinction is on a bit or on an alternative already supplied. □

Lemma 14.12 (Pivots from scalar pair division)**.** *Scalar pair division gives the hypothesis of Lemma 14.11: every finite list $a_0, \dots, a_m$ of points is all zero or has a pivot. No omniscience principle and no choice is used.*

Proof. By induction on m. For one entry, divide the pair $(a_0, 0)$. The first alternative says $a_0 = 0$. The second says $a_0 \neq 0$, and then $p = 0$ is a pivot with $q_0 = 1$. The third would need $0 \neq 0$.

For $a_0, a_1, \dots, a_{m+1}$, the induction hypothesis applies to $a_1, \dots, a_{m+1}$. If these are all zero, treat a_0 as before: either the whole list is zero, or 0 is a pivot with coefficients $1, 0, \dots, 0$. If they have a pivot p with coefficients $q_1, \dots, q_{m+1}$, divide the pair (a_0, a_p). Both zero contradicts $a_p \neq 0$. If $a_0 \neq 0$ and $a_0 d = a_p$, then 0 is a pivot with coefficients $1, dq_1, \dots, dq_{m+1}$, since $a_0 dq_j = a_p q_j = a_j$. If $a_p \neq 0$ and $a_p d = a_0$, then p is a pivot of the whole list with coefficient d at 0. □

Theorem 14.13 (Basis selection is scalar pair division)**.** *The following are equivalent:*

1. *every finite family of vectors in every $\mathbb{R}^n_\delta$ has a basis selection;*
2. *scalar pair division.*

Neither direction uses an omniscience principle or a choice principle.

Proof. Lemmas 14.12 and 14.11 give the first statement from the second.

Conversely, let a and d be points, and take a basis selection b of the two vectors (a) and (d) of $\mathbb{R}^1_\delta$. If neither is selected, spanning by the empty selection gives $a = 0$ and $d = 0$. If only (d) is selected, then $d \neq 0$ by the first statement of Lemma 14.11, and spanning gives $a = cd$ for some c; this is the third alternative. If only (a) is selected, the second alternative follows in the same way. If both are selected, the combination $d \cdot (a) + (-a) \cdot (d)$ vanishes, so independence gives $-a = 0$, while $a \neq 0$ because (a) is selected; this case cannot occur. The four cases are the four values of two bits. □

Rank and nullity add nothing: a basis selection of the columns already determines a basis of the kernel by an explicit formula.

Theorem 14.14 (Rank–nullity is scalar pair division)**.** *A basis selection of the columns completes to rank–nullity data, and rank–nullity is again scalar pair division.*

1. *If b is a basis selection of the columns of an $n \times m$ matrix A, then there are vectors $K_0, \dots, K_{m-1}$ that form, with b, rank–nullity data for A.*
2. *Rank–nullity is equivalent to scalar pair division.*

Neither statement uses an omniscience principle or a choice principle.

Proof. For the first statement, spanning gives for each column j coefficients $C_{j,l}$ with $\mathrm{col}_j = \sum_{b_l=1} C_{j,l}\,\mathrm{col}_l$. These are finitely many existence statements, and induction on m combines them into one table C; no choice principle is involved. Put

$$K_j = e_j - \sum_{b_l=1} C_{j,l}\, e_l.$$

Then $AK_j = \mathrm{col}_j - \sum_{b_l=1} C_{j,l}\,\mathrm{col}_l = 0$. For j with $b_j = 0$, the coordinate of K_j at an unselected index j' is 1 if $j' = j$ and 0 otherwise, since the subtracted sum lives on selected indices.

Independence. If $\sum_{b_j=0} c_j K_j = 0$, its coordinate at an unselected j' is $c_{j'}$, so every $c_{j'}$ is 0.

The kernel. Let $Ax = 0$, and put $y = x - \sum_{b_j=0} x_j K_j$. At an unselected index j', $y_{j'} = x_{j'} - x_{j'} = 0$. Moreover $Ay = Ax - \sum_{b_j=0} x_j AK_j = 0$. Since y vanishes off the selection, $Ay = \sum_{b_l=1} y_l \operatorname{col}_l$, and independence of the selected columns gives $y_l = 0$ for every selected l. So $y = 0$, and $x = \sum_{b_j=0} x_j K_j$.

For the second statement, scalar pair division gives a basis selection of the columns of every matrix (Theorem 14.13), and the first statement completes the data. Conversely, the first clause of rank–nullity data, for the matrix whose columns are $v_0, \dots, v_{k-1}$, is a basis selection of that family, and Theorem 14.13 gives scalar pair division. □

14.4 Linear algebra costs exactly WLPO

Scalar pair division mixes two tasks: deciding which points are zero, and dividing. The first is WLPO. The second can be stated on its own, and Section 14.6 prices it.

WLPO also gives, for every point x, $x \le 0$ or $x \ge 0$. Indeed the third row of Table 12.2 decides $x \le 0$ or $\neg(x \le 0)$, and $\neg(x \le 0)$, which is $\neg\neg(x > 0)$, excludes $x < 0$. So WLPO gives LLPO, by Corollary 12.18, with no choice.

Theorem 14.15 (Scalar pair division splits)**.** *The following are equivalent:*

1. *scalar pair division;*
2. *WLPO and dominated division by nonzero points;*
3. *WLPO and dominated division.*

No choice is used. By Theorems 14.13 and 14.14, basis selection for every finite family and rank–nullity are each equivalent to WLPO together with dominated division.

Proof. From the first statement. Divide the pair (x, x). The first alternative gives $x = 0$, and the other two give $x \neq 0$. So every point satisfies $x = 0 \vee x \neq 0$, which is WLPO by Corollary 12.18. Now let $a \neq 0$ and $|b| \le |a|$, and divide the pair (a, b). The first alternative contradicts $a \neq 0$; the second is the required quotient; the third gives $b \neq 0$ and $bc = a$, and Lemma 14.7 gives d with $ad = b$.

From the second. Let a and b be points, and let WLPO decide $a = 0 \vee a \neq 0$ and $b = 0 \vee b \neq 0$. If both are zero, the first alternative holds. If $a = 0$ and $b \neq 0$, then $b \cdot 0 = a$; if $a \neq 0$ and $b = 0$, then $a \cdot 0 = b$. If both are nonzero, LLPO, which WLPO gives, compares $|a|$ and $|b|$ by the fourth row of Table 12.2. If $|b| \le |a|$, dominated division by the nonzero point a gives the second alternative; if $|a| \le |b|$, division by b gives the third.

The third statement. Assume WLPO and dominated division by nonzero points, and let $|b| \le |a|$. If $a = 0$, then $|b| \le 0$, so $|b| = 0$, $b = 0$, and $c = 0$ works. If $a \neq 0$, the quotient c with $ac = b$ has $|c| \le 1$ by Lemma 14.6. Conversely, dominated division gives dominated division by nonzero points by forgetting the bound. □

Theorem 14.16 (Linear algebra costs exactly WLPO)**.** *Consider the three assertions: scalar pair division; every finite family of vectors in every $\mathbb{R}_\delta^n$ has a basis selection; and rank–nullity.*

1. *Each of them implies WLPO. No choice is used.*
2. *WLPO implies each of them. The proof uses unique choice: in each division it performs, one table of bits, each bit the answer to a WLPO decision, is collected into a sequence. No countable choice is used.*

Under the reading of maps of this book, which contains unique choice, each of the three assertions is therefore equivalent to WLPO.

Proof. The first part is Theorem 14.15 for scalar pair division, and Theorems 14.13 and 14.14 reduce the other two assertions to it.

For the second part, assume WLPO. We first show: *if $0 \le a \le b$, then $bc = a$ for some c in $[0,1]$.* Here b may be zero. Put $g(c) = bc - a$. Item 4 of Proposition 12.19 gives a natural number $K \ge 1$ with $b \le K$. Since $|g(x) - g(y)| = b\,|x - y|$, points of $[0,1]$ within $1/(K(k+1))$ of each other have values within $1/(k+1)$, so g is uniformly continuous on $[0,1]$; and $g(0) = -a \le 0 \le b - a = g(1)$. For each pair (n, j) of natural numbers, the third row of Table 12.2 decides $g(j/2^n) \le 0$ or $\neg\big(g(j/2^n) \le 0\big)$. So the relation "$\beta = 1$ and $g(j/2^n) \le 0$, or $\beta = 0$ and $\neg(g(j/2^n) \le 0)$" has exactly one bit β for each pair. Unique choice, part of the reading of maps, turns it into a table $B(n, j)$. If $B(n, j) = 1$, then $g(j/2^n) \le 0$. If $B(n, j) = 0$, then $\neg(g(j/2^n) \le 0)$, which excludes $g(j/2^n) < 0$, so $g(j/2^n) \ge 0$. This is exactly the table that the second part of the proof of Theorem 12.54 obtains from LLPO and countable choice. The rest of that proof, from the recursion defining a_n on, uses no further choice; it gives a root c of g in $[0,1]$, that is $bc = a$.

Next, *dominated division.* Let $|b| \le |a|$. The first step, applied to $0 \le |b| \le |a|$, gives c in $[0,1]$ with $|a|\,c = |b|$. LLPO, which WLPO gives, decides $a \ge 0$ or $a \le 0$, and $b \ge 0$ or $b \le 0$. Choose signs $\varepsilon_a, \varepsilon_b \in \{1, -1\}$ accordingly, so that $|a| = \varepsilon_a a$ and $|b| = \varepsilon_b b$. Then $a(\varepsilon_a \varepsilon_b c) = \varepsilon_b |a| c = \varepsilon_b |b| = b$, and $|\varepsilon_a \varepsilon_b c| = c \le 1$.

Theorem 14.15 now gives scalar pair division, and Theorems 14.13 and 14.14 give the other two assertions. Unique choice entered only in the first step, once for each division, and each time for a relation with exactly one bit for each pair. □

Unique choice is not the only way back. Two stronger principles give linear algebra with no choice at all: LPO, and the sequence form $\mathrm{WLPO}_{\mathrm{seq}}$ of WLPO (Definition 13.8), which says that for every binary matrix $(\alpha_i(n))$ there is a binary sequence f with $f(i) = 1$ exactly when $\alpha_i(n) = 0$ for every n. It names, in one sequence, the WLPO decision for every row.

Proposition 14.17 (Two routes with no choice)**.** *LPO implies scalar pair division, and so does* $\mathrm{WLPO}_{\mathrm{seq}}$*. Neither proof uses a choice principle. Consequently each of LPO and* $\mathrm{WLPO}_{\mathrm{seq}}$ *gives basis selections for every finite family and rank–nullity, with no choice.*

Proof. Assume LPO. It gives WLPO and $\mathrm{MP}_{\mathrm{bin}}$ (Proposition 12.9). Let a and b be points, and decide $a = 0 \vee a \ne 0$. If $a \ne 0$, then $\mathrm{MP}_{\mathrm{bin}}$ makes a apart from 0 (Corollary 12.18), so a has an inverse u, and $a(ub) = b$. If $a = 0$, decide $b = 0 \vee b \ne 0$; both zero is the first alternative, and if $b \ne 0$, its inverse u' gives $b(u'a) = a$.

Assume $\mathrm{WLPO}_{\mathrm{seq}}$. Applied to a matrix all of whose rows equal a given binary sequence, it gives WLPO. As in the proof of Theorem 14.16, it suffices to divide a by b when $0 \le a \le b$, because the rest of that proof uses only WLPO. This is a statement about two given points, so we may fix regular sequences naming a and b. For each pair (n, j), listed as a natural number, the formulas for sums and products give an explicit regular sequence naming $g(j/2^n) = b \cdot j/2^n - a$. Let $\alpha_{(n,j)}(N) = 1$ when its stage N exceeds $2/(N+1)$. $\mathrm{WLPO}_{\mathrm{seq}}$ gives a binary sequence f. If $f(n,j) = 1$, the sequence has no positive stage, and since positivity belongs to the point, $g(j/2^n) \le 0$. If $f(n,j) = 0$, then $\neg(g(j/2^n) \le 0)$, and $g(j/2^n) \ge 0$ as before. The second part of the proof of Theorem 12.54, with f as its table, gives the quotient, and no choice has been made.

Theorems 14.13 and 14.14 give the last sentence. □

14.5 The sign of a nonzero point

Dividing by a point that is only known not to be zero produces a sign. Suppose $x \neq 0$ and $|x|\,c = x$. Lemma 14.8 gives $x \ge 0$ or $x \le 0$, a weak sign. LLPO gives a weak sign for every point, and $\mathrm{MP}_{\mathrm{bin}}$ gives a strict sign for every nonzero point; both principles are standard in constructive analysis [14, 40]. The weak sign of nonzero points alone has its own price.

Definition 14.18 (The disjunctive Markov principle)**.** $\mathrm{MP}^{\vee}$ states that every binary sequence α that has at most one entry equal to 1 and is not all zero satisfies

$$(\forall k,\ \alpha(2k) = 0) \quad \text{or} \quad (\forall k,\ \alpha(2k+1) = 0).$$

Its sequence form $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ states: for every binary matrix $(\alpha_i(n))$ whose rows each have at most one entry 1 and none of which is all zero, there is a binary sequence f such that $f(i) = 1$ gives $\alpha_i(2k) = 0$ for every k, and $f(i) = 0$ gives $\alpha_i(2k+1) = 0$ for every k.

$\mathrm{MP}^{\vee}$ is LLPO restricted to sequences known not to be all zero, and $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ is $\mathrm{LLPO}_{\mathrm{seq}}$ of Definition 13.8 restricted to matrices none of whose rows is all zero, with the same reading of the bits. In constructive reverse mathematics $\mathrm{MP}^{\vee}$ is known as a common weakening of Markov's principle, here $\mathrm{MP}_{\mathrm{bin}}$, and LLPO [43, 44]. In $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ the side named for each row is unique: if both sides of a row were free of events, the row would be all zero.

Theorem 14.19 (The weak sign of a nonzero point costs $\mathrm{MP}^{\vee}$)**.** *The weak sign of nonzero points and* $\mathrm{MP}^{\vee}$ *are related as follows.*

1. *Every point $x \neq 0$ of $\mathbb{R}_\delta$ satisfies $x \ge 0$ or $x \le 0$ if and only if* $\mathrm{MP}^{\vee}$ *holds.*
2. *LLPO implies* $\mathrm{MP}^{\vee}$*, and so does* $\mathrm{MP}_{\mathrm{bin}}$*.*

No choice is used.

Proof. This is the proof of Theorem 12.8 with one hypothesis carried along. Assume $\mathrm{MP}^{\vee}$ and let $x \neq 0$. The claim is a statement about one point, so we may fix a regular dyadic presentation (q_n) of x (Theorem 12.17). Form the sequence e of signed events and the first-event sequence f exactly as in that proof: $e(2k) = 1$ when stage k is a positive witness, $e(2k+1) = 1$ when

it is a negative witness, and $f(j) = 1$ when the bounded search through $e(0), \dots, e(j)$ reports its first event at j. Then f has at most one entry 1. It is not all zero: if it were, e would be all zero, since an entry $e(j) = 1$ gives by bounded search a first event $j' \le j$ with $f(j') = 1$; then x would have neither a positive nor a negative witness, and Lemma 12.10 with Theorem 12.17 would give $x = 0$. So $\mathrm{MP}^\vee$ applies to f. The proof of Theorem 12.8 turns "every even entry of f is 0" into $D(x)$, that is $x \le 0$, and "every odd entry is 0" into $U(x)$, that is $x \ge 0$.

Conversely, let β have at most one entry 1 and not be all zero, and let $j_{\sigma,\beta}$ be its jump presentation with the parity sign σ. If its point were 0, Lemma 12.5 would make β all zero; so the point is not 0, and by hypothesis it is ≥ 0 or ≤ 0, that is $U(j_{\sigma,\beta})$ or $D(j_{\sigma,\beta})$. The first half of the proof of Theorem 12.8 shows that $U(j_{\sigma,\beta})$ forces every odd entry of β to vanish and $D(j_{\sigma,\beta})$ every even entry.

For the second statement, LLPO gives $\mathrm{MP}^\vee$ by dropping a hypothesis. Under $\mathrm{MP}_{\mathrm{bin}}$, a sequence α that is not all zero has an entry $\alpha(n) = 1$; it is the only one, so every entry of the other parity is 0. □

For the real numbers of Bishop's constructive mathematics, which allows countable choice, the corresponding equivalence is known; the real-number statement goes back to Mandelkern [42, 43, 44]. On $\mathbb{R}_\delta$ no choice is needed, because the proof works from one regular sequence naming the given point.

Collecting such signs along a sequence is a different matter, and it is where unique choice enters. The relevant case of unique choice concerns relations that are confirmed by infinitely many decidable tests.

Definition 14.20 (Unique choice of bits for Π^0_1 relations)**.** Let $r(i, b, n) \in \{0, 1\}$ be given for all natural numbers i, n and bits b, and say that the bit b *passes row* i when $r(i, b, n) = 1$ for every n. *Unique choice of bits for* Π^0_1 *relations* states: if for every i exactly one bit passes row i, then some binary sequence f has $r(i, f(i), n) = 1$ for all i and n.

This is the principle of unique choice for the relations "b passes row i", so under the reading of maps of this book it holds.

Lemma 14.21 (Collecting weak sides)**.** $\mathrm{MP}^\vee$ *together with unique choice of bits for* Π^0_1 *relations implies* $\mathrm{MP}^\vee_{\mathrm{seq}}$. *No other choice is used.*

Proof. Let $(\alpha_i(n))$ be a matrix as in Definition 14.18. Put $r(i, 1, n) = 1$ when $\alpha_i(2n) = 0$, and $r(i, 0, n) = 1$ when $\alpha_i(2n + 1) = 0$, with $r = 0$ otherwise. A bit passing row i is a correct side for that row. $\mathrm{MP}^\vee$, applied to row i, gives one. Two different bits cannot both pass, because then row i would be all zero. So exactly one bit passes each row, and unique choice of bits for Π^0_1 relations gives the sequence f. □

14.6 Dividing by a nonzero point

Dominated division, with a divisor that may be zero, has the price of order dichotomy.

Theorem 14.22 (Dominated division costs LLPO)**.** *Consider dominated division, Definition 14.3(b).*

1. *Dominated division implies LLPO. No choice is used.*
2. *Assuming* AC_ω*, LLPO implies dominated division. Countable choice is used once, to tabulate one bit for every dyadic ratio.*

Proof. For a point x, $|x| \le \||x|\|$, so dominated division of x by $|x|$ gives c with $|x|\,c = x$. Lemma 14.8 gives $x \ge 0$ or $x \le 0$. This holds for every point, which is LLPO by Corollary 12.18.

Conversely, assume LLPO and AC_ω, and let $|b| \le |a|$. The function $g(c) = |a|\,c - |b|$ is uniformly continuous on $[0,1]$, as in the proof of Theorem 14.16, and $g(0) = -|b| \le 0 \le |a| - |b| = g(1)$. The second part of Theorem 12.54 gives a root c in $[0,1]$, so $|a|\,c = |b|$; countable choice is used there, once. The signs of a and b, decided by LLPO, turn c into a quotient $\varepsilon_a\varepsilon_b c$ with $a\varepsilon_a\varepsilon_b c = b$ and $|\varepsilon_a\varepsilon_b c| \le 1$, exactly as in the proof of Theorem 14.16. □

When the divisor is only known not to be zero, the bisection of Theorem 12.54 has no bits to work with: LLPO is not available. What a nonzero divisor does provide concerns two points at once. Let $0 \le a \le b$ with $b \neq 0$, and let $l < r$ be two ratios. Then $a \le b\,l$ and $b\,r \le a$ cannot both hold: together they give $b\,r \le b\,l$, and Lemma 14.5 would give $r \le l$. A search through the stages of a regular sequence for a witness that a point is positive stays silent exactly when the point is ≤ 0, because positivity belongs to the point. So the two stage tests, for $a > b\,l$ and for $b\,r > a$, cannot both stay silent; trisection uses exactly this.

Definition 14.23 (Affine trisection control)**.** For natural numbers n and p with $p + 2^n \le 3^n$, the interval $[p/3^n, (p+2^n)/3^n]$ lies in $[0,1]$ and has width $(2/3)^n$; its *trisection points* are

$$l_{n,p} = \frac{3p + 2^n}{3^{n+1}}, \qquad r_{n,p} = \frac{3p + 2^{n+1}}{3^{n+1}}.$$

Affine trisection control states: for all points a, b with $0 \le a \le b$ and $b \neq 0$ there are bits $B(n,p)$, one for every pair of natural numbers, such that whenever $p + 2^n \le 3^n$, $B(n,p) = 1$ gives $b\,l_{n,p} \le a$ and $B(n,p) = 0$ gives $a \le b\,r_{n,p}$.

Read with the quotient $c = a/b$ in mind, the bit says where c may be sought (Figure 14.1): $B(n,p) = 1$ promises $c \ge l_{n,p}$, and $B(n,p) = 0$ promises $c \le r_{n,p}$. The two promises overlap, and both can be true, so the bit is not determined by a and b; for that reason unique choice cannot collect it directly. The table covers every admissible interval, not only those that a search will visit.

Say that a function g on the points of $\mathbb{R}_\delta$ is *continuous at every point of* $[0,1]$ when for every x in $[0,1]$ and every k there is a j such that every y in $[0,1]$ within $1/(j+1)$ of x has $g(y)$ within $1/(k+1)$ of $g(x)$. A function uniformly continuous on $[0,1]$ has this property.

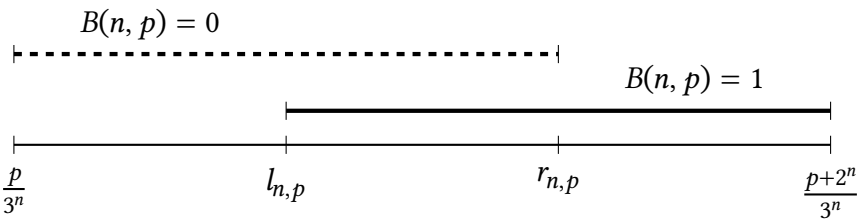


Figure 14.1. One step of trisection. The bit 1 records $b\,l_{n,p} \le a$ and keeps the solid interval $[l_{n,p}, (p + 2^n)/3^n]$; the bit 0 records $a \le b\,r_{n,p}$ and keeps the dashed interval $[p/3^n, r_{n,p}]$. Either kept interval has two thirds of the old width, and the two overlap in the middle third, so a weak side at either of two different points suffices.

Lemma 14.24 (Trisection with weak sides)**.** *Let g be continuous at every point of $[0, 1]$ with $g(0) \le 0 \le g(1)$, and let $B(n, p)$ be bits such that whenever $p + 2^n \le 3^n$, $B(n, p) = 1$ gives $g(l_{n,p}) \le 0$ and $B(n, p) = 0$ gives $g(r_{n,p}) \ge 0$. Then g has a root in $[0, 1]$. No omniscience principle and no choice is used.*

Proof. Define natural numbers p_n by recursion: $p_0 = 0$, and $p_{n+1} = 3p_n + 2^n$ if $B(n, p_n) = 1$, otherwise $p_{n+1} = 3p_n$. Put $\lambda_n = p_n/3^n$ and $\mu_n = (p_n + 2^n)/3^n$. Induction on n gives

$$p_n + 2^n \le 3^n, \qquad \mu_n - \lambda_n = (2/3)^n, \qquad g(\lambda_n) \le 0 \le g(\mu_n),$$

together with $\lambda_n \le \lambda_{n+1}$ and $\mu_{n+1} \le \mu_n$. Indeed, if $B(n, p_n) = 1$, the new interval is $[l_{n,p_n}, \mu_n]$, because $3p_n + 2^n + 2^{n+1} = 3(p_n + 2^n)$, and $g(l_{n,p_n}) \le 0$; the new bound reads $3(p_n + 2^n) \le 3^{n+1}$. If $B(n, p_n) = 0$, the new interval is $[\lambda_n, r_{n,p_n}]$ and $g(r_{n,p_n}) \ge 0$; the new bound reads $3p_n + 2^{n+1} \le 3(p_n + 2^n) \le 3^{n+1}$.

For $m \ge n$ the intervals are nested, so $|\lambda_m - \lambda_n| \le (2/3)^n$. Since $(9/4)^k \ge 1 + 5k/4 \ge k + 1$, the ratios $t_k = \lambda_{2^k}$ satisfy $|t_k - t_l| \le (4/9)^{\min(k,l)} \le 1/(k + 1) + 1/(l + 1)$, so they form a regular sequence. Let t be its point. By Lemma 12.52, item 4, λ_{2^k} is within $1/(k + 1)$ of t. The ratios μ_{2^k} and λ_{2^k} are within $(4/9)^k \le 1/(k + 1)$ of each other, so $\mu_{2(2k+1)}$ is within $1/(2k + 2) + 1/(2k + 2) = 1/(k + 1)$ of t (item 5). Every λ_n lies in $[0, 1]$, so item 2 puts t in $[0, 1]$.

Fix k, and take j as in continuity at t. The point λ_{2^j} of $[0, 1]$ is within $1/(j + 1)$ of t, so $g(\lambda_{2^j})$, which is at most 0, is within $1/(k + 1)$ of $g(t)$. Since k was arbitrary, item 2 gives $g(t) \le 0$. In the same way $g(\mu_{2(2j+1)}) \ge 0$ is within $1/(k + 1)$ of $g(t)$, so $g(t) \ge 0$, and $g(t) = 0$ by item 2 of Proposition 12.19. Every step is a recursion on given bits, a finite computation or an estimate. □

A single point gives such a table for itself, from any sequence naming it.

Lemma 14.25 (A point gives its own trisection table)**.** *Let c be a point of $[0, 1]$. There are bits $B(n, p)$ such that whenever $p + 2^n \le 3^n$, $B(n, p) = 1$ gives $l_{n,p} \le c$ and $B(n, p) = 0$ gives $c \le r_{n,p}$. No omniscience principle and no choice is used.*

Proof. The claim is a statement about the one point c, so we may fix a regular sequence s naming it. For a pair (n, p) let $m = (l_{n,p} + r_{n,p})/2$ and $h = (r_{n,p} - l_{n,p})/2 = 2^n/(2 \cdot 3^{n+1})$, and let $N = 2 \cdot 3^{n+1}$, so that $1/(N + 1) < h$. Compare the ratios s_N and m. By Lemma 12.52, item 4, the

ratio s_N is within $1/(N+1)$ of c. If $s_N \geq m$, put $B(n,p) = 1$; then $c \geq s_N - 1/(N+1) > m - h = l_{n,p}$. Otherwise put $B(n,p) = 0$; then $c \leq s_N + 1/(N+1) < m + h = r_{n,p}$. Each bit is a comparison of two ratios computed from s. □

The last ingredient divides with a sign that the division itself supplies.

Lemma 14.26 (Positive division suffices)**.** *Suppose that for all points $0 \leq a \leq b$ with $b \neq 0$ some c has $bc = a$. Then:*

1. *for all points $d \geq 0$ with $d \neq 0$ and every x with $|x| \leq d$, some r has $dr = x$;*
2. *dominated division by nonzero points holds.*

No omniscience principle and no choice is used.

Proof. For the first statement, $-x \leq |x| \leq d$ and $x \leq |x| \leq d$ give $0 \leq d + x \leq d + d$. Moreover $d + d \neq 0$: if $d + d = 0$, then $d = -d \leq 0$, so $d = 0$. The hypothesis gives c with $(d + d)c = d + x$, and then $d(c + c - 1) = (d + d)c - d = x$.

For the second, let $a \neq 0$ and $|b| \leq |a|$, and put $d = |a|$, so that $d \geq 0$ and $d \neq 0$. The first statement with $x = a$ gives s with $|a|\, s = a$, and Lemma 14.8 gives $a \geq 0$ or $a \leq 0$. The first statement with $x = b$ gives r with $|a|\, r = b$. If $a \geq 0$, then $|a| = a$ and $ar = b$; if $a \leq 0$, then $|a| = -a$ and $a(-r) = b$. No sign was decided by a principle: the weak sign of a came out of a division. □

Theorem 14.27 (Trisection control is division by nonzero points)**.** *The following are equivalent:*

1. *affine trisection control;*
2. *for all points $0 \leq a \leq b$ with $b \neq 0$ some c has $bc = a$;*
3. *dominated division by nonzero points.*

No omniscience principle and no choice is used.

Proof. From the first to the second. Let $0 \leq a \leq b$ with $b \neq 0$, take the bits of affine trisection control, and put $g(c) = bc - a$. It is uniformly continuous on $[0,1]$ (proof of Theorem 14.16), and $g(0) \leq 0 \leq g(1)$. A bit 1 gives $g(l_{n,p}) \leq 0$ and a bit 0 gives $g(r_{n,p}) \geq 0$, so Lemma 14.24 gives a root.

From the second to the third. This is Lemma 14.26.

From the third to the first. Let $0 \leq a \leq b$ with $b \neq 0$. Then $|a| = a \leq b = |b|$, and dominated division by the nonzero point b gives c with $bc = a$; it lies in $[0,1]$ by Lemma 14.10. Lemma 14.25 gives bits for c. Multiplying by $b \geq 0$ turns $l_{n,p} \leq c$ into $b\, l_{n,p} \leq bc = a$, and $c \leq r_{n,p}$ into $a \leq b\, r_{n,p}$. □

Theorem 14.28 (Division by nonzero points costs $\mathrm{MP}^\vee$)**.** *Consider dominated division by nonzero points, Definition 14.3(c), and affine trisection control.*

1. *Dominated division by nonzero points implies* $\mathrm{MP}^\vee$*. No choice is used.*
2. $\mathrm{MP}^\vee_{\mathrm{seq}}$ *implies affine trisection control, and hence dominated division by nonzero points. No choice is used.*
3. $\mathrm{MP}^\vee$*, together with unique choice of bits for* Π^0_1 *relations, implies affine trisection control and dominated division by nonzero points.*
4. $\mathrm{MP}_{\mathrm{bin}}$ *implies dominated division by nonzero points. No choice is used.*

Under the reading of maps of this book, dominated division by nonzero points and affine trisection control are therefore each equivalent to $\mathrm{MP}^\vee$*.*

Proof. The first statement. Let $x \neq 0$. Then $|x| \neq 0$ and $|x| \leq \||x|\|$, so dividing x by the nonzero point $|x|$ gives c with $|x|\,c = x$. Lemma 14.8 gives $x \geq 0$ or $x \leq 0$, and Theorem 14.19 gives $\mathrm{MP}^\vee$.

The second. Let $0 \leq a \leq b$ with $b \neq 0$. This is a statement about two given points, so we may fix regular sequences naming a and b. List the pairs (n, p) as natural numbers i. For an admissible pair, the points $u_i = a - b\,l_{n,p}$ and $v_i = b\,r_{n,p} - a$ are named by explicit regular sequences; let $\alpha_i(N) = 1$ when stage N of the sequence for u_i exceeds $2/(N+1)$, and $\beta_i(N) = 1$ when stage N of the sequence for v_i does. Since positivity belongs to the point, $u_i > 0$ exactly when α_i has an entry 1, and likewise for v_i and β_i. As shown before Definition 14.23, $u_i \leq 0$ and $v_i \leq 0$ cannot both hold; so it is impossible that both α_i and β_i are all zero.

Keep only the first event of the two rows. Let $\gamma_i(2N) = 1$ when N is the least index at which α_i or β_i has an entry 1 and $\alpha_i(N) = 1$, and $\gamma_i(2N+1) = 1$ when N is that least index, $\alpha_i(N) = 0$ and $\beta_i(N) = 1$; all other entries are 0, and each is a bounded search. For a pair that is not admissible, let γ_i have a single entry 1, at position 0. Each row γ_i has at most one entry 1, and none is all zero: if γ_i were all zero, a bounded search would show that α_i and β_i are all zero. So $\mathrm{MP}^\vee_{\mathrm{seq}}$ gives a binary sequence f.

Let i be admissible. If $f(i) = 1$, every even entry of γ_i is 0, and then $\neg\neg(v_i > 0)$. Indeed, suppose β_i were all zero. Then α_i would be all zero too: an entry $\alpha_i(N) = 1$ would give, by bounded search, a least index $N' \leq N$ at which one of the rows has an entry 1, and since $\beta_i(N') = 0$ this would be an even entry $\gamma_i(2N') = 1$. Both rows would be all zero, which is impossible. So $\neg\neg(v_i > 0)$, hence $\neg(v_i < 0)$, that is $a \leq b\,r_{n,p}$, and we put $B(n,p) = 0$. If $f(i) = 0$, the odd entries of γ_i vanish, the same argument gives $\neg\neg(u_i > 0)$, hence $b\,l_{n,p} \leq a$, and we put $B(n,p) = 1$. This is the table of affine trisection control, and Theorem 14.27 gives the division.

The third. Lemma 14.21 gives $\mathrm{MP}^\vee_{\mathrm{seq}}$; apply the second statement.

The fourth. Let $a \neq 0$ and $|b| \leq |a|$. $\mathrm{MP}_{\mathrm{bin}}$ makes a apart from 0 (Corollary 12.18), with an inverse u, and $a(ub) = b$. □

The route through $\mathrm{MP}^\vee$ sharpens the choice used in Theorem 14.16.

Corollary 14.29 (Unique choice for Π^0_1 relations suffices)**.** *WLPO, together with unique choice of bits for Π^0_1 relations, implies scalar pair division, basis selection for every finite family and rank–nullity. No other choice is used.*

Proof. WLPO gives LLPO, which gives $\mathrm{MP}^\vee$ (Theorem 14.19). The third part of Theorem 14.28 gives dominated division by nonzero points, Theorem 14.15 gives scalar pair division, and Theorems 14.13 and 14.14 give the rest. □

14.7 Signs as units

A sign can also be recorded as a unit. For a point x and a bit b, put $\sigma_1 = 1$ and $\sigma_0 = -1$, and say that *b normalizes x* when $|x|\,\sigma_b = x$. A sequence of points is *given by regular sequences* when it is presented as regular sequences $s_0, s_1, \ldots$ naming its terms (Definition 12.39).

Theorem 14.30 (Signed units)**.** *Normalizing bits, one at a time and collected along a sequence, are priced as follows.*

1. *The bit* 1 *normalizes x exactly when $x \geq 0$, and the bit* 0 *exactly when $x \leq 0$. So every point $x \neq 0$ has a normalizing bit if and only if* $\mathrm{MP}^\vee$ *holds, and under WLPO every point has a normalizing bit.*

2. *Let the points $x_0, x_1, \ldots$ be given by regular sequences, with $x_i \neq 0$ for every i. Each of the following, asserted for every such sequence, is equivalent to* $\mathrm{MP}^\vee_{\mathrm{seq}}$*:* signed-unit collection*, a binary sequence f such that $f(i)$ normalizes x_i for every i; and* absolute-quotient collection*, a sequence of points c_i with $|x_i|\,c_i = x_i$ for every i.*

No choice is used.

Proof. For the first statement, $|x| \cdot 1 = x$ says $|x| = x$, which holds exactly when $x \geq 0$, and $|x| \cdot (-1) = x$ says $|x| = -x$, which holds exactly when $x \leq 0$. Theorem 14.19 prices the weak sign of nonzero points, and WLPO gives LLPO, which gives a weak sign for every point.

For the second, let s_i name x_i. *From* $\mathrm{MP}^\vee_{\mathrm{seq}}$ *to signed units.* Form from each s_i the first-event row of the proof of Theorem 14.19, with the stages of s_i above $2/(k+1)$ or below $-2/(k+1)$ as signed events. It has at most one entry 1, and it is not all zero, because $x_i \neq 0$ and a point with neither a positive nor a negative stage is 0 (Lemma F.8). $\mathrm{MP}^\vee_{\mathrm{seq}}$ names a side for every row. If the even entries vanish, $x_i \leq 0$: a positive stage would make the first event a negative stage, and a regular sequence has no stages above $2/(m+1)$ and below $-2/(n+1)$, by the argument of Lemma 12.3. Then the bit 0 normalizes x_i. If the odd entries vanish, $x_i \geq 0$ and the bit 1 normalizes it.

From signed units to $\mathrm{MP}^\vee_{\mathrm{seq}}$. Given a matrix as in Definition 14.18, let x_i be the point of the jump presentation of row i with the parity sign; these are explicit regular sequences. Each $x_i \neq 0$ by Lemma 12.5. A normalizing bit 1 says $x_i \geq 0$, which forces every odd entry of row i to vanish, and a bit 0 says $x_i \leq 0$, which forces every even entry to vanish, as in the proof of Theorem 14.19. So $f(i) = 1$ when the bit is 0, and $f(i) = 0$ when it is 1, is the required sequence.

Assertion on $\mathbb{R}_\delta$	Price	Choice
Scalar pair division (Theorems 14.15 and 14.16)	WLPO	unique choice in the converse
Every finite family in every $\mathbb{R}_\delta^n$ has a basis selection (Theorems 14.13 and 14.16)	WLPO	unique choice in the converse
Rank–nullity (Theorems 14.14 and 14.16)	WLPO	unique choice in the converse
Every matrix has a rank, which is also the rank of its transpose (Theorem G.4)	WLPO	not used
Dominated division (Theorem 14.22)	LLPO	countable choice in the converse
Dominated division by nonzero points; affine trisection control (Theorems 14.27 and 14.28)	$\mathrm{MP}^\vee$	unique choice in the converse
Every $x \neq 0$ has $x \geq 0$ or $x \leq 0$; equivalently, a normalizing bit (Theorems 14.19 and 14.30)	$\mathrm{MP}^\vee$	not used
Signed-unit collection; absolute-quotient collection, for nonzero points given by regular sequences (Theorems 14.30 and G.10)	$\mathrm{MP}^\vee$	unique choice in the converse
Quantitative uniqueness of the roots of $c \mapsto bc - a$ (Theorem G.7)	$\mathrm{MP}_{\mathrm{bin}}$	not used
For 2×2 matrices: a nonzero determinant gives an inverse; an injective matrix is surjective (Theorem G.5)	$\mathrm{MP}_{\mathrm{bin}}$	not used
Every continuous injective map of the line is strictly monotone (Theorem G.9)	$\mathrm{MP}_{\mathrm{bin}}$	not used
k independent vectors of $\mathbb{R}_\delta^n$ have $k \leq n$; a matrix with zero kernel has no more columns than rows (Proposition G.1)	none	not used
A surjective 2×2 matrix is injective (Theorem G.5)	none	not used

Table 14.1. Linear algebra and division on the line built from ratios. "Price" is the principle equivalent to the assertion, and "none" means that the assertion is proved outright. In every priced row the passage from the assertion to the principle uses no choice. "In the converse" means that the named choice principle is used in the passage from the principle back to the assertion; unique choice is part of the reading of maps of this book, and in the rows priced $\mathrm{MP}^\vee$ and in Corollary 14.29 it is needed only for bits of Π^0_1 relations. Each row names the results that prove both directions.

Absolute quotients. Normalizing bits b_i give the quotients $c_i = \sigma_{b_i}$. Conversely, let $|x_i|\, c_i = x_i$ for every i. Each c_i is apart from 0 by Lemma 14.9. For a point c apart from 0 exactly one of $c > 0$ and $c < 0$ holds, and which one is found from any regular sequence naming c by searching for its first stage above $2/(n+1)$ or below $-2/(n+1)$; the search is proved to terminate, because c is apart from 0, and two sequences naming c give the same answer, since a point is not both positive and negative. So the sign is computed from the point itself, with no selection of representatives. If $c_i > 0$, then $x_i = |x_i|\, c_i \geq 0$, and the bit 1 normalizes x_i; if $c_i < 0$, then $x_i \leq 0$ and the bit 0 normalizes it. □

With Theorem G.10 of Appendix G, which shows that $\mathrm{MP}^\vee_{\mathrm{seq}}$ is $\mathrm{MP}^\vee$ together with unique choice of bits for Π^0_1 relations, each collection is equivalent to $\mathrm{MP}^\vee$ under the reading of maps of this book. A collection gives $\mathrm{MP}^\vee$ with no choice; unique choice is used in the converse.

14.8 The prices

Table 14.1 collects the prices of this chapter and of Appendix G. The order and field prices of the line itself are those of Table 12.2.

14.9 Without unique choice

Unique choice is part of the reading of maps of this book, and under that reading the prices of Table 14.1 are final, except that the row priced LLPO uses countable choice in the converse (Theorem 14.22). Now read maps more narrowly: a sequence must be given by an explicit rule, and a search that is proved to terminate counts as one, but a relation with exactly one witness for each number need not define a sequence. This section records which prices survive.

Every result of this chapter marked "no choice" holds over that base, including the passage from each assertion to its principle. Some passages back survive too: LPO and $\mathrm{WLPO}_{\mathrm{seq}}$ still give linear algebra (Proposition 14.17), and $\mathrm{MP}_{\mathrm{bin}}$ and $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ still give division by nonzero points (Theorem 14.28). What is lost is the step from WLPO, or from $\mathrm{MP}^{\vee}$, to a table of bits. Appendix G measures it: $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ is $\mathrm{MP}^{\vee}$ together with unique choice of bits for Π^0_1 relations (Theorem G.10), and a search for a failed test, which would name the surviving bit, costs exactly $\mathrm{MP}_{\mathrm{bin}}$ (Theorem G.11). What remains are two questions.

> *Over a base without unique choice, does WLPO imply scalar pair division, and does* $\mathrm{MP}^{\vee}$ *imply dominated division by nonzero points?*

Theorem G.12 proves the following with no choice. WLPO implies scalar pair division exactly when it implies affine trisection control; in that case Theorems 14.13, 14.14 and 14.16 make basis selection and rank–nullity each equivalent to WLPO with no choice in either direction. A second statement, that scalar pair division is equivalent to WLPO together with unique choice of bits for Π^0_1 relations with no choice in either direction, holds exactly when scalar pair division yields absolute-quotient collections. And WLPO yields unique choice of bits for Π^0_1 relations exactly when the answer to the first question is yes and the second statement holds.

Questions to work through

1. Show directly that a basis selection of the single vector (x) of $\mathbb{R}^1_\delta$ decides $x = 0 \vee x \neq 0$. Which part of Definition 14.1 gives each alternative?

2. Prove that if $x \neq 0$ and $|x|\,c = x$, then $|c| = 1$. Then use Lemma 14.8 to show that the quotient decides $x \geq 0$ or $x \leq 0$ without any bound on c given in advance.

3. In the proof of Theorem 14.16, the bit $B(n, j) = 0$ records $\neg(g(j/2^n) \leq 0)$. Show that this gives $g(j/2^n) \geq 0$, and explain why the relation used there has exactly one bit for each pair, while the bits of Theorem 12.54 need not be unique.

4. Let $0 \leq a \leq b$, $b \neq 0$, and let $l < r$ be ratios. Show that $a \leq bl$ and $br \leq a$ cannot both hold. Why does the argument fail at a single midpoint m, as in bisection?

Chapter 15

Differential equations to certified precision

Take the oscillator $y_0' = y_1$, $y_1' = -y_0$ with $y(0) = (0, 1)$. At time 1 its first coordinate is $\sin 1$. Suppose we are asked for $\sin 1$ to m bits: an integer a with $|\sin 1 - a/2^m| \le 2^{-m}$. The program of this chapter answers the request for 40 bits with 925207132232. To produce it, it takes 32 steps of a Taylor method of order 61, on dyadic numbers with 61 binary digits after the point. Asked for 400 bits, it takes the same 32 steps, now of order 421. The number of steps never changes; only the order and the working precision grow with the request.

Three questions follow. Why can the number of steps stay fixed while the precision grows? What does a request for m bits cost, counted exactly? And what does the integer that comes back certify, if no exact solution of the equation is assumed to exist? Chapter 10 asked for one program for every requested precision and warned that an execution-time claim needs its own construction and bound. Chapters 11 and 12 built the line from rational and dyadic enclosures. This chapter supplies the construction and the bound for a whole class of problems: quadratic and polynomial vector fields with rational data, each with a finite check on one computed trajectory. Its answer is a point of the line $\mathbb{R}_\delta$ given by a program with a proved modulus, an exact count of the arithmetic that program performs, and finally one fixed binary machine that, for every such problem whose check holds at the level named in the request, answers every request for m bits within $K(m+1)^\lambda$ steps. The exponent λ is the same for every problem; only the constant K depends on the problem. Appendix I constructs the machine stage by stage and proves each stage's bound.

15.1 A number is a program with a proof of its precision

The interval protocols of Chapter 10 and the regular sequences of Chapter 12 have moduli of the form $1/(n+1)$. That is enough to name a point. It is too slow for counting work. To add two interval protocols, Chapter 11 reads both operands at request $2n+1$; a sum of sums reads its leaves at about $4n$, and an expression of depth h at about $2^h n$. With a modulus 2^{-n} the same sum needs only two extra bits, and depth h costs $2h$ extra bits. The difference between multiplying and adding precision is what makes a polynomial bound possible.

Definition 15.1 (Certified real)**.** A *certified real* x is a sequence $(a_n)_{n\ge 0}$ of integers such that, for all m, n,

$$|a_m 2^n - a_n 2^m| \le 2^n + 2^m, \qquad \text{that is,} \qquad \left|\frac{a_m}{2^m} - \frac{a_n}{2^n}\right| \le 2^{-m} + 2^{-n}.$$

The integer a_n is the *answer* of x at precision n, and the ratio $a_n/2^n$ is its *stage* n.

The integer form is the definition; the rational form abbreviates it, as in Definition 12.1. A certified real is a program in the sense of Chapter 10 only when its answers are computed by one; every certified real in this chapter is given by an explicit calculation of a_n from n.

Lemma 15.2 (A certified real names a point of the line)**.** *The answers of a certified real form a regular dyadic presentation in the sense of Definition 12.1. Its stages, read as the ratio records $(a_n, 2^n)$, are therefore a regular sequence, and they name a point of $\mathbb{R}_\delta$.*

Proof. Since $n+1 \le 2^n$ for every n, we have $2^{-n} \le 1/(n+1)$. Hence

$$\begin{aligned}|a_m 2^n - a_n 2^m|(m+1)(n+1) &\le (2^n + 2^m)(m+1)(n+1)\\ &\le 2^n 2^m (n+1) + 2^m (m+1) 2^n = 2^{m+n}(m+n+2),\end{aligned}$$

which is (12.1). The remaining assertions are the remarks after Definition 12.14. □

To build certified reals we need one tool. A program usually does not know the exact value it approximates; it knows exact ratios computed from its inputs' own answers. We call such ratios *ideal values*: what the program would return with unlimited precision.

Lemma 15.3 (The ideal-value lemma)**.** *Let (a_n) be integers, let $I(M)$ be ratios for $M \in \mathbb{N}$, and let $T : \mathbb{N} \to \mathbb{N}$ and $C \in \mathbb{N}$. Suppose that*

$$\left|\frac{a_n}{2^n} - I(M)\right| \le 2^{-n} + C\,2^{-M} \qquad \text{whenever } M \ge T(n).$$

Then (a_n) is a certified real.

Proof. Fix m and n. For every $M \ge \max(T(m), T(n))$ the triangle inequality through $I(M)$ gives

$$\left|\frac{a_m}{2^m} - \frac{a_n}{2^n}\right| \le 2^{-m} + 2^{-n} + 2C\,2^{-M}.$$

Since $2^{-M} \le 1/(M+1)$, this bound holds with $2C/(k+1)$ in place of the last term for every k: take $M \ge \max(T(m), T(n), k)$. The Archimedean step of Theorem 12.17 removes it. □

No completed real number appears in this argument. The ideal values are ratios, and the conclusion is an inequality between integers.

Rounding is the only other ingredient. For a ratio q write $\operatorname{near}(q) = \lfloor q + 1/2 \rfloor$ for the nearest integer, halves rounded upward, so that $|\operatorname{near}(q) - q| \le 1/2$. For integers u and k, $\operatorname{near}(u/2^{k+1}) = \lfloor (u + 2^k)/2^{k+1} \rfloor$ is computed by one floor division. For a natural number n write $\operatorname{bl}(n)$ for its number of binary digits, with $\operatorname{bl}(0) = 0$; then $n < 2^{\operatorname{bl}(n)}$.

Proposition 15.4 (Sum and product with additive precision)**.** *Let $x = (x_n)$ and $y = (y_n)$ be certified reals. The negative $-x$ has answers $-x_n$.*

(a) Put $s_n = \operatorname{near}\big((x_{n+2} + y_{n+2})/4\big)$. For $M \ge n+2$,

$$\left|\frac{s_n}{2^n} - \left(\frac{x_M}{2^M} + \frac{y_M}{2^M}\right)\right| \le 2^{-n} + 2 \cdot 2^{-M},$$

so (s_n) is a certified real, the sum $x + y$.

(b) Put $e_x = \mathrm{bl}(|x_0|) + 1$ *and* $e_y = \mathrm{bl}(|y_0|) + 1$. *Then* $|x_M/2^M| \le 2^{e_x}$ *for every* M. *With* $k = n + 3 + e_x + e_y$, *put* $p_n = \mathrm{near}\big(x_k y_k / 2^{2k-n}\big)$. *For every* M,

$$\left|\frac{p_n}{2^n} - \frac{x_M}{2^M}\cdot\frac{y_M}{2^M}\right| \le 2^{-n} + \left(2^{e_x} + 2^{e_y}\right) 2^{-M},$$

so (p_n) *is a certified real, the product* xy.

Proof. (a) Rounding gives $|s_n/2^n - (x_{n+2} + y_{n+2})/2^{n+2}| \le 2 \cdot 2^{-(n+2)} = 2^{-(n+1)}$. The moduli of x and y give $|x_{n+2}/2^{n+2} - x_M/2^M| \le 2^{-(n+2)} + 2^{-M}$, and the same for y. Adding the three bounds gives $2^{-(n+1)} + 2^{-(n+1)} + 2 \cdot 2^{-M}$. The ideal-value lemma with $I(M) = x_M/2^M + y_M/2^M$, $T(n) = n + 2$ and $C = 2$ applies.

(b) The modulus at M and 0 gives $|x_M/2^M - x_0| \le 2^{-M} + 1$, and $|x_0| \le 2^{\mathrm{bl}(|x_0|)} - 1$. So $|x_M/2^M| \le 2^{\mathrm{bl}(|x_0|)} + 2^{-M} \le 2^{e_x}$. Rounding costs $|p_n/2^n - x_k y_k/2^{2k}| \le 2^{-(n+1)}$. For the product of stages, with $p = x_k/2^k$, $q = y_k/2^k$, $p' = x_M/2^M$, $q' = y_M/2^M$,

$$|pq - p'q'| \le |p|\,|q - q'| + |q'|\,|p - p'| \le (2^{e_x} + 2^{e_y})(2^{-k} + 2^{-M}).$$

Finally $2^{e_x} + 2^{e_y} \le 2^{e_x + e_y + 1}$, so $(2^{e_x} + 2^{e_y})2^{-k} \le 2^{-(n+2)}$ and $2^{-(n+1)} + 2^{-(n+2)} \le 2^{-n}$. The ideal-value lemma with $I(M) = (x_M/2^M)(y_M/2^M)$, $T(n) = k$ and $C = 2^{e_x} + 2^{e_y}$ applies. For the negative, the modulus is unchanged by a change of sign. □

The sum asks its inputs for $n + 2$ bits and the product for $n + 3 + e_x + e_y$, where e_x and e_y are read off the inputs' answers at precision 0. Both are n plus a constant. The line's own operations, in Proposition 12.19, read their operands at $2n + 1$ and at $2K(n + 1) - 1$. They compute the same points.

Proposition 15.5 (The certified operations are the line's operations)**.** *Write* $\bar{x}$ *for the point of* $\mathbb{R}_\delta$ *named by a certified real* x. *Then* $\overline{x + y} = \bar{x} + \bar{y}$, $\overline{-x} = -\bar{x}$ *and* $\overline{xy} = \bar{x}\,\bar{y}$, *with the operations of Proposition 12.19 on the right.*

Proof. Let u and v be the stage sequences of x and y. Stage n of the line's sum is $u_{2n+1} + v_{2n+1}$. For $n \ge 1$ we have $2n + 1 \ge n + 2$, and part (a) of Proposition 15.4 with $M = 2n + 1$ puts stage n of $x + y$ within $2^{-n} + 2 \cdot 2^{-(2n+1)} \le 2 \cdot 2^{-n} \le 2/(n + 1)$ of it. For a tolerance $1/(k + 1)$ every $n \ge \max(1, 2k + 1)$ is within it, so the two regular sequences name the same point (Definition 12.14). Stage n of the line's product is $u_j v_j$ for an index $j \ge n$, and part (b) with $M = j$ puts stage n of xy within $2^{-n} + C2^{-j} \le (1 + C)/(n + 1)$ of it, where $C = 2^{e_x} + 2^{e_y}$; Lemma F.8(a) applies. The stages of $-x$ are the negated stages of x, which are the stages of the line's negative. □

A certified real carries its error with it. In the order of the line, $\bar{x}$ and $\eta(a_m/2^m)$ are within 2^{-m} in the sense of Chapter 12: $\bar{x} \le \eta(a_m/2^m) + \eta(2^{-m})$ and $\eta(a_m/2^m) \le \bar{x} + \eta(2^{-m})$. Indeed $2^{-m} = 1/(k+1)$ for $k = 2^m - 1$, and the modulus gives $|a_n/2^n - a_m/2^m| \le 2^{-m} + 2^{-n} \le 2^{-m} + 2/(n+1)$ for every n. This is the stagewise bound of Lemma 12.23 for the stages of x and the constant sequence of $a_m/2^m$, and that lemma gives the two inequalities. A certified real is thus exactly a program that answers every precision request with an integer whose error is proved.

15.2 Carrying every rounding error forward

A long calculation rounds many times. Two tools keep the errors under control: enclosures that carry a radius with every number, and an inequality that propagates the error of a step map through many steps.

Fix a working precision w. A *ball* is a pair (c, r) of integers, read as the dyadic centre $c/2^w$ and radius $r/2^w$.

Definition 15.6 (Ball enclosure)**.** The ball (c, r) *encloses* a ratio p at precision w if $|p - c/2^w| \le r/2^w$; for $p = \langle u; e\rangle$ with $e > 0$ this is the integer inequality $|u2^w - ce| \le re$.

Sums of balls add centres and radii. Products need one floor division and a small correction for the two floors.

Lemma 15.7 (The product of two balls)**.** *If (c_1, r_1) encloses p and (c_2, r_2) encloses q at precision w, then*

$$\left(\left\lfloor \frac{c_1 c_2}{2^w} \right\rfloor,\ \left\lfloor \frac{|c_1| r_2 + |c_2| r_1 + r_1 r_2}{2^w} \right\rfloor + 2\right)$$

encloses pq at precision w.

Proof. Write $p = (c_1 + \varepsilon_1)/2^w$ and $q = (c_2 + \varepsilon_2)/2^w$ with ratios $|\varepsilon_i| \le r_i$. Then

$$2^w pq = \frac{c_1 c_2}{2^w} + \frac{c_1\varepsilon_2 + c_2\varepsilon_1 + \varepsilon_1\varepsilon_2}{2^w}.$$

Let $c = \lfloor c_1 c_2 / 2^w \rfloor$, so that $c_1 c_2/2^w = c + \theta$ with $0 \le \theta < 1$. The second fraction has absolute value at most $Q = (|c_1| r_2 + |c_2| r_1 + r_1 r_2)/2^w < \lfloor Q \rfloor + 1$. Hence $|2^w pq - c| \le \theta + Q < \lfloor Q \rfloor + 2$, which is the claim after division by 2^w. □

Interval arithmetic of this kind goes back to Moore [47]. The solver below needs a different organisation of the same idea: it keeps exact dyadic values and proves, once and for all, how far each rounding can move the result after many steps.

A *closeness relation* on a set V is a relation "a and b are within r", for $a, b \in V$ and ratios r, that is symmetric in a and b, satisfies the triangle law (if a, b are within r and b, c within s, then a, c are within $r + s$) and depends only on the value of r. The examples are ratios with $|a - b| \le r$ and vectors of $\mathbb{Q}_\delta^d$ with the sup norm, $\|a - b\| = \max_i |a_i - b_i| \le r$. A map $\Phi : V \to V$ is *g-Lipschitz* when a, b within r implies $\Phi(a), \Phi(b)$ within gr.

Theorem 15.8 (Discrete Gronwall inequality)**.** *Let Φ be g-Lipschitz on a set with a closeness relation. Suppose that Y_{k+1} is within τ of $\Phi(Y_k)$ and z_{k+1} is within ρ of $\Phi(z_k)$ for every k, and that Y_0 and z_0 are within e_0. Put $E_0 = e_0$ and $E_{k+1} = gE_k + (\tau + \rho)$. Then Y_k and z_k are within E_k for every k. If moreover $g \ge 1$ and $\delta = \tau + \rho$ and e_0 are nonnegative, then*

$$E_k \le g^k e_0 + k\, g^k \delta.$$

Proof. Induction on k. If Y_k and z_k are within E_k, then Y_{k+1} is within τ of $\Phi(Y_k)$, which is within gE_k of $\Phi(z_k)$, which is within ρ of z_{k+1}; the triangle law gives $gE_k + \tau + \rho$. For the bound, $E_{k+1} = gE_k + \delta \le g^{k+1} e_0 + k\, g^{k+1}\delta + \delta$ by induction, and $\delta \le g^{k+1}\delta$ because $g \ge 1$. □

Real vector fields are tame only on a region: a polynomial is Lipschitz on a ball, not on all of space. The error bound then needs a second fact, that the trajectories stay in the region. A margin between an inner region and the region supplies it.

Theorem 15.9 (Gronwall on a region). *Let $D_{\mathrm{in}} \subseteq D$ be subsets of a set with a closeness relation, and let Φ satisfy the Lipschitz condition with constant g for pairs of points of D. Let K be a natural number and μ a ratio such that every point within $r \le \mu$ of a point of D_{in} lies in D. Suppose that*

(i) $z_k \in D_{\mathrm{in}}$ for $k \le K$;

(ii) for $k < K$, Y_{k+1} is within τ of $\Phi(Y_k)$ whenever $Y_k \in D$, and z_{k+1} is within ρ of $\Phi(z_k)$ whenever $z_k \in D$;

(iii) Y_0 and z_0 are within e_0, and $E_k \le \mu$ for $k \le K$, with E_k as in Theorem 15.8.

Then for every $k \le K$, $Y_k \in D$ and Y_k, z_k are within E_k. If instead both Y_k and z_k are known to lie in D for every $k \le K$, condition (ii) and Y_0, z_0 within e_0 give the same bound for $k \le K$. For sup-norm balls the margin condition holds: if $\|a\| \le M - \mu$ and $\|a - b\| \le r \le \mu$, then $\|b\| \le M$.

Proof. Induction on $k \le K$, carrying both claims. At $k = 0$, Y_0 is within $e_0 = E_0 \le \mu$ of $z_0 \in D_{\mathrm{in}}$, so $Y_0 \in D$. At $k + 1 \le K$, both Y_k and z_k lie in D ($z_k \in D_{\mathrm{in}} \subseteq D$), so the step of Theorem 15.8 applies and puts Y_{k+1} within $E_{k+1} \le \mu$ of z_{k+1}; since $z_{k+1} \in D_{\mathrm{in}}$, the margin puts Y_{k+1} in D. The second form is the same induction without the membership claim. The sup-norm statement is the triangle inequality $\|b\| \le \|a\| + \|a - b\| \le M - \mu + \mu$. □

The theorem asks nothing of the region beyond the margin. Its use below is typical: the reference trajectory z is one that has been computed and checked, and the conclusion transfers that check to every finer computation.

15.3 The number of steps at a fixed order

The simplest way to follow $y' = F(y)$ is Euler's method: from y go to $\mathscr{E}_h(y) = y + hF(y)$. Its error analysis is one application of the Gronwall inequality.

Proposition 15.10 (Global error of Euler's method). *Let $F : \mathbb{Q}_\delta^d \to \mathbb{Q}_\delta^d$ satisfy $\|F(a) - F(b)\| \le L\,r$ whenever $\|a - b\| \le r$, and let $h \ge 0$. If $\|Y_{k+1} - \mathscr{E}_h(Y_k)\| \le \tau$ and $\|z_{k+1} - \mathscr{E}_h(z_k)\| \le \rho$ for every k, and $\|Y_0 - z_0\| \le e_0$, then $\|Y_k - z_k\| \le E_k$ for every k, with E_k as in Theorem 15.8 for $g = 1 + hL$.*

Proof. If $\|a - b\| \le r$, then $\|\mathscr{E}_h(a) - \mathscr{E}_h(b)\| \le \|a - b\| + h\|F(a) - F(b)\| \le (1 + hL)r$. So $\mathscr{E}_h$ is $(1 + hL)$-Lipschitz, and Theorem 15.8 applies. □

Here Y can be any sequence that follows the Euler map up to τ per step, and z the rounded sequence that a program computes. The bound grows like $e_0 + k(\tau + \rho)$ times the growth factor $(1 + hL)^k$. With T/h steps and a one-step defect of order h^2, a bound of this kind is of order h: each further certified bit halves the step and doubles the number of steps. For methods of fixed order this is in general not a weakness of the bound. The next theorem proves it for the classical fourth-order Runge–Kutta method [48], on the simplest equation

there is. It is a classical statement about real numbers: it uses the real exponential function and is proved with classical real analysis.

One Runge–Kutta step of size h for $y' = f(y)$ takes y to $y + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$, where $k_1 = f(y)$, $k_2 = f(y + \frac{h}{2}k_1)$, $k_3 = f(y + \frac{h}{2}k_2)$ and $k_4 = f(y + hk_3)$.

Theorem 15.11 (Fourth-order Runge–Kutta needs exponentially many steps)**.** *Let $K \geq 1$, and let RK_K be the result of K Runge–Kutta steps of size $1/K$ for $y' = y$ from $y(0) = 1$, in exact real arithmetic. Then*

$$e - \mathrm{RK}_K \geq \frac{1}{120K^4}.$$

Consequently, if $|e - \mathrm{RK}_K| \leq 2^{-n}$, then $2^n \leq 120K^4$.

Proof. For $f(y) = y$ the four stages are $k_1 = y$, $k_2 = y(1 + \frac{h}{2})$, $k_3 = y(1 + \frac{h}{2} + \frac{h^2}{4})$ and $k_4 = y(1 + h + \frac{h^2}{2} + \frac{h^3}{4})$, so one step multiplies y by

$$b = 1 + h + \frac{h^2}{2} + \frac{h^3}{6} + \frac{h^4}{24},$$

the Taylor polynomial of order four of e^h. Hence $\mathrm{RK}_K = b^K$ with $h = 1/K$. All terms of the exponential series are nonnegative for $h \geq 0$, so $e^h \geq b + c$ with $c = h^5/120$. The binomial expansion, whose terms are nonnegative, and $b \geq 1$ give

$$e = (e^h)^K \geq (b + c)^K \geq b^K + Kb^{K-1}c \geq \mathrm{RK}_K + Kc,$$

and $Kc = K \cdot K^{-5}/120 = 1/(120K^4)$. If the error is at most 2^{-n}, then $1/(120K^4) \leq 2^{-n}$, which is the second claim. □

So n correct bits require at least $(2^n/120)^{1/4}$ steps. The way out is to keep the step fixed and let the order grow with the request. A Taylor method of order N with step $2^{-\sigma}$, where 2^σ is large compared with the field's growth, truncates with error of order 2^{-N}. Taking $N = m$ plus a constant gives m bits with a fixed number of steps. The next two sections make this exact.

15.4 Taylor coefficients of a quadratic field

Every polynomial system can be rewritten as a quadratic one (Section 15.7), so we work with fields $y' = c + Ay + B(y, y)$ on $\mathbb{Q}_\delta^d$. The analysis uses only bounds, which we make part of the data.

Definition 15.12 (Quadratic field)**.** A *quadratic field* on $\mathbb{Q}_\delta^d$ consists of a vector c, maps $A : \mathbb{Q}_\delta^d \to \mathbb{Q}_\delta^d$ and $B : \mathbb{Q}_\delta^d \times \mathbb{Q}_\delta^d \to \mathbb{Q}_\delta^d$, and natural numbers γ, α, β, such that

(i) $\|c\| \leq \gamma$;

(ii) $\|u\| \leq P$ implies $\|Au\| \leq \alpha P$, and $\|u - u'\| \leq r$ implies $\|Au - Au'\| \leq \alpha r$;

(iii) $\|u\| \leq P$ and $\|v\| \leq P'$ imply $\|B(u, v)\| \leq \beta PP'$;

(iv) $\|u-u'\|\le r$, $\|v-v'\|\le s$, $\|v\|\le P$ and $\|u'\|\le P'$ imply $\|B(u,v)-B(u',v')\|\le \beta(rP+P's)$.

For a natural number $M\ge 1$, the *majorant ratio* on the ball of radius M is $R=\gamma+\alpha+\beta M+1$.

When A is linear and B bilinear, a power series $y(t)=\sum_k a_k t^k$ solves $y'=c+Ay+B(y,y)$ exactly when its coefficients satisfy, by comparing the coefficients of t^k,

$$(k+1)\,a_{k+1}=c\,[k=0]+Aa_k+\sum_{i=0}^{k}B(a_i,a_{k-i}),$$

where $[k=0]$ is 1 for $k=0$ and 0 otherwise. This recurrence is finite data for every k. A program that computes it must round, so we build the rounding into the definition.

Definition 15.13 (Coefficient table)**.** For a sequence $f_0,f_1,\dots$ of vectors put

$$\operatorname{next}(f,k)=\frac{1}{k+1}\Big(c\,[k=0]+Af_k+\sum_{i=0}^{k}B(f_i,f_{k-i})\Big).$$

A *rounding* is a map $\varrho:\mathbb{Q}_\delta^d\to\mathbb{Q}_\delta^d$ with $\|\varrho(u)-u\|\le u_\varrho$ for every u. The *coefficient table* at y with rounding ϱ is $a_0=y$, $a_{k+1}=\varrho(\operatorname{next}(a,k))$. With ϱ the identity these are the *exact* coefficients at y.

For the oscillator, $c=0$, $A(u_0,u_1)=(u_1,-u_0)$ and $B=0$, with $\gamma=0$, $\alpha=1$ and $\beta=0$, so $R=2$ on every ball. The exact table at $(0,1)$ is $(0,1)$, $(1,0)$, $(0,-\frac12)$, $(-\frac16,0)$, $(0,\frac{1}{24})$, …: the Taylor coefficients of $(\sin t,\cos t)$. Its entries are bounded by $1\cdot 2^k$, as the next lemma predicts.

Lemma 15.14 (Majorant of the coefficients)**.** *Let $M\ge 1$, $\|y\|\le M$ and $u_\varrho\le 1$. Then $\|a_k\|\le MR^k$ for every k.*

Proof. Strong induction on k. Suppose $\|a_i\|\le MR^i$ for every $i\le k$, and put $X=MR^k\ge 1$. Then $\|c\,[k=0]\|\le\gamma$, $\|Aa_k\|\le\alpha X$, and each of the $k+1$ terms satisfies $\|B(a_i,a_{k-i})\|\le\beta MR^i\,MR^{k-i}=\beta MX$. So

$$\|\operatorname{next}(a,k)\|\le\frac{\gamma+\alpha X+(k+1)\beta MX}{k+1}\le\gamma+\alpha X+\beta MX\le X(R-1),$$

using $\gamma\le\gamma X$. Rounding adds at most $u_\varrho\le 1\le X$, so $\|a_{k+1}\|\le XR=MR^{k+1}$. □

The second estimate compares two tables. It contains two facts in one: the coefficients depend Lipschitz-continuously on the starting point, and rounding perturbs them in a controlled way.

Lemma 15.15 (The difference law)**.** *Let a and a' be the coefficient tables at y and y' with roundings ϱ and ϱ', where $0\le u_\varrho,u_{\varrho'}\le 1$, $\|y\|,\|y'\|\le M$ and $\|y-y'\|\le e$ with $e\ge 0$. Then*

$$\|a_k-a'_k\|\le E(k+1)R^k,\qquad E=e+u_\varrho+u_{\varrho'},$$

for every k.

Proof. Strong induction on k; both tables satisfy Lemma 15.14. At $k = 0$ the bound is $\|y - y'\| \le e \le E$. Suppose it holds for every $i \le k$. The constant terms of $\mathrm{next}(a, k)$ and $\mathrm{next}(a', k)$ cancel, $\|Aa_k - Aa'_k\| \le \alpha E(k+1)R^k$, and by law (iv) of Definition 15.12

$$\|B(a_i, a_{k-i}) - B(a'_i, a'_{k-i})\| \le \beta\big(E(i+1)R^i \cdot MR^{k-i} + MR^i \cdot E(k-i+1)R^{k-i}\big) = \beta ME(k+2)R^k.$$

Summing the $k+1$ terms and dividing by $k+1$,

$$\|\,\mathrm{next}(a,k) - \mathrm{next}(a',k)\| \le \alpha ER^k + \beta ME(k+2)R^k \le E(k+2)R^k(R-1).$$

The two roundings add at most $u_\varrho + u_{\varrho'} \le E \le E(k+2)R^k$, so $\|a_{k+1} - a'_{k+1}\| \le E(k+2)R^{k+1}$. □

With $\varrho = \varrho'$ the identity, the lemma is the Lipschitz law of the exact coefficients. With $y = y'$ and ϱ' the identity, it bounds the effect of rounding every coefficient.

A Taylor step of order N and step $2^{-\sigma}$ sums the table:

$$S_N(f) = \sum_{k=0}^{N} 2^{-\sigma k} f_k.$$

The step must be small compared with $1/R$. We ask for $8R \le 2^\sigma$.

Lemma 15.16 (Taylor sums)**.** *Let $8R \le 2^\sigma$ and $g = 1 + 4R/2^\sigma$.*

(a) If $E \ge 0$ and $\|f_k - f'_k\| \le E(k+1)R^k$ for every $k \le N$, then $\|S_N(f) - S_N(f')\| \le gE$.

(b) If $\|f_k\| \le MR^k$ for every k, then $\|S_{N+j}(f) - S_N(f)\| \le M2^{-N}$ for every j.

Proof. (a) We prove by induction on N the sharper bound $E\big(1 + 4R(2^N - 1)/2^{\sigma+N}\big)$. For $N = 0$ it reads $\|f_0 - f'_0\| \le E$. The term added at $N+1$ is at most $2^{-\sigma(N+1)}E(N+2)R^{N+1}$. First, $(N+2)2^{N+1} \le 4 \cdot 8^N$, by induction on N, since $(N+3) \le \frac{3}{2}(N+2)$. Second, $(8R)^N \le 2^{\sigma N}$. Together they give $(N+2)R^{N+1}2^{\sigma+N+1} \le 4R\,2^\sigma(8R)^N \le 4R\,2^{\sigma(N+1)}$, so the new term is at most $E \cdot 4R/2^{\sigma+N+1}$, and

$$\frac{4R(2^N-1)}{2^{\sigma+N}} + \frac{4R}{2^{\sigma+N+1}} = \frac{4R(2^{N+1}-1)}{2^{\sigma+N+1}}.$$

Finally $(2^N - 1)/2^N \le 1$.

(b) Since $2R \le 2^\sigma$, each added term satisfies $\|2^{-\sigma k} f_k\| \le M(R/2^\sigma)^k \le M2^{-k}$. The terms of index $N+1$ to $N+j$ add up to at most $M(2^{-N} - 2^{-N-j}) \le M2^{-N}$. □

Write $\Phi_N(y) = S_N(a(y))$ for the exact Taylor map of order N, with $a(y)$ the exact table at y. For a working precision w, write $[q]_w$ for $\mathrm{near}(2^w q)/2^w$, applied coordinatewise to vectors; it moves each coordinate by at most $2^{-(w+1)}$. The *computed step* is

$$\Psi_{N,w}(y) = \big[S_N(a^{(w)}(y))\big]_w,$$

where $a^{(w)}(y)$ is the table at y with rounding $[\cdot]_w$: every coefficient and the final sum are rounded to w binary digits.

Lemma 15.17 (One step). *Let $8R \le 2^\sigma$, $g = 1 + 4R/2^\sigma$ and $M \ge 1$.*

(a) If $\|y\|, \|y'\| \le M$ and $\|y - y'\| \le r$ with $r \ge 0$, then $\|\Phi_N(y) - \Phi_N(y')\| \le gr$.

(b) If $\|y\| \le M$ and $N_0 \le N$, then

$$\|\Psi_{N,w}(y) - \Phi_{N_0}(y)\| \le 2^{-(w+1)} + 2^{-(w+1)}g + M2^{-N_0}.$$

Proof. (a) Lemma 15.15 without rounding gives $\|a_k(y) - a_k(y')\| \le r(k+1)R^k$; apply Lemma 15.16(a). (b) Split the difference in three. The final rounding costs $2^{-(w+1)}$. Rounded against exact coefficients, Lemma 15.15 with $e = 0$, $u_\varrho = 2^{-(w+1)}$ and $u_{\varrho'} = 0$, followed by Lemma 15.16(a), costs $2^{-(w+1)}g$. Order N against order N_0 for the exact coefficients costs $M2^{-N_0}$, by Lemmas 15.14 and 15.16(b). □

15.5 A fixed number of steps of growing order

Everything the solver does at precision m is fixed by the problem and m through a short schedule.

Definition 15.18 (Problem, schedule and trajectories). A *problem* consists of a quadratic field, a natural number $M \ge 1$ (the radius of a sup-norm ball), an initial point $y_0 \in \mathbb{Q}_\delta^d$ and a natural end time T. Its *schedule* is

$$R = \gamma + \alpha + \beta M + 1, \qquad \sigma = \mathrm{bl}(4R) + 1, \qquad S = T2^\sigma,$$
$$G = \left\lfloor \frac{(2^\sigma + 4R)^S}{2^{\sigma S}} \right\rfloor + 1, \qquad \kappa = \mathrm{bl}\big(G(1 + 2S(M+2))\big) + 2, \qquad N(n) = n + \kappa.$$

The step is $2^{-\sigma}$, the number of steps is S, and κ is the number of extra bits. The *trajectory at level n* is

$$y_0^{(n)} = [y_0]_{N(n)}, \qquad y_{k+1}^{(n)} = \Psi_{N(n),N(n)}\big(y_k^{(n)}\big) \ :$$

Taylor order and working precision both equal $N(n)$. The problem satisfies the *region certificate at level n_0* if

$$\big\|y_k^{(n_0)}\big\| \le M - 2^{-(n_0+2)} \qquad \text{for every } k \le S.$$

Since $4R < 2^{\mathrm{bl}(4R)}$, we have $8R \le 2^\sigma$, so the lemmas of Section 15.4 apply with $g = 1 + 4R/2^\sigma \le \frac{3}{2}$. The growth factor satisfies $g^S = (2^\sigma + 4R)^S/2^{\sigma S} < G$. None of these numbers depends on the precision. The region certificate is a finite list of comparisons of ratios: it concerns the $S + 1$ vectors of one computed trajectory and is checked by computing them. It is not an assumption about an exact solution.

For the oscillator with $M = 2$ and $T = 1$: $R = 2$, $\sigma = \mathrm{bl}(8)+1 = 5$, $S = 32$, $G = \lfloor (40/32)^{32} \rfloor + 1 = 1263$, and $\kappa = \mathrm{bl}(1263 \cdot 257) + 2 = 21$. The level-0 trajectory, with order and precision 21, passes the region certificate at level 0, so every level is covered by the results below. At $m = 40$ the solver runs level 40, with order and precision 61; these are the numbers of the opening.

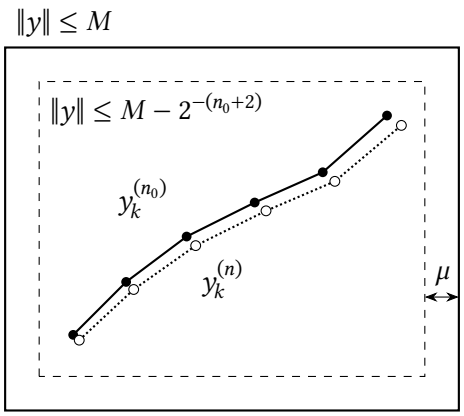


Figure 15.1. The region certificate in two dimensions, where the sup-norm balls are squares. The trajectory at level n_0 (solid, filled points) has been computed and stays in the inner square. A finer level $n \ge n_0$ (dotted, open points) stays within $E_k \le \mu = 2^{-(n_0+2)}$ of it, so it stays in the outer square, where the Taylor step is Lipschitz (Theorem 15.20).

Lemma 15.19 (The error budget)**.** *Let $n \le n'$. Let τ and ρ be the right-hand sides of Lemma 15.17(b) for a step at level n' and at level n, compared with $\Phi_{N(n)}$:*

$$\tau = 2^{-(N(n')+1)}(1+g) + M2^{-N(n)}, \qquad \rho = 2^{-(N(n)+1)}(1+g) + M2^{-N(n)},$$

and let $e_0 = 2^{-(N(n')+1)} + 2^{-(N(n)+1)}$. Then the bound E_k of Theorem 15.8 with these g, τ, ρ and e_0 satisfies $E_k \le 2^{-(n+2)}$ for every $k \le S$.

Proof. Because $g \le 2$, each of τ and ρ is at most $3 \cdot 2^{-(N(n)+1)} + M2^{-N(n)} \le (M+2)2^{-N(n)}$, and $e_0 \le 2^{-N(n)}$. By the second part of Theorem 15.8, with $g^k \le g^S < G$ and $k \le S$,

$$E_k \le G\,2^{-N(n)} + SG \cdot 2(M+2)\,2^{-N(n)} = G\big(1 + 2S(M+2)\big)2^{-(n+\kappa)}.$$

By the definition of κ, $G(1 + 2S(M+2)) < 2^{\kappa-2}$, so $E_k < 2^{-(n+2)}$. □

Theorem 15.20 (Every finer level stays in the region)**.** *If the region certificate holds at level n_0 and $n \ge n_0$, then $\|y_k^{(n)}\| \le M$ for every $k \le S$.*

Proof. The geometry is that of Figure 15.1. Apply Theorem 15.9 with D the ball of radius M, D_{in} the ball of radius $M - 2^{-(n_0+2)}$, margin $\mu = 2^{-(n_0+2)}$, and $\Phi = \Phi_{N(n_0)}$, which is g-Lipschitz on D by Lemma 15.17(a). The reference trajectory is $z = y^{(n_0)}$; it lies in D_{in} by the certificate and follows Φ up to ρ by Lemma 15.17(b). The other trajectory is $Y = y^{(n)}$; while $Y_k \in D$ it follows Φ up to τ by Lemma 15.17(b) with $N_0 = N(n_0) \le N(n)$. The two starting points are roundings of y_0, so they are within e_0. Lemma 15.19 with $n_0 \le n$ gives $E_k \le 2^{-(n_0+2)} = \mu$ for $k \le S$. □

Theorem 15.21 (Two levels agree at the end time)**.** *If the region certificate holds at level n_0 and $n_0 \le n \le n'$, then*

$$\big\| y_S^{(n')} - y_S^{(n)} \big\| \le 2^{-(n+2)}.$$

Proof. By Theorem 15.20 both trajectories stay in the ball of radius M for $k \le S$. The second form of Theorem 15.9, with $\Phi = \Phi_{N(n)}$, $Y = y^{(n')}$, $z = y^{(n)}$ and the defects of Lemma 15.17(b), bounds their distance at step S by E_S, and Lemma 15.19 bounds E_S by $2^{-(n+2)}$. □

The solution value is now a certified real, read off the trajectory at the right level.

Theorem 15.22 (The solution value as a certified real)**.** *Let the region certificate hold at level n_0, and let $i < d$. Put*

$$a_m = \operatorname{near}\Big(2^m \big(y_S^{(\max(m,n_0))}\big)_i\Big).$$

Then $|a_m/2^m - (y_S^{(\max(M',n_0))})_i| \le 2^{-m}$ whenever $m \le M'$, and (a_m) is a certified real x_i. We call x_i, and the point $\bar{x}_i$ of $\mathbb{R}_\delta$ that it names, coordinate i of the solution value at time T.

Proof. Rounding costs $2^{-(m+1)}$. Theorem 15.21 with $n = \max(m, n_0) \le n' = \max(M', n_0)$ costs at most $2^{-(\max(m,n_0)+2)} \le 2^{-(m+2)}$. The sum is at most $3 \cdot 2^{-(m+2)} < 2^{-m}$. The ideal-value lemma with $I(M') = (y_S^{(\max(M',n_0))})_i$, $T(m) = m$ and $C = 0$ gives the certified real. □

No exact solution, no Lipschitz constant and no completed real number is assumed. The modulus of x_i follows from the four bound laws and one finite check. Every proof in Sections 15.1, 15.2, 15.4 and 15.5 is a computation with ratios, an induction or the Archimedean step; none uses a choice principle or an omniscience principle. For the oscillator, the answer of x_0 at precision 40 is 925207132232. Section 15.8 shows that, under classical hypotheses that hold here, x_0 reads $\sin 1$.

15.6 Counting the work

The answer a_m is computed by an explicit finite procedure: round y_0, take S computed steps at level $\max(m, n_0)$, select coordinate i and round. Within a step the order of evaluation is fixed. Row $k+1$ of the table is formed from the vector Af_k, the Cauchy sum $\sum_{i \le k} B(f_i, f_{k-i})$, accumulated term by term from the zero vector, the sum $c\,[k=0] + Af_k$, to which the Cauchy sum is then added, the division by $k+1$ and the rounding. The Taylor sum is accumulated from f_0, adding $2^{-\sigma k} f_k$ for $k = 1, \dots, N$ in turn, and then rounded. We count the arithmetic of this procedure.

The *cost model* charges one unit for each arithmetic operation on ratios: a sum, a product, a quotient by a natural number, a rounding $[\cdot]_w$, and the final rounding to an integer. An operation on vectors of $\mathbb{Q}_\delta^d$ costs d; one evaluation of A costs c_A and one of B costs c_B. Storing and reading values is free, and the size of the operands is not counted here; it is the subject of Proposition 15.26.

Theorem 15.23 (The exact count)**.** *Let the region certificate hold at level n_0, and let $N = N(\max(m, n_0))$. The procedure above returns the answer a_m of Theorem 15.22, and its count is exactly*

$$\operatorname{cost}(m) = 1 + d + S \cdot \operatorname{step}(N), \qquad 2\operatorname{step}(N) = 2N(c_A + 6d) + 2d + (c_B + d)N(N + 1).$$

Moreover

$$\operatorname{cost}(m) \le \big(S(c_A + c_B + 8d) + d + 1\big)(m + n_0 + \kappa + 1)^2 \qquad \text{and} \qquad S \le 16RT.$$

Proof. The procedure follows the definitions of Section 15.5 step for step, so it returns a_m. Row $k + 1$ of a coefficient table costs one evaluation of A, $k + 1$ evaluations of B with $k + 1$ vector additions in the sum (together $(k + 1)(c_B + d)$), two vector additions and one division ($3d$), and the rounding of the row (d). So the rows up to k cost $\operatorname{rows}(k)$ with $\operatorname{rows}(0) = 0$ and $\operatorname{rows}(k + 1) = \operatorname{rows}(k) + c_A + (k + 1)(c_B + d) + 4d$, and induction gives $2\operatorname{rows}(k) = 2k(c_A + 4d) + (c_B + d)k(k + 1)$. The Taylor sum of order N costs N scalings and N additions ($2dN$), and its rounding d; so $\operatorname{step}(N) = \operatorname{rows}(N) + 2dN + d$, which is the displayed formula. The trajectory costs d for the initial rounding and $S\operatorname{step}(N)$ for the steps, and the answer one more rounding.

For the bound, comparing coefficients shows $2\operatorname{step}(N) \le 2(c_A + c_B + 8d)(N + 1)^2$. With $Q = m + n_0 + \kappa + 1 \ge N + 1 \ge 1$, $\operatorname{cost}(m) \le (1 + d)Q^2 + S(c_A + c_B + 8d)Q^2$. Finally $2^{\operatorname{bl}(n)} \le 2n$ for $n \ge 1$, so $2^\sigma = 2 \cdot 2^{\operatorname{bl}(4R)} \le 16R$ and $S = T2^\sigma \le 16RT$. □

The count is quadratic in m because one step of order N forms the Cauchy sums of every row, about $N^2/2$ evaluations of B, while the number of steps does not depend on m. The extra precision κ does not depend on m either. It grows at most linearly in RT: with $a = 4R/2^\sigma$ we have $aS = 4RT$ and $(1 + a)^S \le e^{aS} = e^{4RT}$, so G has at most about $5.8\,RT$ binary digits, and κ adds to these the digits of $1 + 2S(M + 2)$ and two more.

Fields met in practice are given by their coefficients.

Definition 15.24 (Coefficient field)**.** A *coefficient field* of dimension d consists of a vector $c \in \mathbb{Q}_\delta^d$ and, for each coordinate i, a finite list of terms (a, j) and a finite list of terms (b, j, k), with ratios a, b and indices $j, k < d$. It defines

$$(Au)_i = \sum_{(a,j)} a\,u_j, \qquad B(u, v)_i = \sum_{(b,j,k)} b\,u_j v_k,$$

with the bounds $\gamma = \max_i \lceil |c_i| \rceil$, $\alpha = \max_i \sum_{(a,j)} \lceil |a| \rceil$ and $\beta = \max_i \sum_{(b,j,k)} \lceil |b| \rceil$, the sums running over the terms of coordinate i.

Proposition 15.25 (Coefficient fields are quadratic fields)**.** *A coefficient field satisfies the four laws of Definition 15.12. One evaluation of A costs $c_A = 2 \cdot$ (number of A-terms) operations and one evaluation of B costs $c_B = 3 \cdot$ (number of B-terms). Hence, for a problem whose field is a coefficient field and which satisfies the region certificate at level n_0, the answer a_m costs at most*

$$\big(S(c_A + c_B + 8d) + d + 1\big)(m + n_0 + \kappa + 1)^2$$

operations, with $S \le 16RT$.

Proof. Each law is the triangle inequality over a finite sum. For instance, if $\|u - u'\| \le r$, $\|v - v'\| \le s$, $\|v\| \le P$ and $\|u'\| \le P'$, then $|bu_j v_k - bu'_j v'_k| \le |b|\big(|u_j - u'_j|\,|v_k| + |u'_j|\,|v_k - v'_k|\big) \le \lceil|b|\rceil(rP + P's)$, and summing over the terms of coordinate i gives at most $\beta(rP + P's)$. The bounds for c and A are the same argument with one factor. A term of A is evaluated with one product and one addition, a term of B with two products and one addition. Insert c_A and c_B into Theorem 15.23. □

Unit cost per operation is only half of a cost statement: the numbers could grow. They do not grow faster than the precision allows. A ratio is stored as an integer numerator over a positive integer denominator, without reduction, and we bound both.

Proposition 15.26 (Operand sizes)**.** *Let a problem have a coefficient field and satisfy the region certificate at level n_0, and let $N = N(\max(m, n_0))$. Let $\mathscr{B}(N)$ be the capacity* (H.1) *of Appendix H; it is a polynomial $q_P N^2 + l_P N + c_P$ whose natural coefficients depend only on the problem. Every ratio that the computation of a_m stores, uses as an operand or produces has numerator and denominator of absolute value at most $2^{\mathscr{B}(N)}$. Moreover*

$$\mathscr{B}(N) \le (q_P + l_P + c_P)\,(m + n_0 + \kappa + 1)^2,$$

so each of these ratios is written with at most $2\mathscr{B}(N) + 3$ binary digits.

The proof, a bookkeeping of numerators and denominators through every operation, is in Section H.1 of Appendix H. Together with Theorem 15.23 it says that m bits cost $O(m^2)$ operations on numbers of $O(m^2)$ digits, for a fixed problem. With any implementation of the four operations whose cost is bounded by a polynomial in the lengths of the operands, the answer a_m therefore costs a number of bit operations bounded by a polynomial in m. Section 15.9 makes this exact for one fixed machine.

15.7 Polynomial fields become quadratic

Reducing a polynomial system to a quadratic one by new variables goes back to Kerner [49]. We introduce one new variable for every monomial of bounded degree.

Definition 15.27 (Polynomial field and its quadratization)**.** A *polynomial field* p on $\mathbb{Q}_\delta^d$ gives each coordinate a finite list of terms (q, ξ), with a ratio q and an exponent vector $\xi \in \mathbb{N}^d$, and $p_i(x) = \sum q\, x^\xi$ over the terms of p_i. Let $D = \max(1, \deg p)$, where $\deg p$ is the largest total degree $|\xi| = \xi_0 + \cdots + \xi_{d-1}$ of a term. The *quadratized field* has one variable w_α for each exponent vector α with $|\alpha| \le D$, the constant monomial included. It is the coefficient field with $c = 0$, no A-terms, and, for the variable w_α, the B-terms

$$(\alpha_i q,\ \alpha - e_i,\ \xi) \qquad \text{for each } i \text{ with } \alpha_i > 0 \text{ and each term } (q, \xi) \text{ of } p_i,$$

where e_i is the ith unit vector. For an initial point x_0, a radius $M \ge 1$ and an end time T, the *quadratized problem* has the initial values $w_\alpha(0) = x_0^\alpha$, radius M and end time T. If it satisfies the region certificate at level n_0, the *polynomial solution* at time T is the certified real x_i of Theorem 15.22 for the variable w_{e_i}.

The construction mirrors the chain rule: for a solution x of $x' = p(x)$,

$$(x^\alpha)' = \sum_i \alpha_i x^{\alpha - e_i} p_i(x) = \sum_i \sum_{(q,\xi)\in p_i} (\alpha_i q)\, x^{\alpha - e_i} x^\xi.$$

Both $\alpha - e_i$, when $\alpha_i > 0$, and ξ have total degree at most D, so both are variables, and no monomial needs to be split. The certified computation uses only the quadratized field; the chain rule enters when the result is compared with a classical solution (Corollary 15.30). Conversely, a solution of the quadratized system that starts at the monomials of x_0 consists of the monomials of a solution of $x' = p(x)$ (Proposition H.11): the new variables add no solutions.

Theorem 15.28 (Dimension and cost of the quadratization)**.** *Let p be a polynomial field with $D = \max(1, \deg p)$ and t_p terms in all.*

(a) The variables are exactly the exponent vectors of total degree at most D; there are at most $(D+1)^d$ of them. Write d_Q for their number.

(b) The quadratized field has $c_A = 0$ and $c_B \le 3 d_Q t_p$.

(c) If the quadratized problem satisfies the region certificate at level n_0, the answer of the polynomial solution x_i at precision m costs at most

$$\big(S(3d_Q t_p + 8d_Q) + d_Q + 1\big)(m + n_0 + \kappa + 1)^2$$

operations, where $S \le 16RT$ and R and κ belong to the quadratized problem.

Proof. (a) List the vectors by recursion on d: for $d = 0$ there is the empty vector; for $d+1$ coordinates take, for each first entry $a \le D$, the vectors (a, ξ) with ξ in the list for d coordinates and degree $D - a$. Induction on d shows that (a, ξ) is listed exactly when $a + |\xi| \le D$, and that the list has at most $(D+1)\cdot(D+1)^d$ entries. (b) The field has no A-terms. The variable w_α receives at most one B-term for each term of each p_i, so at most t_p terms, and $c_B \le 3 d_Q t_p$. (c) Insert (a) and (b) into Proposition 15.25. □

Like the proofs of Sections 15.1, 15.2, 15.4 and 15.5, the proofs of Theorems 15.23 and 15.28 and of Proposition 15.25 are finite computations with ratios and inductions; they use no choice principle and no omniscience principle.

15.8 What the certified point is

Theorem 15.22 produced a point of $\mathbb{R}_\delta$ from finite data and one finite check. Classical analysis has its own object, the solution $y(T)$ of the initial value problem, whenever one exists. The two agree under the hypotheses of the following theorem. It is a classical statement about real numbers: its hypotheses and its proof use the real numbers of classical analysis, their completeness and the calculus of derivatives. None of the results of Sections 15.1 to 15.7 depends on it.

Theorem 15.29 (The certified point is the classical solution value)**.** *Let a problem satisfy the region certificate at level n_0. Suppose that A extends to a real linear map and B to a real bilinear map on $\mathbb{R}^d$, agreeing with A and B on rational vectors, with $\|Au\| \le \alpha\|u\|$ and $\|B(u,v)\| \le \beta\|u\|\,\|v\|$. Suppose that $y : [0,T] \to \mathbb{R}^d$ is continuous, $y(0) = y_0$, y has the right derivative $c + Ay(t) + B(y(t), y(t))$ at every $t \in [0,T)$, and $\|y(t)\| \le M$ on $[0,T]$. Then, for every level $n \ge n_0$,*

$$\left\| y(T) - y_S^{(n)} \right\| \le C_P\, 2^{-n}, \qquad C_P = e^{(\alpha+4\beta M)T}\Big(1 + S\big((\alpha M + \beta M^2)2^{-\sigma} + M + 2\big)\Big),$$

and the answers of the solution value satisfy $|a_m/2^m - y_i(T)| \le 2^{-m}$ for every m and every coordinate i.

The proof is in Section H.2. Its idea is to compare, on each step, the true solution with the Taylor polynomial of the computed point: that polynomial misses the equation by a residual of order 8^{-N}, the classical Gronwall inequality bounds the resulting distance, and the per-step errors are added up as in Section 15.2.

Corollary 15.30 (Polynomial fields)**.** *Let the quadratized problem of a polynomial field p, with initial point x_0, radius M and end time T, satisfy the region certificate at level n_0. Suppose that $x : [0,T] \to \mathbb{R}^d$ is continuous, $x(0) = x_0$, x has the right derivative $p(x(t))$ at every $t \in [0,T)$, and every monomial $x(t)^\alpha$ with $|\alpha| \le D$ has absolute value at most M on $[0,T]$. Then the answers of the polynomial solution x_i satisfy $|a_m/2^m - x_i(T)| \le 2^{-m}$ for every m. The monomial condition holds when $\|x(t)\| \le r$ on $[0,T]$ for some $r \ge 1$ with $r^D \le M$.*

For the oscillator, $y(t) = (\sin t, \cos t)$ satisfies the hypotheses with $M = 2$, so $925207132232/2^{40}$ lies within 2^{-40} of $\sin 1$. The hypotheses of the theorem are the ones the identification uses: a real realization of the field with the same bounds, and a classical solution that stays in the ball of the certificate. The theorem assumes that solution; it does not construct one. Where no classical solution is known to stay in that ball, Theorem 15.29 does not apply as stated, and the certified point is still defined and still computed to every precision.

15.9 One machine for every problem and every precision

Theorem 15.23 counts operations on ratios of unbounded size. The final step of the chapter replaces that count by machine steps on a tape of bits, with one machine for all problems at once. The machine is the plainest one available.

Definition 15.31 (One-tape binary machine)**.** A *control table* is a finite list $H = (h_0, \dots, h_{L-1})$ of instructions, each of one of the forms write(b, s'), left(s'), right(s'), branch(s_0, s_1) and halt, with a bit b and natural numbers s', s_0, s_1, the *control states*. A *configuration* consists of a control state, a head position in $\mathbb{Z}$ and a tape, a map from $\mathbb{Z}$ to $\{0,1\}$; the blank symbol is 0. One step executes h_s for the current state s: write(b, s') writes b in the scanned cell and goes to s'; left(s') and right(s') move the head by one cell and go to s'; branch(s_0, s_1) goes to s_0 if the scanned cell holds 0 and to s_1 if it holds 1. If h_s is halt, or if $s \ge L$, the configuration does not change. On an input word u the machine starts in state 0, with the head on cell 0, u written in cells $0, 1, \dots$ and 0 everywhere else. It *runs from u to v within t steps* if after exactly t steps the current instruction is halt and the whole tape is v written from cell 0, with 0 in every other cell.

Control states are positions in the table, never data, and there are no registers. Because a halted machine stands still, "within t steps" means that it halts at some time at most t, having erased everything it wrote except the output.

Numbers travel in binary. For a word u of bits, the *frame* of u is $1^{|u|}0u$: its length in unary, a separator and the bits. A natural number n is written with its $\mathrm{bl}(n)+1$ lowest binary digits, least significant first, and framed; an integer z is the bit $[z<0]$ followed by the digits of $|z|$, framed.

Definition 15.32 (Requests and answers)**.** Let a problem have a coefficient field of dimension d. The *request* for coordinate i at precision m, with certificate level n_0, is the concatenation of

(i) the *shape*: d; for each coordinate, the number of its A-terms and their indices j; for each coordinate, the number of its B-terms and their index pairs (j,k);

(ii) the *payload*: the number of integers that follow, then the numerator and the positive denominator of every coordinate of y_0, every coordinate of c, every A-coefficient and every B-coefficient, each as a framed integer;

(iii) the framed natural numbers M, T, n_0, i and m.

For a width W and an integer a with $|a|<2^{W+1}$, the *answer word* is the frame of the bit $[a<0]$ followed by exactly $W+1$ binary digits of $|a|$, least significant first. The width used at precision m is $W_m=4\mathscr{B}(N)+4$, with $N=N(\max(m,n_0))$ and $\mathscr{B}$ the capacity of Proposition 15.26.

The request contains the original rational data and nothing computed from them: no schedule, no Taylor order, no working precision, no width and no intermediate value. The machine must compute all of these itself. The answer word is a fixed-width sign-magnitude numeral; by Proposition 15.26 its width is at most a constant times $(m+n_0+\kappa+1)^2$, well above the roughly m significant bits of the answer.

Theorem 15.33 (One machine for every problem and every precision)**.** *There are one control table H and one natural number λ with the following property. Let a problem have a coefficient field, with rational initial point y_0, radius $M\geq 1$ and natural end time T, and let it satisfy the region certificate at level n_0. Then there is a constant K such that, for every coordinate $i<d$ and every precision m, the answer a_m at precision m of the certified real x_i of Theorem 15.22 satisfies $|a_m|<2^{W_m+1}$ (Lemma H.5), and the machine H runs from the request for coordinate i at precision m to the answer word of a_m within $K(m+1)^\lambda$ steps.*

The same H and λ serve every polynomial field. Let p be a polynomial field with rational initial point x_0, radius $M\geq 1$ and natural end time T, whose quadratized problem satisfies the region certificate at level n_0. Then there is a constant K such that, for every $i<d$ and every m, the answer at precision m of the polynomial solution x_i is less than 2^{W_m+1} in absolute value, with W_m the width of the quadratized problem, and H runs from the request of the quadratized problem for the variable w_{e_i} at precision m to the answer word of that answer within $K(m+1)^\lambda$ steps.

Conditions and scope.

Data
Rational coefficients, a rational initial point, a natural radius $M \geq 1$ and a natural end time T, all written in binary in the request.

Certificate
The region certificate at level n_0 is a hypothesis. It is a finite check of the computed trajectory at one level; the machine does not perform it, and the theorem says nothing about problems that fail it. For polynomial fields the machine reads the request of the quadratized problem; it does not quadratize.

Output
The fixed-width sign-magnitude numeral of the integer a_m. Because a_m is an answer of the certified real x_i, $|a_m/2^m - a_n/2^n| \leq 2^{-m} + 2^{-n}$ for every n, and in the order of $\mathbb{R}_\delta$ the point $\bar{x}_i$ and $\eta(a_m/2^m)$ are within 2^{-m}, as shown after Proposition 15.5.

Resources counted
Machine steps of H, including reading the request, all arithmetic, writing the answer and erasing the scratch tape. The bound $K(m+1)^\lambda$ is polynomial in the number m of requested bits, the usual measure for computing real numbers.

Constants
H and λ are fixed once for all problems. K depends on the whole problem (field, dimension, initial point, M, T and n_0) but not on i or m. The exponent λ is a numeral with 55 decimal digits (Theorem I.209): the construction composes the fixed exponents of the stages and makes no attempt to keep it small.

Not claimed
Any bound on how K depends on the size of the problem; any statement about hardware time; that the answer is a prefix of a binary expansion (it is a certified approximation at scale 2^{-m}); and any statement about the logical principles used in the construction.

Identification
The point $\bar{x}_i$ is the classical solution value $y_i(T)$ under the hypotheses of Theorem 15.29, or of Corollary 15.30 for polynomial fields.

The order of the quantifiers is the content of the theorem, in the sense of Chapter 10: there exist H and λ, for every problem with a certificate there exists K, for every i and m the bound holds. Only the constant is chosen after the problem.

Pouly and Graça [45] gave algorithms with guaranteed accuracy for polynomial differential equations, with running times polynomial in the end time, the accuracy and the length of the solution curve, and Bournez, Graça and Pouly [46] characterized polynomial-time computation by solutions of polynomial differential equations of polynomial length. In its dependence on the problem, Theorem 15.33 is weaker than their results: nothing is said about how K grows, the exponent is very large, and a finite region certificate takes the place of a bound on the solution. What it adds is one fixed one-tape binary machine with one exponent for the whole class. In the bit-complexity bound of Pouly and Graça the exponent grows with the

dimension of the system; here one exponent λ serves every problem of every dimension, and the dimension enters only the constant K. It also adds an answer that is a certified point of $\mathbb{R}_\delta$ with a proved modulus, and the exact count of the arithmetic behind it.

How the theorem is proved

The machine is $H = \mathrm{machine}(P)$ for one structured program P, and three lemmas carry the argument from the mathematics of the previous sections to steps of H. The first turns structured programs into control tables without changing the count.

Structured programs are built from skip, write b, left and right by sequencing $P;Q$, a branch $\mathrm{branch}(P_0, P_1)$ on the scanned bit, and a loop $\mathrm{loop}(P)$ that runs P while the scanned bit is 1. Their size is 1 for the four primitives, $|P;Q| = |P| + |Q|$, $|\mathrm{branch}(P_0, P_1)| = 1 + |P_0| + |P_1|$ and $|\mathrm{loop}(P)| = 1 + |P|$. A *counted execution* of P from a head position and tape to another, in t units, is defined by the rules: each primitive takes one unit (skip changes nothing); a sequence adds the units of its parts; a branch takes one unit plus those of the part selected by the scanned bit; a loop on a scanned 0 takes one unit and changes nothing, and on a scanned 1 takes one unit plus those of P and of the loop again. Every test is counted, including the final one of each loop. The compilation $\mathrm{comp}(P, s, x)$, a list of $|P|$ instructions meant for the addresses $s, \dots, s + |P| - 1$ with exit state x, is $[\mathrm{branch}(x, x)]$, $[\mathrm{write}(b, x)]$, $[\mathrm{left}(x)]$ and $[\mathrm{right}(x)]$ for the primitives; $\mathrm{comp}(P, s, s + |P|)$ followed by $\mathrm{comp}(Q, s + |P|, x)$ for $P;Q$; $\mathrm{branch}(s + 1, s + 1 + |P_0|)$ followed by $\mathrm{comp}(P_0, s + 1, x)$ and $\mathrm{comp}(P_1, s + 1 + |P_0|, x)$ for a branch; and $\mathrm{branch}(x, s + 1)$ followed by $\mathrm{comp}(P, s + 1, s)$ for a loop. Finally $\mathrm{machine}(P)$ is $\mathrm{comp}(P, 0, |P|)$ followed by halt.

Lemma 15.34 (The counted compiler)**.**

(a) If P has a counted execution in t units from head h and tape τ to head h' and tape τ', and a control table holds $\mathrm{comp}(P, s, x)$ *at the addresses $s, \dots, s + |P| - 1$, then t steps of that table from state s, head h and tape τ reach state x, head h' and tape τ'.*

(b) If P has a counted execution in t units from head 0 *and the input tape of u to a tape that is v written from cell* 0 *with* 0 *elsewhere, then* $\mathrm{machine}(P)$ *runs from u to v within t steps.*

Proof. (a) Induction on the derivation of the execution. A primitive is one instruction, executed in one step, with exit x (skip is a branch whose two targets are x). For $P;Q$ the table holds $\mathrm{comp}(P, s, s + |P|)$ at s and $\mathrm{comp}(Q, s + |P|, x)$ at $s + |P|$; the induction hypothesis for P and then for Q gives $t_P + t_Q$ steps from s to x. For a branch, the first instruction moves in one step to $s + 1$ or to $s + 1 + |P_0|$ according to the scanned bit, and the hypothesis for the selected part finishes at x. For a loop, the first instruction tests the scanned bit in one step: on 0 it goes to x with nothing changed; on 1 it goes to $s + 1$, the hypothesis for P returns the machine to s, and the hypothesis for the remaining execution of the loop finishes at x. In every case the number of steps is the number of units. (b) Apply (a) with $s = 0$ and $x = |P|$; the instruction at address $|P|$ is halt. □

The second lemma says that the trajectory needs only integers at one working precision, with exactly two roundings per row and step.

Stage	Contract, from its input tape to its output tape
Parse	Move the bits of every frame onto tracks 3 and 4 of the rows (the saved request), marking on track 15 the first row of every frame after the first, and erase the input cells as they are read, except for one mark on the last cell. At most $100000(\ell+1)^3$ steps for a request of length ℓ (Theorem I.42).
Decode	Erase that mark. Write every natural number of the request in unary on tracks 9 and 10: d, the term counts and indices, the payload length, M, T, n_0, i and m. Exponent 8 (Theorem I.58); with Parse, 11 in X (Corollary I.59).
Schedule	From the saved request compute in unary the magnitudes of the payload integers and, by row sums and maxima of the numbers $\lceil\lvert q\rvert\rceil$, the bounds γ, α, β and the capacities s, L, s_0 (Theorem I.62); then, by a fixed program of 51 unary operations, the sum $R=\gamma+\alpha+\beta M+1$, σ, S, G, κ, q_P, l_P and c_P; finally N, $\mathscr{B}(N)$ by Horner's rule and W_m. Exponent $g_q = 23040\cdot 6^{51}+12$ (Theorem I.88).
Initialize	Store the controls and reserve the rows; for each coordinate j compute $\operatorname{near}(2^N y_{0,j}) = \lfloor (2p\,2^N+q)/(2q)\rfloor$ from $y_{0,j}=p/q$ by one division and append its signed word to the binary stream. Exponent 190674 (Theorem I.102).
Coefficients	For order l and coordinate j compute the product D of the denominators of the coefficients of coordinate j and the integer U with $c_j[l=0]+(Af_l)_j+\sum_{i\le l} B(f_i,f_{l-i})_j = U/(Ds^2)$, $s=2^N$. The floor of $(2U+Ds(l+1))/(2Ds(l+1))$, computed in an area of $3W_m$ bits, is $r_{l+1,j}=\operatorname{near}(U/(Ds(l+1)))$; append its signed word to the buffer. Exponent 118 per entry, 120 per table (Theorems I.119, I.121).
Taylor sum	For each coordinate form $T=\sum_{l\le N} 2^{\sigma(N-l)} z_l$ by Horner's rule on the stored entries z_l, in a word of $W_m+\sigma N+N+4$ bits; halve T σN times, rounding down, and add the last bit dropped. This is $\operatorname{near}(\sum_{l\le N} 2^{-\sigma l} z_l)$, rounded once. Write the new vector to the binary stream. Exponent 8 in $d+\sigma+N+W_m+r+1$ (Theorem I.161); one step, 136 (Theorem I.162).
Step loop	Repeat the two previous stages, on the same rows, once for each of the S marks of the step count on track 14. Invariant: the stored vector is Z_k of Lemma 15.35, in the ball of radius M by Theorem 15.20; here the region certificate is used. Exponent 137 (Theorem I.184).
Select	Read i from the saved request; copy the signed word of coordinate i of Z_S to the front of the binary stream and end the stream there; empty the buffer. Exponent 3 (Theorem I.187); with Initialize and the loop, 105 252 600 (Theorem I.188).
Round and emit	Shift the magnitude of $z=(Z_S)_i$ right by $N-m$ places, keeping the last digit dropped and the OR of the earlier ones; add one if that digit is 1 and, for negative z, the OR is 1; add rows; write the sign. This is the signed word of $\operatorname{near}(2^m z/2^N)=a_m$ (Lemma 15.35). Exponent 7 in $m+N+W_m+r+1$ (Theorem I.198).
Frame and erase	Copy the word into the output buffer as marks, a 0 and its bits, which is the answer word; write the buffer into the cells $0,1,\dots$; erase every row found through the headers on track 0, and the root; stop on cell 0. At most $2000(r+1)^2$ steps on r rows, which doubles the exponent (Theorem I.204).

Table 15.1. The stages of the machine of Theorem 15.33 in the order in which they run, with the results of Appendix I that construct them. Unless stated otherwise a bound has the form $K(X+r+1)^g$ of Lemma 15.36: r rows are in use at the start, K depends only on the problem and n_0, and g is the stated exponent.

Lemma 15.35 (The trajectory in integers)**.** *Fix a level n and let $w = N(n)$. Define integer vectors Z_k by $Z_0 = \mathrm{near}(2^w y_0)$, coordinatewise, and let Z_{k+1} be obtained from Z_k as follows: the rows $r_0 = Z_k$ and $r_{l+1} = \mathrm{near}\big(2^w \,\mathrm{next}(f, l)\big)$ with $f_j = r_j/2^w$, for $l < N(n)$, and then $Z_{k+1} = \mathrm{near}\big(2^w S_{N(n)}(f)\big)$. Then $y_k^{(n)} = Z_k/2^w$ for every k, and*

$$a_m = \mathrm{near}\Big(2^m \,\frac{(Z_S)_i}{2^{N(\max(m,n_0))}}\Big)$$

with Z_S computed at level $\max(m, n_0)$.

Proof. For every ratio q, $[q]_w = \mathrm{near}(2^w q)/2^w$. Induction on the row index shows that $r_l/2^w$ is the coefficient $a_l^{(w)}$ of the table at $Z_k/2^w$ with rounding $[\cdot]_w$, and then $Z_{k+1}/2^w = \Psi_{w,w}(Z_k/2^w)$. Induction on k gives the first claim, and the formula for a_m is that of Theorem 15.22. □

The third lemma composes resource bounds. A stage of P starts with r rows of scratch tape in use and ends with r' rows after t units; its contract bounds $t + r' + 1$ by a constant times a power of $X + r + 1$, where $X = m + n_0 + \kappa + 1$ is the scale of the request.

Lemma 15.36 (Composing resource bounds)**.**

(a) Let $N \geq 1$ and $e \geq 1$. If $s + r + 1 \leq A\,N^e$ and $t + r' + 1 \leq K(N + r + 1)^g$, then

$$(s + t) + r' + 1 \leq \big(A + K(A + 1)^g\big) N^{e(g+1)}.$$

(b) If $c \geq 1$, then $(m + c)^\lambda \leq c^\lambda (m + 1)^\lambda$.

(c) Suppose a first stage, started with r rows in use, ends after t_1 units with r_1 rows and $t_1 + r_1 + 1 \leq K_1(X + r + 1)^{g_1}$, and a second stage, started from every tape the first can leave, ends after t_2 units with r_2 rows and $t_2 + r_2 + 1 \leq K_2(X + r_1 + 1)^{g_2}$. Then the two stages in sequence end after $t_1 + t_2$ units, and

$$t_1 + t_2 + r_2 + 1 \leq \big(K_1 + K_2(K_1 + 1)^{g_2}\big)(X + r + 1)^{(g_1+1)(g_2+1)}.$$

Proof. (a) Since $N \leq N^e$ and $r + 1 \leq AN^e$, we have $N + r + 1 \leq (A + 1)N^e$, so

$$t + r' + 1 \leq K(A + 1)^g N^{eg} \leq K(A + 1)^g N^{e(g+1)}.$$

Also $s \leq AN^e \leq AN^{e(g+1)}$. Add. (b) $m \leq mc$, so $m + c \leq c(m + 1)$; raise to the power λ. (c) Put $Y = X + r + 1 \geq 1$. Then $t_1 + r_1 + 1 \leq K_1 Y^{g_1} \leq K_1 Y^{g_1+1}$, and since $X \leq Y$, $t_2 + r_2 + 1 \leq K_2(Y + r_1 + 1)^{g_2}$. Apply (a) with $N = Y$, $s = t_1$, $e = g_1 + 1$, $A = K_1$, $r = r_1$ and $g = g_2$. □

The program P is a sequence of stages, each with a contract of this form. Table 15.1 lists them, and Appendix I constructs every stage program and proves its contract and its bound. The scratch area is the negative half of the tape, arranged in rows of sixteen tracks. The request does not stay where it was written. The first stage moves it into these rows, its bits onto two tracks and a mark at the start of every frame after the first onto a third, and erases the input cells as it reads them, except for one mark on the last cell; the second stage erases that mark,

so that the input cells stay blank until the answer is written there. Control numbers, such as the schedule and the loop counters, are kept in unary as runs of marked cells. Numerical data are kept as signed binary words of the width W_m, and the arithmetic of a step works in areas of $3W_m$ and of $W_m + \sigma N + N + 4$ bits (Proposition I.32).

Proof of Theorem 15.33. Let P be the program of Theorem I.209 in Appendix I, let H be the control table machine(P), and let

$$\lambda = 11\Big(2(g_q + 1)\big(8(762700 \cdot 138 + 2) + 1\big) + 1\Big), \qquad g_q = 23040 \cdot 6^{51} + 12,$$

be the exponent of that theorem. Both are fixed before any problem is given. Let a problem with a coefficient field satisfy the region certificate at level n_0, and let K be the constant of Theorem I.209; it depends on the problem and on n_0, not on i or m. Part (b) of that theorem is the first claim: for every coordinate $i < d$ and every precision m, $|a_m| < 2^{W_m+1}$, and H runs from the request for coordinate i at precision m to the answer word of a_m within $K(m+1)^\lambda$ steps. Its last part is the second claim, with the same H and λ, because the polynomial solution x_i is the certified real of the variable w_{e_i} of the quadratized problem (Definition 15.27).

Theorem I.209 is proved stage by stage along Table 15.1. The first two stages read the request, whose length is at most a constant depending on the problem times X, within a constant times X^{11} units (Corollary I.59). Every later stage meets a contract whose exponent does not depend on the problem, in X and the rows already in use, and Lemma 15.36(c) composes such contracts in sequence. The step loop adds one to the exponent of a step and puts the number S of steps into the constant (Lemma I.19). Its invariant is the vector Z_k of Lemma 15.35, which lies in the ball of radius M by Theorem 15.20, so every integer the machine stores fits the widths of Proposition I.32. Rounding and emitting leaves the signed word of $\mathrm{near}(2^m (Z_S)_i / 2^N)$, which is the answer a_m of Theorem 15.22 by Lemma 15.35 and fits W_m by Lemma H.5. Framing and erasing is quadratic in the rows and doubles the exponent (Lemma I.205). Lemma 15.36(a) then adds the reading of the request, and part (b) turns $X^\lambda = (m + c)^\lambda$, $c = n_0 + \kappa + 1$, into $c^\lambda (m+1)^\lambda$. Finally Lemma 15.34(b) turns the counted execution of P into a run of H with the same number of steps. □

This chapter proves the mathematics that the machine computes: the certified solution value and its modulus (Theorem 15.22), the count of the arithmetic of the ratio procedure (Theorem 15.23) and the size of its numbers (Proposition 15.26), together with the three lemmas above, which turn a structured program into the control table H, reduce the trajectory to integers and compose stage bounds. The machine does not run the ratio procedure of Section 15.6: it keeps integer numerators over the denominator 2^N and sums by Horner's rule, and Lemma 15.35 shows that it computes the same values. Appendix I constructs every stage program of Table 15.1 from a small set of tape routines, proves its contract and its step bound with a fixed exponent, bounds every integer the machine stores (Proposition I.32), and assembles the stages into the program P with its exponent λ and its constant K (Theorem I.209). With these results the proof of Theorem 15.33 is complete.

Questions to work through

1. An expression is built from certified reals by h nested sums. At what precision does Proposition 15.4 ask its leaves when the expression is requested at precision n? Compare with the request $2n + 1$ of the protocol sum of Chapter 11 after h nestings.

2. Write $y' = -y^2$, $y(0) = \frac{1}{2}$, as a coefficient field with $d = 1$, and take $T = 1$ and $M = 1$. Compute γ, α, β, R, σ, S, G and κ. What finite check does the region certificate at level 0 ask for?

3. Use Theorem 15.11 to find a lower bound for the number of Runge–Kutta steps that can give e to 40 bits. Compare with the 32 steps that the Taylor solver uses for the oscillator at every precision. Which feature of the two methods explains the difference?

4. Compute step(N) and cost(40) for the oscillator from Theorem 15.23, and compare with the bound of Proposition 15.25.

5. Quadratize $x' = x^3$. List the variables, the terms of the new field and the initial values, and check the identity behind Definition 15.27 for the variable x^3.

Chapter 16

A calculus of δ's own objects

Take the unit interval with its finite rational nets, and a sequence of rational open intervals that covers every point of it. Chapter 13 shows that finitely many of the intervals always suffice exactly when the decidable fan theorem FAN_Δ holds, with no choice principle in either direction. Can an omniscience principle pay for this? LPO decides, for each binary sequence, whether it has an entry 1, and once unique choice collects its answers for infinitely many sequences into one sequence, it gives the finite subcover (Corollary 13.10). Does LPO give the finite subcover without that collection?

To ask the question we have to say what a proof is allowed to do. The statement quantifies over compact sets of points of $\mathbb{R}_\delta$, over their nets, over covers and over points. Intuitionistic second-order arithmetic (Definition 7.14) quantifies over numbers and sets of numbers. The theories of Theorems 12.22 and 12.63 add sequences of numbers, and the proof of Theorem 12.63 writes a point as a formula with a sequence variable; a set of points or a function on points then becomes a formula, and a statement about all of them becomes a schema. What we want is a calculus whose variables range over δ's objects directly: points of the line as classes of regular sequences, functions on the line, sets of points, compact sets with their nets. In such a calculus the statement above is one sentence, and "LPO derives it" has a definite meaning once the terms of the calculus are fixed.

This chapter builds that calculus and answers the question for its computable form. Its objects live at two stages, an earlier one and a later one, as in the two-stage structures of Chapter 6. Every derivation is sound for this reading (Theorem 16.8), and every derivation of intuitionistic second-order arithmetic translates into a derivation of the calculus (Theorem 16.11). The calculus admits every map of ordinary mathematics as a term. For that reason it has no recursive model (Theorem 16.14), and it derives every true sentence built from equations over carriers that do not change from one stage to the next, LPO, $\mathrm{WLPO}_{\mathrm{seq}}$ and the compactness sentences among them (Proposition 16.15); there "LPO derives it" says only that the statement is classically true. Restricting the terms to maps computed by programs gives the *computable calculus*, which keeps every rule and has a model in which the objects are those with computable names. In the computable calculus we prove two kinds of separation, and both hold even when every instance of excluded middle is added as a base rule. LPO does not derive the sequential form $\mathrm{WLPO}_{\mathrm{seq}}$ of WLPO (Theorem 16.27). And none of the decidable fan theorem, its c-set form $\mathrm{FAN}_{\mathrm{c}}$, decidable weak König's lemma, Heine–Borel for presented compact sets, the positive lower bound, the positive infimum and presented Heine–Cantor is derivable from LPO (Theorems 16.32 to 16.36). These statements are written with equations between computed terms: an atom of an order or of a membership relation is governed by no rule (Remark 16.19), so a statement resting on one could fail to be derivable for that reason alone.

Every one of these failures has the same source. In the model, LPO is true and each question has its answer, but the answers are not collected into a computable function. With LPO the computable calculus decides each question; it cannot collect the answers. That is the dual of

Theorem 12.22, which shows that collecting decides nothing.

As in Chapters 7 and 12, the arguments about a model are carried out in ordinary classical mathematics: they say which statements are true in a model, and they ask nothing of the calculus under study. Where a proof below uses classical reasoning in that sense, we say so at the step.

16.1 Objects that exist at two stages

Chapter 6 read arithmetic at two stages: a structure M_0 at an earlier stage, a structure M_1 at a later stage, and a passage h from the first to the second. An equation or an existential statement was read at the earlier stage alone, while an implication or a universal statement had to hold at the later stage too. The calculus of this chapter uses the same reading for every kind of object, not only for numbers.

Definition 16.1 (Two-stage carriers and their maps)**.** A *carrier* X consists of a set X_0 of objects at the earlier stage, a set X_1 of objects at the later stage, and a map $h_X : X_0 \to X_1$, the *passage*, which says what each earlier object becomes later. A *map* $f : X \to Y$ of carriers is a pair of maps $f_0 : X_0 \to Y_0$ and $f_1 : X_1 \to Y_1$ that commutes with the passage:

$$f_1(h_X(x)) = h_Y(f_0(x)) \qquad \text{for every } x \in X_0.$$

Two maps are equal when both components are. Identities and composition are taken componentwise. A *predicate* P on X is a pair of subsets $P_0 \subseteq X_0$ and $P_1 \subseteq X_1$ that *persists*: $x \in P_0$ implies $h_X(x) \in P_1$.

We write h for every passage when the carrier is clear. The passage need not be injective or onto: two earlier objects may become one, and new objects may appear later. The following carriers are used throughout.

- The *constant carrier* of a set S has S at both stages and the identity as passage. We use it for the natural numbers $\mathbb{N}$, for the set $\mathbb{Q}$ of ratio records of Chapter 4, for finite binary strings and bits, and for the one-point set $\mathbf{1}$, whose carrier is the context of a closed sentence.
- The *product* $X \times Y$ has $X_0 \times Y_0$ and $X_1 \times Y_1$, with the passage $(x, y) \mapsto (h(x), h(y))$. The projections and the pairing $\langle f, g \rangle$ of two maps are maps of carriers.
- The *exponential* Y^X has, at the earlier stage, all maps of carriers $X \to Y$; at the later stage, all functions $X_1 \to Y_1$; and the passage sends a map (f_0, f_1) to f_1. *Evaluation* $\mathrm{ev} : Y^X \times X \to Y$ is $\mathrm{ev}_0((f_0, f_1), x) = f_0(x)$ and $\mathrm{ev}_1(f, x) = f(x)$. A map $k : Z \times X \to Y$ has a *curried* form $\lambda k : Z \to Y^X$: at the earlier stage z goes to the map $\big(x \mapsto k_0(z, x),\ x \mapsto k_1(h(z), x)\big)$, and at the later stage to $x \mapsto k_1(z, x)$.
- The *truth values* Ω have, at the earlier stage, the pairs (p, q) of propositions with p implying q ("true now" implies "true later"); at the later stage, all propositions; and the passage $(p, q) \mapsto q$.

- A predicate P on X cuts out the *subcarrier* with P_0 and P_1 and the restricted passage. Equivalence relations $\sim_0$ on X_0 and $\sim_1$ on X_1 with $x \sim_0 y$ implying $h(x) \sim_1 h(y)$ give the *quotient* carrier $(X_0/\sim_0, X_1/\sim_1)$ with the induced passage, and the quotient map $X \to X/\sim$.

For constant carriers S and T, a map $(f_0, f_1) : S \to T$ has $f_1 = f_0$, because the passages are identities; so both stages of T^S consist of the functions from S to T, and the exponential is again constant. Predicates on X and maps $X \to \Omega$ correspond: P gives its *characteristic map*, which sends $x \in X_0$ to the pair $(x \in P_0,\ h(x) \in P_1)$ and $x \in X_1$ to the proposition $x \in P_1$; persistence is exactly the condition that the pair lies in Ω_0. So the earlier-stage objects of Ω^X, the *classifiers* on X, are the predicates on X, and the later-stage objects are all properties of later objects. A classifier can grow: an object outside it at the earlier stage can be inside it later.

Definition 16.2 (The carriers of δ's objects)**.** The *displays* are the exponential $\mathbb{Q}^{\mathbb{N}}$ of constant carriers, so both stages consist of the sequences of ratio records. The *regular carrier* $\mathscr{R}$ is the subcarrier of the regular sequences of Definition 12.14, at both stages. The *line* $\mathscr{L}$ is the quotient of $\mathscr{R}$ by the relation $\approx$ of naming the same point, at both stages. The carriers of *functions on the line* and of *sets of points* are the exponentials $\mathscr{L}^{\mathscr{L}}$ and $\Omega^{\mathscr{L}}$.

Regularity is one condition on a sequence, so it persists. Both stages of $\mathscr{L}$ are $\mathbb{R}_\delta$ itself, read through representatives at each stage, with no choice of a representative for a family of points. The function carrier contains every function on $\mathbb{R}_\delta$ at the later stage, and the subset carrier every set of points; at the earlier stage a set of points is a pair of sets with the earlier one contained in the later one.

The structure just described is the Kripke semantics of intuitionistic logic on the frame with two worlds, an earlier one below a later one, with domains that may change from one world to the next [53, 4]. In the language of categories it is the category of presheaves on the two-element chain; products, exponentials and Ω are its products, exponentials and object of truth values. We use none of that theory. Every fact needed below is checked directly from the definitions.

16.2 Terms, formulas and their reading

A formula of the calculus lives in a *context*, a carrier G whose points are the values of its free variables. A closed sentence has the one-point context $\mathbf{1}$. A context $G \times X$ has one more variable, of type X, than G.

Definition 16.3 (Terms)**.** For carriers G and X, the *terms* $t : G \vdash X$ are generated by four constructions:

- $\mathrm{map}(f)$, for every map of carriers $f : G \to X$;
- the pair $\langle t, u \rangle : G \vdash X \times Y$ of $t : G \vdash X$ and $u : G \vdash Y$;
- the application $t\,u : G \vdash Y$ of $t : G \vdash Y^X$ to $u : G \vdash X$;

- the abstraction $\lambda t \,:\, G \vdash Y^X$ of $t \,:\, G \times X \vdash Y$.

Each term is read as a map $[\![t]\!] \,:\, G \to X$:

$$[\![\mathrm{map}(f)]\!] = f, \quad [\![\langle t, u\rangle]\!] = \langle [\![t]\!], [\![u]\!]\rangle, \quad [\![t\,u]\!] = \mathrm{ev} \circ \langle [\![t]\!], [\![u]\!]\rangle, \quad [\![\lambda t]\!] = \lambda [\![t]\!].$$

The *graph* of t is the map $\mathrm{gr}(t) = \langle \mathrm{id}, [\![t]\!]\rangle \,:\, G \to G \times X$.

Variables are not a separate kind of term. The newest variable of the context $G \times X$ is the term $\mathrm{map}(\pi_2)$, and an older variable is reached through the first projection $\pi_1 \,:\, G \times X \to G$. Substituting a term t for the newest variable is reindexing along $\mathrm{gr}(t)$.

Definition 16.4 (Formulas)**.** The *formulas* in context G are generated by ten constructions:

- the atom $R(t)$, for every predicate R on a carrier X and every term $t \,:\, G \vdash X$;
- $\top$ and $\bot$;
- the equation $t = u$, for terms $t, u \,:\, G \vdash X$;
- $A \wedge B$, $A \vee B$ and $A \to B$;
- the quantifiers $\forall_X A$ and $\exists_X A$, for a carrier X and a formula A in context $G \times X$;
- the *reindexing* $f^* A$ in context D, for a map $f \,:\, D \to G$ and a formula A in context G.

We write $\neg A$ for $A \to \bot$.

The quantifiers range over any carrier: numbers, sequences, points of the line, functions on it, its subsets, truth values. Each formula is read as a predicate on its context. Write $g \Vdash_0 A$ when the earlier object $g \in G_0$ belongs to the reading of A, and $g \Vdash_1 A$ when the later object $g \in G_1$ does. The clauses are those of Table 16.1. For a pair (g, x) in a product context we write $(g, x) \Vdash_i A$.

Lemma 16.5 (Readings persist)**.** *For every formula A in context G and every $g \in G_0$: if $g \Vdash_0 A$, then $h(g) \Vdash_1 A$. So the reading of every formula is a predicate on its context.*

Proof. By induction on A, for all contexts at once. For an atom, $[\![t]\!]_0(g) \in R_0$ gives $h([\![t]\!]_0(g)) \in R_1$ by persistence of R, and this point is $[\![t]\!]_1(h(g))$ because $[\![t]\!]$ commutes with the passage. For an equation apply h to both sides and use the same commutation. The cases $\top$, $\bot$, $\wedge$ and $\vee$ follow from the induction hypotheses. For $A \to B$ and $\forall_X A$ the second half of the earlier clause is the later clause at $h(g)$. For $\exists_X A$, a witness $x \in X_0$ with $(g, x) \Vdash_0 A$ gives $(h(g), h(x)) \Vdash_1 A$ by the induction hypothesis, so $h(x)$ is a later witness. For $f^* A$, $f_0(g) \Vdash_0 A$ gives $h(f_0(g)) \Vdash_1 A$, and $h(f_0(g)) = f_1(h(g))$. □

For formulas P and Q in context G we say that P *entails* Q, written $P \vDash Q$, when every $g \in G_0$ with $g \Vdash_0 P$ has $g \Vdash_0 Q$ and every $g \in G_1$ with $g \Vdash_1 P$ has $g \Vdash_1 Q$. A *two-stage truth value* is a pair of propositions A_0 and A_1 with A_0 implying A_1, that is an element of Ω_0; it is the atom $\underline{A}$ of the closed context whose reading is A_0 at the earlier stage and A_1 at the later one.

Formula	earlier stage, $g \in G_0$	later stage, $g \in G_1$
$R(t)$	$[\![t]\!]_0(g) \in R_0$	$[\![t]\!]_1(g) \in R_1$
$\top$, $\bot$	always, never	always, never
$t = u$	$[\![t]\!]_0(g) = [\![u]\!]_0(g)$	$[\![t]\!]_1(g) = [\![u]\!]_1(g)$
$A \wedge B$, $A \vee B$	both, at least one of $g \Vdash_0 A$, $g \Vdash_0 B$	both, at least one at stage 1
$A \to B$	$g \Vdash_0 A$ implies $g \Vdash_0 B$, and $h(g) \Vdash_1 A$ implies $h(g) \Vdash_1 B$	$g \Vdash_1 A$ implies $g \Vdash_1 B$
$\forall_X A$	$(g, x) \Vdash_0 A$ for every $x \in X_0$, and $(h(g), x) \Vdash_1 A$ for every $x \in X_1$	$(g, x) \Vdash_1 A$ for every $x \in X_1$
$\exists_X A$	$(g, x) \Vdash_0 A$ for some $x \in X_0$	$(g, x) \Vdash_1 A$ for some $x \in X_1$
$f^* A$	$f_0(g) \Vdash_0 A$	$f_1(g) \Vdash_1 A$

Table 16.1. The two-stage reading of formulas. An implication and a universal statement at the earlier stage also look ahead to the later stage; the other constructions are read at one stage.

An example. Let $q : \mathscr{R} \to \mathscr{L}$ be the quotient map. The sentence

$$\forall_{\mathscr{L}} \exists_{\mathscr{R}} \big(\mathrm{map}(q \circ \pi_2) = \mathrm{map}(\pi_2 \circ \pi_1) \big)$$

says that every point of the line is the class of a regular sequence. At the later stage it says this of the points of $\mathbb{R}_\delta$. At the earlier stage it says the same of the earlier points and also, through the second half of the universal clause, of the later ones. Its existential is read at one stage at a time: one representative for one point, never a choice of representatives for a family. The sentence is derived from rule 9 of Table 16.2 by one universal introduction.

16.3 Derivations and the twenty-nine base rules

A judgment of the calculus has one hypothesis formula: $P \vdash Q$, with P and Q in the same context. A list of hypotheses is their conjunction. The judgment $\top \vdash Q$ in the one-point context derives the sentence Q.

Definition 16.6 (Derivations)**.** Given a family of *base rules*, each a judgment $P \vdash Q$, a *derivation* is a finite tree built by the following twenty constructions, in which $\pi = \pi_1 : G \times X \to G$ and A, B, C, P, Q are formulas of the appropriate contexts. A *derivation in the calculus* is a derivation from the base rules of Table 16.2.

(1) a base rule $P \vdash Q$;

(2) $P \vdash P$;

(3) $P \vdash \top$;

(4) from $P \vdash \bot$ infer $P \vdash Q$;

(5) from $P \vdash Q$ and $Q \vdash C$ infer $P \vdash C$;

(6) from $P \vdash A$ and $P \vdash B$ infer $P \vdash A \wedge B$;

(7–8) from $P \vdash A \wedge B$ infer $P \vdash A$, and infer $P \vdash B$;

(9–10) from $P \vdash A$ infer $P \vdash A \vee B$; from $P \vdash B$ infer $P \vdash A \vee B$;

(11) from $P \vdash A \vee B$, $P \wedge A \vdash C$ and $P \wedge B \vdash C$ infer $P \vdash C$;

(12) from $P \wedge A \vdash B$ infer $P \vdash A \to B$;

(13) from $P \vdash A \to B$ and $P \vdash A$ infer $P \vdash B$;

(14) from $\pi^* P \vdash A$, with A in context $G \times X$, infer $P \vdash \forall_X A$;

(15) from $P \vdash \forall_X A$ infer $P \vdash \mathrm{gr}(t)^* A$, for a term $t : G \vdash X$;

(16) from $P \vdash \mathrm{gr}(t)^* A$ infer $P \vdash \exists_X A$, for a term $t : G \vdash X$;

(17) from $P \vdash \exists_X A$ and $\pi^* P \wedge A \vdash \pi^* C$ infer $P \vdash C$;

(18) from $P \vdash Q$ infer $f^* P \vdash f^* Q$, for a map $f : D \to G$;

(19) $P \vdash t = t$;

(20) from $P \vdash t = u$ and $P \vdash \mathrm{gr}(t)^* A$ infer $P \vdash \mathrm{gr}(u)^* A$, for A in context $G \times X$.

No construction admits a formula because it is true, and none carries a proof of its own soundness. The base rules need a few abbreviations. Over any context G, 0 is the term of the constant map with value 0, and $S(t)$, $t + u$ and $t \cdot u$ are the terms of the maps that compute the successor, the sum and the product of the values of t and u at each stage. The map $\mathrm{succ} : G \times \mathbb{N} \to G \times \mathbb{N}$ sends (g, n) to $(g, n + 1)$. For a formula A in context $G \times \mathbb{N}$,

$$\mathrm{Ind}(A) := \mathrm{gr}(0)^* A \wedge \forall_{\mathbb{N}}\big(A \to \mathrm{succ}^* A\big).$$

The membership predicate $\in$ on Ω is $\{(p, q) : p\}$ at the earlier stage and $\{q : q\}$ at the later stage. In the context $G \times \Omega^X \times X$, let c be the classifier variable and x the newest variable, let Mem be the atom $\in(c\,x)$, and let $\mathrm{fg} : G \times \Omega^X \times X \to G \times X$ forget the classifier. For a formula A in context $G \times X$,

$$\mathrm{Comp}(A) := \exists_{\Omega^X} \forall_X \big((\mathrm{Mem} \to \mathrm{fg}^* A) \wedge (\mathrm{fg}^* A \to \mathrm{Mem})\big).$$

For $s : G \vdash \mathscr{R}$ write $[s]$ for the term of $q \circ [\![s]\!]$, and $s \approx t$ for the atom of the predicate "names the same point" at the pair $\langle s, t \rangle$. For terms $r : G \vdash \mathscr{L}$ and $s, t : G \vdash \mathscr{R}$ let

$$\mathrm{Surj}(r) := \exists_{\mathscr{R}}\big([\mathrm{map}(\pi_2)] = \mathrm{map}([\![r]\!] \circ \pi_1)\big),$$
$$\mathrm{Eff}(s, t) := ([s] = [t] \to s \approx t) \wedge (s \approx t \to [s] = [t]).$$

Finally, a map $f : D \to G$ *lifts* to $f \times \mathrm{id} : D \times X \to G \times X$.

The eighteen reindexing rules take the place of a syntactic substitution lemma. Reindexing of $\top$ and of $\bot$ needs no rule of its own. For a map $f : D \to G$, $f^* \top \vdash \top$ is construction (3). For the converse, derive $\top \vdash 0 = 0$ in context D by (19), pass to $f^*(0 = 0)$ by rule 19, and cut with the reindexing of $0 = 0 \vdash \top$. In the same way $f^* \bot$ gives $f^*(S(0) = 0)$, rule 18 turns it into the equation $S(0) = 0$ in context D, and rule 1 with (13) gives $\bot$; the converse is (4). Here and below we use that $\mathrm{map}([\![t]\!] \circ f)$ is the same term as the corresponding term over D whenever the two maps are equal, as they are for 0 and $S(0)$.

	Rule	Judgment (P any formula of the context)
1	successor	$P \vdash S(t) = 0 \to \bot$
2	injectivity	$P \vdash S(t) = S(u) \to t = u$
3–4	addition	$P \vdash t + 0 = t$; $\quad P \vdash t + S(u) = S(t + u)$
5–6	multiplication	$P \vdash t \cdot 0 = 0$; $\quad P \vdash t \cdot S(u) = t \cdot u + t$
7	induction	$\mathrm{Ind}(A) \vdash \forall_{\mathbb{N}} A$, for every formula A in every context $G \times \mathbb{N}$
8	comprehension	$P \vdash \mathrm{Comp}(A)$, for every carrier X and formula A in context $G \times X$
9	representatives	$P \vdash \mathrm{Surj}(r)$, for $r : G \vdash \mathscr{L}$
10	effectiveness	$P \vdash \mathrm{Eff}(s, t)$, for $s, t : G \vdash \mathscr{R}$
11	computation	$P \vdash (\lambda t)\, u = \mathrm{map}(\llbracket t \rrbracket \circ \mathrm{gr}(u))$, for $t : G \times X \vdash Y$ and $u : G \vdash X$
12–13	identity	$\mathrm{id}^* A \vdash A$; $\quad A \vdash \mathrm{id}^* A$
14–15	composition	$g^* f^* A \vdash (f \circ g)^* A$, and back
16–17	atoms	$f^* R(t) \vdash R(\mathrm{map}(\llbracket t \rrbracket \circ f))$, and back
18–19	equations	$f^*(t = u) \vdash \mathrm{map}(\llbracket t \rrbracket \circ f) = \mathrm{map}(\llbracket u \rrbracket \circ f)$, and back
20–25	connectives	$f^*(A \circ B) \vdash f^* A \circ f^* B$, and back, for $\circ$ each of $\wedge$, $\vee$, $\to$
26–29	quantifiers	$f^* \forall_X A \vdash \forall_X (f \times \mathrm{id})^* A$, and back; the same for $\exists_X$

Table 16.2. The base rules of the calculus. Rules 1–7 are the arithmetic rules and induction of Chapter 5, in every context and for formulas of every type. Rule 8 is comprehension at every type. Rules 9 and 10 say that the line is the quotient of the regular sequences. Rule 11 computes an application of an abstraction. Rules 12–29 say that reindexing commutes with every construction of formulas.

Remark 16.7 (What a term may be)**.** A term $\mathrm{map}(f)$ may carry *any* map of carriers, and an atom $R(t)$ any predicate. The calculus is therefore an interpreted language: its symbols are the objects of ordinary mathematics themselves, and two terms $\mathrm{map}(f)$ and $\mathrm{map}(g)$ are the same term exactly when $f = g$. A derivation is still a finite tree, and every step is one of the twenty constructions or one of the twenty-nine rules; no formula is admitted because it is true. But terms are identified by their maps, so every true identity of maps is derivable by reflexivity. If $\llbracket t \rrbracket = \llbracket u \rrbracket$, then $\mathrm{map}(\llbracket t \rrbracket \circ \mathrm{id})$ and $\mathrm{map}(\llbracket u \rrbracket \circ \mathrm{id})$ are the same term, so construction (19) derives their equation from any hypothesis P; rule 19 at the identity map turns it into $P \vdash \mathrm{id}^*(t = u)$, and rule 12 gives $P \vdash t = u$. What is *not* the case is that the derivable sentences form a list produced by a program. Three consequences shape the rest of the chapter. An identity between maps proved in ordinary mathematics makes two terms literally equal, which the translation of Section 16.5 uses at its atoms. A term may carry a map that no program computes, which is why the calculus has no recursive model (Section 16.6). And since a term may carry any map, every sentence built from equations over constant carriers that is true in ordinary mathematics is derivable (Proposition 16.15).

16.4 Every derivation holds at both stages

Theorem 16.8 (Soundness of the calculus)**.** *If the calculus derives $P \vdash Q$, then $P \vDash Q$: every earlier object of the context that satisfies P at the earlier stage satisfies Q there, and every later object that satisfies P at the later stage satisfies Q there. More generally, every derivation from a*

family of base rules each of which is an entailment is an entailment. The proof uses no omniscience or choice principle.

Proof. The proof has two steps. Step 1 shows that each of the twenty constructions turns entailments into an entailment, whatever the base rules, provided each base rule is an entailment. Step 2 shows that each of the twenty-nine base rules is an entailment. Induction on the derivation then gives the theorem. Throughout, $\pi = \pi_1 : G \times X \to G$, i is a stage, and we assume $g \Vdash_i P$.

Step 1. Hypothesis, truth, cut, the conjunction rules, the disjunction introductions and reflexivity are immediate from the clauses of Table 16.1. Falsehood is never satisfied, so construction (4) holds vacuously. For disjunction elimination, $g \Vdash_i A \vee B$; if $g \Vdash_i A$, then $g \Vdash_i P \wedge A$ and the second premise gives $g \Vdash_i C$, and the case of B is the same.

Implication introduction, from $P \wedge A \vdash B$. At the later stage, if $g \Vdash_1 A$, then $g \Vdash_1 P \wedge A$, so $g \Vdash_1 B$. At the earlier stage the clause has two halves. If $g \Vdash_0 A$, then $g \Vdash_0 P \wedge A$ and $g \Vdash_0 B$. If $h(g) \Vdash_1 A$, then $h(g) \Vdash_1 P$ by Lemma 16.5, so $h(g) \Vdash_1 P \wedge A$ and $h(g) \Vdash_1 B$. *Implication elimination* uses the first half of the clause at the earlier stage, and the clause itself at the later stage.

Universal introduction, from $\pi^* P \vdash A$. Since $\pi_i(g, x) = g$, every (g, x) satisfies $\pi^* P$ at stage i. At the later stage this gives $(g, x) \Vdash_1 A$ for every $x \in X_1$. At the earlier stage it gives $(g, x) \Vdash_0 A$ for every $x \in X_0$, and persistence gives $h(g) \Vdash_1 P$, hence $(h(g), x) \Vdash_1 A$ for every $x \in X_1$: both halves of the clause. *Universal elimination* at a term t: the clause at the object $x = [\![t]\!]_i(g)$, which lies in X_i, says $(g, [\![t]\!]_i(g)) \Vdash_i A$, and that is $g \Vdash_i \mathrm{gr}(t)^* A$. *Existential introduction* at t uses the same object as witness. *Existential elimination*: the first premise gives $x \in X_i$ with $(g, x) \Vdash_i A$; then $(g, x) \Vdash_i \pi^* P \wedge A$, so $(g, x) \Vdash_i \pi^* C$, which says $g \Vdash_i C$.

Reindexing along f: $g \Vdash_i f^* P$ means $f_i(g) \Vdash_i P$, so $f_i(g) \Vdash_i Q$, which is $g \Vdash_i f^* Q$. *Substitution of equals*: the first premise says $[\![t]\!]_i(g) = [\![u]\!]_i(g)$, and replacing one by the other in $(g, [\![t]\!]_i(g)) \Vdash_i A$ gives $(g, [\![u]\!]_i(g)) \Vdash_i A$.

In Step 1, persistence is used at implication introduction and universal introduction, and nowhere else.

Step 2. Fix g and a stage. In rules 1–6 and 8–11 the hypothesis P plays no role. In rule 7 the hypothesis is $\mathrm{Ind}(A)$, and in rules 12–29 it is one side of an identity of readings.

Rules 1–6. Their conclusions are equations, or implications between equations, of natural numbers that hold at every point: $n + 1 \neq 0$; $m + 1 = n + 1$ gives $m = n$; $m + 0 = m$; $m + (n + 1) = (m + n) + 1$; $m \cdot 0 = 0$; $m \cdot (n + 1) = m \cdot n + m$. At the earlier stage an implication has two halves, and both are the same arithmetic fact.

Rule 7, induction. Let $g \Vdash_0 \mathrm{Ind}(A)$. Then $(g, 0) \Vdash_0 A$. The earlier half of the universal clause and the first half of the implication clause give, for every n, that $(g, n) \Vdash_0 A$ implies $(g, n + 1) \Vdash_0 A$; ordinary induction on n gives $(g, n) \Vdash_0 A$ for every n. For the later half of the conclusion, $(h(g), 0) \Vdash_1 A$ by persistence (the passage of $\mathbb{N}$ is the identity), and the later half of the universal clause gives the step $(h(g), n) \Vdash_1 A \Rightarrow (h(g), n + 1) \Vdash_1 A$; induction again gives every n. At the later stage the rule is ordinary induction.

Rule 8, comprehension. Let A be in context $G \times X$, and let $\chi : G \times X \to \Omega$ be the characteristic

map of the reading of A. At the earlier stage take the classifier $c = (\lambda\chi)_0(g)$. For $x \in X_0$ it gives the pair $\big((g,x) \Vdash_0 A,\ (h(g),h(x)) \Vdash_1 A\big)$, so the atom Mem holds at (g,c,x) exactly when $(g,x) \Vdash_0 A$. After the passage it is the property $x \mapsto \big((h(g),x) \Vdash_1 A\big)$ of later objects, so Mem holds at $(h(g),h(c),x)$ exactly when $(h(g),x) \Vdash_1 A$. Both implications in $\mathrm{Comp}(A)$ are therefore identities, in both halves of every earlier clause. At the later stage take the property $x \mapsto \big((g,x) \Vdash_1 A\big)$.

Rule 9, representatives. At each stage every element of a quotient is the class of some element, so $\mathrm{Surj}(r)$ holds: it is one existential statement about one point.

Rule 10, effectiveness. At each stage two regular sequences have the same class exactly when they name the same point, since the quotient is taken by that equivalence relation.

Rule 11, computation. By the definitions of evaluation and currying, $\mathrm{ev} \circ \langle \lambda[\![t]\!], [\![u]\!] \rangle$ and $[\![t]\!] \circ \langle \mathrm{id}, [\![u]\!] \rangle$ are the same map at both stages, so the equation holds everywhere.

Rules 12–29. Each side is read as the same predicate, by unfolding the clauses of Table 16.1. The only cases that need a remark are $f^*(A \to B)$ and $f^*\forall_X A$, whose earlier clauses mention $h(f_0(g))$ on one side and $f_1(h(g))$ on the other. These are equal because f commutes with the passage.

No step decides a proposition or chooses an object for a family. □

Corollary 16.9 (Derived sentences hold at both stages)**.** *If the calculus derives $\top \vdash \underline{A}$ for a two-stage truth value $A = (A_0, A_1)$, then A_0 and A_1 both hold.*

Proof. The one point of $\mathbf{1}$ satisfies $\top$ at both stages, so it satisfies $\underline{A}$ at both stages by Theorem 16.8. □

16.5 Intuitionistic second-order arithmetic inside the calculus

Intuitionistic second-order arithmetic, HAS, is a calculus about numbers and sets of numbers (Definition 7.14). Its sets become classifiers on $\mathbb{N}$, which may grow from one stage to the next. We write HAS with positions for variables, as Chapter 5 does for its language: $x_0, x_1, \ldots$ are the number variables and $X_0, X_1, \ldots$ the set variables, each counted from the innermost binder of its own kind, and the eigenvariable conditions of Definition 7.14 become liftings.

Definition 16.10 (The translation of HAS)**.** Let $\mathbb{S} = \Omega^{\mathbb{N}}$ be the carrier of sets of numbers and $E = \mathbb{N}^{\mathbb{N}} \times \mathbb{S}^{\mathbb{N}}$ the carrier of *environments*. An environment is a number valuation ρ, giving the value $\rho(k)$ of x_k, and a set valuation σ, giving the classifier $\sigma(j)$ of X_j. A number term t of HAS becomes the term $\ulcorner t \urcorner : E \vdash \mathbb{N}$ of the map that evaluates t at ρ, at each stage; X_j becomes the term $\ulcorner X_j \urcorner : E \vdash \mathbb{S}$ of $(\rho,\sigma) \mapsto \sigma(j)$. Let $\mathrm{push}_{\mathbb{N}} : E \times \mathbb{N} \to E$ put a new number at position

0 of ρ and move the others up, and $\mathrm{push}_{\mathbb{S}} : E \times \mathbb{S} \to E$ do the same for σ. The translation A°, a formula in context E, is

$$(t = u)^\circ = \big(\ulcorner t \urcorner = \ulcorner u \urcorner\big), \qquad (t \in X_j)^\circ = \in\big(\ulcorner X_j \urcorner \ulcorner t \urcorner\big), \qquad \bot^\circ = \bot,$$
$$(A \wedge B)^\circ = A^\circ \wedge B^\circ, \quad (A \vee B)^\circ = A^\circ \vee B^\circ, \quad (A \to B)^\circ = A^\circ \to B^\circ,$$
$$(\forall x\, A)^\circ = \forall_{\mathbb{N}}\, \mathrm{push}^*_{\mathbb{N}} A^\circ, \quad (\exists x\, A)^\circ = \exists_{\mathbb{N}}\, \mathrm{push}^*_{\mathbb{N}} A^\circ,$$
$$(\forall X\, A)^\circ = \forall_{\mathbb{S}}\, \mathrm{push}^*_{\mathbb{S}} A^\circ, \quad (\exists X\, A)^\circ = \exists_{\mathbb{S}}\, \mathrm{push}^*_{\mathbb{S}} A^\circ.$$

A list of hypotheses becomes the conjunction of its translations: $[\,]^\circ = \top$ and $(A, \Gamma)^\circ = A^\circ \wedge \Gamma^\circ$.

The translation is a construction on syntax. It asks nothing about the truth of any formula, and it contains no negative translation: HAS is already intuitionistic.

Theorem 16.11 (HAS compiles into the calculus)**.** *For every list Γ of formulas and every formula A of HAS, if HAS derives $\Gamma \vdash A$, then the calculus derives $\Gamma^\circ \vdash A^\circ$.*

The derivation is built from the HAS derivation, node by node. It uses no soundness and no truth: the only facts of ordinary mathematics it uses are equalities of maps, the substitution lemma for number terms, which make two terms of the calculus the same term. Two tools carry the bookkeeping.

Write $A \equiv B$ when the calculus derives both $A \vdash B$ and $B \vdash A$. This relation is an equivalence, and it is a congruence for $\wedge$, $\vee$, $\to$, the quantifiers and reindexing, by the introduction and elimination constructions (Appendix J.1). Rules 12–29, together with the derived rules for $\top$ and $\bot$, push a reindexing through every construction up to $\equiv$, and $g^* f^* A \equiv (f \circ g)^* A$.

The second tool says that reindexing a translation along two maps of environments that agree on what a formula reads gives equivalent formulas. Call the following *operations* on HAS formulas: substituting a number term s at number position k; lifting the number positions by d at cutoff c (Definition 5.2); and renaming the set positions by a map r of indices, which under a set binder fixes 0 and moves $j + 1$ to $r(j) + 1$. Each has a counterpart on environments: insert the value of s at position k of ρ; delete the d positions starting at c; replace σ by $j \mapsto \sigma(r(j))$. The substitution lemma for number terms, proved by induction on the term, says that evaluating $O(t)$ at ρ is evaluating t at the counterpart of ρ, for each operation O.

Lemma 16.12 (Coherence of reindexing and syntax)**.** *Let O be one of these operations and let $f, g : D \to E$ be maps of carriers such that, at both stages and at every object, the environment given by g is the counterpart of O applied to the environment given by f. Then $f^*(OA)^\circ \equiv g^* A^\circ$ for every formula A of HAS.*

Proof. By induction on A, for all operations O and all D, f and g at once, since under a binder the operation changes. For an equation, rule 18 turns $f^*(\ulcorner Ot \urcorner = \ulcorner Ou \urcorner)$ into the equation of the terms $\mathrm{map}(\ulcorner Ot \urcorner \circ f)$ and $\mathrm{map}(\ulcorner Ou \urcorner \circ f)$. By the substitution lemma, $\ulcorner Ot \urcorner \circ f$ and $\ulcorner t \urcorner \circ g$ are the same map, so these are literally the terms $\mathrm{map}(\ulcorner t \urcorner \circ g)$ and $\mathrm{map}(\ulcorner u \urcorner \circ g)$, and rule 19 gives $g^*(\ulcorner t \urcorner = \ulcorner u \urcorner)$. Both directions are these two rules in the other order. A membership is the same with rules 16 and 17, the set position read through the renaming. For $\bot$ use the derived rules for $\bot$; for $\wedge$, $\vee$ and $\to$ push f^* and g^* inside by rules 20–25 and apply the induction hypotheses with the congruence.

For a number quantifier, rules 26–29 and the composition rules give $f^*(\forall_{\mathbb{N}}\mathrm{push}_{\mathbb{N}}^*(O'A)^\circ) \equiv \forall_{\mathbb{N}}(\mathrm{push}_{\mathbb{N}} \circ (f \times \mathrm{id}))^*(O'A)^\circ$, where O' is O moved under the binder: substitute at $k+1$ the lifted s, or lift at cutoff $c+1$; a renaming of set positions is unchanged. The maps $\mathrm{push}_{\mathbb{N}} \circ (f \times \mathrm{id})$ and $\mathrm{push}_{\mathbb{N}} \circ (g \times \mathrm{id})$ satisfy the hypothesis for O': the new number sits at position 0 on both sides and the old positions move up by one on both sides. The induction hypothesis and the congruence give $\equiv \forall_{\mathbb{N}}(\mathrm{push}_{\mathbb{N}} \circ (g \times \mathrm{id}))^* A^\circ$, which is $\equiv g^*(\forall x\, A)^\circ$ by the same rules backwards. A set quantifier is the same with $\mathrm{push}_{\mathbb{S}}$: a number operation passes the binder unchanged, and a renaming passes it in the lifted form described above. The existential quantifiers are identical with rules 28 and 29. □

Four instances are used. With $f = \mathrm{id}$ and $g = \mathrm{push}_{\mathbb{N}} \circ \mathrm{gr}(\ulcorner t \urcorner)$, substitution at 0 gives

$$A(t)^\circ \equiv \mathrm{gr}(\ulcorner t \urcorner)^* \mathrm{push}_{\mathbb{N}}^* A^\circ, \tag{16.1}$$

after rules 12–15. With $f = \mathrm{push}_{\mathbb{N}}$ and $g = \pi_1$, the lifting by one at cutoff 0 gives $\mathrm{push}_{\mathbb{N}}^*(A\uparrow)^\circ \equiv \pi_1^* A^\circ$, and conjunct by conjunct the same for a list Γ. With $f = \mathrm{push}_{\mathbb{S}}$ and $g = \pi_1$, the renaming $j \mapsto j+1$ gives $\mathrm{push}_{\mathbb{S}}^*(A^+)^\circ \equiv \pi_1^* A^\circ$, where A^+ moves every free set position up by one. And the renaming that sends 0 to j and $k+1$ to k, with $f = \mathrm{id}$ and $g = \mathrm{push}_{\mathbb{S}} \circ \mathrm{gr}(\ulcorner X_j \urcorner)$, gives the instance of a set variable, $A(X_j)^\circ \equiv \mathrm{gr}(\ulcorner X_j \urcorner)^* \mathrm{push}_{\mathbb{S}}^* A^\circ$.

Proof of Theorem 16.11. By induction on the HAS derivation. We give the cases that carry the argument; the others are in Appendix J.2.

Hypotheses. If A occurs in Γ, then A° is a conjunct of Γ°, reached by conjunction eliminations and cuts.

Implication introduction. The HAS premise is $A, \Gamma \vdash B$, so the induction hypothesis derives $A^\circ \wedge \Gamma^\circ \vdash B^\circ$, while construction (12) wants $\Gamma^\circ \wedge A^\circ \vdash B^\circ$. Cut with the derivation of $\Gamma^\circ \wedge A^\circ \vdash A^\circ \wedge \Gamma^\circ$ by the conjunction rules, then apply (12). Disjunction elimination is repaired in the same way.

Universal introduction. The premise is $\Gamma\uparrow \vdash A$, whose translation is derived. Reindexing along $\mathrm{push}_{\mathbb{N}}$ gives $\mathrm{push}_{\mathbb{N}}^*(\Gamma\uparrow)^\circ \vdash \mathrm{push}_{\mathbb{N}}^* A^\circ$, and the second instance above replaces the left side by $\pi_1^*\Gamma^\circ$. Construction (14) gives $\Gamma^\circ \vdash \forall_{\mathbb{N}}\mathrm{push}_{\mathbb{N}}^* A^\circ$, which is $\Gamma^\circ \vdash (\forall x\, A)^\circ$.

Universal elimination at a term t. Construction (15) with the term $\ulcorner t \urcorner$ gives

$$\Gamma^\circ \vdash \mathrm{gr}(\ulcorner t \urcorner)^* \mathrm{push}_{\mathbb{N}}^* A^\circ,$$

and (16.1) turns it into $\Gamma^\circ \vdash A(t)^\circ$.

Substitution of equals. The premises translate to $\Gamma^\circ \vdash \ulcorner t \urcorner = \ulcorner s \urcorner$ and $\Gamma^\circ \vdash A(t)^\circ$. By (16.1) the second is $\Gamma^\circ \vdash \mathrm{gr}(\ulcorner t \urcorner)^* \mathrm{push}_{\mathbb{N}}^* A^\circ$; construction (20) with the formula $\mathrm{push}_{\mathbb{N}}^* A^\circ$ gives the same with s, and (16.1) again gives $\Gamma^\circ \vdash A(s)^\circ$.

The six arithmetic axioms. The translation of $S(t)$ is the term $S(\ulcorner t \urcorner)$, of $t + s$ the term $\ulcorner t \urcorner + \ulcorner s \urcorner$, and so on, as maps. So each axiom translates to an instance of rules 1–6.

The induction axiom $\forall X(0 \in X \to \forall x(x \in X \to S(x) \in X) \to \forall x(x \in X))$. Let B be the atom "the newest number belongs to the newest set" in context $E \times \mathbb{S} \times \mathbb{N}$. By the coherence lemma at the atoms, the three parts of the matrix translate, up to $\equiv$, into $\mathrm{gr}(0)^* B$, $\forall_{\mathbb{N}}(B \to \mathrm{succ}^* B)$

and $\forall_{\mathbb{N}} B$. Rule 7 at B gives $\mathrm{Ind}(B) \vdash \forall_{\mathbb{N}} B$, and two conjunction steps and two implication introductions turn it into $P \vdash \mathrm{gr}(0)^* B \to \forall_{\mathbb{N}}(B \to \mathrm{succ}^* B) \to \forall_{\mathbb{N}} B$, for every hypothesis P; construction (14) over $\mathbb{S}$ and the equivalences just named finish the case.

Comprehension $\exists X \forall x (x \in X \leftrightarrow A^+)$, where A^+ is A with its set positions moved up so that X does not occur in it. The outer quantifier is a quantifier over $\mathbb{S} = \Omega^{\mathbb{N}}$, the classifier quantifier of Comp with $X = \mathbb{N}$. Under the two binders, $(x \in X)^\circ$ is the atom Mem, because the two maps that read the classifier at the number are equal; and $(A^+)^\circ$ is $\equiv \mathrm{fg}^* \mathrm{push}^*_{\mathbb{N}} A^\circ$, by the coherence lemma for the shift of set positions composed with forgetting the classifier. So the translation is $\equiv \mathrm{Comp}(\mathrm{push}^*_{\mathbb{N}} A^\circ)$, an instance of rule 8. □

Remark 16.13 (Arithmetic proofs inside the calculus)**.** Lemma 7.16 makes every posit-free derivation of Chapter 6, with induction on every formula, a derivation in HAS of the same judgment, and Theorem 16.11 then compiles it into the calculus. So every theorem of Heyting arithmetic, read as above, is derivable in the calculus. For a theorem φ of Peano arithmetic, Theorem 20.39(1) gives a closed term of type $\mathrm{Ty}(\varphi^N)$, and Theorems 20.36 and 20.32 turn it into a posit-free derivation of the negative translation φ^N, exactly as in the proof of the second part of Theorem 20.39; the two steps above then make $(\varphi^N)^\circ$ derivable in the calculus. Each composition is immediate from the results cited.

Composed with Theorem 16.8, the compiler shows that the translation of every theorem of HAS holds at both stages. The main use of the compiler is the other direction: the calculus contains HAS, so whatever the calculus fails to derive, HAS fails to derive in translation.

16.6 The full calculus has no recursive model

The separations of Chapter 12 were proved in a model whose sequences are the computable ones. The same cannot be done for the calculus as it stands, because its terms include every map, and some maps are not computed by any program. One such map, placed in one instance of the rule of computation, already rules out every reading of the right kind.

Let $r : \mathbb{N} \to \mathbb{N}$ be the function with $r(i) = 1$ when program i halts on input 0 and $r(i) = 0$ otherwise. It is defined in ordinary mathematics by a case distinction; no program computes it [21]. The map $((), n) \mapsto r(n)$ from $\mathbf{1} \times \mathbb{N}$ to $\mathbb{N}$ is a map of constant carriers, so

$$h_r = \lambda\, \mathrm{map}\big(((), n) \mapsto r(n)\big)$$

is a closed term of type $\mathbb{N}^{\mathbb{N}}$. For each n let $\bar{n}$ be the closed term $\mathrm{map}(() \mapsto n)$, the numeral. Rule 11 at $\bar{n}$ derives

$$h_r\, \bar{n} = \mathrm{map}\big(() \mapsto r(n)\big),$$

and the right side is the numeral $\overline{r(n)}$: the two terms carry the same map, so they are the same term.

Theorem 16.14 (No recursive reading of the full calculus)**.** *There is no pair of assignments V, giving a natural number $V(t)$ to every closed term t of type $\mathbb{N}$, and J, giving a computable function $J(t) : \mathbb{N} \to \mathbb{N}$ to every closed term t of type $\mathbb{N}^{\mathbb{N}}$, such that*

(1) $V(\bar{n}) = n$ for every numeral;

(2) $V(t\,\bar{n}) = J(t)(n)$ for every closed t of type $\mathbb{N}^{\mathbb{N}}$ and every n;

(3) V gives the two sides of every closed instance of rule 11 of type $\mathbb{N}$ the same value.

The proof reasons in a classical metatheory.

Proof. Suppose V and J exist. For every n, by (2), (3) at the instance above, and (1),

$$J(h_r)(n) = V(h_r\,\bar{n}) = V\big(\overline{r(n)}\big) = r(n).$$

So r is the computable function $J(h_r)$. Then program i halts on input 0 exactly when the computable function $J(h_r)$ is 1 at i, which would decide the halting problem [21]. This is impossible. □

The same freedom of terms makes the full calculus derive every true statement of a common kind. Call a formula *equational over constant carriers* when it contains no atom, every equation in it compares two terms whose carrier is constant, and every quantifier in it ranges over a constant carrier. Numbers, sequences of numbers, binary strings, bits, binary sequences, decidable sets of strings, points of the line, functions on the line, ratio covers, presented sets and presented maps all live on constant carriers; truth values and sets of points do not.

Proposition 16.15 (The full calculus derives every true equational sentence)**.** *Reason in a classical metatheory. Let A be a formula of the calculus in a context G that is equational over constant carriers, and suppose that $g \Vdash_1 A$ for every later object $g \in G_1$. Then the calculus derives $P \vdash A$ for every formula P in context G. In particular it derives, from $\top$, the sentences LPO, $\mathsf{WLPO_{seq}}$, SPD and $\mathsf{AR} = \mathsf{SPD} \to \mathsf{WLPO_{seq}}$ of Section 16.8, the seven compactness sentences $\mathsf{FAN}_\Delta^=$, $\mathsf{FAN}_c^=$, $\mathsf{WKL}^=$, $\mathsf{HB}^=$, $\mathsf{PLB}^=$, $\mathsf{PINF}^=$ and $\mathsf{HC}^=$ of Section 16.10, and the implication from each of $\mathsf{HB}^=$, $\mathsf{PLB}^=$, $\mathsf{PINF}^=$ and $\mathsf{HC}^=$ to $\mathsf{FAN}_\Delta^=$, all with their presentations erased.*

Proof. The derivation is built in Appendix J.3; here is its plan. The *indicator* of a formula A in context G is the map $\chi_A : G \to \mathbb{N}$ whose later component sends g to 1 when $g \Vdash_1 A$ and to 0 otherwise, and whose earlier component is the later one composed with the passage. It is defined by a case distinction in the metatheory, and it is a map of carriers because the passage of $\mathbb{N}$ is the identity. Write 1 for the term $S(0)$. By induction on A, the calculus derives $\chi_A = 1 \vdash A$ and $A \vdash \chi_A = 1$ for every formula equational over constant carriers. Five facts carry the induction:

(i) two terms of a constant carrier with the same later values carry the same map, so their equation is derivable from any hypothesis (Remark 16.7);

(ii) substitution of equals, construction (20), turns a derived equation $a = b$ and a true identity between two maps evaluated at a into the same identity evaluated at b;

(iii) rule 1 refutes an equation between two terms whose values are 0 and 1;

(iv) rule 7 derives $\forall_{\mathbb{N}} x\big(x = 0 \vee \exists_{\mathbb{N}} y(x = S(y))\big)$, which splits an indicator into the cases 0 and 1;

(v) at a quantifier over a constant carrier X, a witness or a counterexample chosen in the metatheory for each later object of the context is a map into X, and so a term.

If A holds at every later object, then χ_A and 1 have the same later values, so $P \vdash \chi_A = 1$ is derivable by (i), and a cut with $\chi_A = 1 \vdash A$ gives $P \vdash A$.

The sentences named are equational over constant carriers, as their definitions show, and each holds at the later stage of the one-point context, where after erasure every quantifier ranges over all objects of ordinary mathematics. LPO is excluded middle applied to each sequence. $\mathsf{WLPO}_{\mathsf{seq}}$ holds because the function that marks the vanishing rows is now in the range of its sequence quantifier. SPD holds because excluded middle decides whether $a = 0$, and a point $a \neq 0$ has an inverse (Lemma F.11, with the stage N found as in the proof of Theorem 16.27); AR then holds because its conclusion does. The later readings of the compactness sentences are the statements of Chapter 13 for all presented sets, covers and maps (Section 16.10). In the metatheory the full fan theorem holds by König's lemma. Proposition 13.4 gives FAN_c and FAN_Δ from it, Theorem 12.72 gives weak König's lemma from LLPO and countable choice, Theorems 13.25 and 13.28 and Corollary 13.30 give Heine–Borel, the positive lower bound and the positive infimum from FAN_Δ, and Theorem 13.32 gives Heine–Cantor for presented sets from the full fan theorem. An implication with a true conclusion is true. □

So in the full calculus, for sentences of this kind, "A derives B" says only that $A \to B$ is true in ordinary classical mathematics: a derivation of $A \vdash B$ makes $A \to B$ true at the later stage by Theorem 16.8, and conversely the proposition supplies one. LPO derives $\mathsf{WLPO}_{\mathsf{seq}}$ and every compactness sentence of Section 16.10 there, because all of them are classically true. The question of the chapter's opening therefore has its answer in the computable calculus.

The obstruction to a recursive reading is not a missing rule but a term that is present. Two facts show exactly where it lies.

Remark 16.16 (What a recursive reading can and cannot keep)**.** A closed term of type $\mathbb{N}^{\mathbb{N}}$ agrees with some computable function exactly when the function it carries is computable. And rule 11 has a recursive reading at every computable body: if $t : \mathbf{1}\times\mathbb{N} \vdash \mathbb{N}$ carries a computable function, one computable function gives the values of the right sides $\mathrm{map}([\![t]\!] \circ \mathrm{gr}(\bar{n}))$ of all its instances at numerals. A second obstruction comes from the existential rule. For every closed term t of type $\mathbb{N}^{\mathbb{N}}$ the calculus derives $\exists_{\mathbb{N}^{\mathbb{N}}}(\mathrm{map}(\pi_2) = \mathrm{map}([\![t]\!] \circ \pi_1))$, by existential introduction at t, reflexivity and rule 19. A reading in which this quantifier ranges over computable sequences and equality means equality would make the sequence of t computable, and $t = \mathrm{map}(() \mapsto r)$ is a closed sequence term whose sequence is not.

So a model in which functions are computable must restrict the terms. The next section does this by a property of each term, not by a list of forbidden ones.

16.7 The computable calculus and its model

The restriction keeps the carriers, the readings, the twenty constructions and the twenty-nine rules, and asks one more thing of each term and each map: that a program compute it on names.

Definition 16.17 (Names and tracked maps)**.** A *presentation* of a carrier X is a relation "e names x" between natural numbers e and later objects $x \in X_1$; a later object is *named* when some number names it. Only the later stage carries names. We use the following presentations, with $\{e\}(a)$ the output of program e on input a and $(e)_0, (e)_1$ the components of the pair numbered e, as in the proof of Theorem 12.22.

- On $\mathbb{N}$, e names n when $e = n$. On binary strings, bits and ratio records, e names an object when e is its standard code.
- On Ω and on $\mathbf{1}$, every number names every object.
- On $X \times Y$, e names (x, y) when $(e)_0$ names x and $(e)_1$ names y.
- On Y^X, e names a function $f : X_1 \to Y_1$ when, for every name a of every $x \in X_1$, the run $\{e\}(a)$ halts with a name of $f(x)$.
- On $\mathscr{R}$, e names a regular sequence s when $\{e\}(n)$ halts with the code of the ratio record s_n for every n. On $\mathscr{L}$, e names a point when it names a regular sequence of which the point is the class.

A map $f : X \to Y$ of presented carriers is *tracked* when one program c sends every name of every later object to a name of its image: for every $x \in X_1$ and every name a of x, $\{c\}(a)$ halts with a name of $f_1(x)$.

Every presentation of a truth value carries no information about it: the number 0 names both the true and the false proposition. This is deliberate. A name of a classifier is then a program that halts on every name of its argument, and nothing more.

The identity, composition, projections, pairing, evaluation, currying, the arithmetic operations of $\mathbb{N}$, constant maps with a named value, the quotient map $\mathscr{R} \to \mathscr{L}$ and every map into Ω are tracked: evaluation by a universal program, currying by computing a program for a function of the second argument from the first, and a map into Ω by the program that returns 0 (Appendix J.4).

Definition 16.18 (The computable calculus)**.** Its contexts and quantifier domains are presented carriers, with products presented by pairs. Its terms are the terms of Definition 16.3 whose readings are tracked, and its reindexing maps are tracked maps. Its formulas are built by the ten constructions of Definition 16.4 from these terms and maps, with every predicate allowed as an atom. Its derivations are built by the twenty constructions of Definition 16.6 from the twenty-nine base rules of Table 16.2, with tracked terms and maps as their parameters; in rule 11, the map $[\![t]\!] \circ \mathrm{gr}(u)$ is tracked by composing the trackers of t and u. *Erasing* the presentations and the trackers sends each formula and each derivation of the computable calculus to a formula and a derivation of the calculus.

The atoms are not restricted, and every rule is kept. What changes is that a term must come with a program. The term h_r of the previous section is not a term of the computable calculus (Proposition 16.21(d)).

Remark 16.19 (Atoms that no rule governs)**.** Only three kinds of base rule act on atoms: rule 8, through the membership predicate $\in$ on Ω; rule 10, through the predicate $\approx$ on $\mathscr{R} \times \mathscr{R}$; and the reindexing rules 16 and 17, which move any atom along a map. An atom of a predicate other than $\in$ and $\approx$ is therefore governed by no rule. Choose, for some of these predicates R, a formula φ_R in the context X of R, and replace every atom $R(t)$ by the reindexing $[\![t]\!]^*\varphi_R$. This replacement maps derivations of the computable calculus to derivations: the constructions and the other rules are schematic in formulas, and an instance of rule 16 or 17 becomes an instance of the composition rules 14 and 15. Take $\varphi_R = \bot$ for every predicate R, other than $\in$ and $\approx$, that holds at every later object. A derivation of $\top \vdash R(t)$ for such a predicate would become a derivation of $\top \vdash [\![t]\!]^*\bot$, which Theorem 16.22 below excludes. So even the closed atom of the true relation $0 \leq 0$ is not derivable from $\top$. Sentences whose content rests on such relations, such as the order of the numbers or of the line, the length of a string or membership in a presented set, are therefore written with equations between terms of tracked maps (Section 16.10). The sentences of Section 16.8 already contain no atom.

Definition 16.20 (The model of named objects)**.** For a formula A of the computable calculus in context G and a later object $g \in G_1$, define $g \Vdash A$ by recursion on A:

- $g \Vdash R(t)$ when $[\![t]\!]_1(g) \in R_1$, and $g \Vdash t = u$ when $[\![t]\!]_1(g) = [\![u]\!]_1(g)$;
- $\top$, $\bot$, $\wedge$, $\vee$ and $\to$ are read classically;
- $g \Vdash \forall_X A$ when $(g, x) \Vdash A$ for every *named* $x \in X_1$, and $g \Vdash \exists_X A$ when $(g, x) \Vdash A$ for some named $x \in X_1$;
- $g \Vdash f^*A$ when $f_1(g) \Vdash A$.

P entails Q in the model when every named $g \in G_1$ with $g \Vdash P$ has $g \Vdash Q$.

This is a Tarskian model in the spirit of Henkin's general models of higher types [52]: truth is classical, and the quantifiers at each type range over a chosen part of the domain, here the named objects. In it, sequences of numbers range over the computable functions and points of the line over the points with computable names, as in the model of Theorem 12.63; truth values range over all propositions and sets of points over all sets of points, since every one of them is named.

The countermodels below assemble standard ingredients of recursive analysis. On the constant carriers of numbers, sequences, strings and decidable sets, the model of named objects is the ω-model REC of the computable sets, read with classical logic, and that weak König's lemma fails there is standard [76, Chapter VIII]. The dead cover of Section 16.10 covers the computable points of $[0, 1]$ by rational intervals with no finite subcover, as the singular coverings of Zaslavskiĭ and Ceĭtin do [77], and the tent function of that section, continuous at the computable points of $[0, 1]$ and not uniformly continuous there, is the corresponding example for Heine–Cantor in recursive constructive mathematics [40, Chapter 3].

Proposition 16.21 (The named objects)**.** *In the presentations of Definition 16.17:*

(a) a function $f : \mathbb{N} \to \mathbb{N}$ is named in $\mathbb{N}^{\mathbb{N}}$ exactly when it is computable, and the same holds for binary sequences and for maps from binary strings to bits;

(b) a point of $\mathscr{L}$ is named exactly when it is the class of a regular sequence whose entries are computed by a program; in particular the point $\eta(q)$ of every ratio record q is named;

(c) every tracked map from Ω to $\mathbb{N}$ takes the same value at every proposition;

(d) the map $((), n) \mapsto r(n)$ of Section 16.6 is not tracked.

Proof. (a) If e names f, then $\{e\}(n)$ halts with $f(n)$ for every n, so a program computes f; conversely a program for f names f, since the name of a number is the number. For strings and bits compose with the computable coding. (b) is the definition; the constant sequence of q is computed by a program that ignores its input. (c) Let c track $f : \Omega \to \mathbb{N}$. The number 0 names every proposition p, so $\{c\}(0)$ halts with a name of $f_1(p)$, which is $f_1(p)$ itself; the run has one output, so $f_1(p) = \{c\}(0)$ for every p. (d) A tracker c would send the name of $((), n)$ with components 0 and n to $r(n)$, so r would be computable. □

Theorem 16.22 (Soundness of the computable calculus)**.** *If the computable calculus derives $P \vdash Q$, then P entails Q in the model of named objects.*

Proof. Two facts about names are used. (N1) If g is named and t is a term of the computable calculus, then $[\![t]\!]_1(g)$ is named: run the tracker on a name of g; the same holds for tracked maps. (N2) A pair (g, x) is named exactly when g and x are, by pairing and unpairing names.

The twenty constructions are handled as in Step 1 of the proof of Theorem 16.8, at the later stage only, with the following changes. Universal introduction uses (N2): if g is named and x is named, then (g, x) is named and satisfies $\pi^* P$, so it satisfies A. Universal elimination and existential introduction use the object $[\![t]\!]_1(g)$, which is named by (N1). Existential elimination uses (N2), and reindexing uses (N1) for the map. Since implication is read at one stage, persistence is not needed.

For the base rules: rules 1–6 are arithmetic truths. For rule 7, every number is named, so a quantifier over named numbers is a quantifier over all numbers, and the rule is ordinary induction. For rule 8, the property $x \mapsto \big((g, x) \Vdash A\big)$ is a later object of Ω^X, and it is named, by any program that halts with 0 on every input, because 0 names every proposition; with it as the classifier, both implications of $\mathrm{Comp}(A)$ are identities. For rule 9, a name of a point names a regular sequence whose class is that point, and that sequence is named. Rules 10, 11 and 12–29 are identities at the later stage, as in Theorem 16.8. □

Corollary 16.23 (The computable calculus is consistent)**.** *The computable calculus does not derive $\top \vdash \bot$ in the one-point context.*

Proof. The point of $\mathbf{1}$ is named and satisfies $\top$, and nothing satisfies $\bot$. □

Theorem 16.22 is the tool for every separation below. To show that a sentence B is not derivable from a sentence A, it is enough to find that A is true and B false at the point of **1**. Implication introduction and elimination show that $A \vdash B$ is derivable exactly when $\top \vdash A \to B$ is, so the two forms of a separation stand or fall together (Question 3); and a derivation of $\top \vdash B$ gives one of $A \vdash B$ by a cut with $A \vdash \top$. Each separation below is stated in all three forms, $\top \vdash B$, $\mathsf{LPO} \vdash B$ and $\top \vdash \mathsf{LPO} \to B$.

The connectives of this model are read classically. So $A \vee \neg A$ is true at every point for every formula A, and every omniscience principle stated for one sequence at a time, such as the sentence LPO of Section 16.8, is true. The decision is made for each instance separately, by the classical mathematics in which the model is described, and no program makes it. A principle that asks for infinitely many decisions collected into one sequence, such as $\mathsf{WLPO}_{\mathsf{seq}}$, can be false there (Theorem 16.27). The separations below therefore do not come from a missing decision. They come from the quantifiers: a sequence, a path or a point exists in the model only if a program names it. Comprehension always produces a classifier on $\mathbb{N}$, but no tracked map reads a number off a classifier: composed with the tracked map $p \mapsto (n \mapsto p)$ from Ω to $\Omega^{\mathbb{N}}$, a tracked map from $\Omega^{\mathbb{N}}$ to $\mathbb{N}$ becomes a tracked map from Ω to $\mathbb{N}$, which takes one value at every proposition by Proposition 16.21(c); so it takes the same value at the empty classifier and at the full one (Question 4).

Remark 16.24 (Excluded middle as a base rule)**.** Every instance $A \vee \neg A$ is true at every point of the model of named objects, because the model reads the connectives classically. So Theorem 16.22 holds for the computable calculus with every judgment $P \vdash A \vee \neg A$, in every context and for all formulas P and A, added as a base rule: the proof needs one more case, and this observation is that case. Every separation below therefore holds in that extension too. A sentence that is false in the model is not derivable there from $\top$, nor from LPO, nor in the form $\top \vdash \mathsf{LPO} \to B$. In the computable calculus itself, the same argument shows that such a sentence is not derivable from LPO together with any finite list of universal closures $\forall_{X_1} \cdots \forall_{X_k}(A \vee \neg A)$ of instances of excluded middle.

Theorem 16.25 (HAS compiles into the computable calculus)**.** *Present the environment carrier E of Definition 16.10 as a product of exponentials; its named environments are those with a computable number valuation, and every set valuation. Then all the terms and maps of the translation are tracked, and every derivation of $\Gamma \vdash A$ in HAS gives a derivation of $\Gamma^\circ \vdash A^\circ$ in the computable calculus. Erasing it gives a derivation, in the calculus, of the translated judgment of Theorem 16.11.*

Proof. The construction is that of the proof of Theorem 16.11, with the same cases and the same coherence lemma; the reindexing maps it uses are $\mathrm{push}_{\mathbb{N}}$, $\mathrm{push}_{\mathbb{S}}$, the successor, projections, graphs of the terms $\ulcorner t \urcorner$ and $\ulcorner X_j \urcorner$, and the maps that forget a classifier. Each is tracked: evaluating a number term at a valuation named by a program is computed from that program, and putting a new entry in front of a valuation is computed from the programs of the entry and of the valuation (Appendix J.4). A set valuation is named by any program that halts on every input with the number of a program that halts on every input, because every number names every truth value. Erasing the presentations sends A° to the translation of Definition 16.10, formula by formula. □

For a translated HAS formula, the model of named objects reads every set variable as an arbitrary set of numbers and every number variable as a number; so it gives each HAS formula its ordinary classical truth, and every principle of Definition 7.18, excluded middle included, is true there. So the model separates none of these principles from one another inside HAS. Its separations concern sentences with quantifiers over sequences, paths, trees and points, which lie outside the image of the translation.

Proposition 16.26 (The model is not a realizability model)**.** *Read the formulas of the computable calculus by Kleene's realizability [50] over the same presentations: e realizes $A \wedge B$ when $(e)_0$ realizes A and $(e)_1$ realizes B; $A \vee B$ when $(e)_0 = 0$ and $(e)_1$ realizes A, or $(e)_0 \neq 0$ and $(e)_1$ realizes B; $A \to B$ when $\{e\}(a)$ halts with a realizer of B for every realizer a of A; $\forall_X A$ at g when $\{e\}(a)$ halts with a realizer of A at (g, x) for every name a of every $x \in X_1$; $\exists_X A$ when $(e)_0$ names some x and $(e)_1$ realizes A at (g, x); f^*A at g when e realizes A at $f_1(g)$; $\top$ always and $\bot$ never; and an atom or an equation by evidence, given by a relation* $\mathrm{Ev}(p, e)$ *between truth values and numbers that is* sound, $\mathrm{Ev}(p, e)$ *implying p, and has evidence for the true proposition. Let θ be the formula $\in(\mathrm{map}(\pi_2)) \vee \neg\in(\mathrm{map}(\pi_2))$ in context $\mathbf{1} \times \Omega$, which says $p \vee \neg p$ of the newest truth value p. Then no number realizes the instance* $\mathrm{Comp}(\theta)$ *of rule 8, whatever the evidence relation, while* $\mathrm{Comp}(\theta)$ *is true in the model of named objects.*

Proof. The second statement is Theorem 16.22 applied to the rule. For the first, first look at θ itself, and call $(e)_0$ the *tag* of a number e. At p true, a realizer of θ must have tag 0: a tag $\neq 0$ would need a realizer of $\neg p$, which would turn evidence for the true proposition into a realizer of $\bot$. At p false, a realizer must have tag $\neq 0$, since tag 0 needs evidence for a false proposition, and soundness excludes it. So no number realizes θ at both values. Both values have tag-correct realizers: the number of the pair $(0, a)$, with a evidence for truth, and the number of the pair $(1, 0)$.

Now suppose e realizes $\mathrm{Comp}(\theta)$. Then $(e)_0$ names a classifier χ on Ω, and $d = (e)_1$ realizes $\forall_\Omega\big((\mathrm{Mem} \to \theta) \wedge (\theta \to \mathrm{Mem})\big)$ at χ. The number 0 names both truth values, so the one run $\{d\}(0)$ yields one number b that realizes the matrix at p true and at p false. Its second component turns the realizers of θ just described into evidence for $\chi(\mathrm{true})$ and for $\chi(\mathrm{false})$; by soundness both propositions hold, so both are the true proposition, and the evidence a for truth is evidence for each. Its first component then sends this one number a, in one run, to one number that realizes θ at p true and at p false. That is impossible. □

The model of named objects decides each instance of a disjunction on its own, which is why it validates the instance; a realizer must make the decision uniformly in the names, and over a presentation in which a name says nothing about a truth value it cannot. Kleene's reading in the proof of Theorem 12.22 was defined for arithmetic with sequences, whose atoms are decided equations. Over these presentations it does not extend to the computable calculus: one instance of rule 8 has no realizer.

16.8 LPO decides each row but cannot collect the rows

Scalar pair division says that for every two points a and b of the line, either both are 0, or $a \neq 0$ and b is a multiple ac of a, or $b \neq 0$ and a is a multiple bc of b; here $x \neq 0$ means $\neg(x = 0)$. Chapter 14 shows that it is equivalent to basis selection and to rank–nullity

(Theorems 14.13 and 14.14), that it implies WLPO (Theorem 14.16), and that it follows from $\mathrm{WLPO}_{\mathrm{seq}}$ (Proposition 14.17): for every binary double sequence there is a binary sequence that records, for every row, whether the row vanishes. Does scalar pair division give back $\mathrm{WLPO}_{\mathrm{seq}}$? With unique choice it does, since WLPO with unique choice gives $\mathrm{WLPO}_{\mathrm{seq}}$ (Corollary 13.10, item 3). Without it the question is about collection: scalar pair division settles one pair of points at a time, while $\mathrm{WLPO}_{\mathrm{seq}}$ asks for one sequence that settles infinitely many rows at once. We answer it in the computable calculus, even with LPO added and with excluded middle as a base rule.

Let Seq be $\mathbb{N}^{\mathbb{N}}$ with its presentation, so that its named objects are the computable functions, and write αn for the application of a sequence term to a number term. Let $\mathrm{Bin}(\alpha)$ be $\forall_{\mathbb{N}}(\alpha n = 0 \vee \alpha n = 1)$, and write $\langle i, n\rangle = (i+n)\cdot(i+n)+n$, the pairing of the proof of Theorem 12.22, a term built from $+$ and $\cdot$. The following are closed sentences of the computable calculus:

$$
\begin{aligned}
\mathsf{LPO} &:= \forall_{\mathsf{Seq}}\alpha\big(\mathrm{Bin}(\alpha) \to \forall_{\mathbb{N}} n\,(\alpha n = 0) \vee \exists_{\mathbb{N}} n\,(\alpha n = 1)\big),\\
\mathsf{WLPO}_{\mathsf{seq}} &:= \forall_{\mathsf{Seq}} A\,\Big(\mathrm{Bin}(A) \to \exists_{\mathsf{Seq}} f\big(\mathrm{Bin}(f) \wedge \forall_{\mathbb{N}} i\,(fi = 0 \leftrightarrow \forall_{\mathbb{N}} n\, A\langle i, n\rangle = 0)\big)\Big),\\
\mathsf{SPD} &:= \forall_{\mathscr{L}} a\,\forall_{\mathscr{L}} b\,\Big((a = 0 \wedge b = 0) \vee \big(\neg(a = 0) \wedge \exists_{\mathscr{L}} c\,(a\cdot c = b)\big)\\
&\qquad\qquad \vee \big(\neg(b = 0) \wedge \exists_{\mathscr{L}} c\,(b\cdot c = a)\big)\Big),\\
\mathsf{AR} &:= \mathsf{SPD} \to \mathsf{WLPO}_{\mathsf{seq}}.
\end{aligned}
$$

The point 0 and multiplication on the line are tracked maps (Appendix J.4). The second sentence is $\mathrm{WLPO}_{\mathrm{seq}}$ of Definition 13.8 with the two bits of f exchanged: here $f(i) = 0$ marks a vanishing row. All four sentences are built from application, arithmetic, equations and the connectives. No atom of them asserts a principle.

Theorem 16.27 (LPO does not collect)**.** *In the computable calculus none of the following is derivable:* $\top \vdash \mathsf{WLPO}_{\mathsf{seq}}$*;* $\mathsf{LPO} \vdash \mathsf{WLPO}_{\mathsf{seq}}$*;* $\top \vdash \mathsf{LPO} \to \mathsf{WLPO}_{\mathsf{seq}}$*;* $\top \vdash \mathsf{AR}$*;* $\mathsf{LPO} \vdash \mathsf{AR}$*;* $\top \vdash \mathsf{LPO} \to \mathsf{AR}$*. The same holds when every instance of excluded middle is added as a base rule. The proof reasons in a classical metatheory.*

Proof. By Theorem 16.22, and by Remark 16.24 when excluded middle is added, it is enough to evaluate the sentences at the point of **1** in the model of named objects; Lemma J.2 unfolds each of them there.

LPO *is true.* Its sequences range over the computable functions (Proposition 16.21(a)). For a binary α, either some $\alpha(n)$ is 1, or none is, and then every $\alpha(n)$ is 0. This is excluded middle for the statement "some $\alpha(n)$ is 1", applied in the metatheory to each α separately; no program decides it uniformly in α.

$\mathsf{WLPO}_{\mathsf{seq}}$ *is false.* By Proposition 16.21(a), the sentence is true exactly when every computable binary A has a computable binary f with $f(i) = 0$ exactly when $A\langle i, n\rangle = 0$ for every n. Let $A\langle i, n\rangle$ be 1 when program i halts on input 0 within n steps, and let A be 0 at the numbers that are not values of the pairing. This is the bounded-halting function h of the proof of Theorem 12.22, read through the pairing: A is computable, since n steps of a run can be simulated and the pairing is inverted by a bounded search. Row i vanishes exactly when

program i never halts on input 0. A computable f as required would decide, for every i, whether program i halts on input 0, which no program does [21].

SPD *is true.* Let a and b be named points. If $a = 0$ and $b = 0$, the first alternative holds. Suppose $a \neq 0$; whether $a = 0$ is decided here by excluded middle in the metatheory. Let s be a regular sequence named by a program, of which a is the class. Some stage N has $|s_N| > 2/(N+1)$: otherwise every stage satisfies $|s_n| \leq 2/(n+1)$ and $a = \eta(0)$ by Lemma 12.15. This existence is again classical. For this N, Lemma F.11 gives the regular sequence $n \mapsto 1/s_{4(N+1)^2(n+1)-1}$, whose point y has $ay = \eta(1)$. A program computes it from the program of s, with N fixed, so y is named. Multiplication is tracked, so $c = y \cdot b$ is named, and $a \cdot c = b$: the second alternative holds. If $a = 0$ and $b \neq 0$, the third alternative holds in the same way. The stage N depends on a, so the inverse is named for each a separately; no program computes it from a name of a.

Conclusion. LPO is true and $\mathsf{WLPO}_{\mathsf{seq}}$ false, so no derivation of $\top \vdash \mathsf{WLPO}_{\mathsf{seq}}$ or of $\mathsf{LPO} \vdash \mathsf{WLPO}_{\mathsf{seq}}$ exists, nor of $\top \vdash \mathsf{LPO} \to \mathsf{WLPO}_{\mathsf{seq}}$, whose conclusion is false at the point. Since SPD is true and $\mathsf{WLPO}_{\mathsf{seq}}$ false, AR is false, and the same argument excludes the last three judgments. □

Proposition 16.28 (The sentence asks the reversal question)**.** *Erase the presentations of* AR*. The resulting sentence of the calculus has, at each of the two stages, the reading: scalar pair division for all points of* $\mathbb{R}_\delta$ *implies* $\mathrm{WLPO}_{\mathrm{seq}}$ *for all binary double sequences.*

Proof. After erasure every quantifier ranges over a constant carrier, so the earlier and the later halves of each clause of Table 16.1 say the same thing, and unfolding the clauses gives the statement. □

So the computable calculus does not derive, even from LPO and with excluded middle as a base rule, the implication from scalar pair division to $\mathrm{WLPO}_{\mathrm{seq}}$; in particular it does not derive that implication outright. In the full calculus the same sentences are derivable outright (Proposition 16.15), so the separation is a fact about computed terms. What fails is collection. In the model, LPO is true and settles each row: for every i there is exactly one bit that records whether row i vanishes. What the model lacks is a computable function with these values, and the calculus has no rule that produces a number function from such a relation. Its comprehension rule produces classifiers, and no tracked map reads a number off a classifier (Proposition 16.21(c), composed with the tracked map $p \mapsto (n \mapsto p)$ from Ω to $\Omega^{\mathbb{N}}$, as in Question 4). Scalar pair division is true in the model for the same reason that LPO is: it is settled pair by pair, without a program.

The result is the dual of Theorem 12.22. There, countable choice collects witnesses and decides nothing. Here, omniscience decides and collects nothing.

16.9 Kleene's tree

The compactness statements fail in the model for one reason, found by Kleene [51, 4]: there is a computable infinite binary tree with no computable path. The proof of Theorem 13.11 uses the same tree to show that excluded middle does not give the fan theorem; we collect its properties in one lemma. We use the trees and paths of Definition 12.71.

Say that program e *has halted by n with output o* when its run on input 0 halts within n steps with output o. *Kleene's tree* T_K consists of the binary strings l such that, for every $e < |l|$ that has halted by $|l|$ with some output o, the digit l_e is 1 if $o = 0$ and 0 if $o \neq 0$. A digit is constrained once its program has halted, and the constraint contradicts the output.

Lemma 16.29 (Kleene's tree)**.**

(a) Membership in T_K is decided by a program.

(b) T_K is a decidable binary tree: it contains the empty string and every initial segment of each of its members.

(c) T_K has a member of every length.

(d) Every computable binary sequence p leaves T_K: $\bar{p}(n) \notin T_K$ for some n.

Proof. (a) Deciding $l \in T_K$ runs each program $e < |l|$ for $|l|$ steps and compares outputs with digits.

(b) The empty string has no constraint. Let $lb \in T_K$ and $e < |l|$. If e has halted by $|l|$ with output o, it has halted by $|l| + 1$ with the same output, so the constraint on digit e holds in lb, and digit e of l is digit e of lb. So $l \in T_K$.

(c) For a length N, let g_N have digit e equal to 1 exactly when program e has halted by N with output 0. Each constraint at length N is met: output 0 gives digit 1, and any other output gives digit 0.

(d) The function $(c, x) \mapsto p(c)$ is computable, so by Kleene's recursion theorem [19, Chapter XII] there is a program e with $\{e\}(x) = p(e)$ for every x. Its run on input 0 halts with output $p(e)$, say within k steps. Let $n = k + e + 1$. Then $e < n$ and e has halted by n with output $p(e)$. If $p(e) = 0$, the constraint demands digit 1 at e; if $p(e) = 1$, it demands digit 0. Either way $\bar{p}(n)$, whose digit at e is $p(e)$, violates it, so $\bar{p}(n) \notin T_K$. □

The tree and its path in ordinary mathematics are a different matter. By Theorem 12.72, LLPO with countable choice gives a path through T_K, and classically one exists. The lemma says only that no program computes one. In the model of named objects, paths are computable functions, so there T_K is an infinite decidable tree without a path.

16.10 Compactness is not derivable in the computable calculus

We now return to the question of the chapter's opening. The statements studied in Chapter 13 become sentences of the computable calculus, and none of them is derivable, even from LPO and with excluded middle as a base rule. Every relation they use, among numbers, strings and points, is written as an equation between terms of tracked maps, so that the rules of the calculus act on it. An atom of the order or of membership would be governed by no rule (Remark 16.19), and a sentence that rested on one would fail to be derivable for that reason alone.

Carriers and tracked maps. Let W be the carrier of binary strings and B that of bits, both constant and presented by codes; $\mathsf{P} = \mathsf{B}^{\mathbb{N}}$ the carrier of binary sequences; and $\mathsf{D} = \mathsf{B}^{\mathsf{W}}$ the carrier of decidable sets of strings, given by their maps to $\{0,1\}$ as in Definition 13.1. By Proposition 16.21(a), the named sequences and the named decidable sets are the computable ones. The carriers of presented sets (Definition 13.22), of ratio covers (Definition 13.5) and of presented uniform maps (Definition 13.27) are constant; each object is the set, cover or map together with its data. A presented set is *named* by a program that computes, from n, the list of programs of the regular sequences of its n-th net. A cover is named by a program that computes the two ratios of its k-th interval from k. A presented uniform map is named when programs name its set, its function on the line, its transformer, read on programs, and its modulus. For Heine–Cantor a *map package* is a presented set with a function on the line, named when both are. The sentences below use the terms of the following tracked maps, whose programs are described in Appendix J.4:

- on numbers, the successor S and the sum $+$;
- on strings, the length $|w|$, the digit w_i at position i (with the value 0 when $i \geq |w|$), the concatenation wt, the string wb with one more digit b, the empty string $\emptyset$ and the bit 1; and the applications $p(i)$ of a binary sequence and $D(w)$ of a decidable set;
- on the line, $\max(x,y)$, $x+y$, $-x$, the points 0 and 1, and the point $\eta(1/(k+1))$ of a number k;
- for a ratio cover c, the endpoints $a_c(k)$ and $b_c(k)$ of its k-th interval, as points of the line;
- for a presented set K, the number $\ell_K(n)$ of entries of its n-th net and the point $y_K(n,i)$ named by the i-th entry, which is 0 when $i \geq \ell_K(n)$; and on numbers $r(n) = 2^n - 1$, so that $1/(r(n)+1) = 2^{-n}$;
- for a presented uniform map or a map package M, its set K_M and the value $f_M(x)$ of its function at a point x; and for a function f on the line, the application $f(x)$.

Definition 16.30 (The relations as equations)**.** In the computable calculus the formulas on the right abbreviate the relations on the left. Each is built from equations, the connectives and quantifiers over $\mathbb{N}$ and W; it contains no atom.

$$
\begin{aligned}
n \leq N &:\equiv \exists_{\mathbb{N}} k\,(n+k=N), \qquad i < n :\equiv \exists_{\mathbb{N}} k\,\big(i + S(k) = n\big),\\
\mathrm{Hit}(D,p,n) &:\equiv \exists_{\mathsf{W}} w\big(|w| = n \wedge \forall_{\mathbb{N}} i\,(i<n \to p(i) = w_i) \wedge D(w) = 1\big),\\
\mathrm{CHit}(D,p,n) &:\equiv \exists_{\mathsf{W}} w\big(|w| = n \wedge \forall_{\mathbb{N}} i\,(i<n \to p(i) = w_i) \wedge \forall_{\mathsf{W}} t\, D(wt) = 1\big),\\
a \leq b &:\equiv \max(a,b) = b, \qquad 0 < y :\equiv \exists_{\mathbb{N}} k\,\big(\eta(1/(k+1)) \leq y\big),\\
a < b &:\equiv \exists_{\mathbb{N}} k\,\big(\eta(1/(k+1)) \leq b + (-a)\big),\\
x \approx_j y &:\equiv x \leq y + \eta(1/(j+1)) \wedge y \leq x + \eta(1/(j+1)),\\
x \in I_k(c) &:\equiv a_c(k) < x \wedge x < b_c(k), \qquad x \in [0,1] :\equiv 0 \leq x \wedge x \leq 1,\\
x \in K &:\equiv \forall_{\mathbb{N}} n\, \exists_{\mathbb{N}} i\,\big(i < \ell_K(n) \wedge x \approx_{r(n)} y_K(n,i)\big).
\end{aligned}
$$

So $\mathrm{Hit}(D,p,n)$ says $D(\bar{p}(n)) = 1$ through its one witness, the initial segment $\bar{p}(n)$, and $\mathrm{CHit}(D,p,n)$ says that $D(\bar{p}(n)t) = 1$ for every string t. A point lies in a presented set when every net has an entry within 2^{-n} of it; since presented sets are closed (Proposition 13.23), this is membership.

Lemma 16.31 (The equations read as the relations)**.** *In the model of named objects, at every named object of the context, each formula of Definition 16.30 holds exactly when the relation it abbreviates holds: $n \le N$; $i < n$; $D(\bar{p}(n)) = 1$; $D(\bar{p}(n)t) = 1$ for every string t; $a \le b$, $0 < y$ and $a < b$ in the order of the line; x and y lie within $1/(j+1)$; x lies in the k-th interval of c; $0 \le x \le 1$; and x is a member of K.*

The proof is in Appendix J.5. With these abbreviations the decidable fan theorem and its c-set form (Definition 13.2) and decidable weak König's lemma (Definition 12.71) are the sentences

$$\begin{aligned}
\mathsf{FAN}_{\Delta}^{=} &:= \forall_{\mathsf{D}} D\Big(\forall_{\mathsf{P}} p\, \exists_{\mathbb{N}} n\, \mathrm{Hit}(D,p,n) \to \exists_{\mathbb{N}} N \forall_{\mathsf{P}} p\, \exists_{\mathbb{N}} n\, \big(n \le N \wedge \mathrm{Hit}(D,p,n)\big)\Big),\\
\mathsf{FAN}_{c}^{=} &:= \text{the same with CHit in place of Hit},\\
\mathsf{WKL}^{=} &:= \forall_{\mathsf{D}} T\big(\mathrm{Tree}(T) \to \mathrm{Inf}(T) \to \exists_{\mathsf{P}} p\, \forall_{\mathbb{N}} n\, \mathrm{Hit}(T,p,n)\big),
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{Tree}(T) &:\equiv T(\varnothing) = 1 \wedge \forall_{\mathsf{W}} w \forall_{\mathsf{B}} b\, \big(T(wb) = 1 \to T(w) = 1\big),\\
\mathrm{Inf}(T) &:\equiv \forall_{\mathbb{N}} n\, \exists_{\mathsf{W}} w\big(|w| = n \wedge T(w) = 1\big).
\end{aligned}$$

The four compactness sentences for presented sets are

$$\begin{aligned}
\mathsf{HB}^{=} &:= \forall K \forall c\Big(\forall_{\mathscr{L}} x\big(x \in K \to \exists_{\mathbb{N}} k\, x \in I_k(c)\big)\\
&\qquad \to \exists_{\mathbb{N}} N \forall_{\mathscr{L}} x\big(x \in K \to \exists_{\mathbb{N}} k\,(k < N \wedge x \in I_k(c))\big)\Big),\\
\mathsf{PLB}^{=} &:= \forall M\Big(\forall_{\mathscr{L}} x\big(x \in K_M \to 0 < f_M(x)\big)\\
&\qquad \to \exists_{\mathbb{N}} k \forall_{\mathscr{L}} x\big(x \in K_M \to \eta(1/(k+1)) \le f_M(x)\big)\Big),\\
\mathsf{PINF}^{=} &:= \forall M\Big(\exists_{\mathscr{L}} x(x \in K_M) \to \forall_{\mathscr{L}} x\big(x \in K_M \to 0 < f_M(x)\big) \to \exists_{\mathscr{L}} m\,\big(\mathrm{Inf}_M(m) \wedge 0 < m\big)\Big),\\
\mathsf{HC}^{=} &:= \forall M\Big(\forall_{\mathscr{L}} x\big(x \in K_M \to \forall_{\mathbb{N}} k \exists_{\mathbb{N}} j\, \mathrm{Ctl}_M(x,j,k)\big)\\
&\qquad \to \forall_{\mathbb{N}} k \exists_{\mathbb{N}} j \forall_{\mathscr{L}} x\big(x \in K_M \to \mathrm{Ctl}_M(x,j,k)\big)\Big).
\end{aligned}$$

Here K ranges over presented sets and c over ratio covers, and M over presented uniform maps in $\mathsf{PLB}^{=}$ and $\mathsf{PINF}^{=}$ and over map packages in $\mathsf{HC}^{=}$. The formula $\mathrm{Inf}_M(m)$ is

$$\forall_{\mathscr{L}} x\big(x \in K_M \to m \le f_M(x)\big) \wedge \forall_{\mathscr{L}} a\Big(\forall_{\mathscr{L}} x\big(x \in K_M \to a \le f_M(x)\big) \to a \le m\Big),$$

which says that m is the infimum of f_M on K_M, and

$$\mathrm{Ctl}_M(x,j,k) :\equiv \forall_{\mathscr{L}} y\big(y \in K_M \to x \approx_j y \to f_M(x) \approx_k f_M(y)\big),$$

so that the premise and the conclusion of $\mathsf{HC}^{=}$ are pointwise and uniform continuity on K_M (Definition 13.12). The forms for $[0,1]$ alone take c over ratio covers and f over the functions

on the line, $\mathscr{L}^{\mathscr{L}}$:

$$\begin{aligned}
\mathsf{HB}^{=}_{[0,1]} &:= \forall c\Big(\forall_{\mathscr{L}} x\big(x \in [0,1] \to \exists_{\mathbb{N}} k\, x \in I_k(c)\big) \\
&\qquad \to \exists_{\mathbb{N}} N \forall_{\mathscr{L}} x\big(x \in [0,1] \to \exists_{\mathbb{N}} k (k < N \wedge x \in I_k(c))\big)\Big), \\
\mathsf{PLB}^{=}_{[0,1]} &:= \forall f\Big(\mathrm{U}(f) \to \forall_{\mathscr{L}} x\big(x \in [0,1] \to 0 < f(x)\big) \\
&\qquad \to \exists_{\mathbb{N}} k \forall_{\mathscr{L}} x\big(x \in [0,1] \to \eta(1/(k+1)) \le f(x)\big)\Big), \\
\mathsf{PINF}^{=}_{[0,1]} &:= \forall f\Big(\mathrm{U}(f) \to \forall_{\mathscr{L}} x\big(x \in [0,1] \to 0 < f(x)\big) \to \exists_{\mathscr{L}} m\,\big(\mathrm{Inf}_f(m) \wedge 0 < m\big)\Big), \\
\mathsf{HC}^{=}_{[0,1]} &:= \forall f\big(\mathrm{C}(f) \to \mathrm{U}(f)\big),
\end{aligned}$$

where $\mathrm{Inf}_f(m)$ is $\mathrm{Inf}_M(m)$ with $x \in [0,1]$ in place of $x \in K_M$ and f in place of f_M, and the uniform and the pointwise continuity of f on $[0,1]$ are

$$\begin{aligned}
\mathrm{U}(f) &:\equiv \forall_{\mathbb{N}} k \exists_{\mathbb{N}} j \forall_{\mathscr{L}} x\Big(x \in [0,1] \to \forall_{\mathscr{L}} y\big(y \in [0,1] \to x \approx_j y \to f(x) \approx_k f(y)\big)\Big), \\
\mathrm{C}(f) &:\equiv \forall_{\mathscr{L}} x\Big(x \in [0,1] \to \forall_{\mathbb{N}} k \exists_{\mathbb{N}} j \forall_{\mathscr{L}} y\big(y \in [0,1] \to x \approx_j y \to f(x) \approx_k f(y)\big)\Big).
\end{aligned}$$

None of these sentences contains an atom. By Lemma 16.31, each of them has in the model of named objects the reading of Lemma J.2: the statement of Chapter 13 with every quantifier restricted to the named objects. Erasing the presentations, the later reading of each of the seven sentences $\mathsf{FAN}^{=}_{\Delta}$, $\mathsf{FAN}^{=}_{c}$, $\mathsf{WKL}^{=}$, $\mathsf{HB}^{=}$, $\mathsf{PLB}^{=}$, $\mathsf{PINF}^{=}$ and $\mathsf{HC}^{=}$ is the corresponding statement of Chapter 13 for all presented sets, covers and maps of ordinary mathematics.

Theorem 16.32 (Fan theorems and weak König's lemma)**.** *Let X be one of $\mathsf{FAN}^{=}_{\Delta}$, $\mathsf{FAN}^{=}_{c}$ and $\mathsf{WKL}^{=}$. In the computable calculus none of $\top \vdash X$, $\mathsf{LPO} \vdash X$ and $\top \vdash \mathsf{LPO} \to X$ is derivable, and the same holds when every instance of excluded middle is added as a base rule. The proof reasons in a classical metatheory.*

Proof. LPO is true in the model of named objects (proof of Theorem 16.27), so by Theorem 16.22, and by Remark 16.24 when excluded middle is added, it is enough to show that each sentence is false there. In the model, quantifiers over binary sequences and over decidable sets of strings range over the computable ones, and Hit, CHit, Tree and Inf read as the relations they abbreviate (Lemmas 16.31 and J.2).

$\mathsf{FAN}^{=}_{\Delta}$. Let $D(l) = 1$ exactly when $l \notin T_K$. By Lemma 16.29(a) D is computable, and by (d) every computable sequence has an initial segment with $D = 1$. Fix N. By (c) some string g of length N lies in T_K; let p be g followed by zeros, a computable sequence. Then $\bar{p}(N) = g \in T_K$, and by (b) every $\bar{p}(n)$ with $n \le N$ lies in T_K. So no N bounds the bar on the computable sequences, and the sentence is false.

$\mathsf{FAN}^{=}_{c}$. With the same D, the c-set of D contains every string with $D = 1$: a string outside T_K has all its extensions outside T_K, by (b). So every computable sequence meets the c-set. A uniform bound for the c-set would, at the empty extension, be a uniform bound for D, which was just refuted.

$\mathsf{WKL}^{=}$. T_K is a named decidable set of strings, a tree by (b), with a member of every length by (c). A path in the model is computable and leaves T_K by (d). □

Proposition 16.33 (The reading of weak König's lemma)**.** *Erasing the presentations of* $\mathsf{WKL}^{=}$ *gives a sentence whose later reading is weak König's lemma for decidable binary trees. That statement is equivalent to* $\mathrm{LLPO}_{\mathrm{seq}}$ *(Definition 13.8), and neither direction uses a choice principle.*

Proof. The first assertion is the unfolding of the clauses, all carriers being constant after erasure; at the later stage the abbreviations Hit, Tree and Inf read as the relations they abbreviate, with every string and every sequence in the range of the quantifiers. The equivalence is Proposition 13.9. □

So weak König's lemma, written with equations, is not derivable in the computable calculus even from LPO and with excluded middle as a base rule, while its later reading is exactly $\mathrm{LLPO}_{\mathrm{seq}}$: LLPO for infinitely many rows at once, with the answers collected into one sequence. With LPO the computable calculus has the decisions and lacks the collection.

The compactness sentences fail on the unit interval, presented as in Chapter 13; its package is named (Appendix J.6). We use the base-four geometry of the proof of Theorem 13.6: for a string s of length n, the number $\kappa(s)$ with base-four digits $3s_0, \dots, 3s_{n-1}$, the Cantor point $c(s) = \kappa(s)/4^n$ in $[0,1]$, the Cantor interval $J(s)$, whose left end is $c(s)$, and the open cells $C(s)$ with ratio endpoints. Three of the facts proved there are used:

(C1) if $|t| \geq |s|$ and $c(t) \in C(s)$, then s is an initial segment of t (fact (b));

(C2) for every regular sequence r naming a point x of $[0,1]$, the *address* a_r, whose digit at each depth compares one stage of r with a ratio, has $x \in C(\bar{a}_r(n))$ for every n (fact (c));

(C3) if r is computed by a program, so is a_r (Appendix J.6).

Fix an enumeration $w_0, w_1, \dots$ of the binary strings. The *dead cover* of T_K has as its k-th interval the empty interval $(0,0)$ if $w_k \in T_K$, and the cell $C(w_k)$ if $w_k \notin T_K$. Its endpoints are ratios computed from k, so the cover is named.

Theorem 16.34 (Heine–Borel for presented sets)**.** *Let X be* $\mathsf{HB}^{=}$ *or* $\mathsf{HB}^{=}_{[0,1]}$*. In the computable calculus none of $\top \vdash X$,* $\mathsf{LPO} \vdash X$ *and* $\top \vdash \mathsf{LPO} \to X$ *is derivable, and the same holds when every instance of excluded middle is added as a base rule. The proof reasons in a classical metatheory.*

Proof. As before, it is enough to show that both sentences are false in the model, at the dead cover and, for $\mathsf{HB}^{=}$, at the named presented set $[0,1]$. By Lemma 16.31, a named point is a member of the presented set $[0,1]$ exactly when $0 \leq x \leq 1$, and it lies in the k-th interval of the cover exactly when it lies between the two ends; so at this cover both sentences say that finitely many of its intervals cover the named points of $[0,1]$.

The dead cover covers every named point of $[0,1]$. Let $x \in [0,1]$ be named, by a program for a regular sequence r. By (C3) the address a_r is computable, so by Lemma 16.29(d) some initial segment $u = \bar{a}_r(n)$ lies outside T_K. By (C2), $x \in C(u)$, and $C(u)$ is the interval of the index of u, a number and so named.

No initial part of the cover suffices. Fix N and put $L = \sum_{k<N}(|w_k| + 1)$, which exceeds $|w_k|$ for every $k < N$. By Lemma 16.29(c) some string t of length L lies in T_K. Its Cantor point $c(t)$ is

a ratio in $[0,1]$, so it is named (Proposition 16.21(b)). Suppose $c(t)$ lay in the interval $k < N$. That interval is not empty, so $w_k \notin T_K$ and $c(t) \in C(w_k)$. Since $|w_k| < L = |t|$, (C1) makes w_k an initial segment of t, and $t \in T_K$ gives $w_k \in T_K$ by Lemma 16.29(b). This contradiction shows that $c(t)$ lies in none of the first N intervals. □

The two positivity statements reduce to Heine–Borel inside the model, through the cover series f_c of Lemma 13.26. That lemma shows that f_c is positive at every point that lies in an interval of c; that $f_c \geq 1/(k+1)$ on a set of points makes the first N intervals cover that set for some N; and that f_c, with the transformer of names it describes and a modulus, is a presented uniform map on every set of points. When the cover and the presented set are named, so is this presented uniform map, with a modulus computed by a program (Appendix J.7).

Theorem 16.35 (Positive lower bounds and positive infima)**.** *Let X be one of $\mathsf{PLB}^{=}$, $\mathsf{PINF}^{=}$, $\mathsf{PLB}^{=}_{[0,1]}$ and $\mathsf{PINF}^{=}_{[0,1]}$. In the computable calculus none of $\top \vdash X$, $\mathsf{LPO} \vdash X$ and $\top \vdash \mathsf{LPO} \to X$ is derivable, and the same holds when every instance of excluded middle is added as a base rule. The proof reasons in a classical metatheory.*

Proof. It is enough to show that the four sentences are false in the model. By Lemma 16.31, $0 < y$, $a \leq b$ and membership read as the relations they abbreviate.

$\mathsf{PLB}^{=}$. Suppose it were true. Let c be the dead cover of T_K and consider the named presented uniform map f_c on the presented set $[0,1]$. By the first part of the proof of Theorem 16.34, every named point of $[0,1]$ lies in an interval of c, so f_c is positive there. The sentence gives k with $1/(k+1) \leq f_c(x)$ for every named $x \in [0,1]$. Applied to the set of named points of $[0,1]$, the lemma gives an N such that the first N intervals cover every named point of $[0,1]$, which the second part of that proof refutes.

$\mathsf{PLB}^{=}_{[0,1]}$. The same argument applies to f_c as a function on the line. It is named, and its modulus $\mu(k) = 2k+1$ (Appendix J.7) makes $\mathrm{U}(f_c)$ true in the model.

$\mathsf{PINF}^{=}$. Suppose it were true; we show that $\mathsf{PLB}^{=}$ would be true. Let M be a named presented uniform map, positive at the named points of K_M. If K_M has no named point, every k is a lower bound on its named points. Otherwise $\mathsf{PINF}^{=}$ gives a named m that is the infimum of f_M on the named points of K_M, with $0 < m$, which gives k with $1/(k+1) \leq m$; so $1/(k+1) \leq m \leq f_M(x)$ there. The case distinction on whether K_M has a named point is made in the metatheory. So $\mathsf{PLB}^{=}$ would hold, which was just refuted.

$\mathsf{PINF}^{=}_{[0,1]}$. In the same way it would give $\mathsf{PLB}^{=}_{[0,1]}$, with no case distinction, since $[0,1]$ has named points. □

Heine–Cantor needs a function rather than a cover. We use the tents of the proof of Theorem 13.13: for a string s of length n, the centre $g(s)$ of its gap, at distance $2^{-(2n+1)}$ from $c(s)$, and the tent $\tau_s(x) = \max(0, 1 - 8 \cdot 4^n |x - g(s)|)$, which vanishes outside the overlap $O(s)$ of the cells of the children of s. The *finite heads* of T_K are $H_N = \sum\{\tau_s : s \in T_K,\ |s| < N\}$. The *tent function* τ of T_K takes at a point x the common value of the heads $H_N(x)$ from some N on, when they become constant, and the value 0 when they do not; this definition by cases is made in the metatheory.

Theorem 16.36 (Heine–Cantor for presented sets)**.** *Let X be $\mathsf{HC}^{=}$ or $\mathsf{HC}^{=}_{[0,1]}$. In the computable calculus none of $\top \vdash X$, $\mathsf{LPO} \vdash X$ and $\top \vdash \mathsf{LPO} \to X$ is derivable, and the same holds when every instance of excluded middle is added as a base rule. The proof reasons in a classical metatheory.*

Proof. We show that $\mathsf{HC}^{=}$ is false in the model at the map package of $[0,1]$ and τ, and $\mathsf{HC}^{=}_{[0,1]}$ at the function τ on the line. By Lemma 16.31, $\approx_j$ reads as closeness within $1/(j+1)$ and membership as membership. Three facts are proved in Appendix J.8.

τ is named. A program computes a name of $\tau(x)$ from every name of every point x of the line: it runs the address digits of the given regular sequence and searches for the first depth n at which the address string lies outside T_K. The search ends on every name, because the address of a computable regular sequence is computable and leaves T_K by Lemma 16.29(d). At that depth all heads from n on agree at x, so $\tau(x) = H_n(x)$, and a name of $H_n(x)$ is computed from n and the name of x by finite arithmetic.

τ is continuous at every named point of $[0,1]$. Such a point lies in the open cell of a string outside T_K, and on that cell τ equals one finite head, a finite sum of tents, which changes by a bounded multiple of the change of its argument.

τ is not uniformly continuous on the named points of $[0,1]$. For every j, a string $s \in T_K$ of length $j+1$ gives the two ratios $c(s)$ and $g(s)$ of $[0,1]$, which are named and lie within $2^{-(2j+3)} \le 1/(j+1)$ of each other, while $\tau(c(s)) = 0$ and $\tau(g(s)) = 1$. These values are not within $1/2$, so at $k = 1$ no j serves.

So the premise of each sentence holds at τ and its conclusion fails. LPO and every instance of excluded middle are true, and Theorem 16.22 with Remark 16.24 finishes the proof. □

These results say that the computable calculus does not derive compactness, written with equations, even with LPO and with excluded middle as a base rule. They do not say that compactness fails. In ordinary mathematics with excluded middle and countable choice, weak König's lemma holds (Theorem 12.72), and with it the decidable fan theorem and Heine–Borel for presented sets (Theorems 13.7 and 13.25); LPO with unique choice already gives the decidable fan theorem (Corollary 13.10); and the full calculus derives all seven sentences for presented sets outright (Proposition 16.15). What the model shows is what a derivation in the computable calculus would need and that calculus lacks: a way to gather, along infinitely many strings or points, answers that LPO supplies one at a time.

16.11 What remains open

A model with two later stages. Every separation of this chapter uses one model, in which truth is classical and the quantifiers range over named objects. A separation of a different kind, such as one that keeps a compactness statement while refuting weak excluded middle, $\neg p \vee \neg\neg p$, needs a model in which a truth value can be undecided. The natural candidate is a finite Kripke structure with a root and two incomparable later stages. To use it for the computable calculus one needs an interpretation of the calculus in that structure that commutes with the ten constructions of formulas, is coherent with the binders, makes every instance of the twenty-nine rules an entailment, and reads the two sentences to be separated as their

counterparts in the structure. Whether such an interpretation exists is open. Two natural attempts fail. The model of named objects, copied to all three stages, meets the first three requirements, but it makes weak excluded middle true. And a classifier at the root built from one truth value at each branch can describe only the sets $\{(p, q) : L(p) \text{ and } R(q)\}$, while excluded middle at the root, true at (true, true) and at (false, false) and false at (true, false), is not of that form: L and R would have to hold at both truth values.

Derivations of the compactness implications. Chapter 13 proves in ordinary mathematics that Heine–Borel for presented sets, the positive lower bound, the positive infimum and presented Heine–Cantor each imply the decidable fan theorem. In the full calculus the erased implications from $\mathsf{HB}^{=}$, $\mathsf{PLB}^{=}$, $\mathsf{PINF}^{=}$ and $\mathsf{HC}^{=}$ to $\mathsf{FAN}_{\Delta}^{=}$ are derivable (Proposition 16.15), as is $\mathsf{FAN}_{\Delta}^{=}$ itself. Whether the computable calculus derives $\top \vdash X \to \mathsf{FAN}_{\Delta}^{=}$ for each X among $\mathsf{HB}^{=}$, $\mathsf{PLB}^{=}$, $\mathsf{PINF}^{=}$ and $\mathsf{HC}^{=}$ is open. The model of named objects cannot settle it: both sides are false there, so each implication is true there.

The full calculus. Proposition 16.15 settles the full calculus for every sentence that is equational over constant carriers: it derives each one that is true in ordinary classical mathematics, so between two such sentences it proves no separation that classical mathematics does not already give. A separation of another kind in the full calculus would have to concern sentences outside this class: sentences with atoms, or with quantifiers over carriers that change between the stages, such as Ω and the sets of points.

Realizability. Proposition 16.26 rules out uniform realizability over presentations in which a name says nothing about a truth value. Whether presentations whose names of truth values carry evidence give a realizability model of the computable calculus is open.

Native prices. $\mathrm{LLPO}_{\mathrm{seq}}$ is not written here as a sentence of the calculus, and the equivalence of Proposition 16.33 is proved in ordinary mathematics, not as a derivation.

Questions to work through

1. In Table 16.1, drop the second half of the earlier clause for implication. Find a formula in the context Ω whose reading is then no longer a predicate, and point to the step of Lemma 16.5 that fails.

2. Show that conditions (1) and (2) of Theorem 16.14 can be met together. Which instance of condition (3) does your assignment violate?

3. Show that a judgment $A \vdash B$ between sentences is derivable exactly when $\top \vdash A \to B$ is, in the calculus and in the computable calculus.

4. Use Proposition 16.21(c) to show that no tracked map from $\Omega^{\mathbb{N}}$ to $\mathbb{N}^{\mathbb{N}}$ sends every classifier on $\mathbb{N}$ to its 0–1 function, the function with value 1 exactly where the classifier holds. Where does the proof of Theorem 16.27 meet the same obstacle?

5. In the proof of Theorem 16.32, the path p made of a member g of T_K of length N followed by zeros has no initial segment of length at most N outside T_K. Explain why p nevertheless leaves T_K, and bound the length of its first initial segment outside T_K in terms of the program given by the recursion theorem.

Part V

Mathematics and finite physical records

Part V connects finite mathematical descriptions with finite physical records. Its theorems concern stated mathematical models of an apparatus: a record space and a reader (Chapter 17); finite operations with a supplied preparation, externally selected controls and a complete reader (Chapter 18); and an interaction that conserves two stated charges, read by a neutral sharp pointer (Chapter 19). Every physical reading of these results is relative to those listed hypotheses. A theorem says what follows when an apparatus satisfies them; that a particular device satisfies them is an identification, which the theorem does not supply. These chapters reason in classical mathematics, about worlds and recorders as about finite operations and conserved charges.

Two results should be read at exactly this scope. Theorem 18.11 approximates every finite instrument, within any tolerance on every retained reference, with externally switched positive-energy pulses; it is not a claim that an autonomous apparatus with bounded controls exists. Theorem 19.5 is a lower bound on the error of every recorder that meets its four conditions; it is not a claim that the bound is attained. Each theorem of Chapters 17 to 19 is followed by a remark that lists its conditions, preparation, controls, resources and identifications, as they apply, and what it does not claim.

Chapter 17

What a finite record can establish

An instrument returns one of two marks. The experimenter wants to know which of three preparations was supplied. If that mark is the whole available record, a reader cannot always give the right answer: two preparations must leave the same mark, and the reader then receives the same information in two cases requiring different answers.

The conclusion changes if the question changes. Perhaps the experimenter only needs to separate one preparation from the other two. Or perhaps the apparatus can perform another test. A limit on recognition must therefore name both the question and the record. Counting possible preparations alone does not tell us which of their differences an observer must recover.

This chapter gives that distinction a precise form. We first prove the collision argument for an arbitrary record. We then construct the record of an adaptive testing procedure and count its possible outcomes. Finally, we replace exact answers by answers with a stated error, and Section 17.7 asks which questions a growing record can confirm, refute or settle, and what deciding them costs. Each argument concerns the information supplied to the reader, except the cost results of that section, which concern the logical principles that a decision requires; physical preparation, retention and readout enter as separate requirements on the apparatus.

17.1 The target, the record and the reader

Let W be the allowed preparations or worlds. A member $w \in W$ contains whatever distinctions the experiment can depend on. We need not reconstruct all of w. Instead, fix a target map

$$\tau : W \longrightarrow Y.$$

The value $\tau(w)$ is the answer the task requires. Two worlds may differ while having the same target value.

Let R be the possible retained records and let $r : W \to R$ give the record actually obtained in each world for the procedure under consideration. A reader d takes a record and either returns an element of Y or reports failure. We write its answer as $d(r)$ when it succeeds.

Definition 17.1 (Exact readout)**.** The procedure gives an exact readout of τ if the reader succeeds on every actual record and

$$d(r(w)) = \tau(w) \qquad \text{for every } w \in W. \tag{17.1}$$

Records that never occur may be rejected.

The order of these maps matters. The apparatus acts on the world to produce a record; the reader acts on the record to return the target. An arbitrary assignment of a different code to every answer does not establish that the apparatus can obtain the assigned code from an unknown preparation. Equation (17.1) is the required connection.

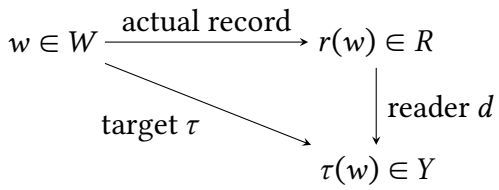


Figure 17.1. Exact readout requires the two routes to agree for every allowed world. A mathematical code supplies the upper route only when the apparatus actually produces it.

Lemma 17.2 (A record cannot separate its own collisions)**.** *If the readout is exact and $r(w) = r(v)$, then $\tau(w) = \tau(v)$.*

Proof. Apply the same reader to the equal records. Exactness identifies its answer on one side with $\tau(w)$ and on the other with $\tau(v)$. Thus

$$\tau(w) = d(r(w)) = d(r(v)) = \tau(v).$$ □

Only distinctions relevant to the target must survive. If the target itself separates all worlds, the lemma says that the record map must be injective. Otherwise, several worlds can share a record without causing an error, provided their required answers agree.

17.2 Several worlds can remain compatible with one record

The map r describes one determined record per world. Some experiments have several possible outcomes, and some records leave several preparations possible. We can state the essential condition without choosing a probability distribution or a deterministic outcome map.

Write $C(s, w)$ when record s is compatible with world w under the stated experimental model. The condition includes the experiment that was performed and every relevant control or calibration assumption. A reader is sound on this relation if, whenever $C(s, w)$ holds, it succeeds and returns $\tau(w)$.

Proposition 17.3 (A sound answer is constant on compatible worlds)**.** *Suppose the reader is sound. If $C(s, w)$ and $C(s, v)$, then $\tau(w) = \tau(v)$. Moreover, if an actual record map satisfies $C(r(w), w)$ for every w, soundness gives exact readout for that map.*

Proof. Soundness says that the single answer $d(s)$ equals both target values, which proves the first assertion. For the second, apply soundness to $s = r(w)$ and the stated actual-world compatibility condition. □

This formulation separates two possible failures. A proposed record may be compatible with worlds requiring different answers. Or the compatibility model may omit the world that actually produced the record. The first failure concerns discrimination; the second concerns the model's relation to the experiment. A proof of the first condition does not repair the second.

The proposition is a necessary condition on a sound reader. It does not construct an effective reader from an arbitrary compatibility relation. Finding the compatible worlds, deciding their common target and making the record readable can each require further work.

17.3 A finite capacity bound

Choose a finite family of allowed worlds w_i, indexed by a finite set I, whose target values are pairwise different. The family need not include every allowed world. Suppose the whole retained record has only M possible values.

Theorem 17.4 (Capacity required by exact recognition)**.** *If the readout is exact on the allowed worlds, then*

$$|I| \leq M.$$

Conditions and scope.

Preparation and record
A finite family of allowed worlds with pairwise different targets, and a record map r giving the actual record in each. The record space must contain everything the reader uses: timing, channel, control settings and the printed mark.

Resources counted
The number M of possible values of the whole retained record.

Identification
Applying the theorem to an apparatus identifies its complete output with r. The theorem itself concerns the map r.

Not claimed
That any apparatus reaches the records, or that M distinct records suffice for a construction.

Proof. Associate i with its actual record $r(w_i)$. If two indices have the same record, Lemma 17.2 gives the same target value. The choice of the family then gives the same index. We have constructed an injection from I into the set of possible records, so the finite cardinal inequality follows. □

The bound concerns distinguishable records, including everything the reader may use. If timing, position, the choice of output channel or a separate control register supplies additional information, it belongs in R. Counting only the printed mark would then count an incomplete record.

No optimality claim is built into the theorem. An injection into a large record space is necessary for exact recovery; it does not show how a particular apparatus reaches the needed records. The construction of a working procedure is a separate positive task.

17.4 Adaptive tests still produce a bounded transcript

An experimenter may choose the next test after seeing the previous result. To count this procedure correctly, we must preserve that dependence. Fix a set A of available tests and an alphabet of $B \geq 1$ possible outcomes. For this section, a test a has a determined outcome

$$o(a, w) \in \{0, \dots, B-1\}$$

in each world w. Fix the procedure in advance. It may adapt to the outcomes, but it receives no information about w other than through those outcomes.

Definition 17.5 (A bounded adaptive test tree)**.** A tree with budget m has one of two forms:

1. Stop and return a specified answer, using none of the remaining tests.
2. If $m > 0$, choose a test $a \in A$ and, for each of its B outcomes, supply a continuation tree with budget $m - 1$.

Execution follows the continuation selected by the observed outcome. Write $\mathrm{run}(T, w)$ for the answer eventually returned by tree T.

The budget decreases at every test, so every execution stops. Different branches can choose different tests and stop at different depths. This is a finite inductive description of the whole policy, rather than a list of independent tests to which adaptation is added afterward.

Choose one padding symbol p from the outcome alphabet. We now construct a transcript $c_T(w)$ with exactly m slots. At a stop node, fill all remaining slots with p. At a test node, write the observed outcome in the first slot and follow it with the continuation's transcript.

The corresponding reader already knows T. At a stop node it returns that node's answer and ignores the remaining slots. At a test node it reads the first symbol, chooses the corresponding continuation, and reads the tail there. Denote this reader by D_T.

Lemma 17.6 (The transcript retains the answer of the procedure)**.** *For every bounded tree and every allowed world,*

$$D_T(c_T(w)) = \mathrm{run}(T, w).$$

Proof. Induct on the tree. At a stop node, both sides are its specified answer. At a test node, the first transcript symbol is $o(a, w)$. The reader therefore chooses exactly the continuation chosen in execution. The remaining transcript is the transcript of that continuation. Apply the induction hypothesis to its smaller tree. □

Padding loses no information needed for this answer. The fixed tree tells the reader when a stop node has been reached. Even if a padding symbol is also a genuine test outcome, the reader knows which slots it is supposed to read from the branch it has followed.

Theorem 17.7 (Capacity of a bounded adaptive procedure)**.** *Suppose a tree performs at most m tests with B possible outcomes per test and returns the correct target for every world in a finite family of N distinct target values. Then*

$$N \le B^m. \tag{17.2}$$

Conditions and scope.

Preparation and control
Each test has a determined outcome in each world, and the adaptive policy is fixed in advance; it learns about the world only through the outcomes.

Resources counted
At most m tests, each with B possible outcomes.

Not claimed
Tests that change the system are covered only after their changing state and the available controls are modelled. The bound B^m is a necessary capacity, not a construction.

Proof. There are B^m strings of length m over the outcome alphabet: each of the m slots has B choices. Lemma 17.6 and correctness of the tree give exact readout from its constructed transcript. Distinct target values therefore give distinct transcripts; apply Theorem 17.4. □

In particular, no such procedure with one binary test can always identify one of three distinct targets: the conclusion would require $3 \le 2$. Adaptation helps choose useful questions, but it cannot make a transcript carry more distinct answers than it has distinguishable values.

This tree model assumes determined outcomes $o(a, w)$ for the supplied world. When a test changes the system, its changing state and the available controls must be represented before applying this construction. The earlier collision and record-capacity theorems apply to any supplied actual record map with exact recovery; the tree theorem adds the particular structure that justifies the bound B^m.

17.5 An error tolerance gives a separation bound

Many tasks ask for an approximation. Let target values lie in a space with a distance dist satisfying symmetry and the triangle inequality. Let the reader return a reconstructed value $d(s)$, and fix an error allowance $\varepsilon \ge 0$. Suppose

$$\operatorname{dist}(\tau(w), d(r(w))) \le \varepsilon \qquad \text{for every allowed } w. \tag{17.3}$$

Lemma 17.8 (A shared record bounds target separation)**.** *If $r(w) = r(v)$, then*

$$\operatorname{dist}(\tau(w), \tau(v)) \le 2\varepsilon.$$

Proof. Both worlds are reconstructed as the same value $y = d(r(w)) = d(r(v))$. The triangle inequality and the two error guarantees give

$$\operatorname{dist}(\tau(w), \tau(v)) \le \operatorname{dist}(\tau(w), y) + \operatorname{dist}(y, \tau(v)) \le 2\varepsilon. \qquad \square$$

Theorem 17.9 (Capacity for separated approximate targets)**.** *Suppose a finite family has N target values, every distinct pair separated by more than 2ε. A readout satisfying* (17.3) *on this family requires at least N distinguishable records. For a bounded adaptive tree as above, it requires $N \le B^m$.*

Conditions and scope.

Conditions
A distance with symmetry and the triangle inequality, an error allowance ε met in every allowed world, and a family whose targets are separated by more than 2ε.

Not claimed
An optimal code, or a geometry beyond the stated distance.

Proof. If two members had the same record, Lemma 17.8 would contradict their strict separation. The record map is therefore injective on the selected family, giving the first bound. For the tree, use its padded transcript and reader. Lemma 17.6 transfers the procedure's error guarantee to this reader, and there are B^m possible transcripts. □

The strict inequality is the relevant margin. An approximation is permitted to merge sufficiently nearby targets. The theorem counts a family whose members are far enough apart that the specified error cannot hide their differences. It supplies a necessary capacity, without assuming a geometry or an optimal code beyond the stated distance and separation.

17.6 Finite length without a common budget

A different question arises when every individual record is finite but there is no single length bound for all records. Such a record family need not be finite. Its binary strings can still be encoded by natural numbers. Define

$$c(\varnothing) = 0, \qquad c(0s) = 2c(s) + 1, \qquad c(1s) = 2c(s) + 2. \tag{17.4}$$

Here $0s$ and $1s$ prepend a symbol to the finite string s.

Lemma 17.10 (Finite binary records have distinct natural-number codes)**.** *If $c(s) = c(t)$, then $s = t$.*

Proof. Induct on s. The empty string has code zero; every nonempty string has positive code, so equality with the empty string forces the other string to be empty. For two nonempty strings, a leading zero gives an odd code and a leading one gives an even code. Equal codes therefore have the same leading symbol. Subtract its offset and cancel the factor two to obtain equal tail codes. The induction hypothesis gives equal tails, hence equal strings. □

Theorem 17.11 (Exact finite-record identification is countable)**.** *Suppose every world yields a finite binary record, a reader exactly recovers its target, and different worlds have different target values. Then the world family injects into the natural numbers and is, classically, countable.*

Conditions and scope.

Conditions
Every world yields a finite binary record; a reader recovers the target exactly; different worlds have different targets.

Not claimed
A uniform length or time bound, or an apparatus that acquires the records. The identification of the physically real collections with those whose members leave exact finite records is not assumed here; the theorem says what follows for a collection that satisfies the conditions.

Proof. Lemma 17.2 makes the record map injective under the target-distinctness assumption. Compose that map with the injective code of Lemma 17.10. The result is an injection $w \mapsto c(r(w))$ into the natural numbers. □

No uniform length bound was used. Conversely, countability alone does not give a uniform record length, a time bound or an apparatus that can acquire the record. This is why the fixed-budget theorem and the countability theorem answer different questions.

The hypothesis is exact identification of the stated target family. A finite record can support a universal mathematical statement when it contains a proof whose rules justify that statement. The record-capacity argument does not count the objects mentioned in a proposition as though the reader had to identify each of them separately. It concerns a task whose different required answers must be distinguished by actual records.

17.7 What a record can confirm, refute or settle

A detector is examined once a second and reports 1 when it has fired and 0 otherwise. Two questions can be put to it. Has it ever fired? A record that ends in a 1 confirms this at once, but no record of zeros refutes it, since the next examination may fire. Will it never fire? Now a single 1 refutes the claim, and no finite record confirms it. Deciding the first question for every binary sequence is LPO and deciding the second is WLPO (Definition 7.8). That the two decisions are two of the omniscience principles is no accident. This section shows which questions finite records can confirm, refute or settle, and what deciding them costs, for any system that is examined again and again.

Worlds, records and recorders

Let W be a set of worlds, and let each world w yield a sequence of examination results $w(0), w(1), \ldots$ in a set A. The *record* of length n is $w{\restriction}n = (w(0), \ldots, w(n-1))$; two worlds *agree up to* n when their records of length n are equal. A *question* is a property Q of worlds. In the detector example W is the set of binary sequences, each read by itself; we call these worlds *binary runs*.

Definition 17.12 (Recorders)**.** A *recorder* is a rule R that assigns to every finite record either an answer, *yes* or *no*, or the mark $\circ$, meaning "read on". Its *answer on* w is $R(w{\restriction}n)$ for the least n at which this is not $\circ$, if there is one. R *confirms* Q when, for every world, $Q(w)$ holds exactly when the answer on w is *yes*; it *refutes* Q when $\neg Q(w)$ holds exactly when the answer on w is *no*; it *settles* Q when it does both. Q is *confirmable*, *refutable* or *settled* when some recorder confirms, refutes or settles it.

Q is Σ_1 when there is a set S of records with $Q(w) \Leftrightarrow \exists n\,(w{\restriction}n \in S)$, Π_1 when there is one with $Q(w) \Leftrightarrow \forall n\,(w{\restriction}n \in S)$, and Δ_1 when it is both. Q is *open* when every world w with $Q(w)$ has an n such that every world agreeing with w up to n satisfies Q, and *closed* when $\neg Q$ is open.

A recorder is any rule. It need not be computable, so the negative results below hold against unlimited computing power. We reason classically about worlds and recorders, as Section 7.12 reasons about its models; the cost theorems further on are proved without excluded middle or choice and say so.

Theorem 17.13 (Records confirm the open questions)**.** *For every set of worlds with its examinations and every question Q: Q is confirmable if and only if it is Σ_1, if and only if it is open; Q is refutable if and only if it is Π_1, if and only if it is closed; and Q is settled if and only if it is Δ_1, if and only if it is both open and closed.*

Conditions and scope.

Conditions
Any worlds and any question; recorders are arbitrary rules on finite records. The proof is classical: excluded middle decides whether a record forces Q. Applied to an experiment, the examinations are its successive outcomes and a recorder is its stopping rule.

Not claimed
That the constructed recorder can be carried out by a device, or any bound on the number of examinations it needs.

Proof. Say that a record r of length n *forces* Q when every world w with $w{\restriction}n = r$ satisfies Q. The *forcing recorder* answers *yes* at a record that forces Q, *no* at one that forces $\neg Q$, and $\circ$ otherwise; a record of an actual world cannot force both.

Let R confirm Q, and let $Q(w)$ hold, so that R answers *yes* first at $w{\restriction}n$. A world agreeing with w up to n has the same records up to that length, so R answers *yes* on it as well, and it satisfies Q. So Q is open. Conversely, let Q be open. If $Q(w)$, some record of w forces Q, so the forcing recorder answers on w, and its answer is *yes*, since no record of w forces $\neg Q$. If its answer on w is *yes*, a record of w forces Q, so $Q(w)$. Hence the forcing recorder confirms Q. If Q is Σ_1 with the set S, it is open, because agreeing worlds share their records; if Q is open, it is Σ_1 with S the set of records that force Q.

The three statements about refutation follow by applying these to $\neg Q$, with the answers exchanged and S replaced by its complement. A recorder that settles Q confirms and refutes it, so Q is open and closed; if Q is open and closed, the forcing recorder, which is the same rule for Q and $\neg Q$, confirms and refutes it. □

The theorem is the characterization, in formal learning theory, of the hypotheses that a method can verify, refute or decide with certainty [64, Chapter 4].

On binary runs, "ever fires" is Σ_1 and open, and "never fires" is Π_1 and closed. Neither is settled: the run of zeros agrees up to n with the run whose only 1 is at place n.

The zero test and the event test

The two detector questions are the model cases: the zero test reduces to every question that no recorder confirms, and the event test to every question that no recorder refutes, in the following sense.

Definition 17.14 (Finite-information maps and reductions)**.** A map Φ from worlds W' to worlds W has *finite information* when each examination of $\Phi(w')$ is fixed by finitely many examinations of w': for every w' and k there is an n such that $\Phi(v')(k) = \Phi(w')(k)$ for every v' agreeing with w' up to n. The *zero test reduces to* Q when a finite-information map Φ from binary runs has $Q(\Phi(\alpha)) \Leftrightarrow \forall n\, \alpha(n) = 0$; the *event test reduces to* Q when one has $Q(\Phi(\alpha)) \Leftrightarrow \exists n\, \alpha(n) = 1$. LLPO *reduces to* Q when one has, for every α with at most one entry 1, that $Q(\Phi(\alpha))$ gives $\alpha(2k) = 0$ for every k and $\neg Q(\Phi(\alpha))$ gives $\alpha(2k+1) = 0$ for every k.

The price proofs of Chapter 12 have this form: the jump construction of Theorem 12.6 turns a binary sequence into a point whose stages depend on finitely many entries.

Theorem 17.15 (What a record cannot settle)**.** *For every set of worlds and every question Q: Q is confirmable exactly when the zero test does not reduce to it; Q is refutable exactly when the event test does not reduce to it; and Q is settled exactly when neither test reduces to it.*

Conditions and scope.

Conditions
Any worlds and any question; the reductions are arbitrary finite-information maps and need not be effective. The proof is classical and chooses, for each n, a world witnessing the failure of openness at n.

Not claimed
That a reduction can be computed from a description of Q.

Proof. First, openness pulls back along a finite-information map Φ: if Q is open and $Q(\Phi(w'))$, choose n with every world agreeing with $\Phi(w')$ up to n in Q, and then, by finite information for $k < n$, an m such that every v' agreeing with w' up to m has $\Phi(v')$ agreeing with $\Phi(w')$ up to n. On binary runs "all zero" is not open, as noted above. So if the zero test reduces to Q, then Q is not open, and not confirmable, by Theorem 17.13.

Conversely, suppose Q is not open. Then some w_0 with $Q(w_0)$ has, for every n, a world y_n that agrees with w_0 up to n and fails Q. Let $\Phi(\alpha) = w_0$ when α is all zero, and $\Phi(\alpha) = y_N$ when the first entry 1 of α is at place N. Then $Q(\Phi(\alpha)) \Leftrightarrow \forall n\, \alpha(n) = 0$. For finite information at examination k: if α has its first 1 at N, every run agreeing with α up to $N+1$ has the same image; if α is all zero, every run agreeing with it up to $k+1$ is all zero or has its first 1 at some $N' > k$, and then $y_{N'}(k) = w_0(k)$.

The event test reduces to Q exactly when the zero test reduces to $\neg Q$, with the same map; this gives the statement about refutation. By Theorem 17.13, a question is settled exactly when it is confirmable and refutable. □

These reductions correspond to those of formal learning theory [64, Chapter 5].

On binary runs, questions that a record settles are as simple as they can be.

Theorem 17.16 (Settled questions about binary runs are finite)**.** *A question Q about binary runs is settled if and only if there is a number N such that any two runs that agree up to N have the same answer.*

Conditions and scope.

Conditions
All binary runs as worlds, and any question about them. The passage from a settling recorder to the number N is the one place in this section where compactness enters, through Kőnig's lemma for the binary tree [15] in its classical form.

Not claimed
A computable bound N, or a bound for other sets of worlds: on the promised runs of Theorem 17.17 no bound exists.

Proof. If N exists, every record of length N forces Q or forces $\neg Q$, so Q is open and closed. Conversely, let R settle Q. Call a finite binary string *undecided* when R gives $\circ$ at it and at every initial piece of it. The undecided strings form a tree: every initial piece of an undecided string is undecided. If there were undecided strings of every length, the tree would be infinite, and Kőnig's lemma would give a run α all of whose records are undecided. Then R would have no answer on α, whereas a recorder that settles Q answers *yes* on α if $Q(\alpha)$ and *no* otherwise. So for some N no string of length N is undecided: R answers on every run by the record of length N, and runs agreeing up to N receive the same answer. □

What deciding a question costs

The cost statements below are implications between decisions and principles, proved without excluded middle or choice, like the prices of Chapter 12; the statements about recorders beside them are classical, as before, except the hypothesis of Theorem 17.17(1), which must be proved without excluded middle.

Theorem 17.17 (The binary Markov principle costs time, not information)**.**

1. *If a recorder R is given, and it is proved without excluded middle that R settles Q, then* $\mathrm{MP}_{\mathrm{bin}}$ *implies that $Q(w)$ or $\neg Q(w)$ holds for every world w.*

2. *That cost is attained. On the promised runs, the binary runs that are not all zero, the question "the first entry* 1 *is at an even place" is settled, and its decision for every promised run is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$.

3. *No recorder settles that question within a bound fixed in advance: there are no recorder R and number N such that R settles it and answers on every promised run by the record of length N.*

Conditions and scope.

Logic

Part 1 and the equivalence in part 2 use no excluded middle, no choice principle and no omniscience principle besides $\mathrm{MP}_{\mathrm{bin}}$: whether a recorder answers at a given record is decided. The hypothesis of part 1 must be constructive; a question that is open and closed only classically can cost more: for a binary sequence β, the constant question "β is all zero" is classically open and closed on any set of worlds, yet deciding it at a world for every β is equivalent to WLPO (Definition 7.8), which intuitionistic second-order arithmetic does not derive from $\mathrm{MP}_{\mathrm{bin}}$ (Table 7.2). Part 3 and the settledness in part 2 are proved classically.

Not claimed

That every settled question costs exactly $\mathrm{MP}_{\mathrm{bin}}$; a question whose truth value is decided outright costs nothing.

Proof. (1) Fix w, and let $a(n) = 1$ when R answers at $w{\restriction}n$ and $a(n) = 0$ otherwise. The sequence a is not all zero: if it were, R would have no answer on w, so $Q(w)$ would fail, since R confirms Q, and $\neg Q(w)$ would fail, since R refutes Q. $\mathrm{MP}_{\mathrm{bin}}$ gives n with $a(n) = 1$, and a search through $a(0), \dots, a(n)$ finds the first answer. If it is *yes*, then $Q(w)$; if it is *no*, then $\neg Q(w)$.

(2) The recorder that reads on until the first 1 and then answers with its parity settles the question, since a promised run has a first 1. Given $\mathrm{MP}_{\mathrm{bin}}$, a promised run has an entry 1, a search finds the first one, and its parity decides the question. Conversely, suppose the question is decided on every promised run, and let α be not all zero. Decide it for α and for the shifted run α' with $\alpha'(0) = 0$ and $\alpha'(n + 1) = \alpha(n)$, which is also promised. If either first 1 is at an even place, α has an entry 1. If neither is, then every $\alpha(k)$ is 0: if $\alpha(k) = 1$, a search finds the first 1 of α at some $N \leq k$, and N even contradicts the first decision while N odd puts the first 1 of α' at the even place $N + 1$. This contradicts the promise, so absurdity gives an entry 1 in this case too.

(3) Suppose R and N exist. The runs whose only 1 is at $N + 1$ and at $N + 2$ agree up to $N + 1$, so R answers on both by the record of length N, and with the same answer. Their first entries 1 have different parities. □

So a question settled by a given recorder, constructively, costs at most the binary Markov principle, the guarantee that a search which cannot fail succeeds, and a settled question can cost exactly that. A recorder enacts that search by waiting, and what the waiting costs is time without a bound. The stronger principles cannot be bought this way. The principle paid here is $\mathrm{MP}_{\mathrm{bin}}$; the restricted schema of Theorem 6.9 costs nothing under induction on every formula, because a quantifier-free body comes with a bound.

Proposition 17.18 (The other prices are not paid by records)**.**

1. *Deciding "all zero" for every binary run is WLPO, and deciding "some entry is* 1*" for every binary run is LPO.*
2. *No recorder has an answer on every binary run with at most one entry* 1 *such that the answer is* yes *only when every even place is* 0*, and* no *only when every odd place is* 0*.*
3. *If the zero test, the event test or LLPO reduces to Q, then deciding Q for every world implies WLPO, LPO or LLPO respectively, and Q is not settled.*

The implications between decisions and principles in parts 1 and 3 use no choice principle.

Proof. (1) The first is the definition of WLPO. For the second, LPO gives a decision of "some entry is 1"; conversely, if it is decided for α, its failure gives $\alpha(n) = 0$ for each n, since $\alpha(n)$ is 0 or 1. (2) On the run of zeros such a recorder must answer, at some record of length n. If the answer is *yes*, the run whose only 1 is at $2n$ has the same record of length n, receives the same answer, and has a 1 at an even place; if it is *no*, use the place $2n + 1$. (3) A decision of Q at $\Phi(\alpha)$ is a decision of the reduced question at α. Theorem 17.15 covers the first two cases. If R settled Q and LLPO reduced to Q through Φ, then, because openness pulls back along Φ, the question $Q \circ \Phi$ about binary runs would be settled, and a recorder settling it would contradict part 2. □

Verdicts that may be revised

An experiment may also report a verdict after every examination and revise it. Such experiments reach further, and their reach is again exact.

Definition 17.19 (Guessing rules)**.** A *guessing rule* g assigns a verdict, *yes* or *no*, to every finite record. Its verdicts *settle on* b along w when $g(w{\restriction}n) = b$ for all large n. The rule *confirms Q in the limit* when $Q(w)$ holds exactly when its verdicts settle on *yes*, and *decides Q in the limit* when its verdicts settle on *yes* along every w with $Q(w)$ and on *no* along every other w. Q is Σ_2 when there are sets $S_0, S_1, \ldots$ of records with $Q(w) \Leftrightarrow \exists m\, \forall n\, (w{\restriction}n \in S_m)$, Π_2 when there are sets with $Q(w) \Leftrightarrow \forall m\, \exists n\, (w{\restriction}n \in S_m)$, and Δ_2 when it is both.

Theorem 17.20 (Revisable verdicts confirm exactly the Σ_2 questions)**.** *For every set of worlds and every question Q:*

1. *some guessing rule confirms Q in the limit if and only if Q is Σ_2, and some guessing rule decides Q in the limit if and only if Q is Δ_2;*
2. *every settled question is decided in the limit, and on binary runs the question "some entry is* 1*" is decided in the limit but not settled;*
3. *on binary runs, read as a census in which $\alpha(n) = 1$ records a region not recorded before, finiteness, $\exists N \forall n \geq N \alpha(n) = 0$, is Σ_2 and not Π_2, and no guessing rule decides it in the limit;*
4. *for observed orbits $x : \mathbb{N} \to \mathbb{Z}$, each read by itself, boundedness, $\exists B\, \forall t\, |x(t)| \leq B$, is Σ_2 and not Π_2, and no guessing rule decides it in the limit.*

Conditions and scope.

Conditions
Any worlds and any question; guessing rules are arbitrary rules on finite records, and the proof is classical.

Identification
That a cosmological census is a stream of new-region events, or an observed orbit a stream of recorded integers, and that every binary run is a possible census (every integer sequence a possible orbit), is a modelling assumption that the theorem does not supply.

Not claimed
A computable guessing rule, or a bound on when its verdicts settle.

Proof. (1) If g confirms Q in the limit, Q is Σ_2 with S_m the set of records that have verdict *yes* or length less than m. Conversely, let Q be Σ_2. Say that a candidate m *survives* a record r of length n when $m \le n$ and every initial piece of r, of every length up to n, lies in S_m. A candidate $m \le n$ that does not survive a record of length n survives none of its extensions. Let g answer *yes* at a nonempty record when it has a least surviving candidate and that candidate is also the least survivor of the record with its last entry removed, and *no* otherwise. If $Q(w)$, let c be the least m with every record of w in S_m. Each smaller candidate fails at some record, so from some length on the least survivor along w is always c, and the verdicts settle on *yes*. If the verdicts settle on *yes*, the least survivor is eventually constant, say a; then a survives every record of w, and $Q(w)$.

If g decides Q in the limit, it confirms Q in the limit, and with the verdicts exchanged it confirms $\neg Q$; and Q is Π_2 exactly when $\neg Q$ is Σ_2, with the complementary sets. Conversely, let Q and $\neg Q$ be Σ_2 with sets S_m and T_m. Run the candidates S_m at the even numbers $2m$ and T_m at the odd numbers $2m + 1$, and let the verdict be *yes* when the least survivor is even, or when there is none. The least candidate that survives every record of w exists and is even exactly when $Q(w)$, and the least survivor is eventually that candidate.

(2) If R settles Q, the rule that repeats the answer of R so far, and says *yes* before it has one, decides Q in the limit. The rule "*yes* once a 1 has appeared" decides "some entry is 1" in the limit; it is not settled, as noted after Theorem 17.13.

(3) Finiteness is Σ_2 with S_N the records whose entries at places $\ge N$ are 0. Suppose g decides it in the limit. Build strings $\sigma_0 \subseteq \sigma_1 \subseteq \cdots$: σ_0 is empty, and since σ_k followed by zeros is finite, g eventually says *yes* along it, at some record $\sigma_k 0^t$; put $\sigma_{k+1} = \sigma_k 0^t 1$. The run that extends every σ_k has infinitely many entries 1, yet g says *yes* at each of its records $\sigma_k 0^t$, so its verdicts do not settle on *no*. If finiteness were Π_2, it would be Δ_2 and decided in the limit by part 1.

(4) Boundedness is Σ_2 with S_B the records whose entries have absolute value at most B. For a census α let $x_\alpha(0) = 0$, and $x_\alpha(t + 1) = t + 1$ if $\alpha(t + 1) = 1$ and $x_\alpha(t)$ otherwise. This orbit is bounded exactly when the census is finite, and its record of length n is computed from that of α; so a rule deciding boundedness in the limit would decide finiteness, against part 3. As there, boundedness is not Π_2. □

These classes go back to the limiting recursion of Gold and Putnam [61, 62] and underlie formal learning theory [63, 64], where the topological reading of verifiable and refutable

hypotheses is standard. Here the rules are arbitrary, so the theorems concern information, not computation.

The exact value of a measured constant

A measured constant reaches us as a stream of readings at increasing precision. We take this as the measurement model: a *measured world* is a classical real number v, its value, together with rational readings $r_0, r_1, \ldots$ satisfying $|v - r_n| < 2^{-n}$ for every n; the n-th examination yields r_n. A real number y is *consistent* with a record $(r_0, \ldots, r_{n-1})$ when $|y - r_k| < 2^{-k}$ for every $k < n$.

Theorem 17.21 (An exact value can be refuted, never confirmed)**.** *In the measurement model, let c be a real number.*

1. *The question "the value is exactly c" is refutable, not confirmable and not settled, and the zero test reduces to it.*
2. *A record of a measured world establishes exactly that the value is consistent with it. The values of the measured worlds with a given record are exactly the values consistent with it; and if c is consistent with the record of length n of a measured world, then for some $\delta > 0$ every y with $|y - c| < \delta$ is the value of a measured world with the same record.*
3. *Every record consistent with the golden ratio φ is consistent with $\varphi + \varphi^{-K}$ for all large K, and these are pairwise different values, all different from φ.*

Conditions and scope.

Conditions
The measurement model, with strict tolerances; they make the consistent values an open set.

Identification
That a laboratory constant reaches experiment as such a world is the measurement model itself, which the theorem does not supply. An exact value cannot come from the record; the theorem does not say where one comes from or that any constant has one.

Not claimed
That any particular constant has the value c, or anything about measurements whose readings do not fit the model.

Proof. (2) The values consistent with a record form a finite intersection of open intervals, hence an open set; so if c is consistent, so is every y with $|y - c| < \delta$ for some $\delta > 0$. A measured world with that record has a consistent value, by its tolerances. Conversely, a consistent y is the value of the world whose readings begin with the record and continue with rationals within 2^{-k} of y.

(1) If the value v of a world differs from c, choose n with $2 \cdot 2^{-n} < |v - c|$. A world with the same reading r_n and value c would give $|v - c| \le |v - r_n| + |r_n - c| < 2 \cdot 2^{-n}$. So the record of length $n + 1$ forces the failure of the question, which is therefore closed and refutable by Theorem 17.13. It is not open: every record of a world with value c is, by part 2, the record of

a world with value $c + \delta/2$. So it is not confirmable, not settled, and the zero test reduces to it, by Theorem 17.15.

(3) Take δ from part 2 for $c = \varphi$. Since $0 < \varphi^{-1} < 1$, the numbers φ^{-K} are positive, decrease strictly and fall below δ. □

A finite agreement with a derived value therefore never separates it from its neighbours. Measurement can refute the identification of a constant with a derived value; it cannot confirm it.

The verdict of a law at finite precision

The same distinction appears inside $\mathbb{R}_\delta$, where no measurement model is needed. Let Σ be a set of states and call maps $\Sigma \to \mathbb{R}_\delta$ quantities. A *law* is built from atoms $x = y$, $x \le y$ and $x < y$ between quantities by $\wedge$, $\vee$ and conjunctions and disjunctions over $i < n$. At a state σ, the law *holds* when it is true with its atoms read in $\mathbb{R}_\delta$. It *passes the test at tolerance* $1/(k+1)$ when it is true after each atom $x \le y$ or $x < y$ is replaced by $x < y + 1/(k+1)$, and each $x = y$ by $x < y + 1/(k+1) \wedge y < x + 1/(k+1)$. It is *refuted* when an atom fails with a margin: $y < x$ refutes $x \le y$ and $x < y$, and $x < y \vee y < x$ refutes $x = y$; a conjunction is refuted when a conjunct is, and a disjunction when every disjunct is. A law is *closed* when it is built from equations and $\le$ by conjunctions only, and *strict* when its atoms are all strict.

Theorem 17.22 (The verdict of a law)**.** *Let ψ be a law, σ a state and k a natural number.*

1. *ψ passes the test at tolerance $1/(k+1)$ at σ, or it is refuted there. If it holds, it passes every test; if it is refuted, it does not hold.*
2. *If ψ is closed, it holds at σ exactly when it passes the test at every tolerance.*
3. *Passing a test, and being refuted, each hold exactly when some entry of a binary sequence is 1. If ψ is closed, it holds exactly when every entry of a binary sequence is 0; if it is strict, it holds exactly when some entry of a binary sequence is 1.*
4. *Let the two quantities of an atom take every pair of values in $\mathbb{R}_\delta$. Deciding the atom at every state is WLPO for $x = y$ and $x \le y$, and LPO for $x < y$. That every failure is a refutation is $\mathrm{MP}_{\mathrm{bin}}$ for $x = y$ and $x \le y$, and false for $x < y$. The law $x \le y \vee y \le x$ holds at every state exactly under LLPO, although it passes every test at every state.*

Parts 1 to 3 use no omniscience principle and no choice principle, and each equivalence in part 4 is proved in both directions without choice.

Conditions and scope.

Identification
Reading a physical law as such a law, and a test at finite precision as the tolerance test, identifies the physical quantities with points of $\mathbb{R}_\delta$; the theorem does not supply that identification.

Not claimed

A price for laws with quantifiers over infinite families, or for any particular physical law.

Proof. (1) Induction on the law. For an atom $x \le y$ or $x < y$, cotransitivity (Proposition 12.19) applied to $y < y + 1/(k+1)$ at x gives $x < y + 1/(k+1)$ or $y < x$; an equation uses this twice. A finite conjunction or disjunction of laws each of which passes or is refuted passes or is refuted, by finitely many case distinctions. The other two statements follow atom by atom: $x \le y < y + 1/(k+1)$ gives $x < y + 1/(k+1)$, and $y < x$ contradicts both $x \le y$ and $x < y$.

(2) If $x < y + 1/(k+1)$ for every k, then $y < x$ is impossible, since a positive $x - y$ exceeds some $1/(k+1)$ by Proposition 12.19; so $x \le y$. Two such inequalities give an equation, and conjunctions pass to their conjuncts.

(3) Fix regular sequences naming the values of the quantities at σ. A strict inequality $u < v$ says that the sequence for $v - u$ has a stage s_N with $s_N > 2/(N+1)$, a comparison of ratios decided at each N; so each atom of a test or of a refutation, and $x < y$ itself, holds exactly when some entry of a binary sequence is 1. Two such forms give one for their conjunction and for their disjunction, by searching both sequences up to each stage, and the same holds for finite families. A closed law holds exactly when it is not refuted, because $x \le y$ is $\neg(y < x)$ and $x = y$ is $x \le y \wedge y \le x$; and the negation of a form with some entry 1 is the form with every entry 0.

(4) Because the quantities take every pair of values, each assertion at every state is the same assertion for all points of $\mathbb{R}_\delta$, priced in Table 12.2; the rows used are the decisions of $x = y$, $x \le y$ and $x < y$, the row priced at $\mathrm{MP}_{\mathrm{bin}}$ that turns $\neg(y \le x)$ into $x < y$ and $x \ne y$ into $x < y \vee y < x$, and the dichotomy. For $x < y$, a state with $x = y$ fails the law and is not refuted, since $y < y$ is impossible. Every test of the dichotomy passes: by part 1, $x \le y$ passes it or $y < x$, and then $y < x + 1/(k+1)$. □

A finite-precision test of a law always has an outcome, and a closed law is the conjunction of its tests. Exactness is what costs: deciding an equation is WLPO, and the step from "not exactly true" to "visibly false" is the binary Markov principle $\mathrm{MP}_{\mathrm{bin}}$, not the restricted schema of Theorem 6.9, which costs nothing under induction on every formula.

17.8 The physical task left by the mathematics

The record results expose a concrete obligation for a physical construction. Specify the allowed preparations and the target. Show how the procedure produces the record, how it is retained, and how its reader returns the required answer within the stated error. Account for all information available to the reader and for the resources the procedure uses. The next two chapters take up that positive construction.

A counting theorem can reject a proposed capacity before an apparatus is built. A large enough capacity does not complete the apparatus. Likewise, an exact logical principle from Chapter 12 concerns which conclusions follow from an infinite presentation; a finite physical record requires an account of how the relevant information reaches the reader. Both questions have exact mathematical forms. Keeping their premises visible is what makes the connection useful.

Questions to work through

1. In Lemma 17.2, identify the assumption that fails if the reader also receives an unrecorded control setting. How should the record space be repaired?

2. Follow the transcript reader through a tree that stops before using its full budget. Explain why the padding symbol need not be distinguishable from every genuine test outcome.

3. Prove Theorem 17.4 for a selected family that is smaller than the whole world space. Which map must be injective?

4. Identify the uses of symmetry and the triangle inequality in Lemma 17.8. Explain why the separated-family theorem uses a strict margin.

5. Compare the hypotheses of Theorems 17.7 and 17.11. Explain why finite length for each record does not supply the same conclusion as a fixed common budget.

6. Show that the question "the run has at least two entries 1" about binary runs is confirmable and not refutable. Give a finite-information map that reduces the event test to it, and say what deciding the question for every run costs.

7. In part 2 of Theorem 17.17, a decision of the question for α alone is not enough to find an entry 1 of α. Which case does it leave open, and how does the shifted run close it?

8. A measured world has the readings $r_0 = 1$, $r_1 = 3/2$ and $r_2 = 13/8$. Which values are consistent with this record? Is the golden ratio φ? Find the largest δ for which every y with $|y - \varphi| < \delta$ is consistent, and the least K for which $\varphi + \varphi^{-K}$ is consistent.

Chapter 18

The construction of finite operations

A description of an apparatus must say more than which mark it prints. The apparatus can change its input, return a quantum system as well as a mark, and act on something correlated with another system. A construction that reproduces the marks while destroying those correlations may perform a different operation.

We will follow a finite instrument through four descriptions. First come the matrices that specify its action. Those matrices give an explicit reversible operation on a larger space. Finite sequences of elementary operations approximate that reversible operation. Finally, positive energy operators enact the required sequence, with its duration, error and scalar energy contribution retained.

There is a second question along the way: can a finite procedure find the approximating sequence from a description of the instrument? We prove an equivalence between two effective descriptions. One supplies rational approximations to the instrument's matrices; the other supplies finite operation sequences with a prescribed accuracy. The proof uses a search with rational certificates, so it never asks an exact equality question about an arbitrary real number. The chapter reasons in classical mathematics; the search, for instance, is shown to end because a sequence that passes its test exists.

The first calculation needs only multiplication of two-by-two matrices. Use it to keep the purpose of the larger construction in view: we must retain both the mark and the returned system. The later proofs use the finite spectral theorem to factor positive matrices and take unitary logarithms. The final control argument uses matrix exponentials and their derivatives. The matrix sections of Appendix C explain how these tools enter, including the difference between an isometry into a larger space and a unitary on that whole space.

18.1 An operation with a retained outcome

Fix positive integers a, b, s. The input has a basis states, the returned system has b basis states, and the classical mark has s possible values. The matrices in this chapter have complex entries. Write $K^{\dagger}$ for the conjugate transpose of K, and I for an identity matrix on the space indicated by the equation. In Part V the italic letters I and U always denote an identity matrix and a unitary matrix; the sans-serif I of Chapter 9 is an iterator, and the sans-serif U of Section 20.3 is a universal program (Appendix N).

A state is a positive semidefinite matrix ρ with trace one. Positivity means $v^{\dagger}\rho v \geq 0$ for every vector v. This includes the state $vv^{\dagger}$ of a unit vector and also statistical mixtures of such states. The trace condition fixes the total probability.

The same marks can conceal a different operation

Consider a system with two basis states and two possible marks, 0 and 1. The matrices

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad K_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{18.1}$$

select the first and second basis states. Write an input state as $\rho = \left(\begin{smallmatrix} p & z \\ \bar{z} & q \end{smallmatrix}\right)$, so $p, q \geq 0$ and $p + q = 1$. Multiplication gives the two output blocks

$$K_0 \rho K_0^\dagger = \begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix}, \qquad K_1 \rho K_1^\dagger = \begin{pmatrix} 0 & 0 \\ 0 & q \end{pmatrix}.$$

Their traces are p and q: the probabilities of the two marks. Their sum has trace one because $K_0^\dagger K_0 + K_1^\dagger K_1 = I$.

Now keep K_0 but replace the second matrix by

$$L_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

It still satisfies $K_0^\dagger K_0 + L_1^\dagger L_1 = I$, and

$$L_1 \rho L_1^\dagger = \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix}.$$

The probability of mark 1 is still q. For every input, both devices therefore print the same marks with the same probabilities. But after mark 1, the first device leaves the second basis state and the second device returns the first. On input $\rho = K_1$, mark 1 is certain and the returned states are different.

A description that keeps only the two probabilities cannot distinguish these devices. An instrument keeps the output blocks as well. The general definition below adds several possible branches for each mark; the example has one branch per mark.

Definition 18.1 (A finite instrument)**.** Choose a finite branch set Z and matrices $K_{yz} : \mathbb{C}^a \to \mathbb{C}^b$, where y is the retained mark. Require

$$\sum_{y,z} K_{yz}^\dagger K_{yz} = I_a. \tag{18.2}$$

The output is the block diagonal matrix whose yth block is

$$\mathcal{J}_y(\rho) = \sum_z K_{yz} \rho K_{yz}^\dagger.$$

The trace of this block is the probability of mark y. The mark is retained even when the branch label z is not read.

Every block is positive, since $u^\dagger K \rho K^\dagger u = (K^\dagger u)^\dagger \rho (K^\dagger u) \geq 0$. Taking the trace and using (18.2) shows that the traces of the output blocks sum to one. Thus the formula describes both the returned system and the probabilities of all marks.

Different lists of branch matrices can describe the same instrument. To avoid treating a choice of branch decomposition as extra information, form one matrix for each outcome:

$$C_y[(u,i),(v,j)] = \sum_z K_{yz}[u,i]\,\overline{K_{yz}[v,j]}. \tag{18.3}$$

Here i, j are input indices and u, v are output indices. These are the *Choi matrices* [23], with output indices written first. They determine the instrument because

$$\mathcal{I}_y(\rho)[u,v] = \sum_{i,j} C_y[(u,i),(v,j)]\rho[i,j].$$

Proposition 18.2 (Valid matrix descriptions)**.** *A family C_y describes a finite instrument precisely when*

$$C_y \geq 0 \quad \textit{for every } y, \qquad \sum_{y,u} C_y[(u,i),(u,j)] = \delta_{ij}. \tag{18.4}$$

For such a family, one can take $Z = \{1,\dots,b\}\times\{1,\dots,a\}$.

Proof. Given branch matrices, arrange their entries into columns indexed by z. Equation (18.3) is the product of this column matrix with its adjoint, so it is positive. Summing its output diagonal gives the transpose of $\sum_{y,z} K_{yz}^\dagger K_{yz}$, which is I_a.

Conversely, a positive finite matrix has a positive square root. This can be constructed by taking the nonnegative square roots of its eigenvalues in an orthonormal eigenbasis. Let $S_y = C_y^{1/2}$, and reshape its zth column into a b by a matrix:

$$K_{yz}[u,i] = S_y[(u,i),z].$$

The product $S_y S_y^\dagger = C_y$ gives (18.3). The normalization in (18.4), with i, j interchanged, gives (18.2). These are all the required conditions. □

This is a finite-dimensional existence construction. An exact choice of eigenvectors is not yet a procedure for obtaining an operation from successively better rational data. We will obtain that procedure by a different route.

18.2 An explicit reversible extension

The next construction is a finite-dimensional dilation: it realizes an operation as part of a reversible one on a larger space. Dilation theory originates with Stinespring [22]; here the block formula makes the action on the entire enlarged space explicit.

Imagine keeping every branch label for the moment. Define the rectangular matrix V by stacking the branch matrices:

$$V[(u,y,z),i] = K_{yz}[u,i].$$

Equation (18.2) says $V^\dagger V = I_a$. Thus V preserves lengths and inner products. It embeds the input into the larger space D with basis labels (u,y,z).

We need a reversible operation on a whole space, including states outside this embedded input. Rather than leave its action there unspecified, write it down.

Theorem 18.3 (A reversible extension with its unused space retained). *If $V^\dagger V = I_a$, put $P = VV^\dagger$. On $\mathbb{C}^a \oplus D$ define*

$$U_V = \begin{pmatrix} 0 & V^\dagger \\ V & I_D - P \end{pmatrix}. \tag{18.5}$$

Then $U_V^\dagger = U_V$, $U_V^2 = I$, and

$$U_V(v, 0) = (0, Vv).$$

Reading the output and mark coordinates, while retaining the complete branch sum, gives exactly the original instrument.

Conditions and scope.

Conditions
A matrix V with $V^\dagger V = I_a$, here the stacked branch matrices of a valid instrument, and an input state placed in the summand $\mathbb{C}^a$.

Not claimed
A device that enacts U_V, or a preparation that places the input in $\mathbb{C}^a$. The theorem is an identity between matrices; Theorem 18.11 supplies the register, controls and reader that approximate U_V.

Proof. The matrix P is self-adjoint and $P^2 = V(V^\dagger V)V^\dagger = P$. Also $(I_D - P)V = 0$ and $V^\dagger(I_D - P) = 0$. Therefore block multiplication gives

$$U_V^2 = \begin{pmatrix} V^\dagger V & V^\dagger(I_D - P) \\ (I_D - P)V & VV^\dagger + (I_D - P)^2 \end{pmatrix} = \begin{pmatrix} I_a & 0 \\ 0 & I_D \end{pmatrix}.$$

Self-adjointness follows by transposing the blocks and conjugating them. Multiplication by $(v, 0)$ gives $(0, Vv)$. Its (u, y, z) component is $K_{yz}v$ at output index u, so summing the unread branches produces the instrument's yth block. The same calculation applies to mixed input states by matrix multiplication and linearity. □

The extension is a concrete answer to the question, "What happens on the rest of the space?" Its complement is part of the operation. The input preparation must place the state in the designated input summand; that requirement is not supplied by the block identity itself.

18.3 A fixed register and a complete reader

Use a register of n eight-state cells. Its basis is the set of functions from $\{1, \dots, n\}$ to $\{0, \dots, 7\}$, so its dimension is 8^n. The choice of register can depend on a, b, s while remaining fixed for every valid instrument of those dimensions.

There is a convenient way to place a unitary in a determinant-one matrix. If U acts on a space E, use

$$\tilde{U} = U \oplus U^\dagger \oplus I_R. \tag{18.6}$$

This is unitary, has determinant $\det(U)\det(U^\dagger) = 1$, and agrees with U on the first summand. Choose $n \geq 2$ with $8^n \geq 2\dim E$, take $\dim R = 8^n - 2\dim E$, and fix a bijection of basis labels. Such an n exists; the deliberately generous choice $n = 2\dim E + 2$ suffices. Applied to (18.5), this supplies one layout for all instruments with the chosen dimensions.

An approximation to $\tilde{U}$ need not keep the input entirely within the intended output rows. The reader must still account for the whole output. Split the rows into normal rows (u, y, z) and complementary rows ℓ. Choose a default output u_0 and mark y_0. Each normal row retains its own output and mark. Each complementary row has a separate unread branch and is assigned to (u_0, y_0).

Proposition 18.4 (Complete readout preserves normalization)**.** *Let W be the columns of a unitary selected by the fixed input embedding. The branch matrices obtained by the above reader satisfy*

$$\sum_{y,z'} K_{yz'}^\dagger K_{yz'} = W^\dagger W = I_a,$$

where z' ranges over both normal and complementary branches. For the ideal placed unitary, the complementary branches vanish and its Choi matrices are the specified C_y.

Proof. Each row of W occurs exactly once in the sum of branch Gram matrices. Normal rows occur with their own output labels. A complementary row occurs at the default output and mark, with its own branch label. Expanding a matrix entry of the sum therefore gives exactly the same row sum as $W^\dagger W$. Distinct columns of a unitary are orthonormal. For the ideal operation, Theorem 18.3 puts the selected columns in the normal rows with entries $K_{yz}[u, i]$ and zero elsewhere. □

No outcome has been discarded or conditionally renormalized. Leakage into an unintended row is an error in the instrument, and the comparison below counts it.

18.4 An error bound that keeps a reference system

Suppose the input is correlated with a retained reference Q of any finite dimension. Its joint state τ is a positive trace-one matrix on $Q \otimes \mathbb{C}^a$. Define the joint output by retaining the reference, the returned system and the mark. Its yth diagonal block has entries

$$\mathcal{J}_C^Q(\tau)[(y, r, u), (y, r', v)] = \sum_{i,j} C_y[(u, i), (v, j)]\tau[(r, i), (r', j)]. \tag{18.7}$$

Blocks with different marks are zero. The branch factorization proves positivity, since this is a sum of $(I_Q \otimes K_{yz})\tau(I_Q \otimes K_{yz})^\dagger$.

For a Hermitian matrix X, write $\|X\|_1$ for the sum of the absolute values of its eigenvalues. We say that two instruments differ by at most ε on retained references if

$$\|\mathcal{J}_C^Q(\tau) - \mathcal{J}_D^Q(\tau)\|_1 \leq \varepsilon \tag{18.8}$$

for every finite Q and every joint state τ. This convention has no factor of one half. It requires no independence between input and reference.

Lemma 18.5 (Entry errors control the joint output)**.** *Suppose C, D are valid instruments and, for every entry and mark, $|C_y[p,q] - D_y[p,q]| \le \eta$, with $\eta \ge 0$. Then their error on retained references is at most*

$$sab^2\eta. \tag{18.9}$$

The bound does not grow with the reference dimension.

Proof. For a matrix A of size ab with entries bounded in absolute value by η, the triangle inequality and Cauchy–Schwarz give

$$|v^\dagger A v| \le \eta\Big(\sum_p |v_p|\Big)^2 \le ab\eta \sum_p |v_p|^2.$$

Take $A = C_y - D_y$, which is Hermitian. With $r = ab\eta$, this proves $rI \pm (C_y - D_y) \ge 0$ for every y.

Apply the positive output construction (18.7) to these two families. If X is the output difference and F is the output of the family with every block equal to rI, then $F + X \ge 0$ and $F - X \ge 0$. Thus, for every unit vector v, $|v^\dagger X v| \le v^\dagger F v$. Choose an orthonormal eigenbasis of X and sum this inequality. It gives $\|X\|_1 \le \operatorname{tr} F$.

The diagonal sum in (18.7), applied to the family rI, is $\operatorname{tr} F = sbr \operatorname{tr}\tau = sbr$. Substituting $r = ab\eta$ proves (18.9). The only reference contribution was its joint state inside $\operatorname{tr}\tau = 1$. □

Let m be the number of normal plus complementary unread branch labels in the fixed complete reader. If two unitaries are within η in operator norm, every corresponding entry differs by at most η. Their entries also have absolute value at most one. The elementary identity

$$x\bar{y} - u\bar{v} = (x-u)\bar{y} + u\overline{(y-v)}$$

therefore bounds each product difference by 2η. A Choi entry sums over m branches. Applying the lemma gives the useful implication

$$\|U - W\| \le \eta \quad \Longrightarrow \quad \text{joint output error} \le L\eta, \qquad L = 2msab^2. \tag{18.10}$$

All dimensions and the branch count are fixed before the target instrument and requested tolerance.

18.5 Two descriptions of finite operations

We now distinguish a finite alphabet used to describe operations from the continuous controls used to enact them. A cell label has three binary coordinates. On these coordinates use the Hadamard matrix

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

the phase matrix $\operatorname{diag}(1, i)$, and the Toffoli permutation that flips one coordinate when two selected control coordinates are both one. The target differs from each control, but the controls

may coincide; that case is controlled-NOT. A word is a finite list of such gates with their coordinate labels. Matrix multiplication assigns a unitary to each word.

Two properties of this alphabet do different jobs. First, a word's matrix and complete Choi readout admit effective rational approximations. Every entry is obtained by finitely many sums and products in $\mathbb{Q}(\sqrt{2}, i)$; certified rational enclosures for $\sqrt{2}$, propagated through those finite operations, give any requested entry accuracy. Second, on a fixed register with at least two cells, these words are dense in its special unitary group. The finite gate and matrix-generation arguments are developed in Appendix A. We need the existence of an approximating word here, without a runtime estimate for finding it.

Together with the explicit dilation and fixed placement, density gives a word whose complete Choi readout is within any positive entry tolerance of any valid target C. Indeed, approximate the placed unitary in operator norm and use the product estimate preceding (18.10).

Definition 18.6 (An effective instrument description)**.** An ordinary name of C is one finite procedure which, on input k, returns the full array of rational complex entries $q_y^{(k)}[p, q]$ and guarantees

$$|q_y^{(k)}[p, q] - C_y[p, q]| \le 2^{-k}.$$

A word name is one finite procedure which, on input k, returns a word whose complete readout differs from C by at most 2^{-k} on every finite retained reference. Its register and layout remain fixed as k varies.

The procedures are required to finish at every requested precision for the instrument they name. A mere assertion that a suitable word exists for each k does not yet give a word name.

18.6 Searching with a margin

To find a word, enumerate all finite words and test a rational certificate. For complex rationals z, w, use

$$d_1(z, w) = |\operatorname{Re}(z - w)| + |\operatorname{Im}(z - w)|.$$

This is decidable by rational arithmetic, and

$$|z - w| \le d_1(z, w) \le 2|z - w|. \tag{18.11}$$

The left inequality is the triangle inequality for real and imaginary parts; each part has absolute value at most $|z - w|$, proving the right. Test every Choi entry, not just the outcomes most likely on one input.

For requested output error 2^{-k} put

$$M = sab^2, \qquad h = k + M + 3, \qquad r = 2^{-h}, \qquad e = 2^{-(h+3)} = r/8. \tag{18.12}$$

Obtain target data q at entry accuracy e. For each candidate word, compute its complete readout data q_W at the same accuracy. Accept the first word for which $d_1(q_W[y, p, q], q[y, p, q]) \le r$ at every entry.

Lemma 18.7 (Soundness and termination of the search)**.** *For a valid named target, the search terminates. Every accepted word has joint output error at most 2^{-k}.*

Proof. If a word is accepted, its exact Choi entries differ from those of the target by at most

$$e + r + e = \tfrac{5}{4}r \le 2r.$$

The three terms are word-evaluation error, accepted rational discrepancy and target-naming error. Lemma 18.5 bounds the joint output error by $M(2r)$. Since $M \le 2^M$,

$$M(2r) \le 2^M\, 2\, 2^{-(k+M+3)} = 2^{-k}/4 \le 2^{-k}.$$

For termination, density supplies a word whose exact Choi entries are within e of the target. Its rational data and the target data are then within $3e = 3r/8 < r/2$ in absolute complex value. By (18.11), their rational discrepancy is less than r, so this word passes. Every individual test finishes, and enumeration eventually reaches a passing word. □

The slack has a purpose. Testing exact equality of two limits would not give a terminating decision. Here the acceptance test is finite, and density supplies a word strictly inside its tolerance. A nearby rejected candidate simply causes the search to continue.

18.7 Recovering the matrix description

The reverse direction also needs the retained reference. Suppose an operation is known to approximate an instrument on every joint input. Use a reference of dimension a and the normalized state

$$\tau_a = \frac{1}{a}\,\Omega\Omega^{\dagger}, \qquad \Omega = \sum_{i=1}^{a} e_i \otimes e_i. \tag{18.13}$$

Its trace is one, and it is positive. Substitution in (18.7) gives

$$\mathcal{J}_C^{Q}(\tau_a)[(y,i,u),(y,j,v)] = \frac{1}{a} C_y[(u,i),(v,j)].$$

Thus this one joint input exposes every Choi entry, with the known factor $1/a$.

Lemma 18.8 (Joint output control recovers every entry)**.** *If C, D have retained-reference error at most ε, then $|C_y[p,q] - D_y[p,q]| \le a\varepsilon$ at every entry.*

Proof. Apply the error bound to (18.13). Every entry of a Hermitian matrix has absolute value at most its trace norm: expand it in an orthonormal eigenbasis, use the triangle inequality, and bound each product of eigenvector coordinates by one. The displayed probe identity therefore gives $|C_y[p,q] - D_y[p,q]|/a \le \varepsilon$. □

Theorem 18.9 (Equivalence of effective finite descriptions)**.** *For fixed positive a, b, s and a valid finite instrument C, an ordinary name exists if and only if a word name exists. The forward translation can be carried out by one procedure fixed for those dimensions and their register layout, taking the ordinary name as an input.*

Conditions and scope.

Conditions
Fixed dimensions a, b, s, a valid instrument C, and a supplied name; both names are required to finish at every precision.

Not claimed
A procedure deciding whether a proposed description is valid, or an efficient search bound. The theorem concerns descriptions; it constructs no apparatus.

Proof. Given an ordinary name, run the rational search at each requested precision. Lemma 18.7 proves both termination and the required joint output accuracy. The instructions for evaluation, enumeration and testing are fixed; they call the supplied name at the precision in (18.12). This is the stated uniform forward translation.

Conversely, given a word name and requested entry accuracy 2^{-k}, call the word name at precision $j = k+(a+1)+3$. Evaluate that word's Choi entries to accuracy 2^{-j}. The probe lemma bounds the difference between its exact entries and the target entries by $a2^{-j}$. The triangle inequality gives total entry error at most $(a+1)2^{-j} \le 2^{-k}$. The last inequality follows from $a+1 \le 2^{a+1}$ and the extra factor 2^{-3}. This constructs a single ordinary naming procedure. □

The equivalence concerns descriptions of the whole instrument. It neither decides whether an arbitrary proposed description is valid nor gives a general efficient search bound. Validity and the supplied name are the premises under which termination has been proved.

18.8 Enacting a sequence with positive energy

The word description is now connected to a physical evolution. We state the controls as well as the energy they require. On the native register, the continuous elementary operations are cyclic three-state rotations within a cell and the two orderings of the adjacent-cell coupling

$$E = \frac{I + iX}{\sqrt{2}}, \qquad X = S \otimes F.$$

Here S is diagonal, with

$$S = \mathrm{diag}(1, -1, -1, 1, -1, 1, 1, -1).$$

The operator F swaps each even label with the following odd label. Both S and F are Hermitian involutions. Each elementary operation is e^G for an explicitly specified skew-Hermitian generator G. The rotations have continuously selectable real angles. These controls generate the full traceless skew-Hermitian matrix algebra on registers of at least two cells; the bracket argument and passage to finite words appear in Appendix A. Consequently their finite words approximate every special unitary on that register.

Fix a positive background energy operator $H \geq 0$, independently of the target operation. Let $\hbar > 0$ and choose a pulse duration $t > 0$. For a skew-Hermitian generator G, use the full energy operator

$$H_{G,t} = H + \frac{\hbar}{t}\big(iG + \|G\|I\big). \tag{18.14}$$

Since iG is Hermitian and its eigenvalues are at least $-\|G\|$, the parenthesized operator is positive. Hence $H_{G,t} \geq 0$. The scalar term is part of the energy operator even when a particular readout is insensitive to its phase.

The evolution for duration t is $e^{-itH_{G,t}/\hbar}$. The scalar part commutes with everything, so this equals

$$e^{-i\|G\|}\exp\big(G - itH/\hbar\big).$$

For skew-Hermitian matrices A, B, differentiating $e^{sA}e^{(1-s)B}$ and integrating from zero to one gives $\|e^A - e^B\| \leq \|A - B\|$: both exponential factors in the integrand are unitary and have norm one. Therefore

$$\big\|e^{-itH_{G,t}/\hbar} - e^{-i\|G\|}e^G\big\| \leq t\|H\|/\hbar. \tag{18.15}$$

For a word w of ℓ pulses, subtract the two products one factor at a time. Every other factor has norm one. The errors add, giving

$$\|U_{w,t} - e^{-ic(w)}W_w\| \leq \ell t\|H\|/\hbar, \qquad c(w) = \sum_{G \text{ in } w} \|G\|. \tag{18.16}$$

Here W_w is the ideal word matrix and $U_{w,t}$ is its actual evolution under the full pulse energies.

18.8.1 Retaining the phase

One can also match a target's full scalar phase. This matters when claiming an operator approximation, even though a common phase cancels in the quadratic instrument readout. A pair of inverse cyclic rotations leaves the ideal word unchanged and contributes 2θ to $c(w)$ when $\theta \geq 0$. For a desired phase e^{-ir}, choose

$$z = \left\lceil \frac{c(w) - r}{2\pi} \right\rceil, \qquad \theta = \frac{r + 2\pi z - c(w)}{2}. \tag{18.17}$$

The defining inequalities for the ceiling give $0 \leq \theta < \pi$. Appending the inverse pair makes the total scalar action $r + 2\pi z$, so the phase is exactly e^{-ir}. Both added pulses retain H and their positive scalar baseline.

Every finite unitary has a Hermitian logarithm: diagonalize the unitary, choose a real argument for each unit-modulus eigenvalue, and conjugate the diagonal arguments back. Subtracting the mean trace separates this logarithm into a scalar part and a traceless part. Approximate the latter by the native controls, and apply (18.17) to the former. This supplies an ideal word with the requested phase to any positive operator tolerance.

Theorem 18.10 (Finite positive-pulse realization)**.** *Fix a native register with at least two cells, $H \geq 0$, and $\hbar > 0$. For any unitary target U, error tolerance $\delta > 0$ and total duration bound $T > 0$, there is a finite native word and a positive common pulse duration t such that its full energy operators are positive, $\ell t < T$, and its actual evolution differs from U by less than δ in operator norm. The construction allows the amplitudes in* (18.14) *and externally selected switching.*

Conditions and scope.

Control
Pulses chosen from the native generators and switched by an external controller, each with the amplitude $\hbar/t$ of (18.14).

Resources counted
A fixed background energy $H \geq 0$, the scalar action $\hbar c(w)$ of (18.19), and total duration below T.

Not claimed
A bound on drive strength, which shrinking t would violate; an autonomous controller; or a model of actuator work and switching equipment.

Proof. Choose the ideal word and phase correction with error less than $\delta/2$, and let ℓ be its full length, including the two phase pulses. Set

$$t = \min\left\{\frac{\delta\hbar}{2(\ell\|H\| + 1)}, \frac{T}{\ell + 1}\right\}. \tag{18.18}$$

Both quantities are positive. The first gives $\ell t\|H\|/\hbar \leq \delta/2$, while the second gives $\ell t \leq \ell T/(\ell + 1) < T$. Equation (18.16) and the ideal approximation error then give a total error less than δ. Positivity was proved directly from (18.14). □

Shorter duration does not make the controls free. Their strength includes the factor $\hbar/t$. The integrated scalar baseline is exactly

$$\sum_{G \text{ in } w} t\frac{\hbar}{t}\|G\| = \hbar c(w). \tag{18.19}$$

The duration cancels; the action remains. This equality records that baseline contribution, rather than a complete model of actuator work, switching equipment or total experimental cost. A bound on available drive strength would add a constraint that Theorem 18.10 does not impose.

18.9 The finite bridge and its physical interfaces

Combine the constructions in their actual order (Figure 18.1). Positivity and normalization give branch matrices. Stacking gives an isometry. The block matrix gives a reversible extension. Adjoint padding puts it on a fixed register, complete readout counts every row, and the pulse construction approximates the placed unitary.

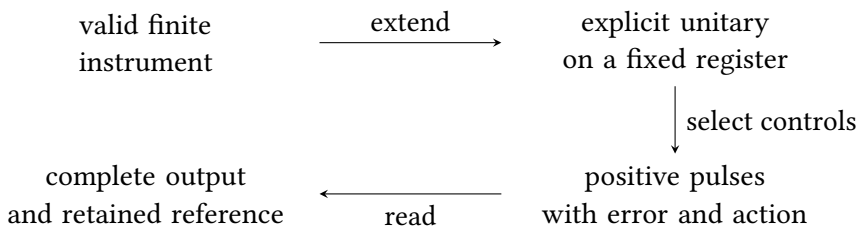


Figure 18.1. The finite bridge retains the whole operation. Input preparation, selectable controls and the complete reader are physical interfaces required by the construction. Error is compared on the returned system, mark and every finite retained reference.

Theorem 18.11 (Realization of a complete finite instrument)**.** *Fix positive dimensions a, b, s and default output and mark labels. There is a finite native register and a fixed input and complete-output layout, chosen before the target instrument, with the following property. For any valid C, any fixed positive background energy H on that register, $\hbar > 0$, tolerance $\varepsilon > 0$ and duration bound $T > 0$, the positive-pulse construction produces a complete instrument whose retained-reference error is at most ε, with total pulse duration less than T. It retains the scalar baseline* (18.19)*. It uses the selectable controls, input preparation and complete readout described above.*

Conditions and scope.

Preparation
The input is placed in the designated input summand of the fixed register layout.

Control
Externally selected positive-energy pulses, as in Theorem 18.10.

Reader
The complete readout of Proposition 18.4, which counts every normal and complementary row.

Resources counted
The background energy H, the scalar action $\hbar c(w)$, and total duration below T.

Identification
Applying the theorem to a laboratory apparatus identifies its cells, generators, preparation and reader with the native register, its controls and the complete readout.

Not claimed
An autonomous apparatus with bounded controls; access to every physical carrier; or a construction of the source fixture of Section A.8, whose Hamiltonian is an input.

Proof. Choose the register and layout using the fixed branch size ab and the dilation space; neither depends on the entries of C. Construct its ideal placed unitary U_C as above. Let $L = 2msab^2$ be the constant of the complete reader in (18.10). Apply Theorem 18.10 with operator tolerance $\eta = \varepsilon/(L+1)$ and duration bound T. Its output is unitary, so the complete reader remains normalized. Its error on every retained reference is at most $L\eta \le \varepsilon$. The ideal reader gives C exactly, and every pulse retains its positive energy and baseline contribution. □

Section A.8 gives a concrete record-and-field assembly to which the pulse theorem applies. It retains an old source on $14+2N$ cells, adds three complete record–reference pairs and two field banks, and specifies the added holding and edge energies. Its full register has $20+2N+4a+4b$ eight-state cells. Positivity follows on the whole register, including coordinates unused by the field code.

The old source's positive interior Hamiltonian is an explicit input to that assembly. The appendix does not derive the material source fixture or a microscopic actuator for it. Thus the application to the native two-flux source retains this supplied Hamiltonian, its fixed external fluxes and the same preparation, readout and control interfaces. The on-page construction establishes the record-and-field extension and the full-register transport needed here. It does not replace an unprovided source construction by a claim of universal physical access.

The conclusion has two complementary strengths. The naming equivalence shows that effective matrix descriptions and effective finite operation descriptions determine the same valid finite instruments. The pulse construction supplies explicit positive energy operators and carries their error through the complete reader. Establishing an autonomous apparatus with bounded controls, or establishing that every possible physical carrier belongs to this finite class, requires additional constructions. Those questions concern the remaining interfaces of this particular result.

Questions to work through

1. In the proof of Theorem 18.3, identify the exact use of $V^{\dagger}V = I$. Explain the role of $I_D - VV^{\dagger}$.

2. Expand one entry of the Gram sum for the complete reader. Explain why two complementary rows must not be combined into a single coherent branch before taking their Gram matrices.

3. Reproduce the three error terms in an accepted rational certificate and the three terms used to prove that some word is accepted. Why are these different arguments?

4. Trace the input-dimension factor a in Lemma 18.8. Identify the normalization that produces it.

5. Verify the two duration inequalities in (18.18) and the baseline identity (18.19). State which resource grows when the duration shrinks.

Chapter 19

Conservation and the cost of keeping a record

A recorder can have enough possible marks and still be unable to produce the right one. Imagine asking whether the first of two cells is occupied and the second is empty. The apparatus must couple to those cells, leave an outcome in a readable pointer, and obey its conservation laws throughout the interaction. These are requirements on one operation. They cannot be satisfied by choosing the record and the dynamics independently.

We will prove a quantitative limit for this question. Suppose the source is two native eight-state cells and the recorder supplies r more such cells, together with any finite neutral memory. Suppose the interaction exactly conserves the specified total clock energy and total occupation, and the final yes/no pointer is a sharp event in the neutral memory. If its yes probability differs from the desired one by at most e for every source state, then

$$e \geq \frac{1}{2(r^2+3)}. \tag{19.1}$$

The recorder's cells may be correlated with one another and with its neutral memory. The initial source and the complete recorder must be separate. We will state these conditions precisely before using them. The chapter reasons in classical mathematics, as Chapter 18 does.

The proof has three physical steps. Conservation restricts the event that can be read at the end of the interaction. Accuracy bounds the noise of that event. A particular source state then makes the restriction incompatible with arbitrarily small noise unless the recorder supplies sufficient charge spread. The constants in (19.1) come from a four-entry vector and a bound on a sum of r projections.

The conservation argument belongs to the line of work begun by Wigner and Araki and Yanase [26, 27], with quantitative measurement-error bounds developed by Ozawa [28]. The native charges and explicit witness below determine the particular bound here.

19.1 The event, the interaction and the pointer

Work first with an arbitrary finite-dimensional source space $\mathscr{H}_S$ and recorder space $\mathscr{H}_R$. The inner product is conjugate-linear in its first argument, and A^* is the adjoint of an operator A, the conjugate transpose written $A^\dagger$ in Chapter 18. A *sharp event* is an orthogonal projection: an operator P satisfying $P^* = P = P^2$. For a unit source vector x, the desired yes probability is $\langle x, Px\rangle$.

Prepare the recorder in a unit vector $p \in \mathscr{H}_R$. Inserting that preparation gives the isometry

$$J : \mathscr{H}_S \longrightarrow \mathscr{H}_S \otimes \mathscr{H}_R, \qquad Jx = x \otimes p.$$

The interaction is a unitary U on the full tensor product. (As in Chapter 18, the italic U is a unitary matrix and I an identity matrix; they are unrelated to the universal program U and the iterator I.) A sharp pointer B is read after the interaction. Its actual yes probability is

$$\langle UJx, BUJx\rangle.$$

Pull this final event back to the initial time:

$$Z = U^*BU, \qquad M = J^*ZJ. \tag{19.2}$$

The operator Z is a projection. The compressed operator M is an effect, meaning $0 \le M \le I$, but it need not be a projection. Its expectation is exactly the probability the apparatus produces:

$$\langle x, Mx\rangle = \langle UJx, BUJx\rangle.$$

Compression describes the given preparation and interaction. It does not supply a new event that the apparatus is free to choose.

Definition 19.1 (Uniform event accuracy)**.** The recorder has error at most e for the event P if

$$\big|\langle x, (M-P)x\rangle\big| \le e \qquad \text{for every unit } x \in \mathscr{H}_S. \tag{19.3}$$

The expression inside the absolute value is real. The preparation p, interaction U and pointer B are fixed across these source inputs.

Uniformity matters. An apparatus designed for one known state need not answer the same question on an unknown state. Our lower bound concerns a single recorder that meets (19.3) for all inputs.

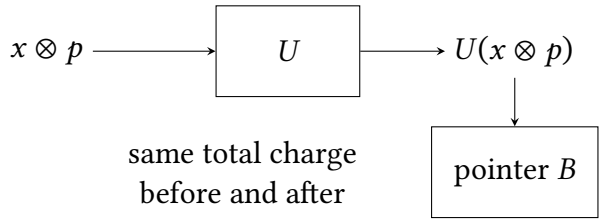


Figure 19.1. The pointer is read after the conserving interaction. Pulling it back gives the event $Z = U^*BU$ on the initial source and recorder.

19.2 Accuracy controls the noise of a sharp event

The discrepancy between the actual event and the ideal one can be measured on the full state, before any compression. Put $Q = P \otimes I$, so that $QJ = JP$, and define the noise operator $E = Z - Q$. On an input Jx, its squared norm is

$$\|EJx\|^2 = \langle x, (P + M - MP - PM)x\rangle. \tag{19.4}$$

Indeed, expand $(Z-Q)^2 = Z + Q - ZQ - QZ$, then use $Z^2 = Z$, $Q^2 = Q$, $J^*J = I$ and $QJ = JP$.

The following identity is what gives the useful constant. Let $D = M - P$. Expansion using $P^2 = P$ yields

$$P + M - MP - PM = (I-P)D(I-P) - PDP. \tag{19.5}$$

Thus the two contributions to the noise involve orthogonal parts of the same input.

Lemma 19.2 (Probability error bounds projection noise)**.** *If* (19.3) *holds, then*

$$\|(Z-Q)Jx\|^2 \le e\|x\|^2 \qquad \text{for every } x \in \mathscr{H}_S.$$

Proof. Rescaling a nonzero vector in (19.3) gives $|\langle y, Dy\rangle| \le e\|y\|^2$ for every y; the zero vector satisfies the same inequality directly. Apply this bound to $(I-P)x$ and Px in (19.5). Then

$$\begin{aligned}\|(Z-Q)Jx\|^2 &\le e\|(I-P)x\|^2 + e\|Px\|^2\\ &= e\|x\|^2.\end{aligned}$$

The last equality is the orthogonal decomposition given by P. □

The orthogonal decomposition is essential to this estimate. Bounding the two terms separately by $e\|x\|^2$ would lose a factor of two.

19.3 What exact conservation does to the pulled-back event

Let A and R be Hermitian charge operators on the source and recorder. Their total charge is

$$C = A \otimes I + I \otimes R. \tag{19.6}$$

Suppose the interaction conserves it exactly, $UC = CU$, and the final pointer commutes with it, $BC = CB$. The pulled-back event also commutes:

$$ZC = U^*BUC = U^*BCU = U^*CBU = CZ.$$

A pointer that is neutral for this charge supplies the second commutation condition. It is an actual requirement on the readout.

For Hermitian operators E, C, any vector ξ and any real number c,

$$\langle \xi, [E,C]\xi\rangle = \langle E\xi, (C-cI)\xi\rangle - \langle (C-cI)\xi, E\xi\rangle.$$

Here $[E,C] = EC - CE$. The scalar c cancels. The triangle and Cauchy–Schwarz inequalities give

$$|\langle \xi, [E,C]\xi\rangle|^2 \le 4\|E\xi\|^2\|(C-cI)\xi\|^2. \tag{19.7}$$

Take $E = Z - Q$. Since $[Z,C] = 0$, we have $[E,C] = -[Q,C]$. Combining (19.7) with Lemma 19.2, for a unit source vector x, gives

$$|\langle Jx, [Q,C]Jx\rangle|^2 \le 4e\|(C-cI)Jx\|^2. \tag{19.8}$$

This is the required measurement constraint. A nonzero commutator on the left forces either noise or charge spread on the right.

19.4 The native source and its two charges

We now specify the cells. One cell has basis $e_0, \dots, e_7$. Use the parity operator S and adjacent-pair flip F from Chapter 18. They obey

$$S^2 = F^2 = I, \qquad SF = -FS.$$

Consequently iSF is a Hermitian involution, and

$$N = \frac{I - iSF}{2} \tag{19.9}$$

is the occupation projection. The four-step cyclic shift Q_4 sends e_j to $e_{j+4 \bmod 8}$. It is another Hermitian involution. Write

$$P_c = \frac{I - Q_4}{2}. \tag{19.10}$$

It projects onto the odd Fourier modes. We will use its sum across cells as a charge count.

For completeness, let W be the normalized Fourier matrix

$$W_{jk} = \frac{\omega^{jk}}{\sqrt{8}}, \qquad \omega = e^{-2\pi i/8},$$

and define the dimensionless clock $K = W\mathrm{diag}(0, 1, \dots, 7)W^*$. The native clock energy is $H_0 = \varepsilon_0 K$, with $\varepsilon_0 > 0$. The value of this positive scale will cancel from the argument. The other conserved native quantity is N.

For any one-cell operator T, let

$$\mathscr{L}(T) = \sum_{j=1}^{r+2} T^{(j)} \otimes I_{\mathscr{H}_0}$$

act on the two source cells, r recorder cells and neutral memory $\mathscr{H}_0$. Superscript (j) means that T acts in slot j and identities act in the other slots.

Lemma 19.3 (The two native charges conserve the count)**.** *If U commutes with $\mathscr{L}(H_0)$ and $\mathscr{L}(N)$, then it commutes with $\mathscr{L}(P_c)$.*

Proof. Since $\varepsilon_0 \neq 0$, energy conservation implies commutation with $\mathscr{L}(K)$. Appendix B gives an explicit finite expression for P_c using I, K, N, complex linear combinations and commutators. It derives that expression from the entries of the specified eight-state matrices.

Operators on different slots commute. Hence $\mathscr{L}$ is linear and

$$\mathscr{L}([T_1, T_2]) = [\mathscr{L}(T_1), \mathscr{L}(T_2)].$$

It sends I to $(r+2)I$, which commutes with U. Commutation with U is preserved by sums, scalar multiples and commutators. Applying these facts in order to the finite expression proves the assertion. □

The count is therefore derived from the two specified conservation laws. It has not been added as a third conservation assumption. The calculation in the appendix is needed for that distinction.

On the two source cells, set

$$A = P_c \otimes I + I \otimes P_c, \qquad P = N \otimes (I - N). \tag{19.11}$$

The event P asks precisely whether the first cell is occupied and the second is empty. The count A and the event P do not commute. We next exhibit a source vector that quantifies their incompatibility.

19.5 A four-entry witness inside the actual source

The necessary calculation fits in a two-dimensional subspace of each cell. Define the eight-by-two matrix

$$V = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -i & i \\ 0 & 0 \\ 0 & 0 \\ -1 & -1 \\ -i & i \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad C_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{19.12}$$

Its columns have norm one and are orthogonal. Substitution in (19.9) and (19.10) gives

$$V^*V = I, \qquad P_cV = VC_2, \qquad NV = VN_2.$$

Thus $V \otimes V$ is an isometric embedding of a four-dimensional space into the actual two-cell source, and it intertwines the source operators with

$$A_2 = C_2 \otimes I + I \otimes C_2, \qquad P_2 = N_2 \otimes (I - N_2).$$

These are restrictions of the stated native matrices, with the same inner products and expectations.

Use the coordinate order $00, 01, 10, 11$, and let

$$v_a = (i, 0, a, i)^T \qquad (a \in \mathbb{R}).$$

Direct multiplication gives

$$\|v_a\|^2 = a^2 + 2, \tag{19.13}$$

$$(A_2 - I)v_a = (a/2, i, i, a/2)^T, \tag{19.14}$$

$$\langle v_a, (A_2 - I)v_a \rangle = 0, \tag{19.15}$$

$$\|(A_2 - I)v_a\|^2 = a^2/2 + 2, \tag{19.16}$$

$$\langle v_a, [P_2, A_2]v_a \rangle = 2ia. \tag{19.17}$$

For example, $P_2 v_a = (0, 0, a, 0)^T$. The two terms in the last commutator expectation are $a^2 + ia$ and $a^2 - ia$, explaining their difference. In (19.15), the four contributions are $-ia/2, 0, ia, -ia/2$ and cancel.

Choose $a = \sqrt{2}$ and return to the source:

$$\psi = \frac{1}{2}(V \otimes V)v_{\sqrt{2}}.$$

Equations (19.13)–(19.17) then give

$$\begin{aligned} \|\psi\| = 1, \qquad & \langle \psi, (A - I)\psi \rangle = 0, \\ \|(A - I)\psi\|^2 = \frac{3}{4}, \qquad & |\langle \psi, [P, A]\psi \rangle|^2 = \frac{1}{2}. \end{aligned} \tag{19.18}$$

This one vector will suffice. We need no assumption that every source state has a large commutator expectation.

19.6 How much charge spread can the recorder supply?

Let the recorder consist of r eight-state cells and a finite neutral memory. Its space and count are

$$\mathscr{H}_R = (\mathbb{C}^8)^{\otimes r} \otimes \mathscr{H}_0, \qquad R = \sum_{j=1}^{r} P_c^{(j)} \otimes I_{\mathscr{H}_0}.$$

No product decomposition of $p \in \mathscr{H}_R$ is assumed. Its cells and memory may share arbitrary correlations.

Lemma 19.4 (A projection-sum budget)**.** *For any r orthogonal projections $R_1, \dots, R_r$ on the same space and any unit vector p,*

$$\left\| \left(\sum_{j=1}^{r} R_j - \frac{r}{2} I \right) p \right\|^2 \le \frac{r^2}{4}.$$

The projections need not commute with one another.

Proof. For each projection, $(R_j - I/2)^2 = I/4$. Hence $\|(R_j - I/2)p\| = \|p\|/2 = 1/2$. The triangle inequality gives

$$\left\| \sum_{j=1}^{r} (R_j - I/2)p \right\| \le \sum_{j=1}^{r} \|(R_j - I/2)p\| = r/2.$$

Squaring proves the claim. The empty sum also satisfies it. □

Applied to the recorder count, this proves

$$\|(R - rI/2)p\|^2 \le r^2/4. \tag{19.19}$$

The estimate already allows coherent and correlated recorder states. It has the quadratic scale required by that freedom. An independent-cell variance calculation would be a different premise and cannot be inserted here.

The complete initial state is $\psi \otimes p$. For $C = A \otimes I + I \otimes R$, we have

$$(C - (1+b)I)(\psi \otimes p) = (A - I)\psi \otimes p + \psi \otimes (R - bI)p.$$

The two terms are orthogonal because $\langle \psi, (A-I)\psi \rangle = 0$. The source is a separate tensor factor from the whole recorder, even when the recorder is internally correlated. Using (19.18),

$$\|(C - (1+b)I)(\psi \otimes p)\|^2 = \frac{3}{4} + \|(R - bI)p\|^2. \tag{19.20}$$

Also, $[P \otimes I, C] = [P, A] \otimes I$. Thus the left side of (19.8), on this input, is exactly $1/2$. We obtain the state-dependent inequality

$$\frac{1}{2} \le e\big(3 + 4\|(R - bI)p\|^2\big) \qquad \text{for every real } b. \tag{19.21}$$

19.7 The two-charge recorder theorem

We can now collect the physical requirements in one statement. A neutral memory carries the identity action of both native charges: it supplies storage and a pointer, but contributes no charge term.

Theorem 19.5 (Finite native recorder bound)**.** *Let the source be two native cells, with the event $P = N \otimes (I - N)$. Let the recorder consist of r native cells and an arbitrary finite neutral memory $\mathscr{H}_0$. Assume:*

1. *The initial state is $x \otimes p$, where p is one fixed unit recorder vector and x ranges over every unit source vector.*
2. *The interaction U is unitary on the whole source–recorder space and exactly conserves both $\mathscr{L}(H_0)$ and $\mathscr{L}(N)$.*
3. *The final pointer is $B = I_{\text{source}} \otimes I_{\text{recorder cells}} \otimes D$, where $D = D^* = D^2$ acts on the neutral memory.*
4. *For every unit source vector x,*

$$\big|\langle U(x \otimes p), BU(x \otimes p)\rangle - \langle x, Px\rangle\big| \le e.$$

Then

$$e \ge \frac{1}{2(r^2+3)}.$$

Conditions and scope.

Preparation
One fixed recorder vector p, separate from the source, and every unit source vector x; the recorder's cells and memory may be correlated with one another.

Conservation and pointer
Exact conservation of the two native charges $\mathscr{L}(H_0)$ and $\mathscr{L}(N)$ by the whole interaction, and a sharp pointer acting on the neutral memory alone.

Resources counted
The number r of charge-carrying recorder cells. The size of the neutral memory is unrestricted and does not enter.

Accuracy
Uniform event-probability error e over all source inputs.

Identification
Applying the bound to a device identifies its clock energy and occupation with the two native charges.

Not claimed
That the bound is attained, that a conserving recorder of this event exists for any r, or that other tasks obey the same constant.

Proof. Lemma 19.3 shows that U commutes with the total count C. The pointer B commutes with C because they act in different tensor factors. Thus $Z = U^*BU$ is a projection commuting with C. Its compression $M = J^*ZJ$ satisfies the uniform event bound.

Apply (19.21) with $b = r/2$, then Lemma 19.4. Accuracy implies $e \geq 0$, so the resulting inequalities have the indicated direction:

$$\frac{1}{2} \leq e\big(3 + 4\|(R - rI/2)p\|^2\big) \leq e(3 + r^2).$$

Divide by the positive number $r^2 + 3$. □

For every finite r, the lower bound is positive. An exact recorder of this event cannot satisfy all four conditions. A larger neutral memory alone does not remove the bound: its dimension never enters the charge budget. Increasing the number of charged cells relaxes the obstruction, but this inequality alone does not construct a recorder attaining it.

19.8 What the resource is, and what the result establishes

The resource appearing in the proof is the recorder's spread in the conserved count. Equation (19.21) retains the actual prepared state; (19.1) replaces that spread by a bound depending only on the number of native cells. The reference cells can exchange this count with the source while the total remains fixed. The neutral pointer reports an outcome without supplying that exchange itself.

The argument belongs to the conservation-law constraints on measurement associated with Wigner, Araki and Yanase. Its specific content here is the native realization: two stated charges imply the count conservation, an explicit embedded source vector fixes the constants, and the projection budget allows arbitrary internal recorder correlations. The statement is about uniform probability accuracy for one specified event. It is not a claim that every task has the same error bound or that the displayed constant is attainable by a conserving apparatus.

There is no conflict with the controlled constructions of Chapter 18. That chapter permits selectable pulses and records their positive energy operators. The theorem here requires an interaction conserving two exact total charges and a neutral sharp pointer. The control supply belongs in the complete physical account when those conditions are imposed. An operation that can be described and approximated with supplied controls need not be available to a closed recorder with a fixed finite charge resource.

The result sharpens the question raised by finite readout. We must ask both which distinctions a record can carry and which interactions can produce it. A physical construction answers the second question only after its preparation, conserved quantities, reference supply and readout have been specified together.

Questions to work through

1. Expand $(Z - Q)^2$ and derive (19.4). Identify every use of idempotence.
2. Verify (19.5). Explain why the two squared norms in Lemma 19.2 add to $\|x\|^2$.
3. Multiply the matrices in (19.12) to check $V^*V = I$, $P_cV = VC_2$ and $NV = VN_2$. Then check the four-entry witness.
4. Derive (19.20). Locate the exact step that uses the initial separation between source and recorder.
5. Trace where each of the four assumptions of Theorem 19.5 enters its proof. Which part of the proof permits correlations inside the recorder?

Chapter 20

What has been forced

Return to the first tally. One arrival leaves a mark; another arrival adds another. We wanted to know what survives when the marks change their shape, their material or their names. We now have several answers, because the question contains several requirements.

If a record preserves every finite step, its arithmetic structure is fixed up to the unique map that preserves the start and the step. If a reader identifies some records, the identifications must be studied as part of that reader. If a procedure answers a question at every requested precision, one rule must cover all those requests. And if an apparatus produces the record, its interaction must actually carry the required information into the available output.

These conclusions were reached by different arguments. Keeping the arguments connected, with their conditions intact, tells us what is unavoidable and what a further construction must provide.

20.1 The arithmetic structure is unique

The starting construction has two clauses: a start, and one further step. Every record is finite and has been obtained by those clauses. Structural induction follows that construction. It supplies the first proofs, including the facts that a step is never the start and that equal stepped expressions have equal predecessors. In the arithmetic language of Part II, each of these two facts, and induction, supplies statements that the other rules cannot (Section 5.7).

Now give any system a designated start and a step operation. There is exactly one map from the finite distinction orbit into that system that preserves those two operations. Theorem 2.3 constructs it: send the start to the designated start, then follow each step. Uniqueness comes from the same finite construction.

That map need not preserve every distinction. A clock that returns to a previous state is one possible target. Faithfulness and generation are the conditions that turn the unique map into an isomorphism. The target step must be injective, its start must not be a step, and every target element must be generated from the start. Under those conditions, no two finite records merge and no target element is omitted. This is Theorem 2.7.

Addition and multiplication are then fixed by their recursive clauses. Their algebraic laws follow by induction; they were not separately chosen to make the tally resemble familiar arithmetic. Conversely, a commutative semiring structure on the orbit in which adding its unit is the step must be the constructed one, and the orbit maps uniquely into every commutative semiring (Section 3.4). The reading into ordinary natural numbers preserves those operations and reflects equality. Signed differences and rational ratios extend the construction through explicit equivalence relations. Their operations are legitimate because changing a representative leaves the result in the same class. Each extension is also determined by its task (Section 4.9). The signed records complete the orbit's addition; the constructed

integers map uniquely into every commutative ring, and the constructed rationals into every field of characteristic zero; and each carries exactly one order compatible with its arithmetic. The line built from ratios in Chapter 12 is determined by its task in the same way, once countable choice is granted: every Archimedean ordered field maps into it in exactly one way that preserves the operations and the order (Theorem 12.38).

The necessity here has a precise form. A step system is faithful and generated exactly when its canonical reading is a bijection (Proposition 2.8), so any faithful generated tally has the same structure. Any operation satisfying the stated recursive clauses agrees with the constructed one; so does any associative operation for which adding 0 changes nothing and adding 1 takes one step (Theorem 3.10). The argument uses finite expressions, equality and induction in the mathematics with which the construction is studied. That induction ranges over all properties. Restricted to the formulas of an arithmetic language, it admits systems with elements that no tally reaches. Naming those resources makes the uniqueness claim exact.

20.2 Identifying records and extending logic are visible choices

A reader may retain less than the full tally. Chapter 7 describes identifications compatible with addition. Once a decidable congruence comes with an exhibited collision between distinct records, its form is forced: an initial segment followed by a repeating period. The classification proves that every collision has this form and that the least period controls all later identifications.

Changing the information supplied with the congruence, or the answer demanded of it, changes the price, and each price is exact. With a decision procedure and an exhibited collision the classification costs nothing. Replacing the collision by the assertion that the congruence is not equality costs Markov's principle MP, for decidable predicates (Definition 6.1). Keeping only the decision procedure, the parity of the index costs LLPO, the question whether the congruence is equality costs WLPO, and the full alternative costs LPO. With no decision procedure the alternative costs excluded middle. Theorems 7.6, 7.7, 7.9 and 7.10 prove these rows, and Theorem 7.11 gives the same prices for readers. In each case the reverse construction matters: the catalogue itself recovers the additional principle. These are equivalences, not merely observations that one familiar proof happens to use a stronger rule. The six prices are also different from one another. For each pair of principles, a model of intuitionistic second-order arithmetic satisfies one and refutes the other, and in each model every version of the classification has the truth value of its principle (Theorem 7.38). That system does not prove MP, so the classification from negative nontriviality genuinely needs it. Among the index–period congruences, exactly those of index zero extend to the signed records, each in one way, and the prices do not change there, provided the parity question is asked of the period: every index is zero, so its parity costs nothing, while the parity of the period costs exactly LLPO (Section 7.8).

The proof ledger of Chapter 6 makes a related distinction at the level of derivations. It records which of the specified additional rules occur, and combination retains every resource used by a premise. Soundness states which principles suffice to interpret the result, and interpreting any derivation needs at most the limited principle of omniscience in the metatheory, applied

to the predicates that formulas define; for the derivations with an EM or LPO entry, taken together, that is also the least assumption that serves. The rules themselves have exact prices. With induction on every formula, every instance of the MP rule, which is restricted to quantifier-free formulas, has a posit-free derivation, because a quantifier-free formula is eventually constant in its searched variable, and the LPO rule, which applies to every formula, derives every instance of excluded middle. The three names then mark only two strengths. Without induction, or with induction on quantifier-free formulas only, the LPO and MP rules are independent and EM is equivalent to the two together. The induction entry stays part of the accounting because these prices depend on it. The collapse concerns these arithmetic rules, not the principles about arbitrary decidable predicates priced in Chapters 7 and 12.

Chapter 9 goes further by separating rules of repetition inside a specified language. Repeating a fixed update, performing an indexed recursion and eliminating into an index-dependent family are different resources there. The countermodels satisfy the weaker rules while failing the proposed stronger judgment. The additional start discriminator has its own separating argument. These concrete models establish the strict boundaries in Theorem 9.10; the boundaries are attached to that calculus and its expressible judgments. The strongest fragment, $\mathscr{E}^{+}$, is the one compared with Peano arithmetic in Section 20.3.

Changing notation does not remove a logical cost. Where a translation is claimed to preserve a result, the translation must carry the relevant operations, assignments and quantifiers. Chapter 8 does this for numbers and lists of one mark. Every arithmetic formula has the same truth under related assignments. The proof transports witnesses in one direction and back in the other; the two round trips are what make the quantified statement possible.

With only one round trip, less transfers, and Theorem 8.8 with Proposition 8.9 says exactly what. Inside any structure obeying the six arithmetic rules the numerals form a copy of the numbers: quantifier-free statements transfer in both directions, purely existential statements go up and purely universal statements come down. A structure with two looping points shows that no other quantifier prefix transfers: a prefix containing a universal quantifier does not go up, and one containing an existential quantifier does not come down. The statement concerns all the formulas with a given prefix; a single universal sentence such as $\forall x(x + 0 = x)$ goes up because it is one of the six rules. The derivations can also be compared with an independent proof system. With excluded middle allowed and LPO and MP excluded, the rules of Chapter 5 prove exactly the theorems of Peano arithmetic, from every finite context (Theorem 8.11). With no added principle they prove exactly the theorems of Heyting arithmetic, and the posit-free derivations are the translations of its derivations, up to the way the injectivity of the successor is written (Corollary 8.12). Section 20.4 then asks, for each shape of statement, what a classical proof gives without excluded middle.

20.3 Peano arithmetic and a richer language

The first question a mathematician asks of these rules has a complete answer: with excluded middle they are Peano arithmetic, the system most mathematicians mean by arithmetic. A second question follows. Could a richer language, one that names functions and functions of functions, prove facts about numbers that those rules cannot? For sentences built from $=$, $\bot$, $\to$ and $\forall$, it cannot. Every such sentence that the richer language proves has a derivation by

the rules of Chapter 5 with no added principle, and the strongest typed fragment of Chapter 9 proves the translations of exactly the theorems of Peano arithmetic.

The proof runs the functions as programs. A proof about functions becomes a derivation, by the arithmetic rules, about the programs that compute them. Three ingredients make this possible, and the section supplies them in order: arithmetic formulas that describe computations, with derivations of their basic laws; one universal program whose correctness the rules themselves derive; and a reading of every formula about functions as a formula about programs, under which each of the nineteen rules of the richer language becomes a derivation. The central steps are proved here. Lengthy routine cases are proved in Appendix K: order and pairing, the coding of finite tables, the closure laws of computation, the test programs inside the record check, the closure generator, the reading of a variable, and the cases of conversion and of the dependent calculus.

A convention for this section. Every derivation constructed in this section is posit-free: its ledger records none of EM, LPO and MP. Its induction entry is usually true, because the derivations induct on formulas with quantifiers. We write $\Delta \vdash \varphi$ for a posit-free derivation of φ from the finite list of hypotheses Δ. Many results are *derived rules*: whenever a context Δ derives the premises, it derives the conclusion. Terms of the arithmetic language may stand for numbers in such statements; a result stated with free variables holds for every substitution of terms for them, because substitution preserves derivations. As in Chapter 5, we use named variables where the binding is clear.

20.3.1 The classical derivations are Peano arithmetic

Chapter 8 presents first-order *Peano arithmetic*, abbreviated PA, as a classical calculus of its own over the formulas of Definition 5.1: six axioms for the start, the step, addition and multiplication, the rules of natural deduction with equality, induction for every formula, and excluded middle. It keeps no ledger. Write $\Gamma \vdash_{\mathrm{PA}} \varphi$ when it derives φ from Γ. By Theorem 8.11, for every finite context Γ and every formula φ, $\Gamma \vdash_{\mathrm{PA}} \varphi$ holds exactly when the rules of Chapters 5 and 6 derive φ from Γ by a derivation whose ledger records neither LPO nor MP.

The theorem places a familiar system inside the accounting of Chapter 6. Peano arithmetic is what the rules prove when the ledger may record EM and quantified induction but neither of the other two principles. The rest of this section uses it to compare a richer language with both systems, and then asks what each system can prove about its own consistency.

20.3.2 A language of functions

The calculus of Chapter 9 can name more than numbers. We use its simplest part. The *finite types* are generated by

$$A ::= O \mid (A \to B).$$

Terms are variables $x_0, x_1, \ldots$, numbered by their distance from a binder as in Chapter 5, together with 0, $S(t)$, abstractions $\lambda A.\, b$, which bind x_0 in b, applications $f a$, iterations $\mathsf{I}(B, b, g, n)$ and indexed recursions $\mathsf{R}(B, b, g, n)$. The sans-serif I and R are the iterator and indexed recursor of Section 9.2, not the identity matrix I or the reading r of earlier chapters. A *type context* Θ lists

the types of the free variables, newest first. Seven typing rules give every term its type: x_k has the kth type of Θ, counting the newest as the 0th; $0 : O$; $S(t) : O$ when $t : O$; $\lambda A.\, b : A \to B$ when $b : B$ in the context A, Θ; $f a : B$ when $f : A \to B$ and $a : A$; $\mathsf{I}(B, b, g, n) : B$ when $b : B$, $g : B \to B$ and $n : O$; and $\mathsf{R}(B, b, g, n) : B$ when $b : B$, $g : O \to (B \to B)$ and $n : O$.

Two terms are *convertible*, $t \equiv u$, when they are related by the least equivalence relation that is closed under every term constructor and under substitution and contains the five computation equations

$$\begin{gathered} (\lambda A.\, b)\, t \equiv b[t], \\ \mathsf{I}(B, b, g, 0) \equiv b, \qquad \mathsf{I}(B, b, g, S(t)) \equiv g\, \mathsf{I}(B, b, g, t), \\ \mathsf{R}(B, b, g, 0) \equiv b, \qquad \mathsf{R}(B, b, g, S(t)) \equiv g\, t\, \mathsf{R}(B, b, g, t). \end{gathered} \tag{20.1}$$

As in Section 9.2, the intermediate terms of a conversion need not be typed. A *parallel step* $t \Rrightarrow u$ contracts any set of the redexes present in t at once: it is generated by the congruences for every constructor and by the five contractions, in which the parts of the redex may themselves take parallel steps. Write $\Rrightarrow^*$ for a finite chain of parallel steps.

Lemma 20.1 (Convertible terms meet)**.** *If $t \equiv u$, there is a term v with $t \Rrightarrow^* v$ and $u \Rrightarrow^* v$. If t has type A in Θ and $t \Rrightarrow^* v$, then v has type A in Θ.*

Proof. The first assertion is the argument of Lemma 9.12 for these five contractions. Let t° contract every redex present in t, recursively in its parts. Induction on a parallel step $t \Rrightarrow u$ shows $u \Rrightarrow t^\circ$; the substitution case uses that parallel steps commute with substitution. So two parallel steps from one term meet in one step each, finite chains meet by induction on their lengths, and having a common reduct is an equivalence relation that contains the computation equations and is closed under constructors and substitution. It therefore contains conversion. For the second assertion, induct on the parallel step and invert the typing of t at each constructor. In the abstraction contraction the body has type B in A, Θ and the argument has type A, and substituting a term of type A for the bound variable keeps type B. The four orbit contractions retype their right sides from the types of b, g and the count. □

The lemma concerns finite terms. It will be used only to choose which derivation to build.

Formulas and rules. Formulas are truth $\top$, falsehood $\bot$, equations $t =_A u$ between terms of type A, conjunctions $P \wedge Q$, implications $P \to Q$ and universal statements $\forall x{:}\, A\ P$, the last binding x_0 in P. A proof derives a judgment $\Theta; \Xi \vdash P$, in which Ξ is a finite list of hypotheses. Each rule carries the typing data of the terms it mentions; these are finite syntactic premises. Write $t{\uparrow}$, $P{\uparrow}$ and $\Xi{\uparrow}$ for the shift that raises every free variable index by one in a term, a formula or a list of formulas, as in Definition 5.2, $P[t]$ for substitution of t at position zero, written $P(t)$ in that definition, and P^S for the formula that replaces the variable x_0 of P by $S(x_0)$ and keeps every other variable. The nineteen rules are these.

(1) *Hypothesis*: $\Theta; \Xi \vdash P$ when P is a member of Ξ.

(2) *More hypotheses*: from $\Theta; \Xi \vdash P$ infer $\Theta; \Xi' \vdash P$ when every member of Ξ is a member of Ξ'.

(3) *Truth*: $\Theta; \Xi \vdash \top$.

(4) *Falsehood*: from $\Theta; \Xi \vdash \bot$ infer $\Theta; \Xi \vdash P$.

(5–7) *Conjunction*: introduction and the two eliminations.

(8–9) *Implication*: introduction, which discharges the newest hypothesis, and elimination.

(10) *Universal introduction*: from $A, \Theta; \Xi{\uparrow} \vdash P$ infer $\Theta; \Xi \vdash \forall x{:}A\ P$.

(11) *Universal elimination*: from $\Theta; \Xi \vdash \forall x{:}A\ P$ and $t : A$ infer $\Theta; \Xi \vdash P[t]$.

(12) *Reflexivity*: $\Theta; \Xi \vdash t =_A t$ for $t : A$.

(13) *Computation*: $\Theta; \Xi \vdash t =_A u$ when $t, u : A$ and $t \equiv u$.

(14) *Replacement*: from $\Theta; \Xi \vdash t =_A u$ and $\Theta; \Xi \vdash P[t]$ infer $\Theta; \Xi \vdash P[u]$, for P formed in A, Θ.

(15) *Successor is not zero*: $\Theta; \Xi \vdash S(t) =_O 0 \to \bot$.

(16) *Successor injectivity*: from $\Theta; \Xi \vdash S(t) =_O S(u)$ infer $\Theta; \Xi \vdash t =_O u$.

(17) *Function extensionality*: from $A, \Theta; \Xi{\uparrow} \vdash (f{\uparrow})\, x_0 =_B (g{\uparrow})\, x_0$ infer $\Theta; \Xi \vdash f =_{A \to B} g$.

(18) *Induction*: from $\Theta; \Xi \vdash P[0]$ and $O, \Theta; \Xi{\uparrow} \vdash P \to P^S$ infer $\Theta; \Xi \vdash \forall x{:}O\ P$, for every formula P formed in O, Θ, with parameters and quantifiers.

(19) *Substitution*: from $\Theta; \Xi \vdash P$ and a substitution σ that assigns to each variable of Θ a term of its type in Θ', infer $\Theta'; \Xi\sigma \vdash P\sigma$.

Systems of this kind are known as extensional arithmetic in all finite types [5, 6].

Arithmetic inside the language. The language has 0 and S but no addition or multiplication. They are terms defined by iteration:

$$\mathrm{add}(a, b) := \mathsf{I}\big(O, a, \lambda O.\, S(x_0), b\big), \qquad \mathrm{mul}(a, b) := \mathsf{I}\big(O, 0, \lambda O.\ \mathrm{add}(x_0, a{\uparrow}), b\big). \tag{20.2}$$

Addition starts at a and applies the successor b times; multiplication starts at 0 and adds a, b times. Under the binder, $a{\uparrow}$ is a shifted past the new variable. The defining equations are computations: $\mathrm{add}(a, 0) \equiv a$, $\mathrm{add}(a, S(b)) \equiv S(\mathrm{add}(a, b))$, $\mathrm{mul}(a, 0) \equiv 0$ and $\mathrm{mul}(a, S(b)) \equiv \mathrm{add}(\mathrm{mul}(a, b), a)$.

A formula of the arithmetic language built from $=$, $\bot$, $\to$ and $\forall$ is called *negative*. It embeds into the language of functions by the map $A \mapsto A^*$ that keeps variables, 0, S, $\bot$ and $\to$, sends $t + u$ to $\mathrm{add}(t^*, u^*)$ and $t \cdot u$ to $\mathrm{mul}(t^*, u^*)$, sends an equation $t = u$ to $t^* =_O u^*$, and sends $\forall A$ to $\forall x{:}O\, A^*$. If A is closed, A^* is formed in the empty type context. We say that the language of functions *proves* A when it proves A^* from no hypotheses in the empty context.

20.3.3 Computations described by arithmetic

The programs are those of Definition 10.5, with the numerical codes of (10.4): the basic programs zero, succ, left, right have codes 0, 1, 2, 3, and pair(f,g), rec(f,g), $f \circ g$ and search(f) have codes $4\langle u,v\rangle + 4$, $4\langle u,v\rangle + 5$, $4\langle u,v\rangle + 6$ and $4u + 7$ when f, g have codes u, v. Inside a formula a program letter stands for its code. Write id $:=$ pair(left, right), which returns its input.

Order and pairing are formulas of the arithmetic language. Write

$$a < b \ \text{ for } \ \exists k\,(a + S(k) = b), \qquad b \le a \ \text{ for } \ \exists l\,(b + l = a),$$

and let the *pairing graph* be

$$G(a,b,z) \ := \ \big(a < b \wedge z = b \cdot b + a\big) \vee \big(b \le a \wedge z = (a \cdot a + a) + b\big).$$

In $\mathbb{N}$ it says $z = \langle a,b\rangle$ for the pairing (10.3). Formulas use the graph, so no pairing symbol is added to the language.

Lemma 20.2 (Order and pairing in the rules)**.** *There are derivations from no hypotheses of: either $a < b$ or $b \le a$, and not both; for all a, b some z satisfies $G(a,b,z)$; $G(a,b,z)$ and $G(a,b,z')$ imply $z = z'$; $G(a,b,z)$ and $G(c,d,z)$ imply $a = c$ and $b = d$; and every z has some a, b with $G(a,b,z)$.*

The proofs are in Appendix K.1. They follow the arguments of Chapter 10 for the pairing (10.3), written as inductions inside the rules. Two of them, the order split and the decoding of every number, induct on formulas with quantifiers.

A successful computation is a finite table of evaluation facts. A row (k,e,n,y) says that the program with code e returns y on input n with *allowance* k: the allowance bounds the input and decreases at each step of a recursion or a search. A table is coded by two numbers through Gödel's remainder function. Let

$$\mathrm{Rem}(a,d,r) \ := \exists q, s\,\big(a = q \cdot d + r \wedge r + S(s) = d\big), \qquad \beta(A,B,i,w) \ := \mathrm{Rem}\big(A, 1 + (i+1)B, w\big).$$

Then $\beta(A,B,i,w)$ says that the ith entry of the table named by A, B is w. Row i of the table is (k,e,n,y) when its entry is the packed number $\langle k,\langle e,\langle n,y\rangle\rangle\rangle$, which we write $\mathrm{Row}(A,B,i;k,e,n,y)$, through three pairing graphs.

Table 20.1 lists the ten alternatives of the formula $\mathrm{Local}_{N,A,B,i}(K,e,n,y)$: the row at position i obeys one of the evaluation clauses of Definition 10.5, with the values it needs supplied by rows that stand later in the table. Define

$$\mathrm{Indexed}(N,A,B) \ := \forall i < N\, \exists K,e,n,y\,\big(\mathrm{Row}(A,B,i;K,e,n,y) \wedge \mathrm{Local}_{N,A,B,i}(K,e,n,y)\big),$$
$$\mathrm{App}(e,n,y) \ := \exists K,N,A,B,i\,\big(\mathrm{Indexed}(N,A,B) \wedge i < N \wedge \mathrm{Row}(A,B,i;K,e,n,y)\big).$$

The formula $\mathrm{App}(e,n,y)$ says that program e returns y on input n. It is a formula of the arithmetic language built from 0, S, $+$ and $\cdot$, with quantifiers.

Alternative	Conditions on the row (K, e, n, y), with $K = k + 1$ and $n \le k$
zero	$e = 0$, $y = 0$
successor	$e = 1$, $y = n + 1$
left	$e = 2$; $G(a, b, n)$; $y = a$
right	$e = 3$; $G(a, b, n)$; $y = b$
pairing	$e = \mathrm{pair}(f, g)$; $G(a, b, y)$; later rows (K, f, n, a) and (K, g, n, b)
composition	$e = f \circ g$; later rows (K, g, n, a) and (K, f, a, y)
recursion, start	$e = \mathrm{rec}(f, g)$; $G(a, 0, n)$; a later row (K, f, a, y)
recursion, step	$e = \mathrm{rec}(f, g)$; $G(a, m + 1, n)$; later rows $(k, e, \langle a, m\rangle, x)$ and $(K, g, \langle a, \langle m, x\rangle\rangle, y)$
search, stop	$e = \mathrm{search}(f)$; $G(a, m, n)$; $y = m$; a later row $(K, f, n, 0)$
search, continue	$e = \mathrm{search}(f)$; $G(a, m, n)$; $0 < x$; later rows (K, f, n, x) and $(k, e, \langle a, m + 1\rangle, y)$

Table 20.1. The local condition of a row. A later row is a row at a position j with $i < j < N$. Calls of a part of the program keep the allowance K; the two calls that repeat the program, the previous step of a recursion and the rest of a search, use $k = K - 1$. The letters a, b, m, x, f, g are existentially quantified, and every pairing is read through its graph.

Theorem 20.3 (Computation inside arithmetic)**.**

1. *For every program c and all numbers n, y, $\mathrm{App}(c, n, y)$ holds in $\mathbb{N}$ exactly when c returns y on n.*

2. Functionality*: there is a derivation from no hypotheses of* $\mathrm{App}(e, n, y) \to \mathrm{App}(e, n, y') \to y = y'$.

3. Closure*: there are derivations from no hypotheses of the following laws, one for each evaluation clause, with every variable universally quantified:*

$$\mathrm{App}(\mathrm{zero}, n, 0), \quad \mathrm{App}(\mathrm{succ}, n, n + 1), \quad \mathrm{App}(\mathrm{id}, n, n),$$
$$G(a, b, n) \to \mathrm{App}(\mathrm{left}, n, a), \qquad G(a, b, n) \to \mathrm{App}(\mathrm{right}, n, b),$$
$$G(a, b, y) \to \mathrm{App}(f, n, a) \to \mathrm{App}(g, n, b) \to \mathrm{App}(\mathrm{pair}(f, g), n, y),$$
$$\mathrm{App}(g, n, a) \to \mathrm{App}(f, a, y) \to \mathrm{App}(f \circ g, n, y),$$
$$G(a, 0, n) \to \mathrm{App}(f, a, y) \to \mathrm{App}(\mathrm{rec}(f, g), n, y),$$
$$G(a, m + 1, n) \to G(a, m, p) \to G(m, x, v) \to G(a, v, q) \to$$
$$\mathrm{App}(\mathrm{rec}(f, g), p, x) \to \mathrm{App}(g, q, y) \to \mathrm{App}(\mathrm{rec}(f, g), n, y),$$
$$G(a, m, n) \to \mathrm{App}(f, n, 0) \to \mathrm{App}(\mathrm{search}(f), n, m),$$
$$G(a, m, n) \to 0 < x \to G(a, m + 1, p) \to \mathrm{App}(f, n, x) \to$$
$$\mathrm{App}(\mathrm{search}(f), p, y) \to \mathrm{App}(\mathrm{search}(f), n, y).$$

   Here f, g range over codes, and the code of a compound program is given through its pairing graph, as in $e = 4p + 6$ with $G(f, g, p)$.

4. Finite choice*: for every formula $\varphi(i, r)$ there is a derivation of*

$$\forall i\, \exists r\, \varphi(i, r) \to \forall j\, \exists A, B\, \forall i < j\, \exists r \big(\beta(A, B, i, r) \wedge \varphi(i, r)\big).$$

Each derivation is posit-free and records the induction entry.

The proofs are in Appendix K.2. The first assertion lists the nodes of a computation tree with each parent before its children, and conversely rebuilds a tree from an indexed table by induction on the number of rows that follow a selected row. Functionality compares two independent tables by strong induction on that same number, which shrinks along every dependency of the first table; the bounds of the two tables need not agree. Finite choice appends one entry at a time to a coded table, using a Chinese remainder construction whose coefficients are written down explicitly. The closure laws raise the bounds of the premise tables, place them in one table, and add the parent row in front.

20.3.4 A universal program whose correctness the rules derive

How can rules that speak only about numbers check a proof that quantifies over functions? By running the functions. Each term of the language of functions will be compiled into a program, a function will be represented by the code of its program, and it will be applied through one fixed *universal program* U. (The sans-serif U is this program throughout; the italic U of Chapters 18 and 19 is a unitary matrix.) On an input $\langle e, n\rangle$, the program U searches $c = 0, 1, 2, \ldots$ for a number that a fixed *record test* accepts as a record of a computation of program e on input n, and returns the output written in that record.

Every compiled application runs through U, and so does every step of a compiled iteration. The compiled proofs therefore need one property of U as a derivation of the arithmetic rules, not merely as a true statement: if e returns y on n, then U returns y on $\langle e, n\rangle$. That property depends on how the record and its test are built.

A record that carries its justifications. The table formula of Table 20.1 would be a poor test. Its local condition says that suitable rows exist somewhere later in the table, so checking one row would itself require a search, and a derivation about the test would have to follow that inner search. The record removes the inner searches by writing down, for every row, the values its justification uses.

A record therefore has two tables. The first is a computation table (N, A, B). The second, (C, D), holds one *justification* for each row:

$$w_j = \langle q, s, f, g, u, v, x, j_1, j_2\rangle,$$

a list packed by repeated pairing, $\langle x_0, \ldots, x_m\rangle = \langle \cdots \langle\langle x_0, x_1\rangle, x_2\rangle \cdots, x_m\rangle$. Its eight read fields are the predecessor s of the row's allowance, the codes f, g of the parts of the row's program, the components u, v of a paired value the clause builds or decodes, the previous recursion value or the positive test value x, and the positions j_1, j_2 of the rows the justification relies on. The first field q is never read. A clause that needs fewer values ignores the remaining fields.

Definition 20.4 (The explicit row check and a valid record)**.** Let (N, A, B) be a table whose row j is (K, e, n, y). A number w, read as $\langle q, s, f, g, u, v, x, j_1, j_2\rangle$, *passes the explicit check at row* j when $K = s + 1$, $n \le s$, and one of the ten cases of Table 20.2 holds; there $j' \mapsto (k', e', n', y')$ means that $j < j' < N$ and row j' is (k', e', n', y'). A number

$$c = \langle q, C, D, N, A, B, i, K, y\rangle$$

is a *valid record* for $\langle e, n\rangle$ when, for every $j < N$, the w with $\beta(C, D, j, w)$ passes the explicit check at row j, when $i < N$, and when row i is (K, e, n, y). Its *output* is y, its last entry.

Case	Program e	Further conditions
zero	$e = 0$	$y = 0$
successor	$e = 1$	$y = n + 1$
left	$e = 2$	$n = \langle u, v\rangle$, $y = u$
right	$e = 3$	$n = \langle u, v\rangle$, $y = v$
pairing	$e = 4\langle f, g\rangle + 4$	$y = \langle u, v\rangle$, $j_1 \mapsto (K, f, n, u)$, $j_2 \mapsto (K, g, n, v)$
composition	$e = 4\langle f, g\rangle + 6$	$j_1 \mapsto (K, g, n, u)$, $j_2 \mapsto (K, f, u, y)$
recursion, start	$e = 4\langle f, g\rangle + 5$	$n = \langle u, 0\rangle$, $j_1 \mapsto (K, f, u, y)$
recursion, step	$e = 4\langle f, g\rangle + 5$	$n = \langle u, v + 1\rangle$, $j_1 \mapsto (s, e, \langle u, v\rangle, x)$, $j_2 \mapsto (K, g, \langle u, \langle v, x\rangle\rangle, y)$
search, stop	$e = 4f + 7$	$n = \langle u, v\rangle$, $y = v$, $j_1 \mapsto (K, f, n, 0)$
search, continue	$e = 4f + 7$	$n = \langle u, v\rangle$, $x > 0$, $j_1 \mapsto (K, f, n, x)$, $j_2 \mapsto (s, e, \langle u, v+1\rangle, y)$

Table 20.2. The ten cases of the explicit row check. The guard $K = s + 1$, $n \le s$ is common to all of them. They are the ten alternatives of Table 20.1, with every intermediate value and every supporting position written into the justification.

Every condition in the table is an equation, an inequality, or a comparison of one named row with four given numbers. The check is told which rows to look at and which intermediate values to use. It searches for nothing inside a row.

A worked record. The program succ $\circ$ succ has code $4\langle 1, 1\rangle + 6 = 18$, since $\langle 1, 1\rangle = 3$. On input 0 it returns 2. List the parent row first:

$$\text{row } 0 = (2, 18, 0, 2), \qquad \text{row } 1 = (2, 1, 0, 1), \qquad \text{row } 2 = (2, 1, 1, 2).$$

The allowance is 2 because the third row has input 1, and allowance $s + 1$ admits inputs up to s. The justifications are $w_0 = \langle 0, 1, 1, 1, 1, 0, 0, 1, 2\rangle$ and $w_1 = w_2 = \langle 0, 1, 0, 0, 0, 0, 0, 0, 0\rangle$. In w_0, the predecessor is $s = 1$, the part codes are $f = g = 1$, the intermediate value is $u = 1$, and the supporting rows are $j_1 = 1$ and $j_2 = 2$. The composition case holds at row 0: its guard is $2 = 1 + 1$ with $0 \le 1$; rows 1 and 2 are later rows below 3; row 1 is $(K, g, n, u) = (2, 1, 0, 1)$; and row 2 is $(K, f, u, y) = (2, 1, 1, 2)$. Rows 1 and 2 pass the successor case. With (A, B) coding the three packed rows and (C, D) the three justifications, $\langle 0, C, D, 3, A, B, 0, 2, 2\rangle$ is a valid record for $\langle 18, 0\rangle$ with output 2.

The test and the universal program. The checking programs are built from the eight constructors without the search constructor; their constructor expressions are given or described in Appendix K.3. A *test* is a program whose value is read as a verdict, positive for yes. The *zero indicator* ζ returns 1 on 0 and 0 elsewhere. The equality test of two programs adds the two truncated differences of their values and applies ζ; the test $a \le b$ applies ζ to $a \dot{-} b$; conjunction and disjunction of tests multiply and add their values. The *recognizer* Ver receives $\langle\langle e, n\rangle, c\rangle$, reads the fields of c by projections, and returns the conjunction of three tests: the bounded universal test over $j < N$ of the explicit row check at row j, the order test $i < N$, and the comparison of row i with (K, e, n, y). Then

$$T := \zeta \circ \mathrm{Ver}, \qquad \mathsf{U} := \mathrm{right} \circ \mathrm{search}_0(T), \qquad \mathrm{search}_0(T) := \mathrm{search}(T) \circ \mathrm{pair}(\mathrm{id}, \mathrm{zero}).$$

The test T returns 0 to accept and 1 to reject. On input p, the program $\mathrm{search}_0(T)$ tests T at $\langle p, 0\rangle, \langle p, 1\rangle, \ldots$ and returns the first c with $T(\langle p, c\rangle) = 0$; the final right projection returns the last entry of c, which is the output written in the record. In $\mathbb{N}$, by Theorem 20.3(1), U returns y on $\langle e, n\rangle$ exactly when e returns y on n. The compiled proofs need more than that truth: they need derivations.

Theorem 20.5 (The record test is correct)**.** *The rules of Chapter 5 give posit-free derivations, from no hypotheses, of the following two statements.*

Acceptance. *If* $\mathrm{Indexed}(N, A, B)$, $i < N$, $\mathrm{Row}(A, B, i; K, e, n, y)$ *and* $G(e, n, p)$*, then for some* c *and* z *with* $G(p, c, z)$*, the test returns* 0 *on* z*:* $\mathrm{App}(T, z, 0)$.

Record soundness. *If* $G(e, n, p)$*,* $G(p, c, z)$ *and* $\mathrm{App}(T, z, 0)$*, then there are* y *and* q *with* $G(q, y, c)$ *and* $\mathrm{App}(e, n, y)$.

Each derivation records the induction entry and none of EM, LPO and MP.

The proof rests on three lemmas. The first runs the checking programs; it is proved in Appendix K.3 by following the program, using the closure laws of Theorem 20.3 at each constructor and induction along each recursion.

Lemma 20.6 (The row check returns and decides)**.** *There are derivations of the following, for all* X, j, p *such that the field programs return* A, B, N, C, D *on* X *and* $G(X, j, p)$*: the explicit row-check program returns a value on* p*, and that value is positive exactly when some* w *with* $\beta(C, D, j, w)$ *passes the explicit check at row* j *of* (N, A, B)*. The bounded universal test over* $j < N$ *returns a value, which is positive exactly when the row check is positive at every* $j < N$*. The test* T *returns* 0 *or* 1 *on every input.*

Lemma 20.7 (A justification and the local condition say the same thing)**.** *There are derivations of: if* w *passes the explicit check at row* j *of* (N, A, B)*, then row* j *has some coordinates* (K, e, n, y) *satisfying* $\mathrm{Local}_{N,A,B,j}(K, e, n, y)$*; and conversely, if row* j *is* (K, e, n, y) *and* $\mathrm{Local}_{N,A,B,j}(K, e, n, y)$ *holds, then some* w *passes the explicit check at row* j.

Proof. For the first, open the nine fields of w and the coordinates of row j. The guard gives the allowance predecessor and the input bound of the local condition. Each case of Table 20.2

gives the alternative of Table 20.1 with the same name: a named row $j_1 \mapsto (\cdots)$ is a later row, with j_1 as the witness and $j < j_1 < N$ together with the row comparison as its matrix; the paired values are the pairing-graph witnesses; and f, g are the code witnesses. Disjunction elimination over the ten cases and existential introduction finish the derivation, which uses no induction.

For the converse, open the alternative that holds and its witnesses: the allowance predecessor, the part codes, the paired values and the later positions. These are the eight read fields; set a field the alternative does not use to 0. Pairing totality (Lemma 20.2), applied eight times starting from 0, gives $w = \langle 0, s, f, g, u, v, x, j_1, j_2 \rangle$ together with the pairing graphs that the check reads back, and the case with the same name holds of w. □

Lemma 20.8 (Collecting the justifications)**.** *There is a derivation of: if* Indexed(N, A, B)*, then there are C, D such that for every $j < N$ some w with $\beta(C, D, j, w)$ passes the explicit check at row j.*

Proof. Apply finite choice, Theorem 20.3(4), to the formula $\varphi(i, w)$: if $i < N$, then w passes the explicit check at row i. Its premise holds at every i. By the order split, either $i < N$, and then indexedness gives row i with its local condition and Lemma 20.7 gives a justification; or $N \leq i$, and then $i < N$ is refuted by the disjointness of the two order branches, so $w = 0$ serves. Finite choice at $j = N$ gives C, D; for $j < N$, the stored w satisfies $\varphi(j, w)$, and $j < N$ discharges its implication. □

Proof of Theorem 20.5. Acceptance. Lemma 20.8 gives C, D holding a justification for every row below N. Pairing totality packs the record $c = \langle 0, C, D, N, A, B, i, K, y \rangle$ and the input z with $G(p, c, z)$, and supplies every pairing graph used in the packing. On z the field programs return the packed fields, and the projections of p return e and n. At every cell $\langle z, j \rangle$ with $j < N$ the stored justification passes the explicit check, so by Lemma 20.6 the row check is positive there and the bounded universal test is positive on z. The order test for $i < N$ and the comparison of row i with (K, e, n, y) return values, positive by hypothesis, by Lemma K.5. Their product is a positive value $r = d + 1$ of Ver on z, and ζ sends a successor to 0, so the composition law gives App$(T, z, 0)$.

Record soundness. Decode c into $\langle q_0, C, D, N, A, B, i, K, y \rangle$ with the decoder of Lemma 20.2; its last pairing gives $G(q, y, c)$. By Lemma 20.6, Ver returns some value r on z, and ζ returns some b on r, so App(T, z, b). Functionality, Theorem 20.3(2), and the hypothesis App$(T, z, 0)$ give $b = 0$. The zero indicator returns 0 only on a successor, so r is positive. The value r is the product of three test values, and a product of natural numbers is positive exactly when its factors are, so each test value is positive, and Lemmas K.5 and 20.6 read off what each says: $i < N$, row i is (K, e, n, y), and the row check is positive at every $j < N$. Lemma 20.6 gives each such row a passing justification and Lemma 20.7 its local condition. That is Indexed(N, A, B), and with $i < N$ and row i it is App(e, n, y).

Every step either applies a derived lemma or is logic, equality and arithmetic on supplied witnesses. Every case split eliminates a disjunction that has itself been derived. No EM, LPO or MP rule occurs. □

The universal program does not look for the record that acceptance built. It looks for the first number the test accepts, which may be a different record. Two facts keep that difference harmless.

Lemma 20.9 (The search finds a record)**.** *There is a derivation of: if* $G(p, c, z)$ *and* $\mathrm{App}(T, z, 0)$*, then* $\mathrm{search}_0(T)$ *returns on* p *some* $d \le c$*, and* T *returns* 0 *on the* w *with* $G(p, d, w)$*.*

Proof. Prove more, by induction on h in a formula universal in s: for every gap h and every s, if $s + h = c$ and T returns 0 on $\langle p, c\rangle$, then the search started at $\langle p, s\rangle$ returns some $d \le c$ at which T returns 0. Induct on h. If $h = 0$, then $s = c$ and the search-stop law returns s. For $h + 1$, the test returns 0 or 1 at $\langle p, s\rangle$ (Lemma 20.6). If it returns 0, the search stops at s. If it returns 1, the search-continue law passes to $s + 1$, whose gap is h, and the induction hypothesis gives d. At $s = 0$, the identity and zero laws and the pair law give the input $\langle p, 0\rangle$, and the composition law connects $\mathrm{search}_0(T)$ on p with the search started there. □

Lemma 20.10 (Universal application)**.** *The rules of Chapter 5 give a posit-free derivation, from no hypotheses, of*

$$\forall e, n, y, p\ \big(G(e, n, p) \to \mathrm{App}(e, n, y) \to \mathrm{App}(\mathsf{U}, p, y)\big).$$

If program e *returns* y *on input* n*, then* U *returns* y *on* $\langle e, n\rangle$*.*

Proof. Assume $G(e, n, p)$ and $\mathrm{App}(e, n, y)$, and open the application: an allowance K and a table (N, A, B) with $\mathrm{Indexed}(N, A, B)$, $i < N$ and $\mathrm{Row}(A, B, i; K, e, n, y)$.

1. Acceptance gives c, z with $G(p, c, z)$ and $\mathrm{App}(T, z, 0)$.
2. Lemma 20.9 gives d, w with $\mathrm{App}(\mathrm{search}_0(T), p, d)$, $G(p, d, w)$ and $\mathrm{App}(T, w, 0)$.
3. Record soundness at e, n, p, d, w gives y', q with $G(q, y', d)$ and $\mathrm{App}(e, n, y')$.
4. Functionality gives $y' = y$.
5. The projection law gives $\mathrm{App}(\mathrm{right}, d, y')$.
6. The composition law for $\mathsf{U} = \mathrm{right} \circ \mathrm{search}_0(T)$ combines steps 2 and 5 into $\mathrm{App}(\mathsf{U}, p, y')$.
7. Equality substitution with $y' = y$ gives $\mathrm{App}(\mathsf{U}, p, y)$.

The conclusion contains none of the temporary witnesses, so every existential elimination closes. The derivation never compares the record found by the search with the record built by acceptance: record soundness says that any accepted record describes an actual computation, and functionality says that all such computations have one output. □

20.3.5 Equality that follows the type

A function is represented by the code of a program. Two codes need not be equal numbers to represent the same function, so equality at a function type must compare what the programs do. It must also require the computations to return. Define formulas $e \simeq_A f$ by recursion on the type:

$$\begin{aligned} e \simeq_O f &:= e = f, \\ e \simeq_{A\to B} f &:= \forall x, y \left(x \simeq_A y \to \exists u, v \left(\mathrm{App}(e, x, u) \wedge \mathrm{App}(f, y, v) \wedge u \simeq_B v\right)\right), \end{aligned}$$

and call e *admissible* at A when $H_A(e) := e \simeq_A e$. This is *hereditary extensional equality*: at an arrow type it asks for actual outputs on every pair of related inputs, and relates the outputs at the result type. Every number is admissible at O. At an arrow type a code is admissible only when it returns related outputs on related admissible inputs; there is no blanket assumption that programs terminate. Each formula $e \simeq_A f$ is a formula of the arithmetic language, since its clauses use App and quantifiers over numbers only.

Proposition 20.11 (Symmetry and transitivity)**.** *For every finite type A there are derivations from no hypotheses of*

$$e \simeq_A f \to f \simeq_A e \qquad \text{and} \qquad e \simeq_A f \to f \simeq_A g \to e \simeq_A g.$$

Proof. Recurse on A. At O both are the symmetry and transitivity of equality. At $A \to B$, for symmetry take $x \simeq_A y$; symmetry at A gives $y \simeq_A x$; the hypothesis at y, x gives outputs u of e on y and v of f on x with $u \simeq_B v$; return v and u, related by symmetry at B.

For transitivity at $A \to B$, take $x \simeq_A y$. Symmetry and transitivity at A give $y \simeq_A y$. The first hypothesis at x, y gives outputs u of e on x and v of f on y with $u \simeq_B v$; the second at y, y gives outputs v' of f on y and w of g on y with $v' \simeq_B w$. Functionality, Theorem 20.3(2), gives $v = v'$; equality substitution gives $v \simeq_B w$; transitivity at B gives $u \simeq_B w$. Introduce u, w. □

Transitivity at an arrow type is the first place where the derivation, and not just the truth, of functionality is needed: the two hypotheses supply two outputs of f on y, and only a derivation that they are equal lets the argument continue.

Environments. The values of the free variables of a term are stored in one number, its *environment*, with the newest value on the right: an environment for A, Θ is a number R with $G(r, x, R)$, where x is the value of x_0 and r is an environment for Θ. Define the *context condition* and the *relation of two environments* by recursion on the type context:

$$\begin{aligned} C_{[\,]}(R) &:= (0 = 0), \\ C_{A,\Theta}(R) &:= \exists r, x \left(G(r, x, R) \wedge C_\Theta(r) \wedge H_A(x)\right), \\ R \approx_{[\,]} R' &:= (0 = 0), \\ R \approx_{A,\Theta} R' &:= \exists r, r', x, x' \left(G(r, x, R) \wedge G(r', x', R') \wedge r \approx_\Theta r' \wedge x \simeq_A x'\right). \end{aligned}$$

The number below the oldest declared value is unrestricted. A derivation of $C_\Theta(R)$ gives one of $R \approx_\Theta R$: open each level and use its admissibility $H_A(x)$, which is $x \simeq_A x$. We call this the *diagonal* of the context condition.

The reading of a formula. Let $c(t)$ be the program compiled from the term t, to be defined in the next subsection, and write $E_t(R, y) := \mathrm{App}(c(t), R, y)$, a formula about programs and not the value $E_\rho(t)$ of Definition 5.3: the compiled term returns y at the environment R. The *reading* $\mathrm{Rd}_R(P)$ of a formula P at an environment R is the arithmetic formula defined by

$$\begin{aligned}
\mathrm{Rd}_R(\top) &:= (0 = 0), \qquad \mathrm{Rd}_R(\bot) := \bot,\\
\mathrm{Rd}_R(t =_A u) &:= \exists x, y\,\big(E_t(R, x) \wedge E_u(R, y) \wedge x \simeq_A y\big),\\
\mathrm{Rd}_R(P \wedge Q) &:= \mathrm{Rd}_R(P) \wedge \mathrm{Rd}_R(Q), \qquad \mathrm{Rd}_R(P \to Q) := \mathrm{Rd}_R(P) \to \mathrm{Rd}_R(Q),\\
\mathrm{Rd}_R(\forall x{:}\,A\ P) &:= \forall x\,\big(H_A(x) \to \forall z\,(G(R, x, z) \to \mathrm{Rd}_z(P))\big).
\end{aligned}$$

An equation becomes the statement that the two compiled programs return values and that the values are hereditarily equal. A universal quantifier over a type ranges over the admissible numbers, and the body is read at the environment extended by the new value; the pairing graph supplies that environment, so no pairing symbol is needed. If P has no free variables, $\mathrm{Rd}_0(P)$ is a closed arithmetic formula.

Theorem 20.12 (The reading is exact)**.** *Let* $\mathrm{Holds}(P, R)$ *be read directly in* $\mathbb{N}$*: an equation holds when both compiled terms return values that are hereditarily equal in* $\mathbb{N}$*, the connectives are read as themselves, and* $\forall x{:}\,A\ P$ *holds at* R *when* P *holds at* $\langle R, x\rangle$ *for every admissible* x*. Then* $\mathrm{Rd}_R(P)$ *is true in* $\mathbb{N}$ *exactly when* $\mathrm{Holds}(P, R)$*. Likewise* $C_\Theta(R)$ *is true exactly when every variable declared in* Θ *has an admissible value at its type under lookup from* R*.*

Proof. Induct on P. The equation clause has two existential witnesses, and by Theorem 20.3(1) they are exactly the two returned values. The connective clauses follow from the induction hypotheses. For a universal formula, if the reading holds and x is admissible, take $z = \langle R, x\rangle$; the graph holds, and the induction hypothesis gives the body at $\langle R, x\rangle$. Conversely, any z with $G(R, x, z)$ is $\langle R, x\rangle$. The context condition is the same induction on Θ: a witness pair $\langle r, x\rangle = R$ supplies the newest value and, by the induction hypothesis, the older ones. □

The theorem says what the reading means. A proof translation needs more: for every proof, an actual derivation of the reading, for every environment and every hypothesis list. The statement that a computation returns in $\mathbb{N}$ does not supply such a derivation for arbitrary variables. The rest of the section constructs the derivations, rule by rule.

20.3.6 Compiled terms return admissible values

Every term t receives a fixed program $c(t)$, by recursion on the term:

$$\begin{aligned}
c(x_k) &:= \mathrm{right} \circ d_k, \qquad c(0) := \mathrm{zero}, \qquad c(S(t)) := \mathrm{succ} \circ c(t),\\
c(\lambda A.\, b) &:= \Lambda_{c(b)}, \qquad c(fa) := \mathsf{U} \circ \mathrm{pair}(c(f), c(a)),
\end{aligned}$$

$$\begin{aligned}
c(\mathsf{I}(B, b, g, n)) &:= \mathsf{It} \circ \mathrm{pair}\big(\mathrm{pair}(c(g), c(b)), c(n)\big),\\
c(\mathsf{R}(B, b, g, n)) &:= \mathsf{It}^{\mathrm{ind}} \circ \mathrm{pair}\big(\mathrm{pair}(c(g), c(b)), c(n)\big).
\end{aligned}$$

The fixed programs are these. The *dropping programs* $d_0 := \mathrm{id}$ and $d_{k+1} := d_k \circ \mathrm{left}$ remove k entries from an environment, so $c(x_k)$ reads the kth entry. The *iterator* and *indexed recursor* are

$$\mathsf{It} := \mathrm{rec}(\mathrm{right}, \mathrm{st}), \qquad \mathsf{It}^{\mathrm{ind}} := \mathrm{rec}(\mathrm{right}, \mathrm{st}^{\mathrm{ind}}),$$

with steps $\mathrm{st} := \mathsf{U} \circ \mathrm{pair}(\mathrm{left} \circ \mathrm{left}, \mathrm{right} \circ \mathrm{right})$ and $\mathrm{st}^{\mathrm{ind}} := \mathrm{ap}(\mathrm{ap}(\mathrm{left} \circ \mathrm{left}, \mathrm{left} \circ \mathrm{right}), \mathrm{right} \circ \mathrm{right})$, where $\mathrm{ap}(f, h) := \mathsf{U} \circ \mathrm{pair}(f, h)$. They receive $\langle a, m \rangle$ with $a = \langle g, b \rangle$ storing a step code and a seed, and m the count. At count 0 they return the seed b. At count $m + 1$, with previous value x, the step receives $\langle a, \langle m, x \rangle \rangle$: st runs g on x through U, and $\mathrm{st}^{\mathrm{ind}}$ runs g on m and the resulting code on x. The *closure generator* Λ_c of a program c receives an environment R and returns the code of the program $c \circ \mathrm{pair}(\kappa_R, \mathrm{id})$, where κ_R is a program that returns R on every input; it builds this code without running c. Its definition is in Appendix K.4.

A function value thus returns at once, before its body runs. What a proof needs is the behavior of that value.

Lemma 20.13 (What the closure does)**.** *For every program c there is a derivation of: for every R there is an f with $\mathrm{App}(\Lambda_c, R, f)$ such that, for all x, p, y, if $G(R, x, p)$ and $\mathrm{App}(c, p, y)$, then $\mathrm{App}(f, x, y)$.*

The proof is in Appendix K.4. The generator contains no search, so it returns on every input; the code it returns is that of a composition whose first part pairs a quoted constant with the identity, and the quoted constant is shown to return R by induction on R.

Lemma 20.14 (Iterations of related steps are related)**.** *For every finite type B there are derivations of the following, for all g, h, b, b', a, a' with $G(g, b, a)$ and $G(h, b', a')$:*

1. *if $g \simeq_{B \to B} h$ and $b \simeq_B b'$, then for every m there are p, p', x, x' with $G(a, m, p)$, $G(a', m, p')$, $\mathrm{App}(\mathsf{It}, p, x)$, $\mathrm{App}(\mathsf{It}, p', x')$ and $x \simeq_B x'$;*
2. *if $g \simeq_{O \to (B \to B)} h$ and $b \simeq_B b'$, the same holds with $\mathsf{It}^{\mathrm{ind}}$ in place of It.*

Proof. For the first statement, induct on m in the displayed formula, keeping the parameters and hypotheses. At 0, pairing totality supplies p, p', and the recursion-start law, with $\mathrm{App}(\mathrm{right}, a, b)$ from $G(g, b, a)$, gives $\mathrm{App}(\mathsf{It}, p, b)$; likewise for p'. The seeds are related by hypothesis. At $m + 1$, open the induction hypothesis: outputs $x \simeq_B x'$ of the two iterations at count m. The arrow clause of $g \simeq_{B \to B} h$ at x, x' gives outputs s of g on x and s' of h on x' with $s \simeq_B s'$. Pairing totality supplies the packed inputs $\langle a, m + 1 \rangle$, $\langle m, x \rangle$ and $\langle a, \langle m, x \rangle \rangle$, and their counterparts for the second iteration. On $\langle a, \langle m, x \rangle \rangle$ the projections in st return g and x, Lemma 20.10 carries $\mathrm{App}(g, x, s)$ to $\mathrm{App}(\mathsf{U}, \langle g, x \rangle, s)$, and the composition law gives the step's output s. The recursion-step law then gives $\mathrm{App}(\mathsf{It}, \langle a, m + 1 \rangle, s)$. The same argument for the second iteration gives s', and $s \simeq_B s'$. The indexed case is the same induction with two arrow clauses at each step: $g \simeq h$ at the equal counts m gives related codes, and those codes at x, x' give related outputs; the step $\mathrm{st}^{\mathrm{ind}}$ runs both applications through U. □

The induction formula asserts that computations exist, so it contains quantifiers and the ledger records the induction entry. Taking $h = g$ and $b' = b$, the lemma says that an iteration of an admissible step from an admissible seed returns an admissible value at every count.

Theorem 20.15 (Typed evaluation)**.** *Let t have type A in Θ. Every context that derives $R \approx_\Theta R'$ derives*

$$\exists x, x' \ \big(E_t(R, x) \wedge E_t(R', x') \wedge x \simeq_A x'\big).$$

In particular, every context that derives $C_\Theta(R)$ derives $\exists x\,(E_t(R, x) \wedge H_A(x))$.

Proof. Induct on the typing derivation of t; there are seven cases.

Variable x_k. Open the relation $R \approx_\Theta R'$ level by level, k times on each side: it supplies environments r_j, r'_j with $G(r_{j+1}, v_j, r_j)$ and $G(r'_{j+1}, v'_j, r'_j)$, $r_0 = R$, $r'_0 = R'$, and at level k the related entries $v_k \simeq_A v'_k$. The left-projection law gives $\mathrm{App}(\mathrm{left}, r_j, r_{j+1})$, the composition law assembles $\mathrm{App}(d_k, R, r_k)$, and the right-projection law gives $\mathrm{App}(\mathrm{right} \circ d_k, R, v_k)$; the same on the other side. The written-out induction on k is in Appendix K.5.

Zero. The zero law gives output 0 on both sides, and $0 = 0$.

Successor $S(t)$. The induction hypothesis gives outputs $x \simeq_O x'$, that is $x = x'$, of the compiled t; the successor and composition laws give the output $x + 1$ on both sides.

Application fa with $f : A \to B$. The induction hypotheses give outputs $e \simeq_{A\to B} e'$ of the compiled f and $v \simeq_A v'$ of the compiled a. The arrow clause gives outputs y of e on v and y' of e' on v' with $y \simeq_B y'$. Pairing totality gives p with $G(e, v, p)$; the pair law gives $\mathrm{App}(\mathrm{pair}(c(f), c(a)), R, p)$; universal application gives $\mathrm{App}(\mathsf{U}, p, y)$; and the composition law gives $E_{fa}(R, y)$. The other side is the same.

Iteration $\mathsf{I}(B, b, g, n)$. The induction hypotheses give related outputs of the three parts: $g \simeq_{B\to B} g'$ for the step, $v \simeq_B v'$ for the seed, and equal counts $m = m'$. Pairing totality gives a, a' with $G(g, v, a)$ and $G(g', v', a')$. Lemma 20.14 at the count m gives related outputs $y \simeq_B y'$ of It on $\langle a, m\rangle$ and $\langle a', m\rangle$. The pair law, applied twice, and the composition law assemble $E_t(R, y)$ and $E_t(R', y')$ from these pieces. *Indexed recursion* is the same, with the second part of the lemma.

Abstraction $\lambda A.\, b$ with $b : B$ in A, Θ. Lemma 20.13 gives closure codes f at R and f' at R'. It remains to derive $f \simeq_{A\to B} f'$. Take $x \simeq_A x'$, and let $G(R, x, z)$ and $G(R', x', z')$ by pairing totality. Then $z \approx_{A,\Theta} z'$, with witnesses R, R', x, x'. The induction hypothesis for the body gives outputs $y \simeq_B y'$ of $c(b)$ at z and z', and the closure behavior gives $\mathrm{App}(f, x, y)$ and $\mathrm{App}(f', x', y')$. That is the arrow clause.

For the last statement apply the first to the diagonal $R \approx_\Theta R$. The two outputs $x \simeq_A x''$ give $H_A(x)$ by symmetry and transitivity, Proposition 20.11. □

The two-environment form matters. The abstraction case needs the induction hypothesis at two different extended environments z, z', which agree only up to the relation; a statement about one environment at a time would not give the arrow clause.

20.3.7 Substitution moves with the environment

A quantified statement is used at a particular argument only after the argument has been placed in the environment where the body is read. Suppose $P(x, z)$ is a formula and $t(z)$ a term of the type of x. The usual conclusion from $\forall x\, P(x, z)$ is $P(t(z), z)$. The value of z is still the old one; the new value of x must be the value of t in that old environment. Substituting first

and then reading must agree with extending the environment by that value and then reading the body. This subsection derives that agreement for every formula, in both directions.

A *substitution* σ from Θ into Θ' assigns to each variable x_k of Θ, of type B, a term σ_k of type B in Θ'. The substituted term $t\sigma$ and formula $P\sigma$ replace each x_k by σ_k. Under a binder the substitution is *lifted*: the lifted substitution σ^+ keeps the new variable x_0 and sends x_{k+1} to $\sigma_k\uparrow$, the term σ_k shifted past the new variable. This is the positional substitution of Definition 5.2, and the lifting is what prevents capture. Let R be an environment for Θ' and R' one for Θ. Say that σ is *faithful from R to R'* when every variable is matched: for each x_k of Θ, of type B,

$$\exists p, q\, \big(E_{\sigma_k}(R, p) \wedge E_{x_k}(R', q) \wedge p \simeq_B q\big).$$

The term substituted for x_k, evaluated in the new environment, returns a value related to the entry that R' stores for x_k.

A substitution whose terms are variables is a *renaming*. The simplest is the shift $x_k \mapsto x_{k+1}$, which turns a term t of Θ into $t\uparrow$ of A, Θ.

Lemma 20.16 (Renaming)**.** *Let ρ be a renaming from Θ into Θ''. Say that ρ* matches the slots *of R and R' when, for each variable x_k of Θ, of type B, a context derives that some z, z', r, r', v, v' satisfy*

$$\mathrm{App}(d_{\rho(k)}, R, z), \quad \mathrm{App}(d_k, R', z'), \quad G(r, v, z), \quad G(r', v', z'), \quad v \simeq_B v' :$$

dropping $\rho(k)$ entries from R and k entries from R' leaves environments whose newest entries are related. If ρ matches the slots of R and R' and t has type A in Θ, the context derives $\exists p, q\, (E_{t\rho}(R, p) \wedge E_t(R', q) \wedge p \simeq_A q)$. In particular, if it derives $R \approx_\Theta R'$ and $G(R, X, \hat{R})$, then it derives $\exists p, q\, (E_{t\uparrow}(\hat{R}, p) \wedge E_t(R', q) \wedge p \simeq_A q)$, for an arbitrary number X.

Proof. Induct on the typing of t. For a variable x_k, the right-projection law applied to $G(r, v, z)$ and the composition law give $\mathrm{App}(\mathrm{right} \circ d_{\rho(k)}, R, v)$, which is $E_{x_{\rho(k)}}(R, v)$; likewise $E_{x_k}(R', v')$; and $v \simeq_B v'$. Zero, successor, application, iteration and indexed recursion are the corresponding cases of Theorem 20.15, applied to the outputs that the induction hypotheses relate. For an abstraction $\lambda A.\, b$, take $x \simeq_A x'$, and let $G(R, x, \tilde{R})$ and $G(R', x', \tilde{R}')$. The lifted renaming sends x_0 to x_0 and x_{k+1} to $x_{\rho(k)+1}$, and it matches the slots of $\tilde{R}$ and $\tilde{R}'$. At the new position, the identity law gives $\mathrm{App}(d_0, \tilde{R}, \tilde{R})$, and the graphs $G(R, x, \tilde{R})$ and $G(R', x', \tilde{R}')$ hold the related values x and x'. At an old position, the left-projection law gives $\mathrm{App}(\mathrm{left}, \tilde{R}, R)$, and composition with the given $\mathrm{App}(d_{\rho(k)}, R, z)$ gives $\mathrm{App}(d_{\rho(k)+1}, \tilde{R}, z)$, since $d_{\rho(k)+1} = d_{\rho(k)} \circ \mathrm{left}$; the same on the other side; the graphs and the relation are those of the given slot. Every step runs a law forward. The induction hypothesis for the body and Lemma 20.13 give the arrow clause, as in Theorem 20.15. For the second statement take $\rho(k) = k + 1$. Lemma K.6 turns $R \approx_\Theta R'$ into $\mathrm{App}(d_k, R, z)$ and $\mathrm{App}(d_k, R', z')$ with $z \approx z'$, whose opening supplies the graphs and related values; the left-projection law on $G(R, X, \hat{R})$ and composition move the first to $\mathrm{App}(d_{k+1}, \hat{R}, z)$. The number X is never read. □

Theorem 20.17 (Term transport)**.** *Let t have type A in Θ and let σ be a substitution from Θ into Θ'. If a context derives that σ is faithful from R to R', together with $R \approx_{\Theta'} R$ and $R' \approx_{\Theta} R'$, then it derives*

$$\exists p, q\, \big(E_{t\sigma}(R, p) \wedge E_t(R', q) \wedge p \simeq_A q\big).$$

Proof. Induct on the typing of t. The variable case is the faithfulness hypothesis. The zero, successor, application, iteration and indexed-recursion cases are those of Theorem 20.15, applied to the outputs that the induction hypotheses relate. For an abstraction $\lambda A.\, b$, with $b : B$ in A, Θ, Lemma 20.13 gives closure codes f of $c((\lambda A.\, b)\sigma)$ at R and f' of $c(\lambda A.\, b)$ at R'; we derive $f \simeq_{A\to B} f'$. Take $x \simeq_A x'$, and let $G(R, x, z)$ and $G(R', x', z')$. The lifted substitution is faithful from z to z'. At the new position, x_0 reads x at z and x' at z'. At position $k+1$, the lifted substitution has $\sigma_k{\uparrow}$, and Lemma 20.16, applied to the diagonal $R \approx_{\Theta'} R$ and to $G(R, x, z)$, relates $\sigma_k{\uparrow}$ at z to σ_k at R; faithfulness relates σ_k at R to x_k at R'; functionality identifies the two outputs of σ_k at R; and x_{k+1} at z' reads x_k at R'. Transitivity joins the three relations. The relations $z \approx z$ and $z' \approx z'$ follow from $R \approx R$, $R' \approx R'$ and $H_A(x)$, $H_A(x')$, which symmetry and transitivity give from $x \simeq_A x'$. The induction hypothesis for the body relates the outputs y of $c(b\sigma^+)$ at z and y' of $c(b)$ at z', and the closure behavior sends x to y and x' to y'. □

Theorem 20.18 (Formula substitution)**.** *Let P be formed in Θ and let σ be a substitution from Θ into Θ'. Let a context derive that σ is faithful from R to R', together with $R \approx_{\Theta'} R$ and $R' \approx_{\Theta} R'$. Then it derives $\mathrm{Rd}_{R'}(P) \to \mathrm{Rd}_R(P\sigma)$ and $\mathrm{Rd}_R(P\sigma) \to \mathrm{Rd}_{R'}(P)$.*

Proof. Induct on P, constructing both directions together.

Truth and falsehood read the same in both environments.

Equation $t =_A u$. Suppose the reading at R' supplies outputs p of t and q of u with $p \simeq_A q$. Term transport gives outputs p' of $t\sigma$ and q' of $u\sigma$ at R, related to outputs of t and u at R', and functionality identifies those with p and q. Symmetry and transitivity, Proposition 20.11, give the chain

$$p' \simeq_A p \simeq_A q \simeq_A q',$$

which is the reading of $t\sigma =_A u\sigma$ at R. The reverse direction runs the same chain from the other end.

Conjunction transports each component.

Implication $P \to Q$. Assume $\mathrm{Rd}_{R'}(P) \to \mathrm{Rd}_{R'}(Q)$, and to derive $\mathrm{Rd}_R(P\sigma) \to \mathrm{Rd}_R(Q\sigma)$ assume $\mathrm{Rd}_R(P\sigma)$. Transport it *back* to $\mathrm{Rd}_{R'}(P)$ by the reverse direction for P, apply the assumed implication, and transport $\mathrm{Rd}_{R'}(Q)$ *forward* by the forward direction for Q. Discharge the assumption. The reverse direction swaps the two roles. A one-way transport for the antecedent would not provide the first step; this is why both directions are proved together. The faithfulness and diagonal hypotheses stay available under the temporary assumption, because they are members of the context.

Universal statement $\forall y{:}\, B\, Q$, with Q formed in B, Θ. Forward: assume the reading at R'. Take y with $H_B(y)$ and z with $G(R, y, z)$; pairing totality gives z' with $G(R', y, z')$, the same value placed in the other environment. The lifted substitution is faithful from z to z': at the new

position both environments read y, related to itself by $H_B(y)$, and at the old positions the argument of Theorem 20.17 applies. The diagonals $z \approx z$ and $z' \approx z'$ follow from the given ones and $H_B(y)$. Instantiate the assumed reading at y and z' to get $\mathrm{Rd}_{z'}(Q)$, transport the body forward by the induction hypothesis, and generalize over y and z. The reverse direction begins with an arbitrary y and z' on the other side, builds z by pairing totality, and transports the body back. The temporary pairing witness is eliminated before generalization closes. □

The extended environments are the proof of capture avoidance, not just a place to store values: the lifted substitution can be compared position by position only because both environments grow by the same value.

Corollary 20.19 (Shifted hypotheses and related environments)**.** *Let Q and P be formed in Θ.*

1. *From $C_\Theta(R)$, $H_A(x)$, $G(R, x, \hat{R})$ and $\mathrm{Rd}_R(Q)$, a context derives $C_{A,\Theta}(\hat{R})$ and $\mathrm{Rd}_{\hat{R}}(Q\uparrow)$.*
2. *From $R \approx_\Theta R'$ and $\mathrm{Rd}_R(P)$, a context derives $\mathrm{Rd}_{R'}(P)$.*

Proof. For the first, the three premises are the witnesses of $C_{A,\Theta}(\hat{R})$. The shift $x_k \mapsto x_{k+1}$ is a substitution from Θ into A, Θ, faithful from $\hat{R}$ to R by Lemma 20.16 applied to the diagonal of $C_\Theta(R)$, and $Q\sigma = Q\uparrow$. The forward direction of Theorem 20.18 turns $\mathrm{Rd}_R(Q)$ into $\mathrm{Rd}_{\hat{R}}(Q\uparrow)$.

For the second, induct on P. An equation: Theorem 20.15 relates the outputs of each term at R and at R', and the chain of the preceding proof carries the relation between the two outputs across. An implication uses that $\approx_\Theta$ is symmetric, which follows level by level from Proposition 20.11, to move the antecedent back. A universal statement extends both environments by the same admissible value. □

Lemma 20.20 (An environment for a substitution)**.** *Let σ be a substitution from Θ into Θ'. A context that derives $C_{\Theta'}(R)$ derives that some R' satisfies $C_\Theta(R')$ and that σ is faithful from R to R'.*

Proof. Induct on Θ. For the empty context take $R' = 0$. For A, Θ_0, the induction hypothesis for the tail substitution $k \mapsto \sigma_{k+1}$ gives R_0 with $C_{\Theta_0}(R_0)$, the tail being faithful from R to R_0. Typed evaluation of σ_0 at R gives x with $E_{\sigma_0}(R, x)$ and $H_A(x)$, and pairing totality gives R' with $G(R_0, x, R')$; these three give $C_{A,\Theta_0}(R')$. At the new position σ_0 returns x at R and x_0 returns x at R'. At an old position $k+1$, faithfulness of the tail relates σ_{k+1} at R with x_k at R_0, and Lemma 20.16 relates x_k at R_0 with x_{k+1} at R'; functionality and transitivity join them. □

Proposition 20.21 (Universal elimination)**.** *Let t have type A in Θ and let P be formed in A, Θ. A context that derives $C_\Theta(R)$ and $\mathrm{Rd}_R(\forall x{:}\,A\ P)$ derives $\mathrm{Rd}_R(P[t])$.*

Proof. Typed evaluation gives x with $E_t(R, x)$ and $H_A(x)$; pairing totality gives R'' with $G(R, x, R'')$. The reading of the universal statement, at x and R'', gives $\mathrm{Rd}_{R''}(P)$. The substitution of t for x_0, which sends x_{k+1} to x_k, is a substitution from A, Θ into Θ, faithful from R to R'': at the new position t returns x at R and x_0 returns x at R'', related by $H_A(x)$; at an old position Lemma 20.16 and symmetry relate x_k at R with x_{k+1} at R''. The diagonals $R \approx_\Theta R$ and $R'' \approx_{A,\Theta} R''$ come from $C_\Theta(R)$, $H_A(x)$ and $G(R, x, R'')$. The forward direction of Theorem 20.18 gives $\mathrm{Rd}_R(P[t])$. Neither temporary witness occurs in the conclusion. □

20.3.8 Every rule becomes a derivation

Six rules need constructions of their own: computation, replacement, the two successor laws, extensionality and induction. The others follow from what has been proved.

Theorem 20.22 (Conversion is sound)**.** *Let t, u have type A in Θ with $t \equiv u$. A context that derives $R \approx_\Theta R'$ derives $\exists x, y(E_t(R, x) \wedge E_u(R', y) \wedge x \simeq_A y)$.*

Proof. Lemma 20.1 gives a term v with $t \Rrightarrow^* v$ and $u \Rrightarrow^* v$, every term on the two chains having type A in Θ. It suffices to prove the statement for a single parallel step $t \Rrightarrow t'$ at related environments. A chain then follows by transitivity of $\simeq_A$, using $R' \approx_\Theta R'$, which follows from $R \approx_\Theta R'$ by symmetry and transitivity at each level; and t at R is related to u at R' through the common reduct v at R', using symmetry for the chain from u.

For a single step, induct on its derivation. A congruence step rebuilds a constructor around stepped parts; the case of Theorem 20.15 for that constructor, applied to the outputs related by the induction hypotheses, gives the relation. The five contractions follow both compiled programs to related values. We give two; the other three are in Appendix K.6.

Abstraction, $(\lambda A.\, b)\, s \Rrightarrow b'[s']$ with $b \Rrightarrow b'$ and $s \Rrightarrow s'$. At R, the compiled application evaluates the closure code f of $c(b)$ and the value x of s, and runs U on $\langle f, x\rangle$. Typed evaluation of b at $\langle R, x\rangle$, which is related to itself because $R \approx R$ and $H_A(x)$, gives an output y; the closure behavior gives $\mathrm{App}(f, x, y)$; and universal application gives U the output y. So the left side returns y at R, the value of b at $\langle R, x\rangle$. On the right, the induction hypothesis relates x to the value x' of s' at R', so $\langle R, x\rangle \approx \langle R', x'\rangle$, and relates y to the value y' of b' at $\langle R', x'\rangle$. The single substitution of s' is faithful from R' to $\langle R', x'\rangle$, so term transport relates y' to the value of $b'[s']$ at R'. Transitivity finishes the case.

Iteration at a successor, $\mathsf{I}(B, b, g, S(n)) \Rrightarrow g'\, \mathsf{I}(B, b', g', n')$. At R the parts return γ, v and $m + 1$. Lemma 20.14 gives the output x of It at count m; the arrow clause of $H(\gamma)$ gives an output s of γ on x; universal application and the recursion-step law give It the output s at count $m + 1$; and functionality identifies s with the value of the left side. At R' the right side applies γ', related to γ by the induction hypothesis, to the value x' of $\mathsf{I}(B, b', g', n')$, which the congruence case relates to x. The arrow clause of $\gamma \simeq \gamma'$ at $x \simeq x'$ relates s to the output on the right. □

Proposition 20.23 (Replacement, successor laws, reflexivity)**.** *Let a context derive $C_\Theta(R)$.*

1. *For $t : A$ it derives $\mathrm{Rd}_R(t =_A t)$, and for $t \equiv u$ at A it derives $\mathrm{Rd}_R(t =_A u)$.*
2. *For $t, u : A$ and P formed in A, Θ, if it derives $\mathrm{Rd}_R(t =_A u)$ and $\mathrm{Rd}_R(P[t])$, it derives $\mathrm{Rd}_R(P[u])$.*
3. *For $t : O$ it derives $\mathrm{Rd}_R(S(t) =_O 0 \to \bot)$; and for $t, u : O$, if it derives $\mathrm{Rd}_R(S(t) =_O S(u))$, it derives $\mathrm{Rd}_R(t =_O u)$.*

Proof. (1) Typed evaluation and Theorem 20.22, at the diagonal $R \approx_\Theta R$.

(2) The reading of the equation supplies x, y with $E_t(R, x)$, $E_u(R, y)$ and $x \simeq_A y$; symmetry and transitivity give $H_A(x)$ and $H_A(y)$. Pairing totality gives R_1, R_2 with $G(R, x, R_1)$ and $G(R, y, R_2)$; with $C_\Theta(R)$, $H_A(x)$ and $H_A(y)$ these give the diagonals that Theorem 20.18 needs.

The single substitution of t is faithful from R to R_1, so the reverse direction of Theorem 20.18 turns $\mathrm{Rd}_R(P[t])$ into $\mathrm{Rd}_{R_1}(P)$. The environments R_1 and R_2 are related at A, Θ: their old parts are R, related to itself, and their new entries satisfy $x \simeq_A y$. Corollary 20.19(2) gives $\mathrm{Rd}_{R_2}(P)$, and the forward direction for the substitution of u gives $\mathrm{Rd}_R(P[u])$. At a function type x and y may be different codes of one function; the proof never replaces one by the other, which equality substitution would require, and uses only the relation.

(3) Typed evaluation gives the value x of t at R. Assume the reading of $S(t) =_O 0$: outputs a of $c(S(t))$ and b of $c(0)$ with $a = b$. The successor and zero laws give the outputs $x + 1$ and 0, and functionality gives $a = x + 1$ and $b = 0$. So $x + 1 = 0$, and the separation rule of Chapter 5 gives $\bot$. For injectivity, typed evaluation gives the values a of t and b of u; the reading of $S(t) =_O S(u)$ supplies equal outputs, which functionality identifies with $a + 1$ and $b + 1$; the injectivity rule of Chapter 5 gives $a = b$. □

Each successor law uses the corresponding arithmetic rule once, on numbers. The reading adds only evaluation and functionality.

Theorem 20.24 (Extensionality). *Let f, g have type $A \to B$ in Θ. Suppose the premise $(f\uparrow)\, x_0 =_B (g\uparrow)\, x_0$ compiles at every extended environment: every context that derives $C_{A,\Theta}(R_1)$ and the readings at R_1 of the shifted hypotheses derives its reading at R_1. Then every context that derives $C_\Theta(R)$ and the readings at R of the hypotheses derives $\mathrm{Rd}_R(f =_{A\to B} g)$.*

Proof. Typed evaluation gives admissible values v of f and w of g at R. It remains to derive $v \simeq_{A\to B} w$. Take $x \simeq_A x'$; then $H_A(x)$. Let $G(R, x, R_1)$. Corollary 20.19(1) gives $C_{A,\Theta}(R_1)$ and the shifted hypotheses at R_1, so the compiled premise gives outputs a of $(f\uparrow)\, x_0$ and b of $(g\uparrow)\, x_0$ at R_1 with $a \simeq_B b$.

Connect a with the function value. The compiled $f\uparrow$ at R_1 returns a value v_1 related to v, by Lemma 20.16, and the compiled x_0 returns x. The arrow clause of $v_1 \simeq v$ at $x \simeq x$ gives outputs a_1 of v_1 and a_0 of v on x with $a_1 \simeq_B a_0$. Universal application and the pair and composition laws make a_1 an output of the compiled application $(f\uparrow)\, x_0$ at R_1, and functionality gives $a = a_1$. In the same way the compiled $g\uparrow$ returns a value w_1 related to w; the arrow clause of $w_1 \simeq w$ at $x \simeq x'$ gives outputs b_1 of w_1 on x and b_0 of w on x' with $b_1 \simeq_B b_0$; and $b = b_1$. Symmetry and transitivity give $a_0 \simeq_B b_0$: an output of v on x related to an output of w on x'. Generalizing over x, x' gives the arrow clause for v, w. □

Extensionality adds no principle: the reading of $f =_{A\to B} g$ already states that the two functions agree on related arguments. The work is the link between the compiled application of $f\uparrow$ to the new variable, which the premise mentions, and the application of the value of f to an argument, which the arrow clause mentions. Universal application supplies that link. The proof extends R by x only; the relation with x' enters through the arrow clause of w.

Theorem 20.25 (Induction). *Let P be formed in O, Θ. Suppose the base $P[0]$ compiles at every environment of Θ and the step $P \to P^S$ compiles at every environment of O, Θ, in the sense of Theorem 20.24. Then every context that derives $C_\Theta(R)$ and the readings at R of the hypotheses derives $\mathrm{Rd}_R(\forall x{:}O\ P)$, by a derivation whose ledger records the induction entry and no principle.*

Proof. Since $H_O(x)$ is $x = x$,

$$\mathrm{Rd}_R(\forall x{:}O\ P) = \forall x\big(x = x \to \forall z\,(G(R, x, z) \to \mathrm{Rd}_z(P))\big).$$

Call the formula under the first quantifier $\Phi_R(x)$, and induct on x in $\Phi_R(x)$ with the induction rule (5.2).

Zero. Let $G(R, 0, z_0)$; with $C_\Theta(R)$ and $0 = 0$ it gives the diagonal of z_0. The compiled base gives $\mathrm{Rd}_R(P[0])$. The substitution of 0 is faithful from R to z_0, since the compiled 0 returns 0 at R and z_0 stores 0; the reverse direction of Theorem 20.18 gives $\mathrm{Rd}_{z_0}(P)$.

Successor. Assume $\Phi_R(x)$ and let $G(R, x + 1, z')$. Pairing totality gives z with $G(R, x, z)$, and $\Phi_R(x)$ gives $\mathrm{Rd}_z(P)$. The environment z satisfies $C_{O,\Theta}(z)$ and, by Corollary 20.19(1), the shifted hypotheses, so the compiled step gives $\mathrm{Rd}_z(P) \to \mathrm{Rd}_z(P^S)$ and hence $\mathrm{Rd}_z(P^S)$. The formula P^S is P under the substitution that sends x_0 to $S(x_0)$ and keeps the other variables. It is faithful from z to z': the compiled $S(x_0)$ returns $x + 1$ at z, and z' stores $x + 1$; both environments keep R as their older part, which gives the other positions and the diagonals. The reverse direction of Theorem 20.18 gives $\mathrm{Rd}_{z'}(P)$, which is $\Phi_R(x + 1)$ after discharge.

The induction rule gives $\forall x\,\Phi_R(x)$. The formula $\Phi_R(x)$ contains the quantifier $\forall z$, so the ledger records the induction entry; the premises are posit-free, so it records nothing else. □

The source induction may be on an equation, with no quantifier. Its reading always has one, because the reading of a universal statement quantifies over the extended environment. A compiled induction therefore always records the induction entry.

Theorem 20.26 (Proofs compile to derivations)**.** *For every proof of $\Theta; \Xi \vdash P$ in the language of functions and every environment term R, there is a posit-free derivation of $\mathrm{Rd}_R(P)$ from the hypotheses $C_\Theta(R)$ and $\mathrm{Rd}_R(Q)$ for Q in Ξ. In particular, for every proof of a closed formula P from no hypotheses, the rules of Chapter 5 derive $\mathrm{Rd}_0(P)$ by a posit-free derivation.*

Proof. Prove more: for every list of hypotheses Δ that contains $C_\Theta(R)$ and $\mathrm{Rd}_R(Q)$ for every Q in Ξ, there is a derivation of $\mathrm{Rd}_R(P)$ from Δ. Recurse on the proof, one case for each rule.

(1) *Hypothesis*: $\mathrm{Rd}_R(P)$ is a member of Δ; search the finite list for its position.

(2) *More hypotheses*: Δ also contains the readings of the smaller list; recurse.

(3) *Truth*: $0 = 0$ by reflexivity.

(4) *Falsehood*: $\mathrm{Rd}_R(\bot)$ is $\bot$; apply the rule for an impossible case.

(5–7) *Conjunction*: the reading of a conjunction is the conjunction of the readings; apply the same rule.

(8) *Implication introduction*: recurse with $\mathrm{Rd}_R(P)$ added to Δ, then discharge it.

(9) *Implication elimination*: the same rule.

(10) *Universal introduction*: introduce fresh x and z and assume $H_A(x)$ and $G(R, x, z)$. Corollary 20.19(1) gives $C_{A,\Theta}(z)$ and the readings of the shifted hypotheses at z; recurse on the premise at z; discharge and generalize.

(11) *Universal elimination*: Proposition 20.21.

(12) *Reflexivity*: Proposition 20.23(1).

(13) *Computation*: Proposition 20.23(1), by Theorem 20.22.

(14) *Replacement*: Proposition 20.23(2).

(15) *Successor is not zero*: Proposition 20.23(3).

(16) *Successor injectivity*: Proposition 20.23(3).

(17) *Extensionality*: Theorem 20.24, recursing on the premise at each extended environment.

(18) *Induction*: Theorem 20.25, recursing on the base and the step.

(19) *Substitution*: let σ go from Θ into Θ'. Lemma 20.20 gives R' with $C_\Theta(R')$ to which σ is faithful from R. The reverse direction of Theorem 20.18 turns the readings $\mathrm{Rd}_R(Q\sigma)$ of the hypotheses into $\mathrm{Rd}_{R'}(Q)$; recurse on the premise at R'; the forward direction turns $\mathrm{Rd}_{R'}(P)$ into $\mathrm{Rd}_R(P\sigma)$. The witness R' does not occur in the conclusion.

Every case combines posit-free derivations, so the result is posit-free. For the last assertion take the empty type context, no hypotheses and $R = 0$; the hypothesis $C_{[\,]}(0)$ is $0 = 0$, which reflexivity proves. □

The derivation exists for every proof, including proofs that pass through functions of any finite type. It uses induction on quantified formulas and universal application, and none of EM, LPO and MP.

20.3.9 Recovering the arithmetic sentence

For a sentence about numbers, Theorem 20.26 gives the reading, not yet the sentence. The reading of an equation says that two programs return values and that the values are equal. The sentence says that two terms are equal. For addition and multiplication the gap is real: the compiled programs run the iterator, and every step of the iterator runs U. That a compiled sum returns the actual sum is a theorem, proved by induction along the iteration. In this subsection $+$ and $\cdot$ are the operations of the arithmetic language, governed by the four recursive rules of Chapter 5.

Lemma 20.27 (Counting iterations)**.** *There are derivations of the following.*

1. *If g returns $x + 1$ on every x and $G(g, v, a)$, then for all m, p with $G(a, m, p)$ the iterator It returns $v + m$ on p.*

2. *If g returns $x + k$ on every x and $G(g, 0, a)$, then for all m, p with $G(a, m, p)$ the iterator returns $k \cdot m$ on p.*

Proof. For the first, induct on m in $\forall p\,(G(a,m,p) \to \mathrm{App}(\mathsf{It}, p, v+m))$. At 0 the recursion-start law returns the seed v, and the rule $v + 0 = v$ rewrites the output. At $m + 1$, pairing totality supplies the packed inputs of the previous call and of the step; the induction hypothesis gives the previous output $v + m$; the premise on g and universal application give the step output $(v + m) + 1$; the recursion-step law joins them; and the rule $v + S(m) = S(v + m)$ rewrites the output. The second is the same induction with seed 0, step output $x + k$, and the rules $k \cdot 0 = 0$ and $k \cdot S(m) = k \cdot m + k$. □

The premise on g is met by the compiled step closures. The closure of $\lambda O.\, S(x_0)$ at R sends x to the output of $c(S(x_0))$ at $\langle R, x\rangle$, which the successor law and the right projection give as $x + 1$. So by Lemma 20.13 the step code returns $x + 1$ on every x, and Lemma 20.27 gives:

Proposition 20.28 (Compiled addition and multiplication)**.** *There are derivations of: if the compiled s returns v at R and the compiled t returns m at R, then the compiled* $\mathrm{add}(s,t)$ *returns $v + m$ at R; and if the compiled $s\!\uparrow$ returns k at every one-step extension of R and the compiled t returns m at R, then the compiled* $\mathrm{mul}(s,t)$ *returns $k \cdot m$ at R.*

Proof. For addition, the compiled program evaluates the step closure to a code f that returns $x + 1$ on every x, the seed to v and the count to m, packs $\langle f, v\rangle$ and $\langle\langle f, v\rangle, m\rangle$ by pairing totality, and runs It; Lemma 20.27(1) gives the output $v + m$, and the pair and composition laws assemble the compiled computation. For multiplication, the step closure of $\lambda O.\ \mathrm{add}(x_0, s\!\uparrow)$ sends x to the output of the compiled addition at $\langle R, x\rangle$, where x_0 returns x and, by the premise, $s\!\uparrow$ returns k; by the first part that output is $x + k$. The seed returns 0, and Lemma 20.27(2) gives $k \cdot m$. □

Environments now hold terms of the arithmetic language. For a list $\sigma = (\sigma_0, \sigma_1, \ldots)$ of terms, let $\mathscr{L}_n(\sigma, R)$ say that R lists $\sigma_0, \ldots, \sigma_{n-1}$, newest first: $\mathscr{L}_0(\sigma, R) := (0 = 0)$ and $\mathscr{L}_{n+1}(\sigma, R) := \exists r\,(G(r, \sigma_0, R) \wedge \mathscr{L}_n((\sigma_1, \sigma_2, \ldots), r))$.

Theorem 20.29 (Compiled terms return their values)**.** *For every arithmetic term t with variables below n, and every list σ of terms, there is a derivation of $\mathscr{L}_n(\sigma, R) \to \mathrm{App}(c(t^*), R, t[\sigma])$, where $t[\sigma]$ replaces each x_i by σ_i.*

Proof. Recurse on the size of t, for all n and σ at once. A variable x_i is read by i left projections and one right projection from the graphs inside $\mathscr{L}_n(\sigma, R)$, and returns σ_i. Zero returns 0. For $S(t)$ the successor law applied to the output $t[\sigma]$ gives $S(t[\sigma])$. For $t + u$ the recursive derivations give the two outputs and Proposition 20.28 gives their sum. For $t \cdot u$ the same proposition needs the value of $t^*\!\uparrow$ at every one-step extension of R. An extension lists a new entry X followed by σ, and $t^*\!\uparrow = (t\!\uparrow)^*$. The derivation for $t\!\uparrow$, which has the size of t, at $n + 1$ and the list $(X, \sigma_0, \sigma_1, \ldots)$ gives the value $t\!\uparrow[(X, \sigma_0, \ldots)]$, which is $t[\sigma]$. □

Theorem 20.30 (Recovery and embedding)**.** *For every negative formula A with free variables below n, every list σ of terms and every environment term R, there are derivations of*

$$\mathscr{L}_n(\sigma, R) \to \big(\mathrm{Rd}_R(A^*) \to A[\sigma]\big) \qquad \textit{and} \qquad \mathscr{L}_n(\sigma, R) \to \big(A[\sigma] \to \mathrm{Rd}_R(A^*)\big).$$

Proof. Recurse on A, constructing both together.

Equation $t = u$. The reading supplies outputs a of $c(t^*)$ and b of $c(u^*)$ at R with $a = b$. Theorem 20.29 gives the outputs $t[\sigma]$ and $u[\sigma]$ of the same programs; functionality gives $a = t[\sigma]$ and $b = u[\sigma]$; so $t[\sigma] = u[\sigma]$. Conversely, from $t[\sigma] = u[\sigma]$, Theorem 20.29 supplies the two outputs the reading requires, and the equation relates them.

Falsehood. Both sides are $\bot$.

Implication $A \to B$. Given $\mathrm{Rd}_R(A^*) \to \mathrm{Rd}_R(B^*)$ and $A[\sigma]$, the embedding direction for A gives $\mathrm{Rd}_R(A^*)$, the hypothesis gives $\mathrm{Rd}_R(B^*)$, and the recovery direction for B gives $B[\sigma]$. The converse uses recovery for A and embedding for B. Neither direction for the implication can be built from one direction for its antecedent.

Universal statement $\forall A$. The reading is $\forall x(x = x \to \forall z(G(R, x, z) \to \mathrm{Rd}_z(A^*)))$. For recovery, take x; pairing totality gives z with $G(R, x, z)$, which with $\mathscr{L}_n(\sigma, R)$ gives $\mathscr{L}_{n+1}(x\sigma, z)$ for the list $x\sigma = (x, \sigma_0, \sigma_1, \ldots)$. The reading gives $\mathrm{Rd}_z(A^*)$, recovery for A at $n + 1$ gives $A[x\sigma]$, and generalization gives $(\forall A)[\sigma]$. For embedding, take x and z with $G(R, x, z)$; the same extended list is listed by z, and the hypothesis instantiated at x gives $A[x\sigma]$, which embedding turns into $\mathrm{Rd}_z(A^*)$. □

Theorem 20.31 (Arithmetic is recovered from its reading)**.** *For every closed negative formula A, a posit-free derivation proves that the reading $\mathrm{Rd}_0(A^*)$ implies A.*

Proof. Apply the recovery direction of Theorem 20.30 with $n = 0$, any list σ, which leaves the closed A unchanged, and $R = 0$. The premise $\mathscr{L}_0$ is $0 = 0$. □

Theorem 20.32 (Conservativity for negative sentences)**.** *Let A be a closed formula of the arithmetic language built from $=$, $\bot$, $\to$ and $\forall$. If the language of functions proves A, then the rules of Chapter 5 derive A by a posit-free derivation.*

Proof. Theorem 20.26 gives a posit-free derivation of $\mathrm{Rd}_0(A^*)$, and Theorem 20.31 a posit-free derivation that the reading implies A. Implication elimination combines them. By Proposition 6.4, the combined ledger records none of the three principles. □

The theorem allows full induction on both sides: the proof about functions may induct on any formula, and the derivation records the induction entry when it inducts on a quantified one.

20.3.10 The comparison with Peano arithmetic

The restriction to $=$, $\bot$, $\to$ and $\forall$ loses nothing for classical arithmetic. The *negative translation* φ^N of a formula of Definition 5.1 is defined by

$$(t = u)^N := \neg\neg(t = u), \qquad \bot^N := \bot, \qquad (\varphi \to \psi)^N := \varphi^N \to \psi^N,$$
$$(\varphi \wedge \psi)^N := \neg(\varphi^N \to \neg\psi^N), \qquad (\forall\varphi)^N := \forall\varphi^N, \qquad (\exists\varphi)^N := \neg\forall\neg\varphi^N,$$
$$(\varphi \vee \psi)^N := \neg\neg(\neg\varphi^N \to \neg\neg\psi^N).$$

It doubly negates each equation, replaces $\exists x\psi$ by $\neg\forall x\neg\psi$ and $\psi \vee \chi$ by $\neg(\neg\psi \wedge \neg\chi)$, and writes every conjunction, including the one just introduced, as $\neg(\psi \to \neg\chi)$ [4]. The result is negative, and it keeps every term and every binder in place, so φ^N is closed exactly when φ is.

Lemma 20.33 (The classical bridge)**.** *For every formula φ there are derivations from no hypotheses of $\varphi \to \varphi^N$ and $\varphi^N \to \varphi$ whose ledgers record no principle other than EM, and no induction.*

Proof. Recurse on φ, building both derivations together; write a, b for the components and A, B for their translations.

Equation. From $s = t$ and $\neg(s = t)$ follows $\bot$, so $s = t \to \neg\neg(s = t)$. Conversely, EM gives $s = t$ or $\neg(s = t)$, and the second contradicts $\neg\neg(s = t)$. *Falsehood.* The identity.

Implication. Given $a \to b$ and A: back from A to a, then b, then forward to B. Given $A \to B$ and a: forward to A, then B, then back to b. Both directions of the components are used.

Universal statement. Introduce the fresh variable and apply the component.

Conjunction. From $a \wedge b$ the components give A and B, and a hypothesis $A \to \neg B$ would then give $\bot$. Conversely, from $\neg(A \to \neg B)$, EM on A gives A, since in the case $\neg A$ the implication $A \to \neg B$ holds outright and is refuted; likewise for B; the components give $a \wedge b$.

Disjunction. From $a \vee b$, either case refutes $\neg(\neg A \to \neg\neg B)$: from A the antecedent $\neg A$ is contradictory, and from B the consequent $\neg\neg B$ holds. Conversely, EM on $A \vee B$: in the negative case $\neg A$ and $\neg B$ hold, so $\neg A \to \neg\neg B$ fails, contradicting the hypothesis; in the positive case the components give $a \vee b$.

Existential statement. From $\exists x\, a$ the component gives $\exists x\, A$, which refutes $\forall x \neg A$. Conversely, EM on $\exists x\, A$: in the negative case $\forall x \neg A$ follows, contradicting $\neg\forall x \neg A$; in the positive case the component gives $\exists x\, a$.

No LPO or MP rule and no induction occurs. EM occurs only in the backward directions of the equation, conjunction, disjunction and existential cases, and in the forward direction of an implication through the backward derivation of its antecedent. □

The typed calculus. In the calculus $\mathscr{E}^+$ of Chapter 9, a negative formula A is expressed by a type $\mathrm{Ty}(A)$: an equation $t = u$ by the identity type $\mathrm{Id}_O(t^*, u^*)$, with add and mul written as iterations; $\bot$ by the empty type $\mathbf{0}$; an implication by a function type; and $\forall x$ by a product over O. We say that $\mathscr{E}^+$ *proves* φ^N when it has a closed term of type $\mathrm{Ty}(\varphi^N)$.

The two halves of the comparison go in opposite directions. From $\mathscr{E}^+$ to arithmetic, a closed term is read in the language of functions and then compiled. From arithmetic to $\mathscr{E}^+$, a derivation is translated node by node into a term.

Reading a typed term. A term of the dependent calculus is read through its *erasure*, a term of the language of functions. Erasure sends $\prod_{x:A} B$ to $|A| \to |B|$ and every other type to O. On terms it keeps variables, 0, S, abstraction, application, I and R, and sets

$$|\mathrm{refl}_t| = 0, \qquad |J(A, x, Q, d, y, p)| = |d|, \qquad |\mathsf{E}(P, b, g, t)| = \mathsf{R}(|P|, |b|, |g|, |t|),$$

with the erasure of an empty-type elimination at A the fixed term of type $|A|$ that returns 0 after all its arguments. A type T is read as a formula $\mathrm{Adm}_T(v)$ about a term v of type $|T|$:

$$\begin{aligned} \mathrm{Adm}_O(v) &:= \top, & \mathrm{Adm}_{\mathbf{0}}(v) &:= \bot, \\ \mathrm{Adm}_{\mathrm{Id}_A(t,u)}(v) &:= (v =_O 0) \wedge (|t| =_{|A|} |u|), & \mathrm{Adm}_{D(t)}(v) &:= (|t| =_O 0) \wedge (v =_O 0), \\ \mathrm{Adm}_{\prod_{x:A} B}(v) &:= \forall x{:}|A|\,\big(\mathrm{Adm}_A(x) \to \mathrm{Adm}_B(vx)\big). \end{aligned}$$

A context Γ is read as the type context $|\Gamma|$ together with the hypotheses $\mathrm{Adm}(\Gamma)$, one formula $\mathrm{Adm}_A(x)$ for each entry $x{:}A$.

Lemma 20.34 (Reading the dependent calculus)**.** *With or without the discriminator, and in each fragment, if $\Gamma \vdash t : T$, then $|t|$ has type $|T|$ in $|\Gamma|$, and the language of functions proves $\mathrm{Adm}_T(|t|)$ in $|\Gamma|$ from the hypotheses $\mathrm{Adm}(\Gamma)$.*

Proof. The erased type $|T|$ depends only on the shape of T, so a conversion $A \equiv B$ gives $|A| = |B|$, and every typing rule erases to a typing rule of the language of functions; this gives the first assertion by induction on the derivation. For the readings, induct on the derivation, proving at the same time that a substitution from Γ to Δ sends every entry $x{:}A$ of Γ to a term whose reading is provable from $\mathrm{Adm}(\Delta)$.

- A variable reads its hypothesis. Zero and successor have type O, whose reading is $\top$.
- An abstraction is universal and implication introduction; a computation step $(\lambda x.\, b)\, x \equiv b$ and replacement identify the reading of the body. An application is universal elimination at $|u|$ and implication elimination.
- Reflexivity: $0 =_O 0$ and $|x| =_{|A|} |x|$ by reflexivity.
- Identity elimination $J(A, x, Q, d, y, p)$: the reading of $p : \mathrm{Id}_A(x, y)$ gives $|p| =_O 0$ and $|x| =_{|A|} |y|$. Replacement moves the reading of d from x to y and, by symmetry, from 0 to $|p|$. The erased term is $|d|$.
- Empty-type elimination is the rule for falsehood.
- Iteration and indexed recursion: induct on x for the reading of the erased recursion at x, using its two computation equations, and instantiate at the erased count. The dependent eliminator is the same induction for $\mathrm{Adm}_{P(x)}$; its erasure is an indexed recursion.
- Conversion of a type: the reading moves along each step of a conversion between the two types. The steps inside terms are computations and replacements; the discriminator steps exchange the two conjuncts at $D(0) \equiv \mathrm{Id}_O(0,0)$, and at $D(S(t)) \equiv \mathbf{0}$ use the rule *successor is not zero* forward and falsehood backward. The full argument is in Appendix K.7.
- A substitution substitutes into the proof and cuts its hypotheses against the readings of the substituted terms.

The remaining judgments, about contexts, types and equalities, carry no reading. □

For a negative formula A, the *canonical term* of $|\,\mathrm{Ty}(A)|$ is 0 at O and $\lambda x.\, c$ at an arrow type, with c the canonical term of the codomain.

Lemma 20.35 (Types of formulas read as the formulas)**.** *Let A be a negative formula whose embedding A^* is formed in Θ. In Θ, from any hypotheses:*

(i) a proof of A^* gives a proof of $\mathrm{Adm}_{\mathrm{Ty}(A)}$ of the canonical term;

(ii) for every term r of type $|\mathrm{Ty}(A)|$, a proof of $\mathrm{Adm}_{\mathrm{Ty}(A)}(r)$ gives a proof of A^*.

Proof. Recurse on A, proving both parts together. For an equation the reading is $(v =_O 0) \wedge (t^* =_O u^*)$, since erasure leaves the embedded terms unchanged; (i) pairs reflexivity with the given proof, and (ii) takes the second conjunct. Falsehood is unchanged. For $A \to B$: in (i), assume the reading of the domain at x; part (ii) for A gives A^*, the given proof gives B^*, part (i) for B gives the reading of the canonical term of B, and computation $(\lambda x.\, c)\, x \equiv c$ with replacement moves it to the canonical function applied to x. In (ii), apply the given proof to the canonical term of A, whose reading (i) supplies, and use (ii) for B. For $\forall A$ the reading is $\forall x{:}O\,(\top \to \mathrm{Adm}(r\,x))$; eliminate at a fresh x, apply the part for A, and introduce the quantifier again. □

Theorem 20.36 (From the typed calculus to functions)**.** *With or without the discriminator, and in each fragment, if the calculus has a closed term of type $\mathrm{Ty}(A)$ for a closed negative formula A, then the language of functions proves A.*

Proof. Let t be the closed term. Lemma 20.34 at the empty context gives a proof of $\mathrm{Adm}_{\mathrm{Ty}(A)}(|t|)$ from no hypotheses and gives $|t|$ the type $|\mathrm{Ty}(A)|$; Lemma 20.35(ii) turns it into a proof of A^*. □

From derivations to typed terms. The other direction translates a derivation node by node. The translated formulas are stable under double negation, and that stability absorbs the classical steps.

Lemma 20.37 (Stability)**.** *For every arithmetic formula φ, in each fragment, with or without the discriminator, and in every context in which its free variables are terms of type O, a term of type $\neg\neg\,\mathrm{Ty}(\varphi^N)$ gives a term of type $\mathrm{Ty}(\varphi^N)$.*

Proof. Recurse on φ. For an equation, a conjunction, a disjunction and an existential statement, $\mathrm{Ty}(\varphi^N)$ is a negation $\neg X$, and $\lambda h.\, \lambda x.\, h\,(\lambda k.\, k\,x)$ has type $\neg\neg\neg X \to \neg X$. For falsehood, apply the given term of $\neg\neg\mathbf{0}$ to the identity on $\mathbf{0}$. For an implication and a universal statement the type is a function type; map $\neg\neg(X \to Y)$ to $\neg\neg Y$ by evaluation at the given argument, and apply the case for the consequent. □

Theorem 20.38 (From derivations to typed terms)**.** *In $\mathscr{E}^+$, every derivation of a sentence φ from no hypotheses by the rules of Chapters 5 and 6, whatever its ledger, gives a closed term of type $\mathrm{Ty}(\varphi^N)$.*

Proof. Prove more: for every derivation of φ from hypotheses Ξ, every context Γ of the calculus, every assignment of terms of type O in Γ to the free variables, and terms in Γ of the types of the translated hypotheses, there is a term in Γ of type $\mathrm{Ty}(\varphi^N)$. Induct on the derivation.

Node	Term	Uses
hypothesis	the supplied term	
reflexivity	$\lambda k.\, k$ refl	
equality substitution	identity elimination along $\mathrm{Id}_O(t^*, s^*)$, with the type of the body as motive, under $\neg\neg$, then Lemma 20.37	
$\neg(S(t) = 0)$	reverse the path to $\mathrm{Id}_O(0, S(t^*))$ and eliminate it with motive $D(y)$, from $\mathrm{refl}_0 \,:\, D(0)$ to $D(S(t^*)) \equiv \mathbf{0}$ (Theorem 9.3)	discriminator
successor injectivity	identity elimination with the predecessor of Theorem 9.2	indexed recursion
equations of $+$, $\cdot$	refl by the iteration equations, under $\neg\neg$	
induction	$\lambda x{:}O.\, \mathsf{E}(P, b, g, x)$, with P the type of the translated body	dependent eliminator
$\to$, $\forall$, $\wedge$	abstraction and application; $\lambda h.\, h\, a\, b$ for a conjunction	
introductions of $\vee$, $\exists$; eliminations of $\wedge$, $\vee$, $\exists$	the standard terms for the negative encodings; each elimination yields a doubly negated conclusion, removed by Lemma 20.37	
falsehood	empty-type elimination	
EM	the type is $\neg\neg(\neg X \to \neg\neg\neg X)$, inhabited by $\lambda k.\, k\,(\lambda n.\, \lambda h.\, h\, n)$	
LPO	the type is $Z \to \neg\neg(\neg\neg Y \to \neg\neg Y)$, with Z the type of the translated premise, inhabited by $\lambda d.\, \lambda k.\, k\,(\lambda y.\, y)$	
MP	the type is $\neg\neg\neg Y \to \neg Y$, inhabited by $\lambda h.\, \lambda x.\, h\,(\lambda k.\, k\, x)$	

Table 20.3. Translating derivation nodes into terms of $\mathscr{E}^+$. Here X and Y abbreviate the types of the relevant translated formulas. The last column names the resource of Theorem 9.10 a node needs.

Implication introduction extends Γ by the translated antecedent; universal introduction and existential elimination extend it by a variable of type O. The nodes translate as in Table 20.3. For a sentence take the empty context. □

Theorem 20.39 (Peano arithmetic and the dependent calculus)**.** *Let φ be a sentence of the arithmetic language.*

1. *If $\vdash_{\mathrm{PA}} \varphi$, then $\mathscr{E}^+$ proves φ^N.*
2. *If $\mathscr{E}^+$ proves φ^N, then $\vdash_{\mathrm{PA}} \varphi$.*

Proof. For the first assertion, Theorem 8.11 gives a derivation of φ by the rules of Chapters 5 and 6, and Theorem 20.38 translates it into a closed term of type $\mathrm{Ty}(\varphi^N)$.

For the second, Theorem 20.36 gives a proof of $(\varphi^N)^*$ in the language of functions, and Theorem 20.32 a posit-free derivation of φ^N. Lemma 20.33 adds a derivation of $\varphi^N \to \varphi$ whose ledger records only EM, so the combined derivation of φ records neither LPO nor MP, and Theorem 8.11 converts it into a Peano derivation. □

Each hypothesis has a definite role. The dependent eliminator, the indexed recursor and the discriminator are used only in the first assertion, for induction, successor injectivity and $\neg(S(t) = 0)$; the second holds in each fragment, with or without the discriminator. The discriminator cannot be dropped: with it disabled, the truth-value reading of Theorem 9.8 gives every closed type of the form $\mathrm{Id}_O(t, u) \to \mathbf{0}$ the value false, so $\mathscr{E}^-$ cannot even separate 0 from $S(0)$, while Peano arithmetic can. Together, the two assertions say that $\mathscr{E}^+$ proves the translations of exactly the sentences that Peano arithmetic proves.

20.3.11 Consistency, incompleteness and finite sets

The comparison says that $\mathscr{E}^+$ and Peano arithmetic stand or fall together. It does not by itself say that either stands. This subsection proves that both are consistent, shows that Peano arithmetic does not prove the sentence expressing its own consistency and that $\mathscr{E}^+$ does not prove the negative translation of that sentence, and reads two further theories through the same comparison. Subsection 20.3.12 continues with Löb's theorem, Gödel's second theorem for finite extensions of Peano arithmetic and a tower of consistency statements. From Theorem 20.42 on, the arguments use a classical metatheory: the completeness of first-order calculi for their models, proved by Henkin's construction [11], and, for Theorem 20.44(3), set theory with a universe of sets.

Corollary 20.40 (Equiconsistency)**.** *$\mathscr{E}^+$ has a closed term of the empty type exactly when Peano arithmetic derives $\bot$.*

Proof. Apply Theorem 20.39 to the sentence $\bot$, whose negative translation is $\bot$ and whose type is $\mathbf{0}$. □

Theorem 20.41 (Peano arithmetic and $\mathscr{E}^+$ are consistent)**.** *In a metatheory that grants LPO for the predicates that arithmetic formulas define, and in particular in a classical metatheory, no derivation by the rules of Chapters 5 and 6 proves $\bot$ from no hypotheses, whatever its ledger. Hence Peano arithmetic does not derive $\bot$, and $\mathscr{E}^+$ has no closed term of the empty type. Posit-free derivations need no such principle: none of them proves $\bot$.*

Proof. By Corollary 6.15, under that LPO every derivation from no hypotheses is sound in $\mathbb{N}$, and $\bot$ is satisfied by no environment. A Peano derivation of $\bot$ would give such a derivation, by Theorem 8.11, and a closed term of $\mathbf{0}$ in $\mathscr{E}^+$ would give one, by Corollary 20.40. For posit-free derivations, Corollary 6.7 needs no principle at all. □

Theorem 20.44 below gives a second proof, through a set model of a stronger calculus.

The consistency sentence. To ask whether Peano arithmetic proves its own consistency, consistency must be a sentence of arithmetic. Fix the following finite proof checker. It reads proofs in the language of $0, S, +, \cdot, =, \bot, \to$ and $\forall$, with fourteen axiom schemas: $A \to (B \to A)$; $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$; contraposition $(\neg B \to \neg A) \to (A \to B)$; instantiation and distribution for $\forall$; reflexivity and substitution of equals; the six arithmetic axioms of Definition 8.10; and induction for every formula. Its rules are modus ponens and generalization. It is a Hilbert-style presentation of Peano arithmetic. Terms, formulas and

proofs are coded by numbers through the pairing $(a, b) \mapsto 2^a(2b + 1) - 1$, and finite tables through Gödel's β of Section 20.3.3. This is a second pairing, beside the pairing (10.3) used elsewhere in this section: the checker, and the proof of Theorem 20.44 in Appendix L, are written with it. The clauses below read codes only through the values stored in the tables, and fact (D5) in the proof of Theorem 20.42 builds, from a certificate coded with (10.3), the tables of an accepted proof of this checker. Let Con(PA) be the sentence saying that no finite tables satisfy the finitely many clauses that describe an accepted proof of $\bot$; Definition L.1 writes it out, clause by clause. The fourteen schemas and two rules of the checker are sound in $\mathbb{N}$, but that alone does not make Con(PA) true, since tables that satisfy its clauses need not be the decoding of an accepted proof. Theorem 20.44 shows that it is true: its first part derives the sentence in a calculus whose set model, in part (3), makes every derived arithmetic sentence true.

Theorem 20.42 (Gödel's second incompleteness theorem for this checker)**.** *If Peano arithmetic is consistent, it does not derive* Con(PA)*. Since it is consistent (Theorem 20.41), it does not derive* Con(PA)*.*

Standard theorem; the argument for this checker. This is Gödel's theorem [8, Satz XI, p. 196], with the derivability conditions of Hilbert and Bernays [9] in the form given by Löb [10], here for this presentation and this checker. The steps are stated here and proved in full in Appendix M, Sections M.1–M.5. One change of method makes it manageable: each statement that Peano arithmetic must derive is shown true in an arbitrary model of Peano arithmetic and then read back as a derivation. Peano arithmetic is complete for its models, by Henkin's construction with free variables as witnesses: a formula true in every model under every assignment has a derivation. Inside a model the reasoning is ordinary mathematics about a discretely ordered semiring with induction for every formula.

Let $\mathrm{Pr}(x)$ be the formula "some certificate proves the formula coded by x", with certificates finite lists of records coded by β-tables, each record an axiom, a modus ponens node, a generalization node, or a local check of a substitution, a lifting or a shift of terms, by the clauses the checker uses. Write $\ulcorner\varphi\urcorner$ for the numeral of the code of φ. Appendix M names this formula $\mathrm{Prov}(x)$ and writes $\mathrm{Pr}(\varphi)$ for $\mathrm{Prov}(\ulcorner\varphi\urcorner)$. Five facts are proved.

(D1) If Peano arithmetic derives φ, it derives $\mathrm{Pr}(\ulcorner\varphi\urcorner)$: the derivation becomes a closed Hilbert proof, and in every model each axiom has a certificate while modus ponens and generalization extend certificates.

(D2) Peano arithmetic derives $\mathrm{Pr}(\ulcorner\varphi \to \psi\urcorner) \to \mathrm{Pr}(\ulcorner\varphi\urcorner) \to \mathrm{Pr}(\ulcorner\psi\urcorner)$: two certificates concatenate with one modus ponens record.

(D3) Peano arithmetic derives $\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \mathrm{Pr}(\ulcorner\mathrm{Pr}(\ulcorner\varphi\urcorner)\urcorner)$. In a model every element x has the code of the numeral term $S^x(0)$; codes commute with substitution; every formula true in all models has a certificate at every instance by closed terms; and every true bounded formula has a certificate, by induction on bounded formulas, with a bounded universal quantifier handled by induction on its bound.

(D4) There is a sentence G with $G \leftrightarrow \neg\,\mathrm{Pr}(\ulcorner G\urcorner)$ derivable: with $\mathrm{Dg}(x, y)$ the graph of substituting the numeral of x into the formula coded by x, take $\theta(x) := \forall y(\mathrm{Dg}(x, y) \to \neg\,\mathrm{Pr}(y))$ and $G := \theta(\ulcorner\theta\urcorner)$.

(D5) Peano arithmetic derives $\mathrm{Con(PA)} \to \neg \Pr(\ulcorner \bot \urcorner)$: from a certificate of $\bot$ in a model, comprehension over β-sequences builds the tables of an accepted checker proof of $\bot$, so the model refutes Con(PA).

From (D1)–(D4), Gödel's propositional argument derives $\neg \Pr(\ulcorner \bot \urcorner) \to G$. Write P_G for $\Pr(\ulcorner G \urcorner)$. Assume P_G. By (D3), $\Pr(\ulcorner P_G \urcorner)$. By (D4) Peano arithmetic derives $G \to \neg P_G$, so (D1) and (D2) give $\Pr(\ulcorner \neg P_G \urcorner)$, that is, $\Pr(\ulcorner P_G \to \bot \urcorner)$, and (D2) again gives $\Pr(\ulcorner \bot \urcorner)$. So P_G implies $\Pr(\ulcorner \bot \urcorner)$, and $\neg \Pr(\ulcorner \bot \urcorner)$ implies $\neg P_G$, which is G by (D4). If Peano arithmetic derived Con(PA), it would derive G by (D5); then (D1) gives $\Pr(\ulcorner G \urcorner)$ while G gives its negation, and Peano arithmetic would derive $\bot$. □

Proposition 20.43 (A criterion for strictness)**.** *Let σ be a sentence that Peano arithmetic does not derive. Then $\mathscr{E}^+$ does not prove σ^N, so any extension of $\mathscr{E}^+$ that proves σ^N proves a negative arithmetic sentence that $\mathscr{E}^+$ does not.*

Proof. If $\mathscr{E}^+$ proved σ^N, Theorem 20.39 would give a Peano derivation of σ. □

A universe supplies such an extension. Let $\mathscr{F}_U$ extend $\mathscr{E}^+$ by one universe: a type of codes, with codes for O, the empty type, identity types and products, a decoding that sends each code to the type it names, and the orbit eliminator allowed to produce codes. Section 9.2 excludes universes from $\mathscr{E}^+$. In $\mathscr{F}_U$ the discriminator need not be primitive: $D(t)$ is the decoding of a code defined by recursion on t, and its two equations hold by computation; the derivations of $\mathscr{E}^+$ are translated accordingly.

Theorem 20.44 (One universe proves the consistency of arithmetic)**.**

1. *$\mathscr{F}_U$ proves* $\mathrm{Con(PA)}^N$.
2. *Every derivation of $\mathscr{E}^+$ translates into one of $\mathscr{F}_U$, so $\mathscr{F}_U$ proves the negative translation of every sentence that Peano arithmetic derives.*
3. *In set theory with a universe of sets (a transitive set closed under pairs, unions, power sets and images of its members under maps into it, and containing the set of numerals; the sets of rank below a strongly inaccessible cardinal form one), $\mathscr{F}_U$ is sound for a set model in which U is read as that universe: it has no closed term of the empty type, and every arithmetic sentence whose negative translation it proves is true. In particular, in that metatheory, Peano arithmetic is consistent. In a type-theoretic metatheory the model uses sets built over two type universes, the small sets over the lower one.*
4. *$\mathscr{E}^+$ does not prove* $\mathrm{Con(PA)}^N$*, so $\mathscr{F}_U$ proves a closed negative arithmetic sentence that $\mathscr{E}^+$ does not.*

Proof. The full proof is in Appendix L, Sections L.1–L.8; its route is this. For (1), the universe lets an elimination over the orbit produce codes, so $\mathscr{F}_U$ reads every formula code, under an environment, as a type: an equation as the double negation of an identity, $\bot$ as $\mathbf{0}$, implication as a function type and $\forall$ as a product over O. Given tables that satisfy the clauses of Con(PA), induction on the fuel of the checker shows that every certified line has an inhabited reading,

case by case on the fourteen schemas and the two rules; the induction schema is the case that needs the universe, because its motive varies with a code. A certified proof of $\bot$ would therefore inhabit $\mathbf{0}$. For (2), the discriminator becomes the decoding of a code defined by recursion, and Theorem 20.39 supplies the arithmetic. For (3), types are read as sets and U as the universe of small sets; raw conversion preserves the values of expressions that satisfy a safety condition, by parallel reduction and its confluence, and soundness is by induction on derivations. The metatheory of (3) is stronger than that of Theorem 20.41, which already gives the consistency of Peano arithmetic. Part (4) is Proposition 20.43 with Theorem 20.42. □

Finite sets and bounded arithmetic. The comparison also reads finite set theory. Let $\mathrm{Pow}_2(x, p)$ be a formula with free variables among x and p, and read membership as binary digits:

$$x \,\varepsilon\, y \;:\Longleftrightarrow\; \exists p \le y \exists q \le y \exists r < p \,\big(\mathrm{Pow}_2(x, p) \wedge y = (2q + 1)p + r\big).$$

When Pow_2 defines $p = 2^x$, this says that binary digit x of y is 1. *Bounded arithmetic* $I\Delta_0$ consists of the rules with excluded middle and with induction only on bounded formulas, those whose quantifiers have the form $\forall y \le t$ or $\exists y \le t$; its derivations record neither LPO nor MP. A *standard bounded graph* is a bounded formula $\mathrm{Pow}_2(x, p)$ that defines $p = 2^x$ in $\mathbb{N}$ and for which $I\Delta_0$ derives $\mathrm{Pow}_2(0, 1)$; $\mathrm{Pow}_2(x, p) \to \mathrm{Pow}_2(S(x), 2p)$; $\mathrm{Pow}_2(S(x), q) \to \exists p \le q(q = 2p \wedge \mathrm{Pow}_2(x, p))$; that p is a function of x; that p is positive; and $\mathrm{Pow}_2(x, p) \to \forall y \le x \exists p'\ \mathrm{Pow}_2(y, p')$.

Proposition 20.45 (Pair formation gives exponentiation)**.** *For every formula* $\mathrm{Pow}_2(x, p)$*, the rules derive* $\forall x \exists p\ \mathrm{Pow}_2(x, p)$ *from the pairing axiom* $\forall x \forall y \exists z \forall w(w \,\varepsilon\, z \leftrightarrow (w = x \vee w = y))$*, with the empty ledger.*

Proof. Apply the pairing axiom to x and x: some z satisfies $w \,\varepsilon\, z \leftrightarrow (w = x \vee w = x)$ for every w. At $w = x$ the right side holds by reflexivity, so $x \,\varepsilon\, z$, and the conjunct $\mathrm{Pow}_2(x, p)$ of $x \,\varepsilon\, z$ gives it for some p. The derivation uses the hypothesis, the rules for the connectives and quantifiers and reflexivity, and nothing else. □

Theorem 20.46 (Pair formation and exponentiation)**.** *There is a standard bounded graph* Pow_2 *for which* $I\Delta_0$ *derives the pairing axiom from* $\forall x \exists p\ \mathrm{Pow}_2(x, p)$*. With Proposition 20.45,* $I\Delta_0$ *derives each of pair formation and exponentiation from the other.*

Proof. The full proof is in Appendix L, Sections L.9–L.12; its route is this. The classical rules are complete for their models, by Henkin's construction with fresh free variables as witnesses, and a closed induction axiom for a bounded formula has a derivation with one bounded induction, so $I\Delta_0$ derives a sentence from given sentences exactly when it holds in every model of $I\Delta_0$ that satisfies them. The arithmetic is then done in an arbitrary model. The graph states $p = 2^x$ through a certificate of size polynomial in p: a number $N < 2p$ marks with its binary digits the checkpoints $2^{\lfloor x/2^j \rfloor}$, a number $B < p^2$ records the exponents $\lfloor x/2^j \rfloor$ in its digit blocks, and each checkpoint u steps to u^2 or $2u^2$. Every certificate is a smaller certificate with one new top checkpoint, which proves the six properties by strong induction on p over bounded formulas. For pair formation take $z = 2^x + 2^y$ when $x \neq y$ and $z = 2^x$ when $x = y$, and read off the digits by division by powers of 2. □

Finite set theory $\mathrm{ZF}^*_{\mathrm{fin}}$ is the classical first-order theory of $\in$ with extensionality, empty set, pairing, union, power set, separation and replacement for every formula, foundation, the negation of infinity, and transitive containment, $\forall x\, \exists u\, (x \subseteq u \wedge u \text{ transitive})$. The *Ackermann reading* [12] φ^a of a set-theoretic sentence quantifies over numbers and reads $x \in y$ as $x\ \varepsilon\ y$.

Theorem 20.47 (Finite set theory is Peano arithmetic)**.** *Let* Pow_2 *be a formula for which Peano arithmetic derives* $\mathrm{Pow}_2(0,1)$*, upward recursion and functionality, as it does for a standard bounded graph. Then for every sentence* φ *of set theory,* $\mathrm{ZF}^*_{\mathrm{fin}}$ *derives* φ *exactly when Peano arithmetic derives* φ^a*. For a standard bounded graph this holds exactly when* $\mathscr{E}^+$ *proves* $(\varphi^a)^N$*.*

Standard theorem; the argument for these calculi. The full argument is in Appendix M, Section M.6. The theorem is Kaye and Wong's [13, Theorems 3.1, 6.5 and 6.6, pp. 500–504], where $\mathrm{ZF}^*_{\mathrm{fin}}$ is called ZF−inf*; here it is proved for these calculi and every graph with the three properties. Both directions go through models, with completeness proved by Henkin's construction for Peano arithmetic and for the calculus of $\mathrm{ZF}^*_{\mathrm{fin}}$. If $\mathrm{ZF}^*_{\mathrm{fin}}$ derives φ and M is a model of Peano arithmetic, binary membership in M satisfies every axiom: pairing, union, power set, separation and transitive containment are comprehension over the digits below a bound, replacement is collection followed by comprehension, foundation is the least number principle, and the negation of infinity takes the largest digit. So φ^a holds in M. Conversely, the von Neumann numerals of a nonempty model V of $\mathrm{ZF}^*_{\mathrm{fin}}$ form a model of Peano arithmetic, and the map sending n to the set of the images of the digit places of n, defined by course-of-values recursion, is a bijection onto V under which membership becomes binary membership; it is onto by $\in$-induction, which uses transitive containment [13, Proposition 5.6, p. 502]. The last assertion is Theorem 20.39. □

20.3.12 Self-reference: Löb's theorem and a tower of consistency statements

Theorem 20.42 shows that Peano arithmetic does not derive its own consistency. What can it derive about its own provability? Take a sentence H that says of itself that it is provable: $H \leftrightarrow \mathrm{Pr}(\ulcorner H \urcorner)$. Henkin asked whether such a sentence is a theorem. Löb's answer is yes, and his theorem settles a wider question: Peano arithmetic certifies the soundness of its proof of φ, in the form $\mathrm{Pr}(\ulcorner \varphi \urcorner) \to \varphi$, only when it proves φ outright. A second question follows. Peano arithmetic is consistent, so the sentence $\neg\, \mathrm{Pr}(\ulcorner \bot \urcorner)$, which says so, is true, and we may add it as an axiom. The stronger theory proves the consistency of Peano arithmetic. Does it prove its own?

The metatheory is classical, as from Theorem 20.42 on. $\mathrm{Prov}(x)$ is the provability formula of Appendix M.2, written $\mathrm{Pr}(x)$ in the proof of Theorem 20.42. As in that proof, $\ulcorner \varphi \urcorner$ is the code of φ (inside a formula, its numeral), and $\mathrm{Pr}(\ulcorner \varphi \urcorner)$ is the sentence $\mathrm{Prov}(\ulcorner \varphi \urcorner)$, which Appendix M writes $\mathrm{Pr}(\varphi)$. The derivability conditions (D1)–(D5) are those of that proof, proved in Appendix M (Propositions M.4, M.5, M.9, M.6 and M.7). A Σ_1 *sentence* is a sentence $\exists x_1 \cdots \exists x_k\, B$ whose matrix B is bounded: built from equations and $\bot$ by $\wedge$, $\vee$, $\to$ and the bounded quantifiers $\forall y \le t$ and $\exists y \le t$, as in Section L.9. Each $\mathrm{Pr}(\ulcorner \varphi \urcorner)$ is one, since Prov is twelve existential quantifiers over a bounded formula (Appendix M.2).

Finite extensions. For a finite list $\Gamma = (\gamma_1, \dots, \gamma_n)$ of sentences, write $\Gamma \Rightarrow \varphi$ for $\gamma_1 \to (\gamma_2 \to \cdots \to (\gamma_n \to \varphi))$. Say that PA + Γ *derives* φ when Peano arithmetic derives $\Gamma \Rightarrow \varphi$; by implication introduction and elimination this is the same as $\Gamma \vdash_{\mathrm{PA}} \varphi$. Put

$$\mathrm{Pr}_\Gamma(\ulcorner\varphi\urcorner) := \mathrm{Pr}(\ulcorner\Gamma \Rightarrow \varphi\urcorner), \qquad \mathrm{Con}(\mathrm{PA}+\Gamma) := \neg\,\mathrm{Pr}_\Gamma(\ulcorner\bot\urcorner).$$

PA + Γ is *consistent* when it does not derive $\bot$. For Γ empty, Con(PA + Γ) is $\neg\,\mathrm{Pr}(\ulcorner\bot\urcorner)$, which (D5) derives from Con(PA).

Theorem 20.48 (The diagonal lemma)**.** *For every formula $\psi(y)$, whose other free variables are parameters, there is a formula D with the same parameters, a sentence when ψ has no other free variable, such that Peano arithmetic derives $D \to \psi(\ulcorner D\urcorner)$ and $\psi(\ulcorner D\urcorner) \to D$.*

Proof. This is the proof of Proposition M.6 with $\neg\,\mathrm{Prov}$ replaced by ψ. Let $\mathrm{Dg}(x, y)$ be the formula of that proof, put $\theta(x) := \forall y(\mathrm{Dg}(x, y) \to \psi(y))$, and let $D := \theta(\ulcorner\theta\urcorner)$. In every model of Peano arithmetic, $\mathrm{Dg}(\ulcorner\theta\urcorner, y)$ holds exactly for $y = \ulcorner D\urcorner$, by Lemma M.2(c) and (d). So at every assignment of the parameters, D holds exactly when $\psi(\ulcorner D\urcorner)$ does, and Theorem M.1 gives both implications. □

Theorem 20.49 (Löb's theorem)**.** *For every formula φ: if Peano arithmetic derives $\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \varphi$, it derives φ. Hence it derives $\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \varphi$ exactly when it derives φ. The diagonal sentence H of $\mathrm{Prov}(y)$, for which Peano arithmetic derives $H \leftrightarrow \mathrm{Pr}(\ulcorner H\urcorner)$, is a theorem of Peano arithmetic.*

Proof. Apply Theorem 20.48 to $\mathrm{Prov}(y) \to \varphi$: some L has $L \leftrightarrow (\mathrm{Pr}(\ulcorner L\urcorner) \to \varphi)$ derivable. Write P for $\mathrm{Pr}(\ulcorner L\urcorner)$. From $L \to (P \to \varphi)$, (D1) and (D2) twice give $P \to (\mathrm{Pr}(\ulcorner P\urcorner) \to \mathrm{Pr}(\ulcorner\varphi\urcorner))$, and (D3) gives $P \to \mathrm{Pr}(\ulcorner P\urcorner)$; so $P \to \mathrm{Pr}(\ulcorner\varphi\urcorner)$. With the hypothesis $\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \varphi$ this gives $P \to \varphi$, which is L by the diagonal property. Then (D1) gives P, and so φ. If Peano arithmetic derives φ, it derives $\mathrm{Pr}(\ulcorner\varphi\urcorner) \to \varphi$ trivially. For H, the diagonal property gives $\mathrm{Pr}(\ulcorner H\urcorner) \to H$, and the first assertion gives H. □

The theorem is Löb's [10]. At $\varphi = \bot$ it is Gödel's second theorem again: $\mathrm{Pr}(\ulcorner\bot\urcorner) \to \bot$ is $\neg\,\mathrm{Pr}(\ulcorner\bot\urcorner)$, so a derivation of it would give a derivation of $\bot$.

Theorem 20.50 (Gödel's second theorem for finite extensions)**.** *Let Γ be a finite list of sentences.*

1. *If* PA + Γ *is consistent, it does not derive* Con(PA + Γ).
2. *If* PA + Γ *derives* $\mathrm{Pr}_\Gamma(\ulcorner\varphi\urcorner) \to \varphi$*, it derives φ.*

Proof. The proofs of Theorems 20.49 and 20.42 use only these facts about the theory and its provability formula: the theory contains Peano arithmetic and is closed under modus ponens, (D1), (D2) and (D3) hold, and the needed diagonal sentences exist. We check them for PA+Γ and Pr_Γ. The theory is closed under modus ponens, since $(\Gamma \Rightarrow (\varphi \to \psi)) \to ((\Gamma \Rightarrow \varphi) \to (\Gamma \Rightarrow \psi))$ is a tautology, and it contains every theorem of Peano arithmetic, since $A \to (\Gamma \Rightarrow A)$ is one. (D1): if PA + Γ derives φ, then (D1) gives $\mathrm{Pr}(\ulcorner\Gamma \Rightarrow \varphi\urcorner)$. (D2): (D1) and (D2) at the first tautology give $\mathrm{Pr}_\Gamma(\ulcorner\varphi \to \psi\urcorner) \to (\mathrm{Pr}_\Gamma(\ulcorner\varphi\urcorner) \to \mathrm{Pr}_\Gamma(\ulcorner\psi\urcorner))$. (D3): (D3) gives $\mathrm{Pr}(\ulcorner A\urcorner) \to \mathrm{Pr}(\ulcorner\mathrm{Pr}(\ulcorner A\urcorner)\urcorner)$ for $A := \Gamma \Rightarrow \varphi$, and (D1) and (D2) at the second tautology turn $\mathrm{Pr}(\ulcorner\mathrm{Pr}(\ulcorner A\urcorner)\urcorner)$ into $\mathrm{Pr}_\Gamma(\ulcorner\mathrm{Pr}_\Gamma(\ulcorner\varphi\urcorner)\urcorner)$.

Diagonal sentences: let $\mathrm{Prov}_\Gamma(y)$ say that Prov holds at the code of $\Gamma \Rightarrow \eta$, for η the formula coded by y; the code of an implication is computed from the codes of its parts by the pairing of Appendix M.2, which a formula defines. In every model, at the numeral of $\ulcorner \eta \urcorner$ this formula holds exactly when $\mathrm{Pr}_\Gamma(\ulcorner \eta \urcorner)$ does, so Theorem 20.48 at $\neg\,\mathrm{Prov}_\Gamma(y)$ and at $\mathrm{Prov}_\Gamma(y) \to \varphi$, with Theorem M.1, gives a sentence G_Γ, and a formula L with the free variables of φ, such that Peano arithmetic derives $G_\Gamma \leftrightarrow \neg\,\mathrm{Pr}_\Gamma(\ulcorner G_\Gamma \urcorner)$ and $L \leftrightarrow (\mathrm{Pr}_\Gamma(\ulcorner L \urcorner) \to \varphi)$.

Now (2) is the proof of Theorem 20.49 read in PA + Γ. For (1), the propositional argument of Theorem 20.42, read in PA + Γ, derives $\neg\,\mathrm{Pr}_\Gamma(\ulcorner \bot \urcorner) \to G_\Gamma$. If PA + Γ derived $\mathrm{Con}(\mathrm{PA} + \Gamma)$, which is $\neg\,\mathrm{Pr}_\Gamma(\ulcorner \bot \urcorner)$ literally, it would derive G_Γ, then $\mathrm{Pr}_\Gamma(\ulcorner G_\Gamma \urcorner)$ by (D1), while G_Γ gives its negation; so it would derive $\bot$. No counterpart of (D5) is needed. □

For Γ empty, part (1) says that Peano arithmetic does not derive $\neg\,\mathrm{Pr}(\ulcorner \bot \urcorner)$; with (D5) this gives Theorem 20.42 again. To use the theorem we need to know which extensions are consistent. The provability formula answers that in $\mathbb{N}$.

Theorem 20.51 (The provability formula reads correctly in $\mathbb{N}$)**.** *For every formula X and every finite list Γ of sentences:*

1. $\mathrm{Pr}(\ulcorner X \urcorner)$ *is true in* $\mathbb{N}$ *exactly when Peano arithmetic derives X;*
2. $\mathrm{Con}(\mathrm{PA} + \Gamma)$ *is true in* $\mathbb{N}$ *exactly when* PA + Γ *is consistent;*
3. *the reflection instance* $\mathrm{Pr}(\ulcorner X \urcorner) \to X$ *is true in* $\mathbb{N}$ *at every assignment, and Peano arithmetic derives it only if it derives X.*

Proof. (1) If Peano arithmetic derives X, Lemma M.3(d) gives a closed Hilbert derivation of a formula with the code $\ulcorner X \urcorner$, and the construction in the proof of Proposition M.4, carried out in $\mathbb{N}$, gives a certificate. Conversely, take a certificate in $\mathbb{N}$. Decode every number as an expression: tags 2 to 6 as a variable, 0, a successor, a sum and a product, tags 7 to 10 as an equation, $\bot$, an implication and a universal statement, and any other number as the term 0 in a term position and as $\bot$ in a formula position. The decoding is total; no well-formedness is assumed. By strong induction on codes, each locally correct record of kinds 1 to 4 computes its operation on the decoded expressions. Hence the conclusion of each axiom record decodes to an instance of its schema, whatever its parameters, and each schema instance is true in every model of Peano arithmetic. Modus ponens and generalization preserve truth in every model. The premises of a proof node are its children, hence smaller, so strong induction on the code of proof nodes shows that every proof record's conclusion decodes to a formula true in every model. The root's conclusion $\ulcorner X \urcorner$ decodes to $X^\star$, the formula of Lemma M.3 with $\wedge$, $\vee$ and $\exists$ written classically, which is true in exactly the same structures as X. Theorem M.1 gives a derivation of X.

(2) Apply (1) to $X := \Gamma \Rightarrow \bot$.

(3) If $\mathrm{Pr}(\ulcorner X \urcorner)$ is true, Peano arithmetic derives X by (1), and X is true at every assignment, since every derivation of Peano arithmetic is sound in $\mathbb{N}$ in a classical metatheory (Corollary 6.15 and Theorem 8.11). The second assertion is Theorem 20.49. □

So every reflection instance is true, and Peano arithmetic proves exactly those it needs no reflection for. The next fact is the internal form of part (1) for Σ_1 sentences.

Proposition 20.52 (Σ_1 completeness inside Peano arithmetic)**.** *For every Σ_1 sentence σ, Peano arithmetic derives $\sigma \to \mathrm{Pr}(\ulcorner \sigma \urcorner)$.*

Proof. Let σ be $\exists x_1 \cdots \exists x_k\, B$, and let σ hold in a model M with witnesses $a_1, \ldots, a_k$. By Lemma M.8(d), the code of B with the numeral codes of the a_i plugged in has a certificate in M. Existential introduction is an instance of a propositional and quantifier law, so Lemma M.8(b) and modus ponens add the k quantifiers one at a time, as in the proof of Proposition M.9, and give a certificate for $\ulcorner \sigma \urcorner$. Theorem M.1 gives the derivation. □

The tower. Put $\Gamma_0 := ()$ and $\Gamma_{k+1} := (\mathrm{Con}(\mathrm{PA} + \Gamma_k), \Gamma_k)$, and write T_k for $\mathrm{PA} + \Gamma_k$, Pr_{T_k} for Pr_{Γ_k} and $\mathrm{Con}(T_k)$ for $\mathrm{Con}(\mathrm{PA} + \Gamma_k)$. So T_0 is Peano arithmetic and $T_{k+1} = T_k + \mathrm{Con}(T_k)$. Such progressions of theories go back to Turing [71].

Theorem 20.53 (A tower of consistency statements)**.** *For every k:*

1. *T_k is consistent;*
2. *T_{k+1} derives $\mathrm{Con}(T_k)$;*
3. *for every Σ_1 sentence σ, T_{k+1} derives $\mathrm{Pr}_{T_k}(\ulcorner \neg\sigma \urcorner) \to \neg\sigma$;*
4. *T_k does not derive $\mathrm{Con}(T_k)$;*
5. *T_k does not derive $\mathrm{Pr}_{T_k}(\ulcorner \neg(0 = 0) \urcorner) \to \neg(0 = 0)$.*

Proof. (1) Prove by induction on k that every member of Γ_k is true in $\mathbb{N}$ and that T_k is consistent. For $k = 0$ there is no member, and Peano arithmetic is consistent (Theorem 20.41). For the step, $\mathrm{Con}(T_k)$ is true by Theorem 20.51(2) and the induction hypothesis, so every member of Γ_{k+1} is true. If Peano arithmetic derived $\Gamma_{k+1} \Rightarrow \bot$, that formula would be true in $\mathbb{N}$, by soundness as in the proof of Theorem 20.51(3), and its premises are true, so $\bot$ would be true.

(2) $\mathrm{Con}(T_k)$ is the first member of Γ_{k+1}.

(3) Argue in T_{k+1}. Assume $\mathrm{Pr}_{T_k}(\ulcorner \neg\sigma \urcorner)$ and σ. Proposition 20.52 gives $\mathrm{Pr}(\ulcorner \sigma \urcorner)$, and (D1) and (D2) at the tautology $\sigma \to (\Gamma_k \Rightarrow \sigma)$ give $\mathrm{Pr}_{T_k}(\ulcorner \sigma \urcorner)$. (D1) and (D2) at the tautology $(\Gamma_k \Rightarrow \sigma) \to ((\Gamma_k \Rightarrow \neg\sigma) \to (\Gamma_k \Rightarrow \bot))$ give $\mathrm{Pr}_{T_k}(\ulcorner \bot \urcorner)$, against $\mathrm{Con}(T_k)$. So $\neg\sigma$.

(4) Theorem 20.50(1) with (1).

(5) By Theorem 20.50(2), a derivation of the instance would give a derivation of $\neg(0 = 0)$ in T_k, and with reflexivity a derivation of $\bot$, against (1). □

Each level certifies the consistency of the level below and more: by (3), whenever T_k refutes a Σ_1 sentence, T_{k+1} certifies that the refutation is correct. No level certifies either fact about itself, and by (5) not even the single reflection instance at $0 = 0$. Reflection for Σ_1 sentences themselves, rather than their negations, lies higher: the instance at A_k two levels up (Theorem 20.54(3)), and the whole Σ_1 schema for T_k at no level of the tower (Theorem 20.54(4)).

Theorem 20.54 (Reflection at A_k lies two levels up). *For every k, let $A_k := \mathrm{Pr}_{T_k}(\ulcorner \bot \urcorner)$, a Σ_1 sentence, and let $\rho_k := \mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner) \to A_k$ be the Σ_1 reflection instance for T_k at A_k.*

1. *ρ_k is true in $\mathbb{N}$.*
2. *T_{k+1} derives ρ_k exactly when it derives $\mathrm{Con}(T_{k+1})$; so it does not derive ρ_k.*
3. *T_{k+2} derives ρ_k.*
4. *For all k and j, let $\sigma_{k,j}$ be $\mathrm{Con}(T_{k+j-1}) \to (\cdots \to (\mathrm{Con}(T_k) \to A_{k+j}))$, with $\sigma_{k,0} = A_k$. The instance $\mathrm{Pr}_{T_k}(\ulcorner \sigma_{k,j} \urcorner) \to \sigma_{k,j}$ is true in $\mathbb{N}$, and T_j does not derive it.*

Proof. Since $\Gamma_{k+1} \Rightarrow \bot$ is $\neg A_k \to (\Gamma_k \Rightarrow \bot)$, (D1) and (D2) at the two classical tautologies $(\neg A \to (\Gamma \Rightarrow \bot)) \to (\Gamma \Rightarrow A)$ and $(\Gamma \Rightarrow A) \to (\neg A \to (\Gamma \Rightarrow \bot))$ give

$$\vdash_{\mathrm{PA}} A_{k+1} \to \mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner) \qquad \text{and} \qquad \vdash_{\mathrm{PA}} \mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner) \to A_{k+1}. \tag{20.3}$$

(1) If $\mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner)$ is true, Peano arithmetic derives $\Gamma_k \Rightarrow A_k$ (Theorem 20.51(1)); that formula is then true, and the members of Γ_k are true (proof of Theorem 20.53(1)), so A_k is true.

(2) Suppose T_{k+1} derives ρ_k. Argue in T_{k+1}: from A_{k+1}, (20.3) gives $\mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner)$ and ρ_k gives A_k, against the member $\mathrm{Con}(T_k) = \neg A_k$; so $\neg A_{k+1}$, which is $\mathrm{Con}(T_{k+1})$. Conversely, if T_{k+1} derives $\neg A_{k+1}$, then by (20.3) it refutes $\mathrm{Pr}_{T_k}(\ulcorner A_k \urcorner)$, and ρ_k holds. By Theorems 20.50(1) and 20.53(1), T_{k+1} does not derive $\mathrm{Con}(T_{k+1})$.

(3) T_{k+2} has the member $\mathrm{Con}(T_{k+1}) = \neg A_{k+1}$, and the converse direction of (2) applies in T_{k+2}.

(4) Truth is as in (1). For the rest, let Δ list the members of Γ_{k+i} above Γ_k, so that $\Gamma_{k+i} \Rightarrow \varphi$ is $\Delta \Rightarrow (\Gamma_k \Rightarrow \varphi)$, and $\sigma_{k,i}$ is $\Delta \Rightarrow A_{k+i}$. The first half of (20.3) at $k+i$, followed by (D1) and (D2) at the tautology that moves the premises Δ past Γ_k, gives

$$\vdash_{\mathrm{PA}} A_{k+i+1} \to \mathrm{Pr}_{T_k}(\ulcorner \sigma_{k,i} \urcorner).$$

In T_{k+i+1} every member of Δ holds and $\neg A_{k+i}$ holds, so $\neg \sigma_{k,i}$. A derivation of the instance at $\sigma_{k,i}$ in T_{k+i+1} would therefore give $\neg \mathrm{Pr}_{T_k}(\ulcorner \sigma_{k,i} \urcorner)$, hence $\neg A_{k+i+1}$, which is $\mathrm{Con}(T_{k+i+1})$; by Theorems 20.50(1) and 20.53(1) there is none. Take $i := j$. Every member of Γ_j is a member of Γ_{k+j+1}, so T_j derives only what T_{k+j+1} derives, and T_j does not derive the instance. □

Consistency gives reflection for negations of Σ_1 sentences, by the proof of Theorem 20.53(3), and it climbs one level per step. Reflection at the Σ_1 sentence A_k is two steps away. Each $\sigma_{k,j}$ is equivalent in Peano arithmetic to $A_k \vee \cdots \vee A_{k+j}$, since $\mathrm{Con}(T_m)$ is $\neg A_m$, and so to a Σ_1 sentence, obtained by merging the existential quantifiers; and by (D1) and (D2), sentences that Peano arithmetic proves equivalent have equivalent reflection instances. So no level of the tower derives every Σ_1 reflection instance for any level. That Σ_1 reflection, already in this local form, and a fortiori uniform Σ_1 reflection, is stronger than consistency is a classical theme of reflection principles [72].

A true sentence about finite sets. Theorem 20.47 turns the second theorem into a statement about finite sets. Let θ be the set-theoretic sentence that reads $\neg \mathrm{Pr}(\ulcorner \bot \urcorner)$ in the von Neumann natural numbers, with addition and multiplication defined by recursion as in Lemma M.11. The *hereditarily finite sets* are the sets obtained from the empty set by finitely many steps of forming finite sets of sets already obtained; under binary membership they are the natural numbers [12].

Proposition 20.55 (A true sentence of finite set theory)**.** *The sentence θ is true in the hereditarily finite sets and is not derivable in* $\mathrm{ZF}^*_{\mathrm{fin}}$. *For every formula* Pow_2 *as in Theorem 20.47, Peano arithmetic derives $\theta^a \leftrightarrow \neg \mathrm{Pr}(\ulcorner \bot \urcorner)$.*

Proof. Let M be a model of Peano arithmetic. Read through binary membership, it is a model of $\mathrm{ZF}^*_{\mathrm{fin}}$ (Lemma M.10). Define $g(0) := 0$ and $g(n+1) := g(n) + 2^{g(n)}$, the number whose binary digits are those of $g(n)$ together with digit $g(n)$, that is, $g(n) \cup \{g(n)\}$. By induction in M, g is injective and each $g(n)$ is a von Neumann natural; every von Neumann natural u is some $g(n)$, by induction along membership, since a nonzero natural is $v \cup \{v\}$ for a natural v; and the recursive addition and multiplication of the set model satisfy, under g, the recursion equations of $+$ and $\cdot$ in M, so g carries them to $+$ and $\cdot$, again by induction. Hence M satisfies θ^a exactly when it satisfies $\neg \mathrm{Pr}(\ulcorner \bot \urcorner)$, and Theorem M.1 gives derivations of both implications. For $\mathbb{N}$ with a standard bounded graph (Theorem 20.46), binary membership is the membership of the hereditarily finite sets, and $\neg \mathrm{Pr}(\ulcorner \bot \urcorner)$ is true there by Theorem 20.51(2) and Theorem 20.41; so θ is true in the hereditarily finite sets. If $\mathrm{ZF}^*_{\mathrm{fin}}$ derived θ, Peano arithmetic would derive θ^a, by Theorem 20.47, and so $\neg \mathrm{Pr}(\ulcorner \bot \urcorner)$, against Theorem 20.50(1) for Γ empty. □

20.3.13 A small case

What does conservativity say in practice? Consider a proof that passes through all functions on O. In the language of functions, prove

$$L: \qquad \forall f{:}O \to O \;\; \forall x{:}O \;\; \forall n{:}O \quad \mathsf{I}\big(O, fx, f, n\big) =_O f\big(\mathsf{I}(O, x, f, n)\big).$$

Starting one step later is the same as taking one more step at the end. Induct on n. At 0, both sides compute to fx. If the equation holds at n, then

$$\mathsf{I}\big(O, fx, f, S(n)\big) \equiv f\big(\mathsf{I}(O, fx, f, n)\big) = f\big(f(\mathsf{I}(O, x, f, n))\big) \equiv f\big(\mathsf{I}(O, x, f, S(n))\big),$$

by computation, replacement with the induction hypothesis, and computation. Now instantiate f by the successor function $\lambda O.\, S(x_0)$. Addition is iteration of that function, so computation turns the instance into

$$B: \qquad \forall x \;\; \forall n \quad S(x) + n = S(x + n).$$

One induction proved a fact about every function, and the arithmetic fact is one of its instances. The sentence B is built from $=$ and $\forall$.

Theorem 20.32 says that this kind of route never reaches a closed negative sentence beyond the posit-free derivations. For B the direct derivation is visible: Lemma 3.2 proved B by induction on n, and the same induction, written with the rules of Chapter 5, uses the quantifier-free formula $S(x) + n = S(x + n)$ and has an empty ledger.

The compiled proof takes a different path. Its reading of L says: for every number f admissible at $O \to O$, and for all numbers x and n, the programs compiled from the two sides return values at the environment that lists f, x, n, and those values are equal. The source induction was on an equation. The compiled induction, Theorem 20.25, is on the reading of that equation, which asserts that values exist, so it inducts on a formula with quantifiers and the ledger records the induction entry. The instantiation, Proposition 20.21, must show that the code of $\lambda O.\, S(x_0)$ is admissible at $O \to O$; that is the abstraction case of Theorem 20.15, and applying that code runs through U, which is where universal application enters. Recovery, Theorem 20.30, then shows that the programs compiled from $S(x) + n$ and $S(x + n)$ return the actual numbers, and so turns the reading of B into B.

The point generalizes. A proof about numbers may pass through functions. The passage never yields a closed negative statement about numbers that the arithmetic rules could not already derive without an added principle.

20.4 What a classical proof gives without excluded middle

Here is a classical proof that every number is even or odd, that is, of

$$\forall x \exists y \big(x = y + y \ \vee \ x = S(y + y) \big).$$

Suppose some number is neither. Then there is a least such number m, by the least number principle, which excluded middle and induction provide. It is not 0, since $0 = 0 + 0$, so $m = S(k)$ for some k. Being smaller, k is even or odd: $k = y + y$ gives $m = S(y + y)$, and $k = S(y + y)$ gives $m = S(y) + S(y)$. Either way m is even or odd, which is a contradiction.

The argument never names the y that belongs to a given x. Does the proof contain one anyway? And does the theorem have a derivation that avoids excluded middle? Section 20.3 compared Peano arithmetic with a richer typed language through the negative translation. This section answers the two questions inside arithmetic, for each shape of statement. For a statement $\forall x \exists y\, A$ with a quantifier-free matrix A, which in this language has no bounded quantifiers, a classical proof always yields a derivation without excluded middle and a program that computes y from x. For $\forall x \exists y \forall z\, A$ it yields, in general, a strategy against every proposed counterexample, and when A mentions only x, y and z it yields the statement itself. For an arbitrary theorem the price is at most LPO, and one theorem pays it.

Notation. Peano arithmetic is the calculus of Definition 8.10, and $\Gamma \vdash_{\mathrm{PA}} \varphi$ says that it derives φ from the finite list Γ. We write $\Gamma \vdash_{\mathrm{HA}} \varphi$ when φ has a posit-free derivation from Γ; by Corollary 8.12 this is the same as a derivation in Heyting arithmetic, the calculus of Definition 8.10 without excluded middle. Formulas may have free variables. The negative translation φ^N is that of Section 20.3.10, and Γ^N lists the translations of the members of Γ. Negation abbreviates implication to $\bot$, so $(\neg\varphi)^N$ is $\neg\varphi^N$.

20.4.1 Translating a classical derivation step by step

Lemma 20.33 proves $\varphi \leftrightarrow \varphi^N$ with excluded middle. The point of the translation is that φ^N needs no excluded middle once φ has a classical proof. Every atom of φ^N is negated, and

negated formulas obey a law that excluded middle is usually invoked for.

Lemma 20.56 (Translated formulas are stable)**.** *For every formula φ:*

1. *the translation commutes with lifting, with substitution of a term, and with the replacement of x_0 by $S(x_0)$ in the induction rule: $(\varphi\uparrow)^N = \varphi^N\uparrow$ and $(\varphi(t))^N = \varphi^N(t)$;*
2. *$\vdash_{\mathrm{HA}} \neg\neg\varphi^N \to \varphi^N$, by a derivation that uses the rules for the connectives and quantifiers only.*

Proof. (1) The translation keeps every term and every binder in place and changes only connectives. Lifting, substitution and the step act on terms and pass through binders by the same clauses, so each commutes with the translation, by recursion on φ.

(2) Recurse on φ. When φ is an equation, a conjunction, a disjunction or an existential statement, φ^N is a negation $\neg X$, and $\neg\neg\neg X \to \neg X$ is derivable: under $\neg\neg\neg X$ and X, the hypothesis X refutes $\neg X$, which gives $\neg\neg X$, and $\neg\neg\neg X$ turns that into $\bot$. For $\bot$, apply $\neg\neg\bot$ to the identity $\bot \to \bot$. For an implication, from $\neg\neg(\varphi^N \to \psi^N)$ and φ^N derive $\neg\neg\psi^N$: a hypothesis $\neg\psi^N$ refutes $\varphi^N \to \psi^N$, against its double negation; the case for ψ then gives ψ^N. For $\forall\varphi$, from $\neg\neg\forall\varphi^N$ derive, at a fresh variable, $\neg\neg\varphi^N$: a hypothesis $\neg\varphi^N$ refutes $\forall\varphi^N$ by universal elimination at that variable. The case for φ gives φ^N, and universal introduction gives $\forall\varphi^N$. □

The lemma is the arithmetic form of Lemma 20.37. No axiom and no induction occurs in these derivations. In particular, the equations of φ^N need no decision procedure, because each occurs doubly negated.

Theorem 20.57 (The negative translation of a derivation)**.** *For every finite list Γ and every formula φ, free variables allowed: if $\Gamma \vdash_{\mathrm{PA}} \varphi$, then $\Gamma^N \vdash_{\mathrm{HA}} \varphi^N$. In particular every theorem φ of Peano arithmetic has $\vdash_{\mathrm{HA}} \varphi^N$.*

Proof. Recurse on the Peano derivation. In each case the induction hypotheses give derivations, in Heyting arithmetic, of the translated premises from the translated context, and we derive the translated conclusion. Write A, B and C for the translations of the formulas the rule mentions. Lemma 20.56(1) identifies the translation of a lifted context with the lifted translated context, and the translation of an instance with the instance of the translation; so the quantifier rules and induction keep their eigenvariable conditions. Lemma 20.56(2) is cited as *stability*: to derive C, it suffices to derive $\bot$ from the further hypothesis $\neg C$.

Hypotheses are the translated hypotheses. *Axioms.* The translation of $\neg(S(t) = 0)$ is $\neg\neg(S(t) = 0) \to \bot$, and the axiom $\neg(S(t) = 0)$ itself refutes the antecedent. The translation of $S(t) = S(u) \to t = u$ is $\neg\neg(S(t) = S(u)) \to \neg\neg(t = u)$: under $\neg(t = u)$ the axiom gives $\neg(S(t) = S(u))$, against the antecedent. Each equational axiom E gives $\neg\neg E$. *Reflexivity* gives $\neg\neg(t = t)$. *Equality substitution*, from $\neg\neg(t = u)$ and $A(t)$ to $A(u)$: by stability assume $\neg A(u)$; then $t = u$ would give $A(u)$ by substitution, so $\neg(t = u)$, against $\neg\neg(t = u)$.

Implication, the universal quantifier, the impossible case and induction translate to the same rule at the translated formulas. For induction the premises become $A(0)$ and $\forall x(A \to A(S(x)))$ and the conclusion $\forall x\, A$, by Lemma 20.56(1).

Conjunction. The translation of $\varphi \wedge \psi$ is $\neg(A \to \neg B)$. Introduction: A and B refute $A \to \neg B$. Elimination to A: by stability assume $\neg A$; then $A \to \neg B$ holds, because its antecedent is refuted, against the premise. Elimination to B: under $\neg B$ the implication $A \to \neg B$ holds outright.

Disjunction. The translation of $\varphi \vee \psi$ is $\neg\neg(\neg A \to \neg\neg B)$; any derivable formula yields its own double negation. From A, the implication $\neg A \to \neg\neg B$ holds, because A refutes its antecedent. From B, it holds because $\neg\neg B$ follows from B. Elimination with conclusion C: by stability assume $\neg C$. The two branches, discharged, give $A \to C$ and $B \to C$, hence $\neg A$ and $\neg B$. Then $\neg A \to \neg\neg B$ would give $\neg\neg B$ and so $\bot$; this refutes the implication, against the premise.

Existential quantifier. The translation of $\exists\varphi$ is $\neg\forall\neg A$. Introduction from $A(t)$: a hypothesis $\forall\neg A$, instantiated at t, gives $\neg A(t)$ and so $\bot$. Elimination with conclusion C: by stability assume $\neg C$. In the lifted context the branch turns A into the lifted C, hence into $\bot$; implication and universal introduction give $\forall\neg A$, against $\neg\forall\neg A$.

Excluded middle. The translation of $\varphi \vee \neg\varphi$ is $\neg\neg(\neg A \to \neg\neg\neg A)$. The implication inside is an instance of $P \to \neg\neg P$, which is derivable, and so is its double negation.

The construction is a recursion on the finite derivation tree and uses no principle in the metatheory. The derivation it produces inducts on A exactly where the Peano derivation inducts on the formula translated as A, and on no other: stability adds none. □

The translation is due to Gödel [65] and Gentzen [67], who used it in his consistency proof for arithmetic [66]. For a sentence, the conclusion $\vdash_{\mathrm{HA}} \varphi^N$ also follows from Theorem 20.39(1), Theorem 20.36 and Theorem 20.32, through the calculus $\mathscr{E}^+$ and the language of functions. The proof above stays in arithmetic, and it allows hypotheses and free variables. Conversely, a derivation in Heyting arithmetic is a Peano derivation, and Lemma 20.33 returns φ from φ^N with excluded middle. So $\Gamma \vdash_{\mathrm{PA}} \varphi$ holds exactly when $\Gamma^N \vdash_{\mathrm{HA}} \varphi^N$ does.

20.4.2 Two quantifiers: a witness and a program

The translation of $\exists y\, A$ is $\neg\forall y\neg A^N$. It says that a witness cannot be ruled out, not that one is at hand. For a quantifier-free matrix and one searched variable the gap is closed by the restricted Markov rule: every object-language instance $\neg\neg\exists y\, A \to \exists y\, A$ has a posit-free derivation once induction on every formula is allowed (Theorem 6.9(3)). Markov's principle for arbitrary decidable predicates is a different matter (Chapter 7). In this language a quantifier-free formula is built from equations between polynomial terms by $\bot$, $\wedge$, $\vee$ and $\to$; a bounded quantifier is a quantifier.

Lemma 20.58 (Quantifier-free translations)**.** *For every quantifier-free formula A, $\vdash_{\mathrm{HA}} A^N \to \neg\neg A$ and $\vdash_{\mathrm{HA}} \neg\neg A \to A^N$.*

Proof. Recurse on A, deriving both implications together; every step is propositional. For an equation, A^N is $\neg\neg A$. For $\bot$, the formulas $\bot$ and $\neg\neg\bot$ imply each other. For the connectives, write a and b for the components; by the induction hypotheses a^N may be replaced by $\neg\neg a$ and b^N by $\neg\neg b$ inside the translation, and $\neg\neg\neg a$ by $\neg a$. It remains to derive three equivalences. First, $\neg(\neg\neg a \to \neg b)$ is equivalent to $\neg\neg(a \wedge b)$: under $\neg(a \wedge b)$ and $\neg\neg a$, a proof of b would give $\neg a$, so $\neg b$, and the implication holds; conversely, under $\neg\neg(a \wedge b)$ and the implication,

$a \wedge b$ gives $\neg\neg a$, hence $\neg b$, a contradiction. Second, $\neg\neg a \to \neg\neg b$ is equivalent to $\neg\neg(a \to b)$: a refutation of $a \to b$ gives both $\neg b$ and $\neg\neg a$; conversely, under $\neg\neg(a \to b)$, $\neg\neg a$ and $\neg b$, the implication $a \to b$ would give $\neg a$. Third, $\neg(\neg a \to \neg\neg b)$ is equivalent to $\neg a \wedge \neg b$, hence to $\neg(a \vee b)$; negating both sides gives the equivalence of $\neg\neg(\neg a \to \neg\neg b)$ with $\neg\neg(a \vee b)$. □

Theorem 20.59 (Two quantifiers convert at no price)**.** *Let A be quantifier-free, without bounded quantifiers; its free variables other than x and y are parameters.*

1. *If $\vdash_{\mathrm{PA}} \exists y\, A$, then $\vdash_{\mathrm{HA}} \exists y\, A$.*
2. *If $\vdash_{\mathrm{PA}} \forall x\, A$, then $\vdash_{\mathrm{HA}} \forall x\, A$.*
3. *If $\vdash_{\mathrm{PA}} \forall x \exists y\, A$, then $\vdash_{\mathrm{HA}} \forall x \exists y\, A$.*

Proof. For (2), Theorem 20.57 gives $\vdash_{\mathrm{HA}} \forall x\, A^N$, and Lemma 20.58 gives $\forall x \neg\neg A$. Heyting arithmetic decides the quantifier-free A (Lemma E.1(2)); in the case $\neg A$ the impossible case applies, so $\neg\neg A \to A$, and universal introduction gives $\forall x\, A$.

For (1), Theorem 20.57 gives $\vdash_{\mathrm{HA}} \neg\forall y \neg A^N$. We derive $\neg\neg\exists y\, A$ from it. Assume $\neg\exists y\, A$. At a fresh y, a proof of A would give $\exists y\, A$, so $\neg A$; and A^N gives $\neg\neg A$ by the lemma, so $\neg A^N$. Universal introduction yields $\forall y \neg A^N$, a contradiction. By Theorem 6.9(3), the instance $\neg\neg\exists y\, A \to \exists y\, A$ of the restricted Markov rule, with y searched and every other variable a parameter, has a posit-free derivation. Implication elimination gives $\exists y\, A$.

For (3), apply the argument of (1) to the translation $\neg\forall y \neg A^N$ at a fresh x, obtained from $\forall x \neg\forall y \neg A^N$ by universal elimination, and finish with universal introduction. □

No principle is recorded, and none is used in the metatheory. The induction entry is true: the decision of quantifier-free formulas and the derivation of the Markov instance induct on formulas with quantifiers. The third assertion is the case of one searched variable of the standard Π_2 conservativity of Peano arithmetic over Heyting arithmetic [4]; with one variable, Lemma 6.8 bounds the search. Friedman's translation, in the next subsection, gives it a second proof [68], by the argument of Theorem 20.63(1) with $G := \exists y\, A$.

For the parity statement, the third assertion says that Heyting arithmetic proves

$$\forall x \exists y \big(x = y + y \;\vee\; x = S(y + y) \big).$$

The derivation it builds also computes.

Theorem 20.60 (A program for the witness)**.** *Let A be quantifier-free, and suppose $\vdash_{\mathrm{PA}} \forall x \exists y\, A$. There is a program that returns, on every input n, a number whose first component y satisfies A at $x = n$, every parameter of A being read as 0. The proof uses the classical metatheory of Appendix E.3.*

Proof. Theorem 20.59 gives a posit-free derivation d of $\forall x \exists y\, A$ from no hypotheses. Theorem E.9, applied to d, gives a program c_d; let r be its value on $\langle \ulcorner() \urcorner, \ulcorner() \urcorner \rangle$, a realizer of $\forall x \exists y\, A$ at the environment $\rho_{()}$, which gives every variable the value 0. By the clause for the universal quantifier, $\{r\}(n)\!\downarrow v$ for some v that realizes $\exists y\, A$ at $[n, \rho_{()}]$. By the clause for the existential

quantifier, $v = \langle y, q\rangle$ with q realizing A at $[y, [n, \rho_{()}]]$, and by Lemma E.7 A is then true there. The program sought runs r on its input, through the universal program; its code is computed from r by fixing part of an input. □

For the parity statement the witness is unique, so the program returns, on n, a pair whose first component is the integer part of $n/2$. Nothing in the classical proof mentioned halving. The program is assembled from the derivation, node by node, as in the proof of Theorem E.9.

20.4.3 Friedman's translation

For three quantifiers the restricted Markov rule no longer closes the gap. The translation of $\exists y\,\forall z\,A$ is $\neg\forall y\neg\forall z\,A^N$, and the rule removes a double negation only in front of an existential statement with a quantifier-free body. Friedman's idea is to keep the double negations and change the meaning of $\bot$: read it as the goal G we want to prove. A double negation $\neg\neg P$ then becomes $(P^G \to G) \to G$, and it yields G as soon as P^G is shown to imply G.

Definition 20.61 (Friedman's translation)**.** Let G be a formula. The translation φ^G of a formula φ is defined by recursion:

$$(t = u)^G := (t = u) \vee G, \qquad \bot^G := G, \qquad (\varphi \circ \psi)^G := \varphi^G \circ \psi^G \ \text{ for } \circ \in \{\wedge, \vee, \to\},$$
$$(\forall\varphi)^G := \forall\,\varphi^{G\uparrow}, \qquad (\exists\varphi)^G := \exists\,\varphi^{G\uparrow}.$$

Under a binder G is lifted, so no free variable of G is captured; in named notation, the bound variables of φ are kept apart from the free variables of G. For a list Γ, Γ^G lists the translations of its members. In particular $(\neg\varphi)^G$ is $\varphi^G \to G$.

Lemma 20.62 (Friedman's translation preserves derivations)**.** *Let G be a formula and Γ a finite list of formulas.*

1. *If $\Gamma \vdash_{\mathrm{HA}} G$, then $\Gamma \vdash_{\mathrm{HA}} \varphi^G$ for every formula φ.*
2. *If $\Gamma \vdash_{\mathrm{HA}} \varphi$, then $\Gamma^G \vdash_{\mathrm{HA}} \varphi^G$.*
3. *For every quantifier-free A: $A \vdash_{\mathrm{HA}} (A^N)^G$ and $\neg A, (A^N)^G \vdash_{\mathrm{HA}} G$.*

Proof. Like the negative translation, φ^G changes atoms only and keeps terms and binders in place. Lifting and substitution therefore commute with it, provided G is lifted with them: $(\varphi\uparrow)^{G\uparrow} = \varphi^G\uparrow$, and $(\varphi(t))^G = \varphi^{G\uparrow}(t)$, because $G\uparrow$ does not contain x_0; the same holds for the step $x_0 \mapsto S(x_0)$.

(1) Recurse on φ, for every G, every context and every substitution instance of φ at once. An equation becomes $(t = u) \vee G$, given by its right side; $\bot$ becomes G; a conjunction follows from its two components, a disjunction from its left component, and an implication from its consequent. For $\forall\varphi$, lift the context: the lifted context derives $G\uparrow$, the induction hypothesis at $G\uparrow$ gives $\varphi^{G\uparrow}$, and universal introduction concludes. For $\exists\varphi$, the induction hypothesis for the instance $\varphi(0)$ gives $\varphi(0)^G$, which is $\varphi^{G\uparrow}(0)$, and existential introduction at 0 concludes.

(2) Recurse on the derivation, for every G at once. Rules that lift the context use the induction hypothesis at $G\uparrow$, together with the commutation just stated; so universal introduction,

existential elimination and induction translate to the same rule at the translated formulas, with the same eigenvariable conditions. The connective rules, universal elimination and existential introduction are the same rule. The remaining cases use the translation of atoms. Reflexivity becomes $(t = t) \vee G$. An equational axiom E becomes $E \vee G$. The axiom $\neg(S(t) = 0)$ becomes $(S(t) = 0 \vee G) \to G$: in the first case the axiom gives $\bot$ and the impossible case gives G. Injectivity becomes $(S(t) = S(u) \vee G) \to (t = u \vee G)$, by cases. Equality substitution, from $t = u \vee G$ and $\varphi(t)^G$: in the first case substitute, and in the second (1) gives $\varphi(u)^G$. The impossible case, from G to φ^G, is (1).

(3) Recurse on A, proving both statements together and writing a^* for $(a^N)^G$. An equation $t = s$ gives $((t = s \vee G) \to G) \to G$. From $t = s$: a hypothesis $(t = s \vee G) \to G$, applied to the left side, gives G. From $\neg(t = s)$: the map $(t = s \vee G) \to G$ exists, since its first case is impossible and its second is the identity, and the translated formula turns it into G. For $\bot$, $\bot^*$ is G, and both statements are immediate. A conjunction $a \wedge b$ gives $(a^* \to (b^* \to G)) \to G$. From $a \wedge b$ the induction hypotheses supply a^* and b^*. Under $\neg(a \wedge b)$, decide a (Lemma E.1(2)): if $\neg a$, then $a^* \to G$, and if a, then $\neg b$ and $b^* \to G$; either way $a^* \to (b^* \to G)$ holds, and the hypothesis gives G. A disjunction $a \vee b$ gives $(P \to G) \to G$ with $P := (a^* \to G) \to ((b^* \to G) \to G)$. From a, the term $k \mapsto (m \mapsto k\,a^*)$ proves P, and a hypothesis $P \to G$ then gives G; from b, use $k \mapsto (m \mapsto m\,b^*)$. Under $\neg(a \vee b)$ we have $\neg a$ and $\neg b$, hence $a^* \to G$ and $b^* \to G$; applying a proof of P to these two gives G, so $P \to G$ holds, and the translated formula gives G. An implication $a \to b$ gives $a^* \to b^*$. From $a \to b$, decide a: if a, then b and b^*; if $\neg a$, then a^* gives G, and G gives b^* by (1). Under $\neg(a \to b)$, the decision of a gives a, since $\neg a$ would prove $a \to b$; so a^* holds, and with it b^*, while $\neg b$ gives $b^* \to G$. □

The translation is Friedman's [68]. Nothing in the proof uses a principle, in the derivations or in the metatheory.

20.4.4 Three quantifiers: a strategy against counterexamples

What does a classical proof of $\forall x\,\exists y\,\forall z\,A$ provide? Kreisel's answer is a game. A challenger who doubts the statement at x must answer every candidate y with a proposed counterexample z. The statement then promises that some candidate survives its own counterexample. That promise can be met without excluded middle.

Theorem 20.63 (Kreisel's no-counterexample form)**.** *Let $A(x, y, z)$ be quantifier-free with no free variable other than x, y and z, and suppose $\vdash_{\mathrm{PA}} \forall x\,\exists y\,\forall z\,A$.*

1. *For every formula $F(x, y, z)$, which may have further free variables,*

$$\vdash_{\mathrm{HA}} \ \forall x\Big[\forall y\exists z\,F(x,y,z) \ \to \ \exists y\exists z\big(F(x,y,z) \wedge A(x,y,z)\big)\Big].$$

2. *In the classical metatheory of Appendix E.3, there is a program that, on input x, returns a program that turns every realizer of $\forall y\,\exists z\,F$ at x into numbers y and z with $A(x, y, z)$ true, together with a realizer of $F(x, y, z)$; the further free variables of F are read as 0.*

Proof. (1) Theorem 20.57 gives $\vdash_{\mathrm{HA}} \forall x \neg\forall y \neg\forall z\, A^N$, and universal elimination at the variable x gives $\neg\forall y \neg\forall z\, A^N$. Put $G := \exists y \exists z\,(F \wedge A)$. By Lemma 20.62(2), Heyting arithmetic derives the translation of that formula by G:

$$\Big[\forall y\big(\forall z\,(A^N)^G \to G\big)\Big] \to G.$$

It remains to derive the premise from the hypothesis $\forall y \exists z\, F$. Fix y and assume $\forall z\,(A^N)^G$. The hypothesis gives a z with $F(x, y, z)$. Decide the quantifier-free $A(x, y, z)$ (Lemma E.1(2)). If it holds, y and z witness G. If not, instantiate the assumption at this z; Lemma 20.62(3) turns $(A^N)^G$ and $\neg A$ into G. Either way G follows, and implication and universal introduction give the claim.

(2) By (1), the formula of (1) has a posit-free derivation d from no hypotheses. As in the proof of Theorem 20.60, Theorem E.9 gives a realizer r of it at $\rho_{()}$. On x, $\{r\}(x)\!\downarrow v$ with v realizing the implication at $[x, \rho_{()}]$; so for every realizer s of $\forall y \exists z\, F$ there, $\{v\}(s)\!\downarrow w$ with w realizing $\exists y \exists z\,(F \wedge A)$. Then $w = \langle y, \langle z, \langle q_F, q_A\rangle\rangle\rangle$ with q_F realizing F and q_A realizing A at those values, and A is true by Lemma E.7. □

The only decision in the proof is the decision of a quantifier-free formula; no Markov principle and no excluded middle is used. The theorem is Kreisel's no-counterexample interpretation [55, 69], here proved inside Heyting arithmetic through Friedman's translation. Read with F the graph of a function f, it says that for every f defined by a formula some y satisfies $A(x, y, f(y))$. A first-order formula can define only some functions. The same reading in $\mathbb{N}$ reaches all of them, and it allows blocks of variables and parameters.

Theorem 20.64 (The no-counterexample form for every function)**.** *Let A be quantifier-free, with free variables x, y, a block $\bar z = (z_1, \dots, z_n)$ and any further parameters, and suppose $\vdash_{\mathrm{PA}} \forall x \exists y \forall \bar z\, A$. For every assignment of numbers to the parameters, every number x and every function f that sends each number y to an n-tuple $f(y)$, some y satisfies $A(x, y, f(y))$. Hence the least such y, found by testing $y = 0, 1, 2, \dots$ in turn, defines a functional $\Phi(x, f)$ with $A(x, \Phi(x, f), f(\Phi(x, f)))$. The proof uses no excluded middle, no omniscience principle and no choice in the metatheory.*

Proof. For a proposition G of the metatheory, read formulas in $\mathbb{N}$ as in Definition 5.3, except that an equation $t = s$ is read as "t and s have the same value, or G" and $\bot$ is read as G. Write $\rho \Vdash_G \varphi$ for this reading. It is Friedman's translation read in $\mathbb{N}$.

Soundness. If $\Gamma \vdash_{\mathrm{HA}} \varphi$, then every environment that satisfies Γ in this reading satisfies φ, for every G. Induct on the derivation as in Theorem 5.5. The connective and quantifier rules are sound for any such reading, and induction is induction in $\mathbb{N}$. The axiom $\neg(S(t) = 0)$ reads "if $S(t)$ and 0 have the same value, or G, then G", true because the values differ. Injectivity and the equational axioms are true casewise. The reading respects lifting and substitution as satisfaction does; the proof of Lemma 5.4 applies word for word. For equality substitution this settles the case of equal values; in the case G, note that G makes the reading of every formula true, by induction on the formula, which also gives the impossible case.

The reading. Fix the parameters, x and f, and let G be "some y satisfies $A(x, y, f(y))$". By Theorem 20.57 Heyting arithmetic derives $\forall x \neg\forall y \neg\forall \bar z\, A^N$, and by Lemma 20.58 it derives

$A^N \to \neg\neg A$. Read at G, the first says: if every y has the property that $\forall \bar{z}\, A^N$ in this reading implies G, then G. We check that property. Fix y and assume the reading of $\forall \bar{z}\, A^N$. At $\bar{z} := f(y)$ it gives the reading of A^N, hence that of $\neg\neg A$, which is: if the reading of A implies G, then G. For a quantifier-free A the reading of A holds exactly when A is true or G holds, by induction on A, deciding each quantifier-free component by Lemma 6.2. And the reading of A does imply G: it gives "$A(x, y, f(y))$ is true, or G", and in the first case y witnesses G. So G holds.

Every step is a construction or a decision of a quantifier-free formula. The search for the least y terminates because some y succeeds, and each test decides a quantifier-free formula. □

A small case shows the game. For a quantifier-free $B(x, z)$, consider

$$\forall x \exists y \forall z \big(B(x, z) \to B(x, y)\big) :$$

if anything has the property $B(x, \cdot)$, then y has it. A classical proof takes y to be such a thing if there is one, and anything otherwise, without saying which case holds. The no-counterexample form asks less: given, for each candidate y, a proposed counterexample $f(y)$, find a y with $B(x, f(y)) \to B(x, y)$. Two candidates suffice. Decide whether $B(x, f(0)) \to B(x, 0)$ holds. If it does, $y = 0$ wins. If not, $B(x, f(0))$ holds, and $y = f(0)$ wins, since then $B(x, y)$ holds whatever $f(y)$ is. The strategy uses the challenger's own answer and one decision of a quantifier-free formula.

More is true: by Gödel's functional interpretation, the functional Φ can be given by a closed term of Gödel's system T of primitive recursive functionals of finite type. Since Definition 8.10 is the usual presentation of Peano arithmetic, the standard result applies (Gödel [70]; see Kohlenbach [56]); it is not proved here, and nothing below depends on it.

When the matrix mentions only x, y and z, as in the small case, the game can be won outright: a single y defeats every challenger.

Corollary 20.65 (Single-variable statements with three quantifiers)**.** *Let A be quantifier-free with no free variable other than x, y and z.*

1. *If $\vdash_{\mathrm{PA}} \forall x \exists y \forall z\, A$, then $\vdash_{\mathrm{HA}} \forall x \exists y \forall z\, A$. The sentence is true in $\mathbb{N}$, with no principle used in the metatheory. In the classical metatheory of Appendix E.3, a program returns on every input x a number whose first component y satisfies $A(x, y, z)$ for every z.*
2. *If A has no free variable other than y and z, and $\vdash_{\mathrm{PA}} \exists y \forall z\, A$, then $\vdash_{\mathrm{HA}} \exists y \forall z\, A$.*

So no sentence of either shape is a theorem of Peano arithmetic and not of Heyting arithmetic.

Proof. Step 1: one searched variable. For every quantifier-free φ with a searched variable z and any parameters, $\vdash_{\mathrm{HA}} \exists z\, \varphi \vee \forall z\, \neg\varphi$. This is the alternative derived in the proof of Theorem 6.9: by the constancy (6.8), a quantifier-free formula in one variable does not change beyond the bound T, and a bounded search below T decides it.

Step 2: a test formula. Put $F(x, y, z) := \neg A(x, y, z) \vee \forall w\, A(x, y, w)$. Then $\vdash_{\mathrm{HA}} \forall x \forall y \exists z\, F$. Apply Step 1 to $\neg A$, with x and y as parameters. If some z has $\neg A$, it witnesses the left side of

F. If every z has $\neg\neg A$, then every z has A, by the decision of A, and $z := 0$ witnesses the right side.

Step 3. Theorem 20.63(1) at this *F*, with Step 2, gives $\vdash_{\mathrm{HA}} \forall x \exists y \exists z\,(F \wedge A)$.

Step 4. In $F \wedge A$ the conjunct A refutes the left side of *F*, so $\forall w\, A(x, y, w)$ holds. Hence $\vdash_{\mathrm{HA}} \forall x \exists y \forall z\, A$.

Truth in $\mathbb{N}$ follows from Corollary 6.7, which needs no principle. The program comes from Theorem E.9 as in the proof of Theorem 20.60: the realizer at x carries y as its first component, and at each z it realizes the quantifier-free $A(x, y, z)$, which is then true by Lemma E.7. For (2), universal introduction over a variable x that does not occur gives $\vdash_{\mathrm{PA}} \forall x \exists y \forall z\, A$; apply (1) and instantiate x at 0. □

Step 1 needs the innermost quantifier to bind a single variable, because Lemma 6.8 concerns one searched variable. For a block of variables the general form, Theorem 20.64, remains.

20.4.5 What each shape of classical theorem costs

The results of this section fix the price of converting a classical theorem, shape by shape, for the shapes of this language. As in Section 6.7, the price of a derivation is the set of principles its ledger records.

Theorem 20.66 (The conversion prices)**.** *Allow induction on every formula, and let φ be a theorem of Peano arithmetic. Quantifier-free formulas are those of Section 20.4.2, without bounded quantifiers.*

1. *If φ is $\exists y\, A$, $\forall x\, A$ or $\forall x \exists y\, A$ with A quantifier-free, or $\forall x \exists y \forall z\, A$ or $\exists y \forall z\, A$ with A quantifier-free in the variables shown only, then φ has a posit-free derivation. For $\forall x \exists y \forall z\, A$ so does, for every formula F, the no-counterexample form of Theorem 20.63.*
2. *φ has a derivation whose ledger records no principle other than LPO.*
3. *The sentence Λ of the proof of Theorem 6.11 is a theorem of Peano arithmetic. For every set $\mathcal{S}$ of the three principles, Λ has a derivation whose ledger records only principles in $\mathcal{S}$ exactly when $\mathcal{S}$ contains EM or LPO.*

Proof. (1) is Theorems 20.59 and 20.63 and Corollary 20.65. (2) Theorem 8.11 gives a derivation whose ledger records no principle other than EM. Replace each EM leaf by a derivation of the same instance from the base rules and the LPO rule (Theorem 6.10), as in Section 6.7. (3) One LPO rule and universal introduction derive Λ; by Theorem 6.10 the LPO leaf has a derivation from the base rules and the EM rule, and Theorem 8.11 turns that derivation into a Peano derivation. So Λ has a derivation recording only LPO and one recording only EM. If $\mathcal{S}$ contains neither, it is contained in {MP}, and the proof of Theorem 6.11(3) shows that no derivation whose additional rules are MP rules proves Λ: every such derivation is realized, and Λ has no realizer. That argument uses a classical metatheory. □

Table 20.4 collects the rows. Together with Remark 6.12 they say where the step from Heyting to Peano arithmetic is paid for. With neither EM nor LPO, the rules derive exactly the theorems

Theorem of Peano arithmetic	What is derived	Price	Proved in
$\exists y\, A$, $\forall x\, A$	the same formula	none	Theorem 20.59
$\forall x \exists y\, A$	the same formula, and a program for y	none	Theorems 20.59, 20.60
$\forall x \exists y \forall z\, A$, with A in x, y, z only	the same formula and a program for y; the no-counterexample form for every formula F	none	Corollary 20.65, Theorem 20.63
$\exists y \forall z\, A$, with A in y, z only	the same formula	none	Corollary 20.65
$\forall x \exists y \forall \bar{z}\, A$, any block and parameters	the no-counterexample form, read in $\mathbb{N}$, for every function	none	Theorem 20.64
any sentence	the same sentence	LPO alone, or EM alone	Theorems 8.11, 6.10
Λ, the LPO instance at the halting matrix	the same sentence, only with EM or LPO	LPO, attained	Theorem 6.11

Table 20.4. The price of converting a theorem of Peano arithmetic, by shape; A is quantifier-free throughout, with no bounded quantifiers, so that a search over one of its variables is bounded in advance (Lemma 6.8). The rows concern this class, not matrices with bounded quantifiers; the row for $\forall x \exists y\, A$ is the case of one searched variable of the standard Π_2 conservativity. "None" means a posit-free derivation: no EM, LPO or MP rule, with induction on every formula allowed. In the row read in $\mathbb{N}$ it means that the proof uses no principle in the metatheory. The programs for y in the second and third rows are obtained in the classical metatheory of Appendix E.3.

of Heyting arithmetic; with either, exactly those of Peano arithmetic. The difference never shows in the shapes of part (1) of Theorem 20.66. It shows at Λ, whose matrix describes a computation and so needs quantifiers, as the discussion after Theorem 6.9 explains.

20.5 Finite access and a uniform construction differ

Every finite tally has its own finite stage. Compatible readings of these stages assemble into one reading of the whole orbit, uniquely. For a given input, the proof uses only that input's stage. The compatibility hypothesis ensures that increasing the budget cannot change an answer already given.

A program is also a finite object. Chapter 10 stores its expression as a tally and recovers the identical expression. Execution is therefore preserved, including the inputs on which execution is undefined. The recovery law does not require a termination claim.

For an interval protocol, a uniform name has a further obligation: one program must return the correct endpoint pair at every precision. The existence of such a name is equivalent to computability of that endpoint function. A supplied finite table of endpoints always has a lookup program, but the table may change with the requested budget. The quantifiers are different:

$$\forall N \exists c\, \forall n < N \qquad \text{and} \qquad \exists c\, \forall n.$$

Here the suppressed conclusion is the same correct endpoint response in both expressions. Theorem 11.11 proves, in the stated classical real display, that the first requirement does not imply the second. Each finite program determines at most one protocol and hence at most one displayed value. Countably many finite programs cannot name every value in that display.

This leaves the enclosure construction intact. Nested rational intervals with the given width bound determine a unique classical value; overlapping at every precision is exactly equality of that value. The distinction is between this mathematical family and an effective rule producing its members. The completeness and classical comparison used for the display are stated where they enter the construction. Chapter 12 prices that completeness on a line that does not presuppose it.

Exact decisions reveal another boundary. For the regular dyadic presentations of Chapter 12, deciding exact zero is equivalent to WLPO. A witnessed positive-or-zero-or-negative answer is equivalent to LPO. The weaker nonnegative-or-nonpositive answer is equivalent to LLPO. Turning a proof that a quantity is not zero into a sign is equivalent to the binary form $\mathrm{MP}_{\mathrm{bin}}$ of Markov's principle (Definition 7.8), and the comparisons themselves recover LPO as WLPO together with that principle. These principles already priced the recognition catalogue in Chapter 7. The same prices hold on the line built from the ratios of Chapter 4, which is one continuum with the dyadic presentations. Its completeness comes apart in the same way. The limit of a Cauchy sequence costs only a choice of representatives, and nothing at all as a located cut. That choice decides nothing: intuitionistic arithmetic with variables for sequences and countable choice for numbers and for sequences proves none of LPO, WLPO, LLPO and $\mathrm{MP}_{\mathrm{bin}}$ (Theorem 12.22). Excluded middle cannot replace it: with excluded middle and choice only for relations whose witnesses a search can find, and without unique choice, neither the Cauchy limit nor the passage from a located cut to a point is provable (Theorem 12.63). Monotone convergence costs LPO, a least upper bound of a bounded sequence WLPO, and an exact root or an attained maximum of a uniformly continuous function LLPO, while the approximate versions cost nothing. The proofs give explicit encodings in both directions. A sequence of increasingly accurate rational answers does not erase these distinctions.

20.6 The physical bridge has a construction and a constraint

The first physical question is about the information a reader actually receives. If two situations produce the same record, an exact reader must give them the same required answer. This collision argument gives the capacity bound in Chapter 17. A bounded adaptive test tree still produces a bounded transcript, because its branches and stopping rules are included in the construction of that transcript. Allowing an error tolerance changes exact distinction into separation by more than twice that tolerance.

These bounds identify what a record must retain. Chapter 18 constructs an operation that returns a quantum system together with its mark. Positivity and normalization supply branch matrices. Stacking those matrices gives an isometry, and the explicit block dilation extends it to a unitary on the whole stated space. A complete reader retains both the intended rows and their complement. The error estimate includes any finite retained reference, so correlations with a system outside the apparatus are part of the comparison.

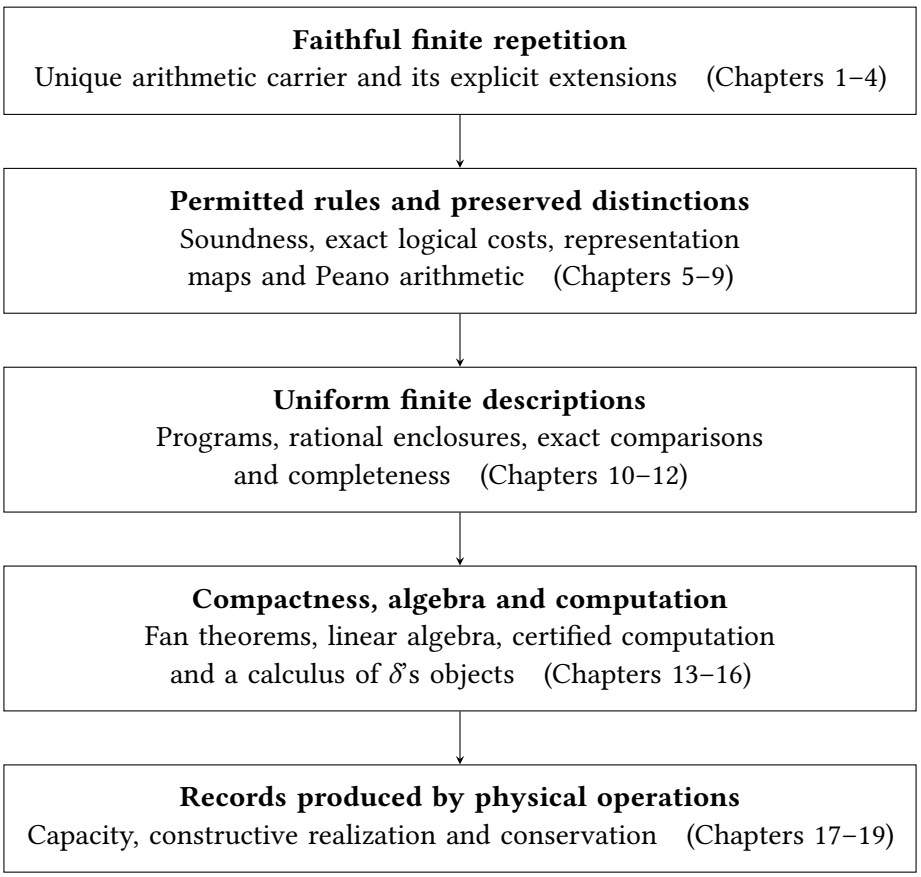


Figure 20.1. A route through the argument. At each stage, the next question requires additional specified objects: a language and its rules, an effective description, or a physical apparatus. The arrows organize the reading; they do not discharge those additional hypotheses.

Two results then connect the descriptions to finite operations. First, ordinary effective matrix names and effective operation-word names can be translated uniformly for fixed dimensions. The search accepts a rational certificate with a strict margin; density ensures that an acceptable word eventually appears. Second, native controls realize the required approximation with positive full energy operators and an explicit scalar action contribution. Input preparation, selectable control pulses, the fixed background holding and the complete reader are all retained in the construction.

The recorder theorem supplies a complementary restriction. For the two-cell event specified in Chapter 19, a whole unitary that conserves both native charges and is followed by a neutral sharp pointer obeys

$$e \geq \frac{1}{2(r^2+3)}.$$

Here r counts the recorder's native charge-carrying cells. Its internal correlations are allowed, while its fixed initial preparation is separate from the unknown source. The proof derives the count conservation from the two native charges, bounds noise by uniform event error, and evaluates an explicit source vector. The size of neutral memory cannot remove the charge bound.

Both results concern actual operations under their stated physical conditions. Supplied

controls and a closed conserving recorder impose different requirements. An apparatus meeting both descriptions would have to include its control supply in the conserving account. The construction does not assert such an autonomous implementation, and the lower bound does not construct an apparatus attaining it. Those are specific tasks left to further construction. Figure 20.1 traces the route of the whole argument.

20.7 What a further claim must show

The book gives several ways to make a necessity claim checkable. Construct the map and prove it unique. Classify every allowed identification. Recover the proposed additional principle from the claim itself. Exhibit a model that satisfies the weaker rules but cannot satisfy the stronger judgment. Compile the proofs of one system into another, then translate their conclusions back. Build the physical operation, then carry its errors and resources through the complete reader.

These methods specify the next work as well. A broader representation claim needs translations preserving the additional structure it names. The comparison with Peano arithmetic in Chapter 8 is one such translation, for a second proof system over the same language. Section 20.3 carries the method to a language of functions, where even the test that runs the compiled programs has its correctness derived by the rules. A uniform construction needs one procedure with the required behavior on every input. A stronger physical realization needs the preparation, controller and reader it assumes to be constructed within the same resource account. Each question can be pursued without replacing the arithmetic, classification and resource results already proved.

The results of this book hold together, about one orbit, with no hypotheses beyond those stated with each theorem. A physical claim that goes further needs an identification of a mathematical object with a physical one, and the book keeps such identifications apart from its theorems. Two concern the results here directly. The identification of *access* says that the physically real collections are exactly those whose members each leave an exact finite record from which a reader recovers them, as in the conditions of Theorem 17.11. That theorem says that such a collection injects into the natural numbers. So access would make every physically real collection inject into the natural numbers, and the classical line of Chapter 11 would not be among them. The converse is not claimed: an injection gives each member the binary numeral of its image, from which the member is read back, but that numeral is a mathematical code, not a record that an apparatus acquires (Section 17.1). The identification of *carrier closure* says that the physically possible finite processes are exactly the valid finite instruments of Chapter 18; with Theorem 18.11 it would give every physically possible finite process an approximation to any tolerance, on every retained reference, by externally switched positive native pulses, with its scalar action recorded. Neither is a theorem, and no theorem of this book assumes either. Further physical consequences are developed in the companion volume under these and other identifications. Parts VI to VIII restate the mathematical content of several of them in this book's own words and prove it, on the line where it holds there and in classical mathematics where they say so, beginning with the cost of a comparison (Chapter 21), and keep each physical identification named beside the theorem it concerns.

The first tally therefore remains visible at the end of the argument. Its start and step determine

a finite construction. Faithfulness determines what must remain distinct. A reader determines what can be identified. A proof exposes the rules needed to pass from those objects to a conclusion. A physical apparatus must enact the required transformation and retain its record. The resulting structure can be followed, checked and used at each of those levels.

Questions to work through

1. Follow the map in Theorem 2.7. Which hypothesis gives injectivity, and which gives surjectivity?

2. Compare the WLPO, LPO and LLPO forms of the recognition catalogue in Theorem 7.9 with Theorems 12.6–12.8, and its Markov form, Theorem 7.7, with Theorem 12.11. For each principle, name the data supplied to the forward construction and the object used in the reverse implication in both settings.

3. Explain how a program can have faithful finite storage while failing to halt on an input. Then explain why a uniform interval name must halt at every precision request.

4. Locate one retained premise in each of the three physical chapters: a condition on the record, a condition on supplied controls, and a condition on conservation. Describe the exact proof step that uses each one.

5. In the small case at the end of Section 20.3, the proof of L inducts on an equation, while its compiled form inducts on the reading of that equation. Identify the quantifiers in that reading, and state which ledger entries the compiled derivation records.

6. Why are the comparison theorems stated for formulas built from $=$, $\bot$, $\to$ and $\forall$? Suppose the finite-type logic had an existential quantifier, read in the same way as the other connectives. Describe what the reading of $\exists y \psi(y)$ would ask a compiled proof to supply, and compare it with what Theorem 20.31 has to recover.

7. Show that the first assertion of Theorem 20.39 fails for $\mathscr{E}^{-}$. Write the type that expresses the translation of $\forall x \neg(S(x) = 0)$, and compute its value under the valuation of Theorem 9.8.

Part VI

The comparison cost and its forcing chain

Part VI turns from what a proof may use to what a physical derivation costs. Recognition Science, as this book restates it, derives the structure of physics from distinction: a cost of comparing two positive quantities, the golden ratio, three spatial dimensions, which it reads off a kept pair of records linked in space, and the constants that follow from them. Each of its steps can be restated on the line $\mathbb{R}_\delta$, where every number is a construction and nothing is decided for free, and then proved or priced.

Chapter 21 proves that the composition law and its calibration leave exactly one cost, $J(x) = \frac{1}{2}(x + x^{-1}) - 1$, with no omniscience principle and no choice, and supplies the functions, the data forms of the omniscience principles and the ledger that the later chapters use. Chapter 22 restates the steps of the forcing chain in the book's own words, in the second column of its price table, and prices each restated step on the line: of the 130 rows of that table (the steps, the second rung's groups of consequences and the constants), 119 hold with no omniscience principle and no choice, ten have an exact price, one is bracketed (below a named principle), and none is open. Chapter 23 proves natively the topological fact on which the dimension step rests: a circle embedded in a sphere can be linked exactly when the sphere has dimension three. Reading a kept pair of records as a linked circle is an identification, which the chain takes as a premise; it is not a theorem of the chapter. Chapter 24 reads the records of a ledger: one account's debits and credits fill a light cone, and one coherently carried binary distinction has the record cone of special relativity; its Section 24.11 climbs from posting toward a clock, the square law, conserved charges and a space of cells, proving what posting forces and naming the premise at each step it does not decide. The last sections of these chapters prove that one bit of calibration with one regularity condition already forces J; that the verdict of the chain holds with no cited premise, and that three of its four brackets are exact prices; that, with the code's hypothesis in its free form, every conclusion of the moral code's main theorem is free on the line; and which values the world's records keep physically real, with the eight-tick comparison at its exact scope.

Parts VI to VIII read maps as rules given with their data, without unique choice (Section 21.1). Each result says whether it is free, what it costs, or whether it is proved in classical mathematics. Physical readings of the objects, such as debit and credit as time and position, are identifications: a theorem states what follows from them and does not supply them.

Chapter 21

The cost of a comparison

Compare a positive quantity with a reference and write $x > 0$ for the ratio. A cost of the comparison is a number $F(x)$ with $F(1) = 0$: comparing a quantity with itself costs nothing. What can such a cost be, if comparing in two steps must cost what the steps cost together? Recognition Science states the requirement as one equation. Composing a comparison x with a comparison y can go forward, to xy, or backward, to x/y, and the two composites together must cost

$$F(xy) + F(x/y) = 2F(x)F(y) + 2F(x) + 2F(y) \qquad \text{for all } x, y > 0. \tag{21.1}$$

In the variables $M = 1 + F$ the requirement reads $M(xy) + M(x/y) = 2M(x)M(y)$: on average over the two directions, composing multiplies the excesses.

Here is a small case. Put $J(x) = \frac{1}{2}(x + x^{-1}) - 1$. Then $J(1) = 0$, $J(2) = \frac{1}{4}$ and $J(4) = \frac{9}{8}$, and at $x = y = 2$ the two sides of (21.1) are $J(4) + J(1) = \frac{9}{8}$ and $2 \cdot \frac{1}{16} + \frac{1}{2} + \frac{1}{2} = \frac{9}{8}$. The quadratic $F(x) = \frac{1}{2}(x-1)^2$ has the same value 0 and the same curvature at $x = 1$, but it fails the law at the same pair: $F(4) + F(1) = \frac{9}{2}$, while $2F(2)^2 + 4F(2) = \frac{5}{2}$. The law at $y = x$ gives the squaring rule $F(x^2) = 2F(x)^2 + 4F(x)$, so the value at 2 fixes the values at 4, 16, 256, ... : from $J(2) = \frac{1}{4}$ it gives $J(16) = 2 \cdot \frac{81}{64} + \frac{9}{2} = \frac{225}{32}$, which is $(16 + \frac{1}{16})/2 - 1$. Squaring moves every ratio toward the reference when it is run backward, and near the reference the cost is fixed by its curvature. This chapter turns that observation into a proof: on the line $\mathbb{R}_\delta$ the composition law and the unit curvature at the reference leave exactly one cost, J, and the proof uses no omniscience principle and no choice principle (Theorem 21.14).

The chapter opens Part VI and provides what Chapters 22 to 30 use. Section 21.1 fixes how maps are read in Parts VI to VIII, without unique choice, and defines the data forms of the omniscience principles that this reading separates from the principles. Section 21.2 builds the logarithm, the sine and the cosine on $\mathbb{R}_\delta$ and proves that the sine's period is 2π, with δ's own π and no choice principle. Section 21.3 proves the forcing theorem, separates the two families that are both called gauges of J, restricts the theorem to the golden orbit and prices the decision of the law for constant costs. Section 21.4 proves, in classical mathematics, that J is the cost of one fair binary distinction. Section 21.5 shows that the recognition ledger is the ring $\mathbb{Z}[\varphi]$, whose prices are compared by an integer sign test, that its history is a sigma-delta encoder of φ^{-3}, and that ledger balances and rates are δ's integers and ratios. Section 21.6 prices the decision whether a given cost obeys the law, and Section 21.7 shows, in classical mathematics, that a comparison which changes a record with certainty costs $J \geq \frac{1}{4}$, less a drift from the holding of the register, when the reference action is one cycle. The last five sections return to the forcing theorem. Section 21.8 shows that one binary distinction, with any one of four mild conditions, fixes the cost, and what the bit alone cannot do. Section 21.9 shows that bounds near the reference give the curvature. Section 21.10 shows that the ratio law is the logarithmic law and prices naming the branch of a cost, deciding its sign and

comparing two costs. Section 21.11 proves the exact values of J that later chapters use, the golden ladders among them, with no principle. Section 21.12 states the cost on ratio records and sorts Recognition Science's five premises for it into the redundant and the independent.

21.1 Reading maps without unique choice

In the first five parts of this book a map is a relation with exactly one output for each input (Appendix C). That reading contains the principle of unique choice: once each input is known to have exactly one output, the outputs form a map (Section 12.9). Section 14.9 and Chapter 16 also worked with a narrower reading, and Parts VI to VIII use that narrower reading throughout.

Here a map is a rule given as data. A map on the points of $\mathbb{R}_\delta$ computes its value from a regular sequence naming the point, and gives the same point for any two sequences that name the same point. A search that is proved to terminate counts as an explicit rule. A proof that every input has exactly one output does not, by itself, produce a map. Every theorem of Parts VI to VIII is proved under this reading. The proofs use a map only by evaluating it, and a rule given as data is also a map under the default reading; so a proof under the narrower reading is also a proof under the default reading, and each theorem stays true there. A statement that quantifies over all maps ranges over more maps under the default reading, and it stays true for the same reason: its proof only evaluates the map it is given, and every map it builds is a rule given as data, hence a map under either reading.

The reading matters for prices. WLPO says of each binary sequence that it is all zero or not. Under the narrower reading this does not give one map that returns the answer for every sequence. We therefore name the data forms.

Definition 21.1 (Data forms of the principles)**.** We use five data forms.

1. *The LPO function* exists when some map d from binary sequences to $\{0,1\}$ satisfies $d(\alpha) = 1$ exactly when $\alpha(n) = 1$ for some n.
2. *The WLPO function* exists when some map d from binary sequences to $\{0,1\}$ satisfies $d(\alpha) = 1$ exactly when $\alpha(n) = 0$ for every n.
3. *The WEM function*, a double-negation decision, exists when some map d from propositions to $\{0,1\}$ satisfies $d(P) = 1$ exactly when $\neg\neg P$.
4. *A zero test on the line* is a map z from $\mathbb{R}_\delta$ to $\{0,1\}$ with $z(x) = 1$ exactly when $x = 0$.
5. *An apartness test on the line* is a map a from $\mathbb{R}_\delta$ to $\{0,1\}$ with $a(x) = 1$ exactly when $x \# 0$, that is, when $x < 0$ or $x > 0$.

Each data form gives its principle with no choice: evaluate the map. The LPO function gives LPO, the WLPO function gives WLPO, the WEM function gives weak excluded middle, WEM, which states $\neg P \vee \neg\neg P$ for every proposition P; a zero test decides $x = 0 \vee x \neq 0$ for every point, which is WLPO by Corollary 12.18; and an apartness test decides $x \# 0 \vee \neg(x \# 0)$, that is, $x > 0 \vee x = 0 \vee x < 0$, for every point, which is LPO by Corollary 12.18. Chapter 28 proves that a zero test exists exactly when the WLPO function does and an apartness test exactly when the LPO function does.

Remark 21.2 (Under unique choice the data forms are the principles)**.** Assume unique choice. Each principle then gives its data form by the same argument: the principle gives every input exactly one bit, since two bits with the same defining property are equal, and unique choice turns that relation between inputs and bits into a map.

1. *WLPO gives the WLPO function.* WLPO says that each binary sequence α is all zero or not, so exactly one bit b has $b = 1$ exactly when $\alpha(n) = 0$ for every n. Unique choice turns this relation into a map d with $d(\alpha) = 1$ exactly when α is all zero.

2. *WLPO gives a zero test.* By Corollary 12.18, WLPO decides $x = 0 \vee x \neq 0$ for each point x, so exactly one bit b has $b = 1$ exactly when $x = 0$. Unique choice turns this relation into a map z on $\mathbb{R}_\delta$, a zero test.

3. *LPO gives an apartness test.* By Corollary 12.18, LPO gives $x > 0 \vee x = 0 \vee x < 0$ for each point x; the outer cases give $x \# 0$ and the middle case excludes it, so exactly one bit b has $b = 1$ exactly when $x \# 0$. Unique choice turns this relation into a map a on $\mathbb{R}_\delta$, an apartness test.

In the same way LPO gives the LPO function and WEM gives the WEM function. Under the default reading of this book each data form is therefore equivalent to its principle, and every theorem of Parts VI to VIII that names a data form is, under that reading, a theorem about the principle. The separations between a principle and its data form, where we state them, are statements about the narrower reading.

Three words recur. A result is *free* when its proof uses no omniscience principle and no choice principle, neither countable nor unique choice. A proof *with no choice principle* may use an omniscience principle that is named as a hypothesis; "T costs exactly P" means that T and P imply each other with no choice principle, and "at most" and "at least" name one of the two directions. A result *proved in classical mathematics* is a theorem about the classical real numbers, or about complex matrices or measures, with excluded middle and the axiom of choice available. Such a result carries no price on $\mathbb{R}_\delta$, and we say so before its statement. A *reading* is a map ρ from $\mathbb{R}_\delta$ to the classical real numbers that preserves sums, products and 1 and satisfies $x < y$ exactly when $\rho(x) < \rho(y)$; statements about readings are classical. Physical readings of the mathematics are named as identifications and kept apart from what is proved.

21.2 Logarithm, sine, cosine and π on the line

Proposition 12.88 constructs the exponential function on $\mathbb{R}_\delta$ at no price. The cost J needs its inverse, and the physics of the later chapters needs the circle functions and π. We construct them as maps in the sense of Section 21.1. We write e^x for $\exp(x)$ and use the facts of Proposition 12.19 without comment.

Lemma 21.3 (Order facts of the exponential)**.** *For all points a, b, c, u, x of $\mathbb{R}_\delta$:*

1. *$e^x > 0$ and $e^x e^{-x} = 1$;*
2. *$1 + u \le e^u$;*
3. *$e^b(a - b) \le e^a - e^b$;*
4. *if $a \le c$ then $e^a \le e^c$, and if $a \le c$ and $b \le c$ then $|e^a - e^b| \le e^c|a - b|$;*
5. *$a < b$ exactly when $e^a < e^b$; in particular $e^a = e^b$ implies $a = b$.*

The proofs are free.

Proof. 1. $e^x e^{-x} = e^0 = 1$, so e^x has an inverse and is positive or negative. It is not negative, since $e^x = (e^{x/2})^2 \ge 0$.

2. For $|h| \le 1$, Proposition 12.88 gives $e^h \ge 1+h-h^2$. If $e^v \ge 1+v-\varepsilon$, then $e^{2v} = (e^v)^2 \ge 2e^v - 1 \ge 1 + 2v - 2\varepsilon$, because $(e^v - 1)^2 \ge 0$. Let $V \ge |u|$ be a natural number and $2^K \ge V$. Applying the doubling step K times to $h = u/2^K$ gives $e^u \ge 1 + u - 2^K h^2 = 1 + u - u^2/2^K \ge 1 + u - V^2/2^K$. This holds for every such K, so $e^u \ge 1 + u$ by Lemma 12.52, item 2.

3. $e^a - e^b = e^b(e^{a-b} - 1) \ge e^b(a - b)$, by item 2 and $e^b > 0$.

4. Suppose $a \le c$ and $e^c < e^a$. Item 3 at the pair (c, a) gives $e^a(c - a) \le e^c - e^a < 0$, so $c - a < 0$ because $e^a > 0$; this denies $a \le c$. Now let $a, b \le c$. Item 3 at (b, a) gives $e^a - e^b \le e^a(a - b) \le e^c|a - b|$, and exchanging a and b gives the other half.

5. If $a < b$, item 3 gives $e^b - e^a \ge e^a(b - a) > 0$. Conversely let $e^a < e^b$, and take a natural number c above a and b. By item 4, $0 < e^b - e^a \le e^c|a - b|$, so $|a - b| > 0$, and Lemma 12.52, item 6, gives $a < b$ or $b < a$; the second would give $e^b < e^a$. If $e^a = e^b$, neither $a < b$ nor $b < a$, so $a = b$. □

A positive point has exactly one logarithm, by item 5. Turning that fact into a map would be unique choice, so we compute the logarithm.

Definition 21.4 (The logarithm)**.** Let $y > 0$, and let s be a regular sequence naming y. Search for the first stage N with $s_N > 2/(N + 1)$; it exists because $y > 0$. By Lemma 12.16, $y \ge 1/(2N + 2)$, and Proposition 12.19, item 4, gives a natural number K_s with $y \le K_s$. Put $c = 2N + K_s + 2$, so that $e^{-c} \le 1/(1 + c) \le y \le 1 + c \le e^c$ by Lemma 21.3, and let

$$f(t) = e^{c(2t-1)} - y \qquad (0 \le t \le 1).$$

Theorem 12.92 computes a root t^* of f in $[0, 1]$ from the data listed below, and we put $\log y = c(2t^* - 1)$.

The data are these. $f(0) \le 0 \le f(1)$. Let A be a natural number with $e^c \le A$. By Lemma 21.3, item 4, $|f(t) - f(t')| \le 2cA|t - t'|$, a modulus of uniform continuity given as a function. If $|f(t)|$ and $|f(t')|$ are at most $1/(\mu + 1)$, then $|e^a - e^{a'}| \le 2/(\mu + 1)$ for $a = c(2t - 1)$ and $a' = c(2t' - 1)$, both at least $-c$. Item 3 of the lemma gives $|e^a - e^{a'}| \ge e^{-c}|a - a'|$: if the right side were larger,

$|a - a'| > 0$, and in the case $a < a'$ we would have $e^{a'} - e^{a} \geq e^{a}(a' - a) \geq e^{-c}(a' - a)$, and symmetrically in the other case. So $|t - t'| \leq A/(c(\mu + 1))$, and $\mu(k) = A(k + 1)$ is a modulus of uniqueness. Finally, the value of f at each point $j/3^n$ is named by a regular sequence computed from n, j and s: the exponential of a ratio is computed from its constant sequence. Every step is a computation from s or a terminating search.

Proposition 21.5 (The logarithm is free)**.** *The logarithm of Definition 21.4 is a map on the positive points of $\mathbb{R}_\delta$: the point* $\log y$ *does not depend on the regular sequence used. For all positive x, y and every point t:*

1. $e^{\log y} = y$*, so every positive point is an exponential;*
2. $\log e^{t} = t$*;*
3. $\log(xy) = \log x + \log y$*;*
4. $x < y$ *exactly when* $\log x < \log y$*.*

No omniscience principle and no choice principle is used.

Proof. Item 1 holds by construction: $f(t^*) = 0$. Two sequences naming y give points with the same exponential, and these are equal by Lemma 21.3, item 5; so log is a map on the points. For item 2, $\log e^{t}$ and t have the same exponential. For item 3, $e^{\log x + \log y} = e^{\log x} e^{\log y} = xy = e^{\log(xy)}$. Item 4 is item 5 of the lemma applied to $\log x$ and $\log y$. □

For $x > 0$ and a point c we write $x^{c} = e^{c \log x}$; for an integer c this is the ordinary power. The circle functions are built from two power series in $s = \theta^2$, exactly as the exponential is built in Proposition 12.88.

Definition 21.6 (Cosine, sine and π_δ)**.** For a point s let

$$\mathrm{C}(s) = \sum_{j \geq 0} \frac{(-s)^{j}}{(2j)!}, \qquad \mathrm{S}(s) = \sum_{j \geq 0} \frac{(-s)^{j}}{(2j+1)!},$$

each the limit of its partial sums, computed on a regular sequence naming s with the modulus given by a natural bound $V \geq |s|$ and the diagonal limit of Theorem 12.25. Put $\cos\theta = \mathrm{C}(\theta^2)$ and $\sin\theta = \theta\,\mathrm{S}(\theta^2)$. For a natural number $m \geq 2$ put

$$\arctan\frac{1}{m} = \sum_{j \geq 0} \frac{(-1)^{j}}{(2j+1)\,m^{2j+1}}, \qquad \pi_\delta = 4\Big(\arctan\frac{1}{2} + \arctan\frac{1}{3}\Big).$$

A *quarter period* is a point q with $0 \leq q \leq 2$ and $\cos q = 0$.

The jth term of each series in s is at most $V^{j}/(2j)!$, and the terms of the arctangent series are at most $2^{-(2j+1)}$, so all three limits come with explicit moduli, and limits are unique, so the values depend only on the point. Theorem 21.8 below shows that there is exactly one quarter period and that it is $\pi_\delta/2$.

Lemma 21.7 (Identities and small-angle bounds)**.** *For all points α, β, θ:*

1. $\cos(-\theta) = \cos\theta$, $\sin(-\theta) = -\sin\theta$, $\cos 0 = 1$ *and* $\sin 0 = 0$;
2. $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ *and* $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$;
3. $\cos^2\theta + \sin^2\theta = 1$, *so* $|\cos\theta| \le 1$ *and* $|\sin\theta| \le 1$;
4. *if* $|\theta| \le 1$, *then* $|\cos\theta - 1| \le \theta^2$, $|\sin\theta - \theta| \le |\theta|^3$ *and* $|\cos\theta - 1 + \theta^2/2| \le \theta^4$;
5. *if* $0 \le \theta \le 2$, *then* $\sin\theta \ge \theta/3$.

The proofs are free.

Proof. 1. The series depend on θ^2 only, and $\theta\,\mathrm{S}(\theta^2)$ is odd.

2. The partial sums $c_N(\theta)$ and $s_N(\theta)$ of the cosine and sine series are the two components of $\sum_{k<2N}(\mathrm{i}\theta)^k/k!$ in pairs of points multiplied by $(a,b)(a',b') = (aa'-bb', ab'+ba')$. As in the proof of Proposition 12.88, part 1, the binomial theorem shows that the product of the sums at α and at β differs from the sum at $\alpha+\beta$ by terms $\alpha^i\beta^j/(i!\,j!)$ with $i,j < 2N \le i+j$, at most a tail of the exponential series at $|\alpha|+|\beta|$ in absolute value. The defect tends to 0, and the two components of the limit are the two formulas.

3. Item 2 at $\alpha = \theta$, $\beta = -\theta$, with item 1.

4. $\cos\theta - 1 = \sum_{j\ge1}(-\theta^2)^j/(2j)!$ has absolute value at most $\theta^2\sum_{j\ge1}1/(2j)! \le \theta^2$, and the same bound for the partial sums passes to the limit (Lemma 12.52, item 2). Likewise $|\sin\theta - \theta| \le |\theta|^3\sum_{j\ge1}1/(2j+1)! \le |\theta|^3$ and $|\cos\theta - 1 + \theta^2/2| \le \theta^4\sum_{j\ge2}1/(2j)! \le \theta^4$.

5. For $0 \le s \le 6$ we have $\mathrm{S}(s) - 1 + s/6 = \sum_{i\ge1}\big(s^{2i}/(4i+1)! - s^{2i+1}/(4i+3)!\big)$, and each bracket is nonnegative because $s \le (4i+2)(4i+3)$. So $\mathrm{S}(s) \ge 1 - s/6$, and for $s = \theta^2 \le 4$ this gives $\mathrm{S}(\theta^2) \ge \frac{1}{3}$ and $\sin\theta = \theta\,\mathrm{S}(\theta^2) \ge \theta/3$. □

Theorem 21.8 (The sine's period is $2\pi_\delta$, with no choice)**.**

1. *For every natural number $m \ge 2$, the point $A = \arctan(1/m)$ satisfies $m\sin A = \cos A$, and $4/(5m) \le A \le 1/m$.*
2. *The point $\pi_\delta/2$ lies in $[\frac{4}{3}, \frac{5}{3}]$ and satisfies $\cos(\pi_\delta/2) = 0$ and $\sin(\pi_\delta/2) = 1$.*
3. *If $0 \le q \le 2$ and $\cos q = 0$, then $4q = 2\pi_\delta$.*

No omniscience principle and no choice principle is used.

Proof. 1, a grid of angles. Fix $m \ge 2$ and $N \ge 1$, and put $h = 1/(mN)$, $t_k = kh$ for $0 \le k \le N$, $\phi_k = h/(1+t_k^2)$ and $\Phi_n = \sum_{k<n}\phi_k$. These are ratios, $0 \le t_k \le 1/m \le \frac{1}{2}$ and $\frac{4}{5}h \le \phi_k \le h$. For $M \ge 1$ let $P_M(t) = \sum_{j<M}(-1)^j t^{2j+1}/(2j+1)$ and $Q_M(t) = \sum_{j<M}(-t^2)^j$, so that $(1+t^2)Q_M(t) =$

$1-(-t^2)^M$. For ratios $t,h\ge 0$ and $n\ge 2$ the binomial expansion and $\binom{n}{i}\le\binom{n}{2}\binom{n-2}{i-2}$ give $|(t+h)^n-t^n-nt^{n-1}h|\le\binom{n}{2}h^2(t+h)^{n-2}$. Hence, for $0\le t\le t+h\le\frac12$,

$$|P_M(t+h)-P_M(t)-hQ_M(t)|\le h^2\sum_{j\ge1} j\,2^{-(2j-1)}=\tfrac89 h^2,$$

since $\binom{2j+1}{2}/(2j+1)=j$. Also $hQ_M(t_k)=\phi_k\bigl(1-(-t_k^2)^M\bigr)$ differs from ϕ_k by at most $h4^{-M}$. Summing over $k<N$, the telescoping sum $P_M(1/m)=\sum_{k<N}\bigl(P_M(t_{k+1})-P_M(t_k)\bigr)$ satisfies $|P_M(1/m)-\Phi_N|\le\frac89 Nh^2+Nh4^{-M}\le 1/N+4^{-M}$. The tail of the arctangent series beyond M terms is at most $\sum_{j\ge M}2^{-(2j+1)}\le 4^{-M}$, so $|A-\Phi_N|\le 1/N+2\cdot4^{-M}$ for every M, and therefore $|A-\Phi_N|\le 1/N$. Since $\frac45 Nh\le\Phi_N\le Nh$, this gives $4/(5m)-1/N\le A\le 1/m+1/N$ for every N, hence the bounds on A.

1, the ray defect. Put $D_k=\sin\Phi_k-t_k\cos\Phi_k$, so $D_0=0$. Fix $k<N$ and write $S=\sin\Phi_k$, $C=\cos\Phi_k$, $t=t_k$, $\phi=\phi_k$, $\cos\phi=1+\gamma$ and $\sin\phi=\phi+\sigma$, with $|\gamma|,|\sigma|\le\phi^2$ by Lemma 21.7, item 4. By the addition formulas,

$$D_{k+1}=Sc'+Cs'-(t+h)(Cc'-Ss'),\qquad c'=\cos\phi,\ s'=\sin\phi.$$

Replacing c' by 1 and s' by ϕ leaves $S-tC+(\phi C+t\phi S-hC)+h\phi S$, and since $h=(1+t^2)\phi$,

$$\phi C+t\phi S-hC=t\phi(S-tC)=t\phi D_k.$$

The replaced terms contribute at most $(1+t+h)(|\gamma|+|\sigma|)\le 4h^2$, and $|h\phi S|\le h^2$. So $|D_{k+1}|\le(1+h)|D_k|+5h^2$. By induction, $|D_k|\le 5kh^2(1+h)^k$: the step is $(1+h)\,5kh^2(1+h)^k+5h^2\le 5(k+1)h^2(1+h)^{k+1}$. With $(1+h)^N\le e^{Nh}=e^{1/m}\le e^{1/2}\le 2$ (because $e\le 3$), $|D_N|\le 10Nh^2=10/(m^2N)\le 3/N$.

1, conclusion. For $|d|\le 1$ the addition formula gives $\sin(b+d)-\sin b=\sin b(\cos d-1)+\cos b\sin d$, so $|\sin(b+d)-\sin b|\le d^2+|d|+|d|^3\le 3|d|$, and likewise for the cosine. With $b=\Phi_N$, $b+d=A$ and $t_N=1/m$,

$$\Bigl|\sin A-\frac1m\cos A\Bigr|\le\frac3N+\frac{3}{2N}+|D_N|<\frac8N.$$

This holds for every N, so $m\sin A=\cos A$.

2. Let $\alpha=\arctan\frac12$ and $\beta=\arctan\frac13$, so $\cos\alpha=2\sin\alpha$ and $\cos\beta=3\sin\beta$. The addition formulas give

$$\sin(\alpha+\beta)=3\sin\alpha\sin\beta+2\sin\alpha\sin\beta=5\sin\alpha\sin\beta,$$
$$\cos(\alpha+\beta)=6\sin\alpha\sin\beta-\sin\alpha\sin\beta=5\sin\alpha\sin\beta,$$

the identity $(2+\mathrm{i})(3+\mathrm{i})=5(1+\mathrm{i})$ in disguise. With $p=\alpha+\beta$, $\pi_\delta/2=2p$ and $\cos 2p=\cos^2p-\sin^2p=0$. By part 1, $\frac25\le\alpha\le\frac12$ and $\frac{4}{15}\le\beta\le\frac13$, so $\frac43\le 2p\le\frac53$. Finally $\sin 2p=2\sin p\cos p=50\sin^2\alpha\sin^2\beta\ge 0$ and $\sin^2 2p=1$, so $(\sin 2p-1)(\sin 2p+1)=0$ with $\sin 2p+1\ge 1$ invertible, and $\sin 2p=1$.

3. Let $0\le a,b\le 2$ with $\cos a=\cos b=0$, and suppose $a<b$. With $u=(a+b)/2$ and $v=(b-a)/2$, the addition formulas give $\cos a-\cos b=\cos(u-v)-\cos(u+v)=2\sin u\sin v$. Here $0<v\le 1$ and $0<u\le 2$, so $\sin u$ and $\sin v$ are positive by Lemma 21.7, item 5, and $\cos a-\cos b>0$, a contradiction. So $\neg(a<b)$, likewise $\neg(b<a)$, and $a=b$. Taking $b=\pi_\delta/2$ from part 2 gives $q=\pi_\delta/2$. Every step is a ratio computation, a stage estimate or an inequality proved by refuting its negation; nothing is chosen. $\square$

Conditions and scope.

Base and premises
The line $\mathbb{R}_\delta$ with its free field operations, and the series of Definition 21.6; nothing else.

Conclusion
Exactly one zero of the cosine lies in $[0, 2]$, it is $\pi_\delta/2$, and four quarter periods make $2\pi_\delta$. Since $\cos q = 0$ and $\sin q = 1$, the addition formulas give $\sin(\theta + q) = \cos\theta$ and $\cos(\theta + q) = -\sin\theta$, so $\sin(\theta + 2\pi_\delta) = \sin\theta$ and $\cos(\theta + 2\pi_\delta) = \cos\theta$.

Not claimed
That π_δ is the least period; no property of the classical number π is used.

Gregory's series for the arctangent and Euler's identity $\pi/4 = \arctan\frac{1}{2} + \arctan\frac{1}{3}$ are classical. What the theorem adds is that both the identity and the quarter period are reached on $\mathbb{R}_\delta$ by explicit estimates, with no limit taken by choice and no decision about a point. The same tool proves Machin's identity (Question 3).

Proposition 21.9 (The classical reading)**.** *In classical mathematics there is exactly one reading* $\rho : \mathbb{R}_\delta \to \mathbb{R}$. *It sends* π_δ *to* π, *the positive root of* $x^2 = x + 1$ *to the golden ratio, and* $\log y$ *to* $\log \rho(y)$ *for every* $y > 0$.

Proof. For a regular sequence s, its stages converge classically; let ρ send the point of s to the limit. Sequences naming the same point have the same limit, sums and products of sequences converge to sums and products, and a positive stage gives a positive limit by Lemma 12.16, while a positive limit gives a positive stage. So ρ is a reading. Any reading fixes the integers, hence the ratios, and for integers a and natural numbers b it satisfies $a < (b+1)\rho(x)$ exactly when $a < (b+1)x$; so $\rho(x)$ is determined by the ratios below x, and the reading is unique. A reading preserves limits with a modulus, because it preserves the bounds $|x - y| \le 1/(k+1)$. So it sends each series of this section to the classical series, π_δ to $4(\arctan\frac{1}{2} + \arctan\frac{1}{3}) = \pi$, and $\log y$ to the classical number whose exponential is $\rho(y)$. The positive root of $x^2 = x + 1$ goes to a positive root of the same equation. □

21.3 The composition law forces J

We now state the law of the opening on $\mathbb{R}_\delta$ and prove that it leaves one cost. It is convenient to work in logarithmic coordinates: a ratio $x > 0$ is e^t with $t = \log x$, the reference $x = 1$ is $t = 0$, and the two composites xy and x/y are $s + t$ and $s - t$.

Definition 21.10 (Comparison costs, the composition law and calibration)**.**

1. A *comparison cost* is any function C from $\mathbb{R}_\delta$ to $\mathbb{R}_\delta$, the cost of the comparison at $t = \log x$. No continuity, monotonicity or other property is assumed.

2. C obeys the *composition law* when

$$C(s + t) + C(s - t) = 2C(s)C(t) + 2C(s) + 2C(t) \qquad \text{for all points } s, t. \tag{21.2}$$

3. C is *calibrated* when $C(0) = 0$ and for every k there is a j such that

$$\left|C(t) - \frac{t^2}{2}\right| \le \frac{t^2}{k+1} \qquad \text{whenever } |t| \le \frac{1}{j+1}. \tag{21.3}$$

   For a point c, C has *curvature* c^2 *at* 0 when $C(0) = 0$ and the same holds with $c^2t^2/2$ in place of $t^2/2$, and *curvature* $-c^2$ with $-c^2t^2/2$.

4. In multiplicative form, a function F on $\mathbb{R}_\delta$ obeys the *multiplicative law* when

$$F(xy) + F(xz) = 2F(x)F(y) + 2F(x) + 2F(y) \quad \text{whenever } x > 0,\ y > 0,\ yz = 1,$$

   and is *calibrated at* 1 when $F(1) = 0$ and for every k there is a j with $|F(x) - (x-1)^2/2| \le (x-1)^2/(k+1)$ whenever $|x-1| \le 1/(j+1)$.

The quotient x/y is written xz with $yz = 1$, so no division map is needed. With $H = 1 + C$ the law (21.2) is d'Alembert's equation $H(s+t) + H(s-t) = 2H(s)H(t)$. On the classical real line its continuous solutions are 0, $\cosh(ct)$ and $\cos(ct)$, and without a regularity condition there are wild solutions [97, 98]. On $\mathbb{R}_\delta$ a function carries no continuity unless it is stated, and the calibration (21.3) is the only regularity we assume: a uniform second-order expansion at the single point 0, with a modulus.

Definition 21.11 (The cost J)**.** For a positive point x put $J(x) = \frac{1}{2}(x + x^{-1}) - 1$, and for every point t put

$$G(t) = \cosh t - 1 = \frac{e^t + e^{-t}}{2} - 1, \qquad \sinh t = \frac{e^t - e^{-t}}{2}.$$

Then $J(e^t) = G(t)$, $J(x) = (x-1)^2/(2x) = J(x^{-1}) \ge 0$, and $x^c = e^{c \log x}$ gives $J(x^c) = G(c \log x)$.

Proposition 21.12 (Facts about $\cosh - 1$)**.** *For every point c:*

1. *$t \mapsto G(ct)$ obeys the composition law;*
2. *$|G(u) - u^2/2| \le 6|u|^3$ and $|G(u)| \le 7u^2$ whenever $|u| \le 1$;*
3. *$G(2w) = 2\sinh^2 w$, so $G \ge 0$;*
4. *G is calibrated, with $j = 6(k+1)$ in* (21.3)*;*
5. *$t \mapsto G(ct)$ has curvature c^2 at 0, and if $M \ge |c|$ is a natural number, $|G(ct)| \le 7M^2t^2$ whenever $|t| \le 1/(M+1)$.*

The proofs are free.

Proof. 1. Put $a = cs$ and $b = ct$, and $A = \cosh a$, $B = \cosh b$. Since $e^{a \pm b} = e^a e^{\pm b}$, $\cosh(a+b) + \cosh(a-b) = 2AB$. The left side of the law is therefore $2AB - 2$, and so is the right side, $2(A-1)(B-1) + 2(A-1) + 2(B-1)$. Only $2 \cdot \frac{1}{2} = 1$ is used.

2. Put $w = u/2$ and $r = \sinh w - w$. Expanding the squares, $\cosh 2w - 1 = 2\sinh^2 w$, so $G(u) - u^2/2 = 2(w+r)^2 - 2w^2 = 4wr + 2r^2$. Next $2r = (e^w - 1 - w) - (e^{-w} - 1 + w)$, and

Proposition 12.88, item 2, bounds each bracket by w^2, so $|r| \le w^2 \le u^2$. Hence $|G(u) - u^2/2| \le 4|u|^3 + 2u^4 \le 6|u|^3$, since $|u| \le 1$, and $|G(u)| \le 6|u|^3 + u^2/2 \le 7u^2$.

3. Shown in item 2.

4. If $|t| \le 1/(6k+7)$, then $6|t|^3 \le 6t^2/(6k+7) \le t^2/(k+1)$.

5. For $|t| \le 1/(M+1)$ we have $|ct| \le 1$, so item 2 gives $|G(ct) - c^2t^2/2| \le 6M^3|t|\,t^2$ and $|G(ct)| \le 7M^2t^2$. Given k, put $j = M + 6M^3(k+1)$; then $|t| \le 1/(j+1)$ gives $6M^3|t| \le 1/(k+1)$. No case distinction on $c = 0$ is made. □

Setting $s = t$ in the law, and using $C(0) = 0$, gives the *duplication law*

$$C(2s) = 2C(s)^2 + 4C(s) \qquad \text{for every point } s. \tag{21.4}$$

The opening used it to pass from $J(2)$ to $J(4)$ and $J(16)$. Run backward, it lets the values near 0 decide everything, and the next theorem makes this exact. Its proof contains the one estimate of the chapter that is not routine: errors that double at every halving step must not compound.

Theorem 21.13 (Duplication determines a cost)**.** *Let f and g be functions on $\mathbb{R}_\delta$ that both obey the duplication law* (21.4)*. Suppose*

(i) for every k there is a j with $|f(t) - g(t)| \le t^2/(k+1)$ whenever $|t| \le 1/(j+1)$, and

(ii) there are natural numbers j_0 and B with $|g(t)| \le Bt^2$ whenever $|t| \le 1/(j_0+1)$.

Then $f(t) = g(t)$ for every point t. The proof is free.

Proof. Put $D = f - g$. Subtracting the two duplication laws,

$$D(2v) = 2D(v)\big(2 + f(v) + g(v)\big). \tag{21.5}$$

A small interval. Take j_1 from (i) at $k = 0$, so that $|D(v)| \le v^2$ for $|v| \le 1/(j_1+1)$. Put $\gamma = 2B+1$, $M = j_0 + j_1 + \gamma$ and $\varepsilon = 1/(M+1)$. For $|v| \le \varepsilon$ we have $v^2 \le \varepsilon$, so $\gamma v^2 \le 1$, and

$$|2 + f(v) + g(v)| = |2 + D(v) + 2g(v)| \le 2 + v^2 + 2Bv^2 = 2 + \gamma v^2 \le 2(1 + \gamma v^2).$$

The bounded product. We claim that for every n and every u with $|2^n u| \le \varepsilon$,

$$|D(2^n u)| \le \big(1 + \gamma(2^n u)^2\big)\, 4^n\, |D(u)|.$$

For $n = 0$ this is clear. For the step put $w = 2^n u$ with $|2w| \le \varepsilon$. By (21.5), the bound on $|2 + f + g|$ and the claim at n, $|D(2w)| \le 4^{n+1}(1 + \gamma w^2)^2 |D(u)|$, and $(1+a)^2 \le 1 + 4a$ for $0 \le a \le 2$ gives $(1 + \gamma w^2)^2 \le 1 + \gamma(2w)^2$. The factors $1 + \gamma(2^i u)^2$ do not multiply up along the doublings: a single factor $1 + \gamma t^2$ covers them all.

Equality near 0. Let $|t| \le \varepsilon$ and fix k. Take j from (i) at $2k+1$, and put $u = t/2^j$. Since $j + 1 \le 2^j$ and $|t| \le 1$, $|u| \le 1/(j+1)$, so $|D(u)| \le u^2/(2k+2)$. The claim with $n = j$ gives

$$|D(t)| \le (1 + \gamma t^2)\, 4^j u^2\, \frac{1}{2k+2} = (1 + \gamma t^2)\frac{t^2}{2k+2} \le \frac{1}{k+1},$$

because $4^j u^2 = t^2$ exactly, $\gamma t^2 \le 1$ and $t^2 \le 1$. As k is arbitrary, $D(t) = 0$ by Corollary 12.24.

The whole line. If $f(u) = g(u)$, the duplication laws give $f(2u) = g(2u)$. For any t take a natural number $A \ge 1$ with $|t| \le A$ and $n = A(M+1)$; then $|t/2^n| \le A/(n+1) < \varepsilon$, so f and g agree at $t/2^n$, and n doublings give $f(t) = g(t)$.

Every inequality above is a closed bound proved from closed bounds, and the only numbers produced are computed from j_0, j_1, B and the given moduli. No comparison of points is decided. □

Theorem 21.14 (The composition law and the calibration force J)**.**

1. *A comparison cost C obeys the composition law and is calibrated if and only if $C(t) = \cosh t - 1$ for every point t.*
2. *Let c be a point. A comparison cost C obeys the composition law and has curvature c^2 at 0 if and only if $C(t) = \cosh(ct) - 1$ for every t.*
3. *A function F obeys the multiplicative law and is calibrated at 1 if and only if $F(x) = J(x)$ for every positive point x.*

No omniscience principle and no choice principle is used. In classical mathematics, moreover: for every reading ρ, every C as in part 1 and all points $x > 0$ and y with $e^y = x$, $\rho(C(y)) = J(\rho(x))$; and for every F as in part 3, $\rho(F(x)) = J(\rho(x))$.

Proof. Part 1 is part 2 at $c = 1$. In part 2, the cost $t \mapsto \cosh(ct) - 1$ obeys the law and has curvature c^2 by Proposition 21.12. Conversely let C obey the law with curvature c^2, and put $g(t) = \cosh(ct) - 1$. Both C and g vanish at 0 and obey the law, so both obey the duplication law. For hypothesis (i) of Theorem 21.13, take the moduli of C and of g at $2k+1$; below both, $|C(t) - g(t)| \le |C(t) - c^2t^2/2| + |c^2t^2/2 - g(t)| \le t^2/(k+1)$. Hypothesis (ii) is Proposition 21.12, item 5, with $B = 7M^2$. So $C = g$.

For part 3 let F obey the multiplicative law and be calibrated at 1, and put $C(t) = F(e^t)$. At $x = e^s$, $y = e^t$ and $z = e^{-t}$ the multiplicative law is the composition law for C, since $e^s e^{\pm t} = e^{s \pm t}$. For the calibration, $C(0) = F(1) = 0$, and for $|t| \le 1$, $|e^t - 1| \le |e^t - 1 - t| + |t| \le 2|t|$. Given k, let j_2 be the modulus of F at $8k+8$, and let $|t| \le 1/(j+1)$ with $j = 2j_2 + 6k + 7$. Then $|e^t - 1| \le 1/(j_2+1)$, so

$$\left| F(e^t) - \frac{(e^t-1)^2}{2} \right| \le \frac{(e^t-1)^2}{8k+9} \le \frac{4t^2}{8k+9} \le \frac{t^2}{2k+2},$$

and $|(e^t-1)^2 - t^2| = |e^t - 1 - t|\,|e^t - 1 + t| \le 3|t|^3$, so $|(e^t-1)^2/2 - t^2/2| \le \frac{3}{2}|t|\,t^2 \le t^2/(2k+2)$. Adding, $|C(t) - t^2/2| \le t^2/(k+1)$. By part 1, $C = \cosh - 1$. For $x > 0$ put $\ell = \log x$ (Proposition 21.5); then $F(x) = C(\ell) = (x + e^{-\ell})/2 - 1$, and $e^{-\ell} = x^{-1}$. So $F(x) = J(x)$.

Conversely, J is calibrated at 1: $J(x) - (x-1)^2/2 = (x-1)^2(1-x)/(2x)$, and for $|x-1| \le \frac{1}{2}$ we have $x \ge \frac{1}{2}$, so $|J(x) - (x-1)^2/2| \le (x-1)^2|x-1| \le (x-1)^2/(k+1)$ when $|x-1| \le 1/(k+2)$. J obeys the multiplicative law: with $yz = 1$, both sides plus 2 equal $\frac{1}{2}(x + x^{-1})(y + z)$, because $(xy)^{-1} = x^{-1}z$ and $(xz)^{-1} = x^{-1}y$.

For the classical statements, a reading ρ preserves products, so $\rho(e^y)\rho(e^{-y}) = 1$ and $\rho(e^{-y}) = \rho(x)^{-1}$; with $\rho(\frac{1}{2}) = \frac{1}{2}$ this gives $\rho(\cosh y - 1) = J(\rho(x))$, and part 1 or part 3 identifies $C(y)$ or $F(x)$. □

Conditions and scope.

Base and premises
The line $\mathbb{R}_\delta$; a cost C that is an arbitrary function, with the composition law (21.2) at all pairs of points and the calibration (21.3) with a modulus.

Conclusion
$C = \cosh - 1$, equivalently $F = J$ on the positive points, with no omniscience principle and no choice principle. The two premises hold for exactly one function.

Identification
That the cost of a physical comparison obeys the law and the calibration is the premise of Recognition Science's derivation of J (Chapter 22). The theorem derives the cost from that premise; it does not supply the premise, and no observation is used.

Not claimed
Any price for deciding whether a given cost obeys the premises; Section 21.6 computes it.

On the classical real line, Recognition Science reaches the same conclusion from the law and the curvature limit $2C(t)/t^2 \to 1$ through the regularity theory of d'Alembert's equation. Theorem 21.14 replaces that theory by the duplication estimate, and the uniform form of the calibration is what the estimate needs. Two remarks fix its scope.

Remark 21.15 (The normalisation is needed and is implied)**.** The constant cost $C \equiv -1$ obeys the law, since $-2 = 2 - 2 - 2$, so some normalisation must exclude it. The second-order clause already implies $C(0) = 0$: at $t = 0$ it reads $|C(0)| \le 0$. Section 21.6 shows that a cost obeying the law has $C(0) = 0$ or $C(0) = -1$, with no omniscience, and that $C(0) = -1$ forces $C \equiv -1$.

The curvature parameter c of part 2 is easily confused with a different family.

Proposition 21.16 (Two gauges, and only one keeps the law)**.**

1. *For every point c, the cost $t \mapsto \cosh(ct) - 1$ obeys the composition law; at $t = \log x$ it equals $J(x^c)$. This is the* exponent family.
2. *For a point λ, the cost $\lambda(\cosh t - 1)$ obeys the composition law if and only if $\lambda^2 = \lambda$.*
3. *If $\kappa\lambda = 2$, then $C = \lambda(\cosh - 1)$ obeys $C(s+t) + C(s-t) = \kappa C(s)C(t) + 2C(s) + 2C(t)$ for all s, t; if it also obeys the composition law, then $\kappa = 2$ and $\lambda = 1$.*

The proofs are free.

Proof. 1. Proposition 21.12, item 1, and $J(x^c) = \cosh(c \log x) - 1$.

2. If $\lambda^2 = \lambda$, multiply the law for $G = \cosh - 1$ by λ and use $2\lambda^2 G(s)G(t) = 2\lambda G(s)G(t)$. Conversely, write the law for λG at $s = t = 1$ and subtract λ times the law for G there: what

remains is $2(\lambda^2-\lambda)G(1)^2=0$. Since $e\geq 2$ by Lemma 21.3, $G(1)\geq \frac{1}{2}e^{-1}>0$, so $G(1)^2$ is invertible and $\lambda^2=\lambda$.

3. $\lambda G(s\pm t)$ summed is $\lambda\big(2G(s)G(t)+2G(s)+2G(t)\big)$, and $2\lambda=\kappa\lambda^2$. If moreover $\lambda^2=\lambda$, then $2\lambda=\kappa\lambda^2=\kappa\lambda=2$, so $\lambda=1$ and $\kappa=2$. □

The *amplitude family* λJ belongs to the law with product coefficient $\kappa=2/\lambda$. Scaling the cost by $\kappa/2$, as in the relation $(\kappa/2)F=J$, leaves the composition law of this chapter unless $\kappa=2$. The family that stays inside the law is the exponent family $J(x^c)$, whose parameter is the curvature c^2 at the reference. The sign of the curvature selects the branch.

Proposition 21.17 (The curvature selects the branch)**.** *Let c be a point. The cost $t\mapsto\cos(ct)-1$ obeys the composition law, has curvature $-c^2$ at 0 and is at most 0. A cost that obeys the law and has curvature $-c^2$ at 0 equals $\cos(ct)-1$ at every point. The proofs are free.*

Proof. By Lemma 21.7, $\cos(a+b)+\cos(a-b)=2\cos a\cos b$, and the computation of Proposition 21.12, item 1, gives the law. Also $\cos u-1=-2\sin^2(u/2)\leq 0$. For a natural number $M\geq|c|$ and $|t|\leq 1/(M+1)$, Lemma 21.7, item 4, gives $|\cos(ct)-1+c^2t^2/2|\leq(ct)^4\leq M^3|t|\,t^2$ and $|\cos(ct)-1|\leq M^2t^2$; the modulus is chosen as in Proposition 21.12, item 5, and Theorem 21.13 applies as in Theorem 21.14. □

Recognition Science's calibration is $+1$, so the oscillating branch is excluded by the sign of a premise, not by an additional assumption.

The golden orbit

The ring $\mathbb{Z}[\varphi]$ carries the ledger of Section 21.5 and the discrete form of the forcing theorem.

Definition 21.18 (The golden ratio and its ring)**.** Let $\varphi=(1+\sqrt{5})/2$ in $\mathbb{R}_\delta$, with $\sqrt{5}$ the nonnegative square root of Proposition 12.19, and $\psi=1-\varphi$. The ring $\mathbb{Z}[\varphi]$ consists of the formal sums $a+b\varphi$ with integers a,b, multiplied by $(a+b\varphi)(a'+b'\varphi)=(aa'+bb')+(ab'+ba'+bb')\varphi$, that is, with $\varphi^2=\varphi+1$; each such sum also names a point of $\mathbb{R}_\delta$. The element $a+b\varphi$ *passes the sign test* when, with $s=2a+b$ and $t=b$ (so that $a+b\varphi=(s+t\sqrt{5})/2$),

$$s\geq 0,\ t\geq 0,\ (s,t)\neq(0,0),\quad\text{or}\quad s<0<t,\ s^2<5t^2,\quad\text{or}\quad t<0<s,\ 5t^2<s^2.$$

The point φ is positive, $\varphi^2=\varphi+1$, $\varphi\psi=-1$, so $\varphi^{-1}=\varphi-1$ and every integer power φ^n lies in $\mathbb{Z}[\varphi]$; also $\varphi^{-3}=2\varphi-3=\sqrt{5}-2$. It is the only positive solution of $x^2=x+1$: if $x>0$ solves it, then $(x-\varphi)(x+\varphi-1)=0$ with $x+\varphi-1>0$.

Lemma 21.19 (The sign test decides the order of $\mathbb{Z}[\varphi]$)**.** *For integers a,b, the point $a+b\varphi$ is positive exactly when $a+b\varphi$ passes the sign test. For every $d\in\mathbb{Z}[\varphi]$ exactly one of the following holds: d passes the sign test, $d=0$, or $-d$ passes the sign test. These are decided by integer arithmetic, and no principle is used.*

Proof. Since $\sqrt{5}$ is irrational, $s^2=5t^2$ holds for integers only at $s=t=0$: otherwise 5 divides s, then t, and the pair can be divided by 5 indefinitely. If $s,t\geq 0$, not both 0, then $s+t\sqrt{5}>0$. If $s<0<t$, then $s+t\sqrt{5}>0$ exactly when $t\sqrt{5}>-s$, and for positive points $u<v$ exactly when

$u^2 < v^2$; so exactly when $5t^2 > s^2$. The third case is symmetric, and if $s, t \le 0$ the point is not positive. The trichotomy follows from integer trichotomy for s, t and $s^2 - 5t^2$. □

For an arbitrary point, deciding $x > 0 \vee x = 0 \vee x < 0$ costs LPO (Corollary 12.18). The lemma says that on $\mathbb{Z}[\varphi]$ it costs nothing.

Proposition 21.20 (J on the golden orbit)**.**

1. *For every integer n, $e^{n \log \varphi} = \varphi^n$.*
2. *Every comparison cost C that obeys the composition law and is calibrated satisfies $2C(n \log \varphi) = \varphi^n + \varphi^{-n} - 2 = 2J(\varphi^n)$, an element of $\mathbb{Z}[\varphi]$, for every integer n.*
3. *A function g on the integers, with values in $\mathbb{Z}[\varphi]$ or in $\mathbb{R}_\delta$, that satisfies*

$$g(m+n) + g(m-n) = g(m)g(n) + 2g(m) + 2g(n), \qquad g(0) = 0, \qquad g(1) = \sqrt{5} - 2, \tag{21.6}$$

is $g(n) = \varphi^n + \varphi^{-n} - 2$.

The proofs are free.

Proof. 1. $e^{\log \varphi} = \varphi$ and $e^{a+b} = e^a e^b$. 2. By Theorem 21.14, $C(n \log \varphi) = \cosh(n \log \varphi) - 1$. 3. The function $\varphi^n + \varphi^{-n} - 2 = 2G(n \log \varphi)$ satisfies (21.6), because $2G$ obeys the law with product coefficient 1 and $\varphi + \varphi^{-1} = 2\varphi - 1 = \sqrt{5}$. Conversely, $m = 0$ gives $g(-n) = g(n)$, and the law at the pair $(k+1, 1)$ gives $g(k+2) = g(k+1)g(1) + 2g(k+1) + 2g(1) - g(k)$ for every $k \ge 0$, so $g(0)$ and $g(1)$ determine g. □

Recognition Science proves part 3 for functions into $\mathbb{Z}[\varphi]$: the discrete composition law with the unit value $\sqrt{5} - 2$ at one golden step has exactly one solution. Part 2 shows that this discrete theorem is Theorem 21.14 restricted to the orbit $\{n \log \varphi\}$, the logarithms of the powers of the golden ratio.

Deciding the law for a constant

The forcing theorem takes the law as a premise. Whether a given cost obeys it is a separate question with a price. The smallest case already costs WLPO.

Proposition 21.21 (Constant costs: deciding the law is WLPO)**.**

1. *The constant cost $C \equiv a$ obeys the composition law exactly when $a^2 + a = 0$.*
2. *The assertion that for every point a the constant cost a obeys the law or does not is equivalent to WLPO, with no choice principle.*
3. *For every point c the law for $t \mapsto \cosh(ct) - 1$ is a free theorem.*

Proof. 1. At $s = t = 0$ the law reads $2a = 2a^2 + 4a$, that is $a^2 + a = 0$; conversely $a^2 + a = 0$ gives $a + a = 2a^2 + 2a + 2a$ at every pair.

2. WLPO decides $x = 0 \vee x \neq 0$ for every point (Corollary 12.18); apply it to $x = a^2 + a$. Conversely, given a point r, decide the law for the constant r^2. It holds exactly when $r^4 + r^2 = 0$, that is when $r^2 = 0$, since both terms are at least 0; and $r^2 = 0$ exactly when $r = 0$. So $r = 0 \vee r \neq 0$ is decided for every r, which is WLPO by Corollary 12.18.

3. Proposition 21.12, item 1. □

21.4 J is the cost of one fair binary distinction

The forcing theorem starts from an arbitrary function. A second route starts from a mechanism. Read a comparison $x > 0$ through one record increment K, an integer drawn from some law, as the power x^K, and take as the cost the expected excess $\mathbb{E}[x^K] - 1$. Which laws give a cost that obeys the composition law? The answer is: exactly the fair choice between two opposite steps, and the unit calibration picks the step ± 1. For the fair sign $S = \pm 1$ with probability $\frac{1}{2}$ each, $1 + J(x) = \mathbb{E}[x^S]$, and the law says $\mathbb{E}[x^S]\,\mathbb{E}[y^{S'}] = \frac{1}{2}\big(\mathbb{E}[(xy)^S] + \mathbb{E}[(x/y)^S]\big)$ for independent fair signs S, S'.

This section is proved in classical mathematics, over the classical real numbers and, for the second theorem, with measure theory. It is a characterisation of J among costs of a stated form, not a result priced on $\mathbb{R}_\delta$. Reading a comparison through an increment as a power is an identification; the theorems say what follows from it. Throughout, the composition law is (21.1) for all classical $x, y > 0$.

Theorem 21.22 (One fair binary distinction, finite support)**.** *In classical mathematics: let S be a finite set of integers and a_k, for $k \in S$, real numbers of any sign with $\sum_k a_k = 1$, and put $C(x) = \sum_{k\in S} a_k x^k - 1$ for $x > 0$.*

1. *C obeys the composition law if and only if there is a natural number c with $C(x) = J(x^c)$ for every $x > 0$.*
2. *In that case $a_k = \frac{1}{2}$ for $k = \pm c$ and $a_k = 0$ for every other $k \in S$ when $c \geq 1$, and $a_0 = 1$, $a_k = 0$ otherwise, when $c = 0$. The weights are nonnegative and the law is the fair choice between c and $-c$.*
3. *For a natural number c, $J(2^c) = \frac{1}{4}$ if and only if $c = 1$. Hence if C obeys the law and $C(2) = \frac{1}{4}$, then $C = J$ on $x > 0$ and the law is the fair step ± 1.*

Proof. Laurent monomials are independent. If $\sum_{k\in S} b_k x^k = 0$ for every $x > 0$, then every $b_k = 0$: with $N = \sum_{k\in S} |k|$ the polynomial $\sum_k b_k X^{k+N}$ vanishes at infinitely many points, so it is 0, and its exponents $k + N$ are distinct.

The law on coefficients. Put $M(x) = 1 + C(x) = \sum_k a_k x^k$. The law for C is equivalent to $M(xy) + M(x/y) = 2M(x)M(y)$. Fix $y > 0$; as a Laurent polynomial in x,

$$M(xy) + M(x/y) - 2M(x)M(y) = \sum_{k\in S} a_k\big(y^k + y^{-k} - 2M(y)\big)x^k,$$

so the law gives $a_k(y^k + y^{-k} - 2M(y)) = 0$ for every $k \in S$ and $y > 0$.

One profile. The weights sum to 1, so some $a_{k_0} \neq 0$, and then $M(y) = \frac{1}{2}(y^{k_0} + y^{-k_0})$ for every $y > 0$. With $c = |k_0|$ this is $C(y) = J(y^c)$. *The weights.* The fair weights w_k of part 2 satisfy $\sum_k w_k x^k = \frac{1}{2}(x^c + x^{-c}) = M(x)$; independence on $S \cup \{c, -c\}$ gives $a_k = w_k$ for $k \in S$. Positivity and fairness are conclusions.

The converse. J obeys the law at all $u, v \neq 0$: both sides plus 2 equal $\frac{1}{2}(u+u^{-1})(v+v^{-1})$. Apply this at $u = x^c$, $v = y^c$.

Part 3. Put $z = 2^c \geq 1$. Then $J(z) = \frac{1}{4}$ exactly when $2z^2 - 5z + 2 = 0$, that is $(2z-1)(z-2) = 0$; $z \geq 1$ excludes $z = \frac{1}{2}$, so $z = 2$ and $c = 1$. Conversely $J(2) = \frac{5}{4} - 1 = \frac{1}{4}$. If $C(2) = \frac{1}{4}$, part 1 gives $C(2) = J(2^c)$, so $c = 1$. □

Theorem 21.23 (One fair binary distinction, beyond finite support)**.** *In classical mathematics: let μ be a probability measure on the real line with $\int e^{st}\, d\mu(t) < \infty$ for every real s, and put $C_\mu(x) = \int x^t\, d\mu(t) - 1$ for $x > 0$.*

1. *C_μ obeys the composition law if and only if $\mu = \frac{1}{2}\delta_c + \frac{1}{2}\delta_{-c}$ for some $c \geq 0$.*
2. *For every real c and every $x > 0$, the law $\frac{1}{2}\delta_c + \frac{1}{2}\delta_{-c}$ has cost $J(x^c)$.*
3. *For real $c \geq 0$, $J(2^c) = \frac{1}{4}$ if and only if $c = 1$.*

Proof. Put $E(s) = \int e^{st}\, d\mu(t)$, so that $C_\mu(e^s) = E(s) - 1$ and $E(0) = 1$. At $x = e^a$, $y = e^b$ the law becomes $E(a+b) + E(a-b) = 2E(a)E(b)$ for all real a, b. At $a = 0$ it gives $E(-b) = E(b)$, and at $a = b = 1$ it gives $E(2) = 2m^2 - 1$ with $m = E(1)$.

The variance argument. Since $e^t e^{-t} = 1$,

$$\left(e^t + e^{-t} - 2m\right)^2 = e^{2t} + e^{-2t} + 2 + 4m^2 - 4me^t - 4me^{-t},$$

and every term is integrable. Integrating, $2(2m^2 - 1) + 2 + 4m^2 - 8m^2 = 0$. A nonnegative function with integral 0 vanishes almost everywhere, so $\cosh t = m$ for μ-almost every t.

Two points. The almost-sure set is not empty; pick t_0 in it and put $c = |t_0|$. Since cosh is even and injective on $[0, \infty)$, μ is carried by $\{c, -c\}$. If $c = 0$, $\mu = \delta_0$. If $c > 0$, write $\mu = p\delta_c + q\delta_{-c}$; then $E(-1) = E(1)$ reads $(p - q)(e^c - e^{-c}) = 0$, so $p = q = \frac{1}{2}$.

Parts 2 and 3. $\int f d\mu = \frac{1}{2}(f(c) + f(-c))$ gives $C_\mu(x) = \frac{1}{2}(x^c + x^{-c}) - 1 = J(x^c)$, and the law follows from the law for J at $u = x^c$, $v = y^c$. Part 3 is the computation of Theorem 21.22 with $z = 2^c \geq 1$, followed by $c \log 2 = \log 2$. □

The hypothesis $c \geq 0$ in part 3 is needed: $J(2^{-1}) = J(\frac{1}{2}) = \frac{1}{4}$, since $J(x) = J(x^{-1})$. Theorem 21.22 assumes no positivity of the weights, and neither theorem assumes symmetry of the law, continuity or a calibration at the reference; the only reference value used is $J(2) = \frac{1}{4}$. Where Theorem 21.14 classifies functions, these theorems classify mechanisms, and both routes end at the same function.

Conditions and scope.

Base and premises
Classical real numbers and, in Theorem 21.23, a probability measure with all exponential moments finite. Proved in classical mathematics.

Conclusion
Among costs $\mathbb{E}[x^K] - 1$, the composition law holds exactly for the fair law on $\pm c$, and the value $\frac{1}{4}$ at $x = 2$ fixes $c = 1$, the cost J.

Identification
Reading a comparison through one increment as a power of the ratio, and the cost as the expected excess.

Not claimed
A version priced on $\mathbb{R}_\delta$.

21.5 The ledger is $\mathbb{Z}[\varphi]$

Recognition Science records recognitions in a ledger. A recognition is summarised by its census: how many independent steps it took and how many closures. Joining a recognition to the one before it is the step that Chapter 22 shows forces the golden ratio. This section proves what that bookkeeping is as mathematics: a census is an element of $\mathbb{Z}[\varphi]$, prices of recognitions are compared by the integer sign test of Lemma 21.19 at no price, the stream of postings is the unique history that a sigma-delta rule produces, and the balances and rates of any ledger are exactly δ's integers and ratios.

Definition 21.24 (Censuses, the join and the posting history)**.**

1. A *census* is a pair (i, c) of natural numbers, not both 0: i independent steps and c closures. The *join* sends (i, c) to $(i + c, i)$. The *base census* is $(0, 1)$, and the kth recognition is the base census joined k times. Its number of *postings* is $i + c$.

2. The *golden value* of (i, c) is $c + i\varphi \in \mathbb{Z}[\varphi]$, and its *price* is the point $i\varphi + c$ of $\mathbb{R}_\delta$.

3. A *posting word* is a finite word in the letters d (a debit, for an independent step) and c (a credit, for a closure); its *balance* is the number of letters d minus the number of letters c. The join acts on words by replacing every d by dc and every c by d. Put $W_0 = \mathsf{d}$ and let W_{k+1} be the join of W_k.

4. Since W_k is an initial segment of W_{k+1}, the words define one infinite *recognition history* h, whose Nth letter is that of any W_k long enough. Write b_N for the balance of its first N letters.

The first words are d, dc, dcd, dcddc and $\mathsf{dcddcdcd}$. Write $F(n)$ for the Fibonacci numbers, $F(0) = 0$, $F(1) = 1$, $F(n + 2) = F(n + 1) + F(n)$, extended by $F(-1) = 1$. By induction, $\varphi^n = F(n)\varphi + F(n - 1)$ and $F(n + 1) - \varphi F(n) = \psi^n$ for $n \geq 0$; the second identity holds because $\psi(F(n+1) - \varphi F(n)) = \psi F(n+1) + F(n)$, using $\psi\varphi = -1$. Since $|\psi| = \varphi^{-1}$ and $F(n) \leq \varphi^n$, it gives

$$\left|\frac{F(n+1)}{F(n)} - \varphi\right| \leq \frac{1}{F(n)^2} \qquad (n \geq 1). \tag{21.7}$$

Theorem 21.25 (The recognition ledger is $\mathbb{Z}[\varphi]$)**.**

1. *The golden value is injective on censuses, and its image is the set of nonzero $a + b\varphi$ with $a, b \geq 0$.*
2. *The join multiplies the golden value by φ, and the kth recognition has golden value φ^k.*
3. *For censuses x and y, the price of x is less than the price of y exactly when the golden value of y minus that of x passes the sign test. Any two prices satisfy exactly one of $<, =, >$, and the case is decided; prices are equal only for equal censuses; the join multiplies the price by φ.*
4. *The kth recognition posts $P(k) = F(k+1)$ times, and $P(k+2)P(k) - P(k+1)^2 = (-1)^k$. No posting count times φ is the next posting count, yet the ratios $P(2n+4)/P(2n+3)$ form a regular sequence that names φ.*

No omniscience principle and no choice principle is used.

Proof. 1. $(i, c) \mapsto c + i\varphi$ reads off both coordinates, and the image is as stated. 2. $\varphi(c + i\varphi) = c\varphi + i(\varphi + 1) = i + (i + c)\varphi$, the golden value of $(i + c, i)$; the base census has value 1. 3. The price is the golden value read as a point, and the map from $\mathbb{Z}[\varphi]$ to $\mathbb{R}_\delta$ preserves sums and products; Lemma 21.19 gives the comparison and the trichotomy, and item 1 gives equality. 4. By induction the kth census is $(F(k), F(k-1))$, with $F(k) + F(k-1) = F(k+1)$ postings, and Cassini's identity $F(n+1)F(n-1) - F(n)^2 = (-1)^n$ [99] at $n = k + 2$ gives the second claim. A nonzero integer times φ has a nonzero φ-coordinate, so it is no integer. Finally $r_n = F(2n+5)/F(2n+4)$ satisfies $|r_n - \varphi| \leq 1/F(2n+4)^2 \leq 1/(2n+2)$ by (21.7), so $|r_m - r_n| \leq 1/(2m+2) + 1/(2n+2)$: the sequence is regular, and the point it names is within $2/(n+1)$ of φ for every n, hence equal to φ. □

Two examples show what the sign test does. Three independent steps cost less than five closures, $3\varphi < 5$: the difference $5 - 3\varphi$ has $s = 7$, $t = -3$ and $5t^2 = 45 < 49$. And eight closures cost less than five steps, $8 < 5\varphi$: here $s = -11$, $t = 5$ and $121 < 125$. A comparison of arbitrary points of $\mathbb{R}_\delta$ costs LPO; a comparison of recognition prices costs nothing, because the prices live in $\mathbb{Z}[\varphi]$.

The history h begins dcddcdcddcddc $\cdots$, the Fibonacci word [100]. Its running balance stays close to the line of slope $\varphi^{-3} = \sqrt{5} - 2 \approx 0.236$, and it is produced by a feedback rule of the kind used in sigma-delta converters [101].

Theorem 21.26 (The history is a sigma-delta encoder of φ^{-3})**.**

(a) $W_{k+2} = W_{k+1}W_k$. The word W_k has $F(k+1)$ debits and $F(k)$ credits, so it is the posting word of the $(k+1)$st recognition; its length is $F(k+2)$ and its balance $F(k+1) - F(k)$.

(b) For every N, the discrepancy $D_N = N\varphi^{-3} - b_N$, an element of $\mathbb{Z}[\varphi]$, satisfies

$$-\frac{2}{\varphi} < D_N < \frac{2}{\varphi^2},$$

so $|N\varphi^{-3} - b_N| < 2$. Each inequality is one sign test.

(c) The running rate b_N/N at $N = F(2n+6)$ equals $-3 + 2F(2n+7)/F(2n+6)$, and these rates form a regular sequence that names φ^{-3}.

(d) For a natural number r, there is an increasing sequence of depths K along which the ratios $P(K+r)/P(K)$ form a regular sequence that names φ^{r}, and there is another along which the ratios $P(K)/P(K+r)$ one that names φ^{-r}.

(e) The Nth posting of h is a debit exactly when $b_N < (N+2)\varphi^{-3}$, and h is the only infinite history with this property.

No omniscience principle and no choice principle is used.

Proof. (a) Write σ for the join on words. It respects concatenation and $\sigma(\mathsf{c}) = \mathsf{d}$, so $W_{k+2} = \sigma^{k+1}(\mathsf{dc}) = W_{k+1}\,\sigma^{k}(\mathsf{d}) = W_{k+1}W_k$. On letter counts σ is the join of censuses, and W_0 has census $(1,0)$, the first recognition. The counts follow from item 4 of Theorem 21.25.

(b) For a word w put $\Delta(w) = |w|\varphi^{-3} - \mathrm{bal}(w)$, which adds over concatenation, and $g_k = \Delta(W_k)$. By (a), $g_{k+2} = g_{k+1} + g_k$. Since $g_0 = \varphi^{-3} - 1 = -2\psi^2$, $g_1 = 2\varphi^{-3} = -2\psi^3$ and $\psi^2 = \psi + 1$, induction gives $g_k = -2\psi^{k+2}$. As $\psi < 0$, the g_k alternate in sign, negative for even k. Put

$$L_k = -\frac{2}{\varphi} + \max(g_k, g_{k+1}), \qquad U_k = \frac{2}{\varphi^2} + \min(g_k, g_{k+1}).$$

We claim that every initial segment p of W_k, the empty one and W_k included, has $L_k \le \Delta(p) \le U_k$. For $k = 0$ and $k = 1$, the segments have Δ equal to 0, g_0 and g_1, and the bounds are $[g_0, 0]$ and $[g_0, g_1]$, using $-2/\varphi = 2\psi$, $2/\varphi^2 = 2\psi^2$ and $1 - \psi^2 = -\psi$. For the step, a segment of $W_{k+2} = W_{k+1}W_k$ either lies in W_{k+1} or is $W_{k+1}p'$ with p' a segment of W_k, and then $\Delta = g_{k+1} + \Delta(p')$. Let k be even, so $g_k < 0 < g_{k+1}$, $g_{k+2} < 0 < g_{k+3}$ and $g_{k+3} < g_{k+1}$. Then $U_{k+2} = U_{k+1}$ and $L_{k+2} \le L_{k+1}$, which covers the first kind; for the second, $\Delta \le 2/\varphi^2 + g_k + g_{k+1} = U_{k+2}$ and $\Delta \ge -2/\varphi + 2g_{k+1} \ge -2/\varphi + g_{k+1} + g_{k+2} = L_{k+2}$. The case of odd k is the mirror image. Every comparison here is between elements of $\mathbb{Z}[\varphi]$, so it is a sign test. Now $|W_N| = F(N+2) \ge N$, so D_N is Δ of an initial segment of W_N; one of g_N, g_{N+1} is positive and the other negative, so $-2/\varphi < D_N < 2/\varphi^2$, and both bounds lie within 2 of 0.

(c) After $F(2n+6)$ postings the record is W_{2n+4}, with balance $F(2n+5) - F(2n+4)$, and $2F(2n{+}7) - 3F(2n{+}6) = F(2n{+}5) - F(2n{+}4)$. By (21.7) the rate is within $2/F(2n{+}6)^2 \le 1/(2n{+}2)$ of $-3 + 2\varphi = \varphi^{-3}$, and the argument of Theorem 21.25, item 4, applies.

(d) For $r \ge 1$, $F(K+r) = F(r)F(K+1) + F(r-1)F(K)$ and $\varphi^r = F(r)\varphi + F(r-1)$ give $F(K{+}r) - \varphi^r F(K) = F(r)\psi^K$, so $|F(K{+}r)/F(K) - \varphi^r| \le F(r)/F(K)^2$ and $|F(K)/F(K{+}r) - \varphi^{-r}| \le F(r)/F(K)^2$. Choose the depths so large that these bounds are at most $1/(2n+2)$ at the nth term, and argue as in (c). For $r = 0$ the ratios are 1.

(e) A debit changes D by $\varphi^{-3} - 1$ and a credit by $\varphi^{-3} + 1$; the interval of (b) has length $2/\varphi + 2/\varphi^2 = 2$. If posting N is a credit, $D_{N+1} < 2/\varphi^2$ gives $D_N < 6 - 4\varphi$; if it is a debit, $D_{N+1} > -2/\varphi$ gives $D_N > 6 - 4\varphi$. Since $6 - 4\varphi = -2\varphi^{-3}$, $D_N - (6 - 4\varphi) = (N+2)\varphi^{-3} - b_N$, which is the rule. A history obeying the rule is determined letter by letter, since the rule at N reads only the first N letters; so it is h. $\square$

The bounds of (b) have length exactly 2, the change of D between a debit and a credit, and this is what makes (e) a rule. The rule is a one-bit quantiser with feedback: it posts a debit whenever

the books have fallen below the target line, with a lead of two postings. Reading postings as physical events, debits as independent steps and credits as closures, is the identification of Definition 21.24; the counts, balances and rates do not depend on the order within a recognition, but the prefix bounds do.

Balances and rates of any ledger

The recognition history is one ledger. For an arbitrary history of postings on one account, the balances and rates are δ's integers and ratios (Definitions 4.3 and 4.9).

Theorem 21.27 (Balances are integers, rates are ratios, and every rate in $[-1, 1]$ is a running rate)**.**

1. *The balance of a posting word w is the signed record $[\#\mathsf{d}(w), \#\mathsf{c}(w)]$ of $\mathbb{Z}_\delta$. Every element of $\mathbb{Z}_\delta$ is a balance, concatenation adds balances, and two words have the same balance exactly when one is turned into the other by swapping adjacent letters and deleting or inserting adjacent pairs* dc*. A balance z and a number n of postings occur together exactly when $|z| \le n$ and $n - z$ is even. The balance forgets the record:* dc *and the empty word have balance* 0 *and different numbers of postings.*

2. *A* rate record *is a word w read over $t+1$ ticks; its value is $\langle \mathrm{bal}(w); t+1\rangle \in \mathbb{Q}_\delta$. Every element of $\mathbb{Q}_\delta$ is a rate record; (w, t) and (w', t') have equal values exactly when $\mathrm{bal}(w)(t' + 1) = \mathrm{bal}(w')(t+1)$; concatenation on one clock adds rates; exchanging the letters negates; replacing each debit of u by v and each credit by v with its letters exchanged, read over $(s + 1)(t + 1)$ ticks, multiplies the rates of (u, s) and (v, t).*

3. *Every point of $\mathbb{R}_\delta$ is named by a regular sequence of rate records.*

4. *For every point x with $-1 \le x \le 1$ there is an infinite history whose running rates b_N/N, read at $N = 6(n + 1)$, form a regular sequence naming x. Conversely, if the running rates of a history, read at any increasing sequence of stages, form a regular sequence, the point it names lies in $[-1, 1]$.*

No omniscience principle and no choice principle is used.

Proof. 1. Counting letters gives the signed record, and $\mathsf{d}^a\mathsf{c}^b$ has balance $[a, b]$. Moves preserve the balance. Conversely, swaps sort a word to $\mathsf{d}^p\mathsf{c}^q$, and deleting pairs leaves d^{p-q} or c^{q-p}, which depends only on the balance. A word with p debits and q credits has $z = p - q$ and $n = p + q$; conversely $p = (n + z)/2$ and $q = (n - z)/2$.

2. The word $\mathsf{d}^a\mathsf{c}^b$ over d ticks has value $\langle [a, b]; d\rangle$. Equality is cross equality of ratio records. In a nested word each debit of u contributes $\mathrm{bal}(v)$ and each credit $-\mathrm{bal}(v)$, so the balance is $\mathrm{bal}(u)\mathrm{bal}(v)$, over $(s + 1)(t + 1)$ ticks.

3. Every ratio record is a rate record, by item 2.

4. Fix a regular sequence q naming x; the proof needs one, and the assertion is a proposition, so nothing is chosen. Post a debit at tick N exactly when $b_N \le Nq_N$, and a credit otherwise. Fix K. Each stage satisfies $|q_K| \le 1 + 1/(K + 1)$, and $|Nq_N - Nq_K| \le N/(K + 1) + 1$ by regularity.

Put $E_N = b_N - Nq_K$. After a debit, $E_N \le N/(K+1) + 1$ and $E_{N+1} = E_N + 1 - q_K$; after a credit, $E_N > -N/(K+1) - 1$ and $E_{N+1} = E_N - 1 - q_K$. By induction from $E_0 = 0$,

$$|b_N - Nq_K| \le 3 + \frac{4N}{K+1} \qquad \text{for all } N \text{ and } K.$$

At $N = 6(n+1)$ this gives $|b_N/N - q_K| \le 1/(2n+2) + 4/(K+1)$ for every K, and since q_K is within $1/(K+1)$ of x, the running rate is within $1/(2n+2)$ of x. These rates therefore form a regular sequence naming x. Conversely, a balance never exceeds the number of postings in absolute value, so every running rate lies in $[-1, 1]$, and so does any point they name (Lemma 12.52, item 2). □

The interval $[-1, 1]$ is forced by one posting per tick. Which history nature runs, and that a measured quantity is a running rate, are identifications the theorem does not make. Chapter 24 builds the light cone on these rates.

21.6 What deciding a cost law costs

Theorem 21.14 takes the composition law as a premise. Checking the premise for a given cost is a decision with a price, and Proposition 21.21 found it already for constants: WLPO. For general costs the price depends on how the cost is given, on whether it is calibrated, and on the reading of maps. This section computes it. The same questions arise for Recognition Science's balance premise on the lattice, and the answers coincide.

Definition 21.28 (Zero tests, aggregation, and how a cost is given)**.**

1. The *function zero test* FZ: for every function H on $\mathbb{R}_\delta$, either $H(x) = 0$ for every x, or not.
2. The *pair zero test* PZ: the same for every function of two points.
3. The *sequence zero test* SZ: for every sequence (y_n) of points, either $y_n = 0$ for every n, or not.
4. *Aggregation* AGG: for every sequence (y_n) of points there is a point z with $z = 0$ exactly when $y_n = 0$ for every n.
5. A cost C is *continuous* when for every point x and every k there is a j with $|C(x) - C(y)| \le 1/(k+1)$ whenever $|x - y| \le 1/(j+1)$. It is *presented* when it is continuous and there are functions of the natural numbers n and j giving regular sequences that name $C(j/3^n)$ and $C(-j/3^n)$.
6. *Deciding the law* for a class of costs means: for every cost C in the class, C obeys the composition law or it does not.

The sequence zero test concerns one sequence of points at a time; it is not the principle $\mathrm{WLPO}_{\mathrm{seq}}$ of Definition 13.8, which decides a sequence of binary rows. Two constructions do most of the work. For a sequence (y_n) of points, its *interpolation* is

$$\iota_y(t) = y_0 + \sum_{k<K} r(t-k)\,(y_{k+1} - y_k), \qquad r(u) = \min\bigl(1, \max(0, u)\bigr),$$

for any natural number $K \geq t$; the terms with $k \geq t$ vanish, so the value does not depend on K. It is continuous, $\iota_y(n) = y_n$ by telescoping, and $\iota_y \equiv 0$ when every y_n is 0. For a function H on $\mathbb{R}_\delta$, the *coded cost* is

$$C_H(t) = \cosh t - 1 + \max(0, t-1)\, H(t-2) + \max(0, -t-1)\, H(t+2).$$

Both ramps vanish on $[-1, 1]$, so C_H agrees with $\cosh -1$ there and hides H outside.

Theorem 21.29 (The price ladder for costs given as functions)**.**

1. *If C obeys the composition law, then $C(0) = 0$ or $C(0) = -1$, and $C(0) = -1$ implies $C(t) = -1$ for every t.*
2. *For every function H, the coded cost C_H is calibrated, and it obeys the composition law exactly when $H(x) = 0$ for every x.*
3. *Deciding the law for all costs implies* FZ*.* FZ *implies* SZ*, and* SZ *is equivalent to WLPO together with* AGG*.*
4. PZ *implies deciding the law for all costs, and* WEM *implies* PZ*.*
5. *Deciding the law for all calibrated costs is equivalent to* FZ*.*

No choice principle is used, and parts 1 and 2 are free. With Theorem 28.36, which derives PZ *from* FZ *with no choice principle, parts 3 and 4 settle the price for all costs: deciding the law for all costs is equivalent to* FZ*, and so to* PZ*, as it is for calibrated costs by part 5; it implies WLPO and follows from* WEM*.*

Proof. 1. At $s = t = 0$ the law gives $2C(0) = 2C(0)^2 + 4C(0)$, so $C(0)(C(0) + 1) = 0$. Cotransitivity of $-1 < 0$ at $C(0)$ gives $-1 < C(0)$, where $C(0) + 1$ is invertible and $C(0) = 0$, or $C(0) < 0$, where $C(0)$ is invertible and $C(0) = -1$. This is a genuine disjunction. If $C(0) = -1$, the law at $(t, 0)$ reads $2C(t) = -2C(t) - 2 + 2C(t)$, so $C(t) = -1$.

2. C_H is calibrated because the calibration only concerns $|t| \leq 1/(j+1) \leq 1$, where $C_H = \cosh -1$. By Theorem 21.14, C_H obeys the law exactly when $C_H = \cosh -1$, that is, when $\max(0, t-1)H(t-2) + \max(0, -t-1)H(t+2) = 0$ for every t. If $H \equiv 0$ this holds. Conversely let x be a point; cotransitivity of $0 < 1$ gives $0 < x$ or $x < 1$. If $0 < x$, read the equation at $t = x + 2$: it says $(x+1)H(x) = 0$ with $x + 1 > 0$, so $H(x) = 0$. If $x < 1$, read it at $t = x - 2$: it says $(1 - x)H(x) = 0$ with $1 - x > 0$. Both cases conclude the same equation, so nothing is chosen.

3. Deciding the law for C_H decides $H \equiv 0$. Given a sequence y, FZ applied to ι_y decides whether every $y_n = \iota_y(n)$ is 0. If SZ holds, constant sequences give $r = 0 \vee r \neq 0$ for every point r, which is WLPO by Corollary 12.18, and the point $z = 0$ or $z = 1$, according to the decision, aggregates. Conversely WLPO decides $z = 0$ for an aggregating point z.

4. A cost obeys the law exactly when the function of two points $(s, t) \mapsto C(s+t) + C(s-t) - 2C(s)C(t) - 2C(s) - 2C(t)$ vanishes, which PZ decides. An assertion that a function vanishes is a conjunction of equations of points, and equations are stable under double negation; WEM gives $\neg P \vee \neg\neg P$, and $\neg\neg P$ gives P.

5. For a calibrated C, the law holds exactly when $C - (\cosh - 1) \equiv 0$, by Theorem 21.14; FZ decides this. Conversely C_H is calibrated, and part 2 turns a decision of its law into a decision of $H \equiv 0$.

For the closing sentence, parts 3 and 4 give that deciding the law for all costs implies FZ and that PZ implies it, and Theorem 28.36 gives PZ from FZ; the proof of that theorem uses parts 3 and 4 and part 1 of Proposition 21.34, and nothing that rests on the closing sentence. Part 3 gives WLPO from FZ, and part 4 gives the decision from WEM. □

The lower bound FZ is the easy half. For the upper bound, Chapter 28 reduces one pair zero test to countably many function zero tests, read along Schoenberg's space-filling curve, and FZ decides countably many at once (Proposition 21.34, part 1). Where FZ itself lies between WLPO and WEM is open; it is question (O2) at the end of this section.

Theorem 21.30 (Continuous and presented costs)**.**

1. *A continuous cost obeys the law exactly when it obeys it at all pairs $(a/3^n, b/3^n)$ with integers a, b and natural n.*
2. *Deciding the law for all continuous costs is equivalent to* SZ*, and so is deciding it for all continuous calibrated costs.*
3. *Deciding the law for all presented costs is equivalent to WLPO, and so is deciding it for all presented calibrated costs.*
4. *The cost* $\cosh - 1$ *is presented, calibrated and obeys the law, so for it the check costs nothing.*

No choice principle is used.

Proof. 1. Fix a triadic $s = a/3^n$. The function $t \mapsto C(s+t)+C(s-t)-2C(s)C(t)-2C(s)-2C(t)$ is continuous at every point and vanishes at every triadic $b/3^m$ with $m \geq n$, since $s = a3^{m-n}/3^m$. The triadic points are dense: every point is within $1/(j+1)$ of one, found from a stage of a regular sequence naming it, inside the proof of an equation. A function continuous at t and vanishing on a dense set vanishes at t. So the defect vanishes for triadic s and every t, and the same argument in s, for fixed t, finishes the proof.

2. An explicit surjection from $\mathbb{N}$ onto the triples (n, a, b) lists the triadic pairs in one sequence, and by part 1 the law is the zero test of the sequence of defects there; SZ decides it. Conversely, for a sequence y the coded cost C_{ι_y} is continuous and calibrated, and it obeys the law exactly when $\iota_y \equiv 0$, that is, when every $y_n = 0$.

3. Let C be presented. The triadic points of depth n are closed under sums and differences, so the defect at $(a/3^n, b/3^n)$ is a ring expression in presented values, and a regular sequence naming it is computed from the given ones. By Lemma 12.15, all these defects vanish exactly when every stage of every one of these sequences lies within $2/(k+1)$ of 0 at its stage k, a single binary sequence after the pairs (defect, stage) are listed. WLPO decides it, and part 1 carries the verdict to all pairs. Conversely, constant costs are presented, and deciding the law for them is WLPO (Proposition 21.21). For calibrated costs use instead C_c for the constant

function c: it is presented, since the exponential of a ratio and the ramps at triadic points are computed, it is calibrated, and it obeys the law exactly when $c = 0$.

4. The exponential of a ratio is computed from its constant sequence, so the values of $\cosh - 1$ at triadic points come with regular sequences; the rest is Proposition 21.12. □

A record of a value is a regular sequence that names it, what an apparatus writes stage by stage. Read that way, part 3 says that a cost recorded at the triadic grid of log-ratios is checked at the price WLPO, the price of checking one binary sequence. The gap between parts 2 and 3 is the difference between SZ and WLPO, which is aggregation.

Theorem 21.31 (Where aggregation comes from)**.**

1. *WLPO together with countable unique choice into* $\{0, 1\}$ *implies* SZ*.*
2. AGG *follows from each of: completeness of* $\mathbb{R}_\delta$ *for sequences with the modulus* $k \mapsto k$ *(every sequence* (F_m) *with* $|F_m - F_n| \le 1/(k+1)$ *for* $m, n \ge k$ *converges); completeness for all Cauchy sequences; least upper bounds of bounded sequences; and a choice of regular sequences naming the terms of each sequence of points, which follows from* AC_ω*.*

No choice principle is used beyond the ones named. So over any of these bases, SZ*, the law for continuous costs, and the balance premise of Theorem 21.32 each cost exactly WLPO.*

Proof. 1. WLPO decides $y_n = 0$ for each n (Corollary 12.18), so each n has exactly one bit b_n with $b_n = 1$ exactly when $y_n = 0$; countable unique choice gives the sequence (b_n), and WLPO applied to $(1 - b_n)$ decides whether every $y_n = 0$.

2. Let $a_i = \min(|y_i|, 1/(i+1))$ and $A_K = \max_{i \le K} a_i$. For $m \ge K$, $A_K \le A_m \le A_K + 1/(K+1)$, since the terms with $i > K$ are at most $1/(K+2)$; so (A_K) has the modulus $k \mapsto k$. Let L be its limit; as (A_K) is nondecreasing, $A_n \le L$ for every n. If every $y_n = 0$, every $A_K = 0$ and $L = 0$. If $L = 0$, then $a_n = 0$ for every n; if $|y_n| > 0$, then $a_n > 0$ as a minimum of positive points, so $|y_n| \le 0$ and $y_n = 0$. So L aggregates. A Cauchy sequence with that modulus is a Cauchy sequence. The least upper bound of $\min(|y_n|, 1)$ is 0 exactly when every $y_n = 0$. With regular sequences $s^{(n)}$ naming the terms, the binary sequence that lists, over all pairs (n, k), whether $|s_k^{(n)}| > 2/(k+1)$ is all zero exactly when every $y_n = 0$ (Lemma 12.15), and its jump point (Lemma 12.5) aggregates. AC_ω chooses the regular sequences. □

The balance premise

Recognition Science's balance premise says that the books close everywhere except at the source. Let a *current* assign to each site x of $\mathbb{Z}^3$ and each direction $i \in \{0, 1, 2\}$ a point $j(x, i)$ of $\mathbb{R}_\delta$, the net rate on the edge from x to $x + e_i$. Its divergence at x is $\sum_i \big(j(x, i) - j(x - e_i, i)\big)$, and the current is *balanced* when its divergence is 0 at every site other than the origin. It is *displayed* when regular sequences naming all its rates are given as one function of (x, i), and *integer* when its rates are integers.

Theorem 21.32 (The balance premise is one sequence zero test)**.**

1. *Deciding balance for every current is equivalent to* SZ*.*
2. *Deciding balance for every displayed current is equivalent to WLPO, and so is deciding it for every integer current.*
3. *Deciding balance for every current is equivalent to deciding it for every integer current together with* AGG*.*

No choice principle is used.

Proof. 1. An explicit surjection $m \mapsto x_m$ from $\mathbb{N}$ onto $\mathbb{Z}^3$, built from the diagonal walk through $\mathbb{N} \times \mathbb{N}$, turns balance into the zero test of the sequence whose mth term is the divergence at x_m, or 0 if x_m is the origin. Conversely, given (y_n), put the rate y_n on the edge leaving $(2n+1, 0, 0)$ in direction 0 and 0 on every other edge. The divergence is y_n at $(2n+1, 0, 0)$, $-y_n$ at $(2n+2, 0, 0)$ and 0 elsewhere, so this current is balanced exactly when every $y_n = 0$.

2. For a displayed current, regular sequences naming the divergences are computed, and as in Theorem 21.30, part 3, balance becomes one binary sequence, which WLPO decides. For an integer current the divergences are integers and WLPO decides directly. Conversely, the current of part 1 built from the constant sequence of the jump point of a binary sequence α, with its dyadic sequence, is displayed and is balanced exactly when α is all zero; with the bits $\alpha(n)$ themselves as rates it is an integer current.

3. Part 1, part 2 and Theorem 21.29, part 3. □

Read as physics, a displayed current is one given with the records of its rates. For such currents the continuum adds nothing to the price of integer postings. A difference could arise only for a current given as a bare function into the line, and then it is aggregation, a matter of completeness or choice, not of omniscience.

Under the default reading, and without it

Unique choice is part of the default reading of this book (Section 21.1). Under it the aggregation gap closes: SZ becomes WLPO. Table 21.1 collects the prices under both readings.

Corollary 21.33 (The prices under unique choice)**.** *Assume unique choice. Then* SZ *is equivalent to WLPO. Deciding the law for all continuous costs, and deciding the balance premise for all currents, each cost exactly WLPO.*

Proof. Countable unique choice into $\{0, 1\}$ is a case of unique choice; apply Theorems 21.31, 21.30 and 21.32. □

For costs given as functions unique choice changes nothing: by Theorem 21.29 the price is FZ under both readings.

More can be said about the bare law along rational lines, and two questions remain.

Table 21.1. The price of deciding a law for every member of a class. "Without unique choice" is the reading of Parts VI to VIII; every entry in that column is proved with no choice principle.

Decision, for every member	Without unique choice	With unique choice	Result
Law, constant costs	WLPO	WLPO	Prop. 21.21
Law, costs given as functions	FZ	FZ	Thm. 21.29, Thm. 28.36
Law, calibrated costs	FZ	FZ	Thm. 21.29
Law, continuous costs, calibrated or not	SZ, that is WLPO and AGG	WLPO	Thm. 21.30, Cor. 21.33
Law, presented costs, calibrated or not	WLPO	WLPO	Thm. 21.30
Law for $\cosh(ct) - 1$	free	free	Prop. 21.21
Balance, every current	SZ	WLPO	Thm. 21.32, Cor. 21.33
Balance, displayed or integer currents	WLPO	WLPO	Thm. 21.32

Proposition 21.34 (The law along rational lines)**.** *Say that C obeys the law* on rational lines *when the law holds at every pair (as, bs) with s a point and a, b integers.*

1. FZ *decides countably many function zero tests at once: for every sequence (H_n) of functions, whether $H_n(x) = 0$ for all n and x.*
2. *For a calibrated cost, the law, the law on rational lines and the duplication law* (21.4) *are equivalent.*
3. *A continuous cost that obeys the law on rational lines obeys the law.*
4. *Deciding the law on rational lines for all costs, and deciding the duplication law for all costs, are each equivalent to* FZ*.*
5. *In classical mathematics, the cost equal to $\cosh x - 1$ at rational x and to 0 elsewhere obeys the law on rational lines but not the law.*

Parts 1 to 4 use no choice principle.

Proof. 1. List the pairs $(n, g) \in \mathbb{N} \times \mathbb{Z}$ explicitly as (n_j, g_j). For a point z let v^z be the sequence whose terms $4j, \dots, 4j + 3$ all equal $H_{n_j}(z - 4j + g_j)$, and put $W(z) = \iota_{v^z}(z)$. On $[4j, 4j + 3]$ the interpolation of a sequence constant on that block is the common value, so $W(z) = H_{n_j}(z - 4j + g_j)$ there. Given n and x, take an integer g with $g - 1 \le x \le g + 2$, from a stage of a regular sequence naming x, take j with $(n_j, g_j) = (n, g - 1)$, and put $z = x + 4j - g + 1$; then $W(z) = H_n(x)$. So $W \equiv 0$ exactly when every $H_n \equiv 0$, and one test of W decides.

2. The line $(1, 1)$ with $C(0) = 0$ is the duplication law, and the duplication law with the calibration gives $C = \cosh - 1$ by Theorem 21.13, as in Theorem 21.14.

3. $a/3^n = a \cdot (1/3^n)$, so the law on rational lines contains the law at triadic pairs, and Theorem 21.30, part 1, applies.

4. Under FZ, part 1 decides the law on rational lines for every cost, a countable family of function zero tests, and the duplication law is one function zero test. Conversely C_H is calibrated, so by part 2 either decision for C_H decides $H \equiv 0$.

5. On the line of the points as with s rational, every point is rational and the cost is $\cosh -1$, which obeys the law. With s irrational, every as with $a \neq 0$ is irrational, and the cost is 0 on the whole line, since $\cosh 0 - 1 = 0$; the cost 0 obeys the law. At $s = \sqrt{2}$ and $t = 1 - \sqrt{2}$, the point $s + t = 1$ is rational and s, t, $s - t$ are not, so the law would give $\cosh 1 - 1 = 0$. □

The cost of part 5 is not continuous, by part 3, and it is defined by cases on rationality, which needs classical logic. Physically read, a comparison of a ratio with its own rational powers is one function zero test, and by Theorem 21.29 so is the full law, even for a cost that jumps between incommensurable ratios.

Conditions and scope.

Base and premises
The narrower reading of maps of Section 21.1; Corollary 21.33 adds unique choice.

Conclusion
Costs calibrated or not: exactly FZ when given as functions (for costs that are not calibrated, with Theorem 28.36), exactly SZ when continuous, exactly WLPO when presented. With unique choice, SZ is WLPO. Recognition Science's own cost is presented and calibrated, and its law is a free theorem.

Not claimed
The open statements below.

Two statements are open. The first is stated over the reading of Parts VI to VIII, without unique choice; the second is open even with unique choice.

(O1) Does WLPO imply AGG? By Theorems 21.29, 21.30 and 21.32, with no choice principle, each of the following is equivalent to a positive answer: SZ is equivalent to WLPO; deciding the law for continuous costs costs exactly WLPO; deciding balance for every current follows from deciding it for integer currents.

(O2) Where does FZ lie between WLPO and WEM?

21.7 A certain change costs a quarter cycle

A comparison is carried out by an apparatus, and the cost $J(x)$ is the action the apparatus spends, in units of a reference action. When can one comparison change a record with certainty, moving one configuration of a register to another with probability one? When the reference action is one full cycle, the answer is a threshold on the cost: $J(x) \geq \frac{1}{4}$, less what the holding of the register contributes (Figure 21.1). This section is proved in classical mathematics, over the classical real and complex numbers; it uses π for the classical number, and the model is stated in full.

Definition 21.35 (Registers, paid words and comparison controllers)**.** In classical mathematics:

1. A *register* of n binary channels has the 2^n configurations as basis; its operators are complex matrices on them, with the operator norm $\|\cdot\|$. A *word* w is a finite list of tokens, each carrying a skew-Hermitian generator G_t. Its *generator cost* is $g(w) = \sum_t \|G_t\|$ and its *drive action* is $\hbar\, g(w)$, for a constant $\hbar > 0$. Among the words are the *phase loops*, which act as the identity up to a phase and have any prescribed generator cost $\theta \geq 0$, and for channels $k \neq l$ the *entangling token*, whose generator is $\frac{\pi}{4}\mathrm{i}X$ with X an involution that sends each configuration to the configuration with channel l flipped, up to a sign.

2. A *holding* is a self-adjoint operator H. The *paid word* runs each token for a time $\varepsilon > 0$ under the holding plus its control: $W = U_{t_m} \cdots U_{t_1}$ with $U_t = \exp\bigl(-\frac{\mathrm{i}\varepsilon}{\hbar}\bigl(H + \frac{\hbar}{\varepsilon}(\mathrm{i}G_t + \|G_t\|)\bigr)\bigr)$. The *drift* of the word is $|w|\,\varepsilon\|H\|$, with $|w|$ the number of tokens.

3. The word makes a *certain change* when $|W_{vu}| = 1$ for some configurations $u \neq v$: the Born weight of moving u to v is 1.

4. A *comparison controller* with reference action $A_0 > 0$ selects a word $\mathrm{sel}(d, r)$ for records $d \geq 0$ and $r > 0$, such that pooled records drive the sum, $\hbar g(\mathrm{sel}(d+e, r)) = \hbar g(\mathrm{sel}(d, r)) + \hbar g(\mathrm{sel}(e, r))$, and a record read against itself drives the reference, $\hbar g(\mathrm{sel}(r, r)) = A_0$. Its *word at ratio* $x > 0$ is $w(x) = \mathrm{sel}(d(x), r(x))$, where $d(x)$ and $r(x)$ are the residual and reference records that Recognition Science's comparison apparatus stores at ratio x; they satisfy $d(x)/r(x) = J(x)$ (Chapter 22).

Each paid token is a phase times $\exp(Y_t)$ with $Y_t = G_t - \frac{\mathrm{i}\varepsilon}{\hbar}H$, a skew-Hermitian generator of norm at most $\|G_t\| + \varepsilon\|H\|/\hbar$. The scalar part $\frac{\hbar}{\varepsilon}\|G_t\|$ of the control is the baseline energy of the token, so the drive action is the time integral of the baseline.

Lemma 21.36 (The drive law and the step bound)**.** *In classical mathematics:*

1. *every comparison controller satisfies $\hbar g(\mathrm{sel}(d, r)) = A_0 d/r$, so its word at ratio x has drive action $A_0 J(x)$;*

2. *for every word, holding, $\hbar, \varepsilon > 0$ and configurations $u \neq v$,*

$$|W_{vu}| \leq \sin\Bigl(\min\Bigl(\frac{\pi}{2},\ g(w) + \frac{|w|\,\varepsilon\|H\|}{\hbar}\Bigr)\Bigr).$$

Proof. 1. Fix r. The function $f(d) = \hbar g(\mathrm{sel}(d, r))$ is nonnegative and additive on $[0, \infty)$, hence nondecreasing, and additivity gives $f(nd) = nf(d)$ for natural n, so $f(qr) = q\,f(r)$ for every rational $q \geq 0$. For $d \geq 0$ and rationals $0 \leq q \leq d/r \leq q'$, monotonicity gives $q\,f(r) \leq f(d) \leq q'\,f(r)$, and letting q and q' tend to d/r gives $f(d) = d\,f(r)/r$. Finally $f(r) = A_0$.

2. For a unit vector p and a skew-Hermitian Y with $\|Y\| = \theta \leq \pi/2$, write $e^Y = \sum_\lambda e^{\mathrm{i}\lambda} P_\lambda$ with spectral projections P_λ and $|\lambda| \leq \theta$; then the real part of $\langle p, e^Y p\rangle = \sum_\lambda e^{\mathrm{i}\lambda}\|P_\lambda p\|^2$ is at least $\cos\theta$. So the angle $d(p, q) = \arccos|\langle p, q\rangle|$ between unit vectors satisfies $d(p, U_t p) \leq \min(\pi/2, \|Y_t\|)$, phases being irrelevant. This angle obeys the triangle inequality: it is the Fubini–Study distance [102]. Hence $d(e_u, We_u) \leq \min(\pi/2, \sum_t \|Y_t\|)$. Since $e_v \perp e_u$ and $\|We_u\| = 1$, $|W_{vu}|^2 \leq 1 - |\langle e_u, We_u\rangle|^2 = \sin^2 d(e_u, We_u)$. □

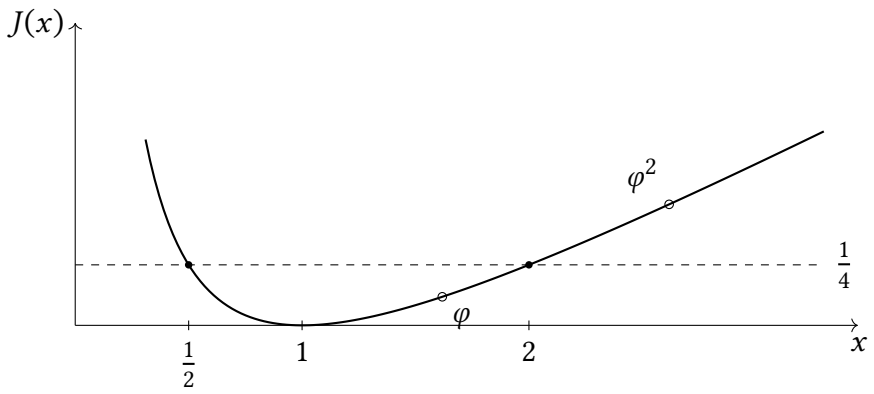


Figure 21.1. The cost $J(x) = \frac{1}{2}(x + x^{-1}) - 1$ and the quarter threshold. At zero holding, with reference action $A_0 = 2\pi\hbar$, a comparison can change a record with certainty exactly when $J(x) \geq \frac{1}{4}$, that is $x \leq \frac{1}{2}$ or $x \geq 2$ (filled dots). A golden step costs $J(\varphi) = \varphi - \frac{3}{2} \approx 0.118$ and falls short; two golden steps cost $\varphi^{-3} \approx 0.236$ and still fall short; $J(\varphi^2) = \frac{1}{2}$ (open dots).

Part 2 is a speed limit of the kind of Mandelstam and Tamm and of Margolus and Levitin [103, 104], in the form of a bound on generator norms along a word. Theorem 24.34 proves part 2 again for the registers of Section 24.9, with overlap angles in place of the Fubini–Study distance, and shows that the bound is attained.

Theorem 21.37 (A certain change needs $J \geq \frac{1}{4}$, less the drift)**.** *In classical mathematics: let a comparison controller have reference action $A_0 = 2\pi\hbar$, let H be a holding, $\varepsilon > 0$ and $x > 0$. If the paid word at ratio x makes a certain change, then*

$$J(x) \geq \frac{1}{4} - \frac{|w(x)|\,\varepsilon\|H\|}{A_0}.$$

Proof. A certain change gives $\sin(\min(\pi/2, \theta)) \geq 1$ in Lemma 21.36, so $\theta \geq \pi/2$, that is $\hbar g(w(x)) \geq \hbar\pi/2 - |w(x)|\varepsilon\|H\|$. By the drive law, $\hbar g(w(x)) = A_0 J(x)$, and $\hbar\pi/2 = A_0/4$. □

If the reference action is m cycles, $A_0 = 2\pi m\hbar$, the same proof gives the threshold $1/(4m)$.

Corollary 21.38 (The threshold, exactly)**.** *In classical mathematics, for comparison controllers with reference action $A_0 = 2\pi\hbar$, as in Theorem 21.37:*

1. *$J(x) \geq \frac{1}{4}$ exactly when $x \leq \frac{1}{2}$ or $x \geq 2$. Moreover $J(2) = \frac{1}{4}$, $J(\varphi) = \varphi - \frac{3}{2} < \frac{1}{4}$, $J(\varphi^2) = \frac{1}{2}$ and $2J(\varphi) = \varphi^{-3} < \frac{1}{4}$.*
2. *At zero holding on $n \geq 2$ channels, some controller's word at ratio x makes a certain change exactly when $x \leq \frac{1}{2}$ or $x \geq 2$. In particular 2 is the least ratio above 1 at which a certain change is possible, and it is attained.*
3. *If $\frac{1}{2} < x < 2$ and a controller's word at ratio x makes a certain change, its drift is positive.*
4. *At zero holding, the words at ratio φ of two controllers, run one after the other, never make a certain change.*

Proof. 1. $J(x) - \frac{1}{4} = (2x-1)(x-2)/(4x)$. Next $\varphi^{-1} = \varphi - 1$ gives $J(\varphi) = \varphi - \frac{3}{2}$, below $\frac{1}{4}$ since $\varphi < \frac{7}{4}$; and $\varphi^{-2} = 2 - \varphi$ gives $J(\varphi^2) = \frac{1}{2}$. Finally $2\varphi - 3 = \varphi^{-3}$, and $\varphi^{-3} = \sqrt{5} - 2 < \frac{1}{4}$ because $5 < \frac{81}{16}$.

2. The theorem at $H = 0$ gives one direction. For the other, let $J(x) \geq \frac{1}{4}$. The controller that selects a phase loop of generator cost $A_0 d/(\hbar r)$ for every pair of records obeys both laws. Change it at the records of ratio x only: select there two entangling tokens on channels $k \neq l$, followed by a phase loop of generator cost $2\pi J(x) - \pi/2 \geq 0$. The drive is unchanged, so both laws still hold. At zero holding the paid word is, up to a phase, $\exp(\frac{\pi}{2}\mathrm{i}X)$ times a phase, and $\exp(\frac{\pi}{2}\mathrm{i}X) = \mathrm{i}X$ because $X^2 = 1$; it moves every configuration to the one with channel l flipped. At $x = 2$ the phase loop has cost 0.

3. Here $J(x) < \frac{1}{4}$, so the theorem forces a positive drift.

4. The two words have generator cost $2A_0 J(\varphi)/\hbar = 2\pi\varphi^{-3} < \pi/2$, so by Lemma 21.36 every off-diagonal entry is below 1. □

Conditions and scope.

Base and premises

Classical complex matrices; the register, the paid word and the controller of Definition 21.35; the apparatus reading $d(x)/r(x) = J(x)$ of Chapter 22.

Conclusion

A certain change costs at least a quarter of the reference cycle, less the drift: $J(x) \geq \frac{1}{4} - |w|\varepsilon\|H\|/A_0$. At zero holding the threshold is exact: $x \notin (\frac{1}{2}, 2)$, attained at gain 2; a golden step, and even two, fall short.

Identification

Four identifications are not derived here: that the physical comparison words are selected by a controller from the stored records; that the reference action is one full cycle, $A_0 = 2\pi\hbar$, which needs a stop at the first return of the phase (with m cycles the threshold is $1/(4m)$), and which is exactly the statement that one comparison against the fundamental clock mode lasts eight ticks (Theorem 24.67); that the comparison and the register share one $\hbar$; and that "certain" means Born weight 1.

Not claimed

Certainty on Recognition Science's own register under its actual holding; Section 24.9 treats that question (Theorem 24.34 and Proposition 24.36).

The binary gain sits exactly at the threshold: comparing a quantity with its double is the cheapest comparison that can change a record with certainty when nothing else moves the register. The golden step, the forced ratio of the ledger, costs less than half of that, and two golden steps still fall short; inside the window $(\frac{1}{2}, 2)$, only the holding of the register can complete a certain change.

21.8 One binary distinction fixes the cost

Theorem 21.14 needs two premises: the composition law, and a calibration at the reference, the uniform expansion $C(t) = t^2/2 + o(t^2)$ with a modulus. The calibration is an analytic statement about infinitely many points near 0. Recognition Science also states one exact number: the cost of a single binary distinction, the comparison of a quantity with its double, is $J(2) = \frac{1}{4}$. Can that one number replace the calibration? This section proves that it can, as soon as the cost is monotone on $t \geq 0$, or has any curvature at all, or is bounded above on one interval, however short and wherever it sits. It also proves that the number alone does not suffice, and that the doubling slice of the law with the number does not either.

Throughout, $t = \log x$ is the logarithmic coordinate of Section 21.3, a cost is an arbitrary function G on $\mathbb{R}_\delta$, and the law is (21.2).

Definition 21.39 (The bit, a signed curvature and the branch-free cost)**.**

1. A cost G *charges the bit* when $G(\ell) = \frac{1}{4}$ for every point ℓ with $e^\ell = 2$.
2. For a point κ, G has *curvature* κ *at* 0 when for every k there is a j with
$$\left|G(t) - \frac{\kappa t^2}{2}\right| \leq \frac{t^2}{k+1} \qquad \text{whenever } |t| \leq \frac{1}{j+1}.$$
At $t = 0$ this gives $G(0) = 0$. Unlike Definition 21.10, the sign of κ is not prescribed.
3. For a point κ the *branch-free cost* is $B_\kappa(t) = \mathrm{C}(-\kappa t^2) - 1$, with $\mathrm{C}(s) = \sum_{j\geq 0}(-s)^j/(2j)!$ the series of Definition 21.6; so $B_\kappa(t) = \sum_{j\geq 1} \kappa^j t^{2j}/(2j)!$.

For $\kappa = c^2$ the series is that of $\cosh(ct) - 1$, and for $\kappa = -c^2$ it is $\cos(ct) - 1$. One formula covers both branches, and evaluating it decides nothing about the sign of κ.

Lemma 21.40 (The branch-free cost)**.** *For every point κ:*

1. *B_κ obeys the composition law, $B_\kappa(0) = 0$, and B_κ has curvature κ at 0;*
2. *a cost that obeys the law and has curvature κ at 0 equals B_κ at every point;*
3. *a function has at most one curvature at 0;*
4. *$B_{c^2}(t) = \cosh(ct) - 1$ and $B_{-c^2}(t) = \cos(ct) - 1$ for every point c.*

Every cost that obeys the law is even, and has $G(0) = 0$ or $G \equiv -1$. The proofs are free.

Proof. 1. The addition formulas of Lemma 21.7 hold for the pair of series $\mathrm{C}(bx^2)$, $x\mathrm{S}(bx^2)$ at every point b, with the same proof, since they are identities of power series in b and x. At (s, t) and $(s, -t)$ the sine terms cancel and $\mathrm{C}(b(s+t)^2) + \mathrm{C}(b(s-t)^2) = 2\mathrm{C}(bs^2)\mathrm{C}(bt^2)$; at $b = -\kappa$ this is d'Alembert's equation for $1 + B_\kappa$, which is the law for B_κ. For $|z| \leq 1$ the terms of $\mathrm{C}(z)$ of index $j \geq 2$ are at most $z^2/2^{j-1}$, so $|\mathrm{C}(z) - 1 + z/2| \leq z^2$, for either sign of z. With $z = -\kappa t^2$ and a natural number $M \geq |\kappa|$, $|B_\kappa(t) - \kappa t^2/2| \leq M^2 t^4$ for $|t| \leq 1/(M+1)$, which is at most $t^2/(k+1)$ once also $|t| \leq 1/(M(k+1)+1)$, since then $M^2 t^2 \leq 1/(k+1)$.

2. Both functions obey the duplication law (21.4), they agree to second order at 0 with a modulus, and $|B_\kappa(t)| \le (M+1)^2 t^2$ near 0; Theorem 21.13 applies.

3. If G has curvatures κ and κ', then at $t = 1/(m+1)$, for m beyond both moduli at k, $|\kappa - \kappa'|\, t^2/2 \le 2t^2/(k+1)$; so $|\kappa - \kappa'| \le 4/(k+1)$ for every k, and $\kappa = \kappa'$.

4. Compare the series term by term.

For the last sentence: the law at $(0,t)$ gives $G(t) + G(-t) = 2G(0)G(t) + 2G(0) + 2G(t)$. Theorem 21.29, part 1, gives $G(0) = 0$, and then $G(-t) = G(t)$, or $G \equiv -1$, which is even. □

Part 2 is Proposition 21.17 and Theorem 21.14 without the case split: the law and a curvature fix the cost as one formula in κ. Which of the two familiar shapes it has is a separate question, priced in Section 21.10.

One positive value fixes the sign of a solution everywhere. Write $g = 1 + G$, so that the law reads $g(s+t) + g(s-t) = 2g(s)g(t)$ and $G(0) = 0$ reads $g(0) = 1$.

Lemma 21.41 (One value fixes the sign)**.** *Let G obey the law with $G(0) = 0$, and let ℓ be a point. If $G(\ell) > 0$, then $G(t) \ge 0$ for every t; if $G(\ell) < 0$, then $G(t) \le 0$ for every t. No regularity is assumed. The proof is free.*

Proof. The law at (x,x) and at (y,y) gives $g(2x) = 2g(x)^2 - 1$, and at $(x+y, x-y)$ it gives $g(2x) + g(2y) = 2g(x+y)g(x-y)$; so $g(x+y)g(x-y) = g(x)^2 + g(y)^2 - 1$. With $A = g(x+\ell)$ and $B = g(x-\ell)$ we have $A + B = 2g(x)g(\ell)$ and $AB = g(x)^2 + g(\ell)^2 - 1$, so

$$\big(g(x+\ell) - g(\ell)g(x)\big)^2 = \Big(\frac{A-B}{2}\Big)^2 = \big(g(\ell)^2 - 1\big)\big(g(x)^2 - 1\big). \tag{21.8}$$

If $G(\ell) > 0$, then $g(\ell)^2 - 1 = G(\ell)(G(\ell)+2) > 0$, so $g(x)^2 \ge 1$ for every x, and $g(x) = 2g(x/2)^2 - 1 \ge 1$, that is $G(x) \ge 0$. If $G(\ell) < 0$, then $g(\ell/2)^2 - 1 = G(\ell)/2 < 0$, and (21.8) at $\ell/2$ gives $g(x)^2 \le 1$, so $g(x) \le 1$ and $G(x) \le 0$. □

Classically this is the dichotomy between the hyperbolic and the oscillating solutions of d'Alembert's equation, which needs no regularity [97]. On the line it needs no decision either: the hypothesis is a strict inequality at one point, and the conclusions are closed.

Theorem 21.42 (One binary distinction fixes the cost)**.** *Let G obey the composition law, have $G(0) = 0$ and charge the bit. Each of the following conditions gives $G(t) = \cosh t - 1$ for every point t:*

(a) G is nondecreasing on $t \ge 0$: $0 \le s \le t$ gives $G(s) \le G(t)$;

(b) G has some curvature κ at 0;

(c) G is bounded above on $[0,1]$: some point M has $G(t) \le M$ for $0 \le t \le 1$;

(d) G is bounded above on one interval $[a, a+d]$ with $d > 0$.

In ratio coordinates: a function F on the positive points with the multiplicative law, $F(1) = 0$ and $F(2) = \frac{1}{4}$, and with any one of the conditions read at $F(e^t)$, is J. No omniscience principle and no choice principle is used.

Proof. The bit. A logarithm ℓ of 2 exists (Proposition 21.5), and $\cosh \ell - 1 = \frac{1}{2}(2 + \frac{1}{2}) - 1 = \frac{1}{4}$. Also $\ell \geq \frac{1}{2}$, since $e^{1/2} \leq 1 + \frac{1}{2} + \frac{1}{4} < 2$ by Proposition 12.88, and the exponential reflects the order (Lemma 21.3, item 5).

The sign. Since $G(\ell) = \frac{1}{4} > 0$, Lemma 21.41 gives $G \geq 0$ everywhere.

The dyadic grid. The duplication law $G(2s) = 2G(s)^2 + 4G(s)$ has at most one nonnegative solution over each value: $2x^2 + 4x = 2y^2 + 4y$ with $x, y \geq 0$ gives $(x - y)(2(x + y) + 4) = 0$, and the second factor is at least 4. Starting from $G(\ell) = \cosh \ell - 1$, induction gives $G(\ell/2^n) = \cosh(\ell/2^n) - 1$ for every n, because both sides are nonnegative solutions over the same value. With $u = \ell/2^n$, the law at $((k + 1)u, u)$ gives the recursion

$$G((k + 2)u) = 2G((k + 1)u)G(u) + 2G((k + 1)u) + 2G(u) - G(ku),$$

which $\cosh - 1$ also obeys. So $G = \cosh - 1$ at every point $k\ell/2^n$ with natural k and n, and by evenness at their negatives.

Squeeze (a). Let $t \geq 0$, $t \leq N$, and $u = \ell/2^n$, so $u \geq 2^{-n-1}$. A finite search with cotransitivity of $<$ at $(k + 1)u < (k + 2)u$ finds j with $ju \leq t \leq (j + 2)u$; no comparison is decided exactly. Monotonicity gives $\cosh(ju) - 1 \leq G(t) \leq \cosh((j + 2)u) - 1$, and cosh is Lipschitz on $[-N - 1, N + 1]$ with a constant E (Lemma 21.3, item 4). So $|G(t) - (\cosh t - 1)| \leq 4Eu \leq 4E\ell/2^n$ for every n, and $G(t) = \cosh t - 1$. For a general t, equality of points is stable under double negation and $\neg\neg(t \geq 0 \vee t \leq 0)$ holds, so the case $t \leq 0$ follows by evenness.

Squeeze (c). Duplication with $G \geq 0$ gives $G(2s) \geq 4G(s)$, so $4^m G(s) \leq M$ whenever $0 \leq 2^m s \leq 1$. For $t \geq 0$ let $d = ju$ be a grid point with $s = t - d \in [0, 2u]$. With $g = 1 + G$, the identity (21.8) at the pair (d, s) reads

$$\big(g(t) - g(s)g(d)\big)^2 = \big(G(s)^2 + 2G(s)\big)\big(G(d)^2 + 2G(d)\big),$$

and $G(s) \leq 4M/4^n$; so $|G(t) - G(d)|$ is at most a constant, depending on M and a bound for t, times 2^{-n}. With the Lipschitz bound of cosh this gives $G(t) = \cosh t - 1$, and the sign of t is handled as in (a).

Curvature (b). Lemma 21.40 gives $G = B_\kappa$, so $B_\kappa(\ell) = \cosh \ell - 1 > 0$. The goal $\kappa = 1$ is stable, so we may split $\neg\neg(\kappa \geq 0 \vee \kappa \leq 0)$. If $\kappa \geq 0$, then $\kappa = c^2$ with $c \geq 0$ (Proposition 12.19), and $\cosh(c\ell) = \cosh \ell$. The function $\cosh - 1$ is strictly increasing on $[0, \infty)$, because $\cosh b - \cosh a = \frac{1}{2}(e^b - e^a)(1 - e^{-(a+b)})$ with both factors positive for $0 \leq a < b$; so it is one-to-one there, $c\ell = \ell$, and $c = 1$ since $\ell > 0$. If $\kappa \leq 0$, then $\kappa = -c^2$ and $B_\kappa(\ell) = \cos(c\ell) - 1 \leq 0$, which is impossible. So $\kappa = 1$ and $G = B_1 = \cosh - 1$.

A local bound (d). Put $c = a + d/2$. For $0 \leq s \leq d/2$ the law at (c, s) gives $2G(s) \leq 2G(s)(G(c) + 1) = G(c + s) + G(c - s) - 2G(c) \leq 2M$, using $G \geq 0$. So $G \leq M$ on $[0, d/2]$, and duplication bounds G on $[0, 2^k d/2]$ for every k. Taking k with $2^k d/2 \geq 1$ gives (c).

Ratio form. Proposition 21.46 below turns the multiplicative law for F into the law for $G(t) = F(e^t)$, and $F(2) = \frac{1}{4}$ is the bit, since every ℓ with $e^\ell = 2$ has $G(\ell) = F(2)$. □

Conditions and scope.

Base and premises
The line $\mathbb{R}_\delta$; a cost that is an arbitrary function with the composition law, $G(0) = 0$, the value $\frac{1}{4}$ at $\log 2$, and one of four mild conditions: monotone on $t \geq 0$, some curvature, a bound on $[0, 1]$, a bound on one interval.

Conclusion
$G = \cosh - 1$, that is $F = J$, with no omniscience principle and no choice principle. No continuity is assumed in (a), (c) or (d), and no value of the curvature in (b).

Identification
That the cost of one binary distinction is $J(2) = \frac{1}{4}$ is Recognition Science's calibration by one bit; the theorem derives the cost from it.

Not claimed
That the bit alone suffices: it does not (Proposition 21.43).

The four conditions cannot all be dropped, and the full law cannot be cut to its doubling slice. In classical mathematics $\mathbb{R}_\delta$ is a complete ordered field, so it has a reading onto the classical line (a bijective reading, which is unique); a classical function on the line is carried to $\mathbb{R}_\delta$ through it, and every premise and conclusion below is read through that bijection.

Proposition 21.43 (What one bit cannot do)**.** *In classical mathematics:*

1. *There is a function G on $\mathbb{R}_\delta$ that obeys the composition law, has $G(0) = 0$ and charges the bit, and differs from* $\cosh - 1$ *at some point.*

2. *There is a function G on $\mathbb{R}_\delta$ that obeys the doubling slice $G(2t) + G(0) = 2G(t)^2 + 4G(t)$ at every t, has $G(0) = 0$, is nondecreasing on $t \geq 0$ and charges the bit, and differs from* $\cosh - 1$ *at some point.*

3. *There is a function G on $\mathbb{R}_\delta$ that obeys the composition law, has $G(0) = 0$ and $|G| \leq 2$ everywhere, and has no curvature at* 0*. Hence the assertion "every cost that obeys the law, vanishes at* 0 *and is bounded on* $[-1, 1]$ *has a curvature" is false, and so is the assertion that for such costs naming the branch* $\cosh(ct) - 1$ *or* $\cos(ct) - 1$ *costs exactly LLPO.*

Proof. Work on the classical line and let $L = \log 2$. Let h be an additive map from the line to the rationals that is bounded on no neighbourhood of 0; it is built from a basis of the line over the rationals [202].

1. Put $a(x) = x + h(x) - (h(L)/L)x$. It is additive, $a(L) = L$, and it is unbounded near 0. The cost $G = \cosh \circ a - 1$ obeys the law, because $\cosh(u + v) + \cosh(u - v) = 2 \cosh u \cosh v$ and $a(s \pm t) = a(s) \pm a(t)$; it has $G(0) = 0$ and $G(L) = \frac{1}{4}$. At a point s with $|s| < 1$ and $|a(s)| > 2$, $G(s) > \cosh 2 - 1 > \cosh s - 1$.

2. Let $f(x) = (x + \lfloor x \rfloor)/2$, nondecreasing with $f(x + 1) = f(x) + 1$, $f(0) = 0$ and $f(\frac{1}{2}) = \frac{1}{4}$. For $t \neq 0$ put $x(t) = \log(|t|/L)/L$, so that $|t| = L\, 2^{x(t)}$ and $x(2t) = x(t) + 1$, and let $\psi(t) = L\, e^{Lf(x(t))}$, $\psi(0) = 0$. Then $\psi(2t) = 2\psi(t)$, ψ is nondecreasing on $t \geq 0$ and $\psi(L) = L$. The cost

$G = \cosh \circ \psi - 1$ obeys the slice, since $\cosh 2u = 2\cosh^2 u - 1$, is nondecreasing on $t \geq 0$ and has $G(L) = \frac{1}{4}$. At $t_0 = Le^{L/2}$, where $x(t_0) = \frac{1}{2}$, $\psi(t_0) = Le^{L/4} < t_0$, so $G(t_0) < \cosh t_0 - 1$.

3. Put $w = \cos \circ h - 1$. It obeys the law, $w(0) = 0$ and $|w| \leq 2$. Suppose it had a curvature κ; then $|w(s)| \leq (|\kappa|/2 + 1)s^2$ for small $|s|$. Take s_0 that small with $|h(s_0)| > 10$; some rational q with $0 < q < 1$ puts $qh(s_0) = h(qs_0)$ in $[\pi/2, 3\pi/2]$ up to sign, so $w(qs_0) \leq -1$, while $(|\kappa|/2 + 1)(qs_0)^2 < 1$. Both named branches have a curvature, by Lemma 21.40, so w is neither; and LLPO holds classically, so the second assertion would make w a branch. □

The bit fixes the cost only along the dyadic grid of $\log 2$ by itself; between the grid points the cost can be reshaped, octave by octave in part 2 and wildly in part 1. A mild regularity condition, of any of the four kinds of Theorem 21.42, closes the gaps. The calibration of Theorem 21.14 is one such condition, and the doubling slice with the calibration already gives J (Proposition 21.34, part 2).

The bit also selects the exponent within the families of Proposition 21.16, and on the line the selection is free.

Proposition 21.44 (The bit and the unit curvature select the exponent)**.** *Let $c > 0$ and $e^\ell = 2$. Then $\cosh(c\ell) - 1 = \frac{1}{4}$ exactly when $c = 1$, and $t \mapsto \cosh(ct) - 1$ is calibrated exactly when $c = 1$. Moreover, for every point t, $\cosh t - 1 = \frac{1}{4}$ exactly when $t = \ell$ or $t = -\ell$, and the disjunction is decided. The proofs are free.*

Proof. $\ell > 0$ and $\cosh - 1$ is one-to-one on $[0, \infty)$, as in the proof of Theorem 21.42(b), so $\cosh(c\ell) = \cosh \ell$ gives $c\ell = \ell$ and $c = 1$. The function $\cosh(ct) - 1$ has curvature c^2, and curvature is unique (Lemma 21.40), so it is calibrated exactly when $c^2 = 1$, that is $c = 1$. Finally $\cosh |t| = \cosh t = \cosh \ell$ gives $|t| = \ell$ (the sign of t enters only inside a negation, and equality is stable), so $|t| > 0$, t is apart from 0, and the sign of t is decided. □

21.9 Bounds force the curvature

The four conditions of Theorem 21.42 are regularity conditions, and they meet the bit. Without the bit, which bounds already force a curvature, and with it, by Lemma 21.40, the closed form B_κ? A bound of quadratic size near 0 does; so does any bound below $\frac{3}{2}$ near 0; and a sign with a bound does. The wild cosine of Proposition 21.43, part 3, bounded by 2, has no curvature.

Theorem 21.45 (Bounds near the reference give the curvature)**.** *Let G obey the composition law with $G(0) = 0$.*

1. *If $|G(t)| \leq Ct^2$ for $|t| \leq 1$, for a point C, then the dyadic estimates $b_m = 4^m G(2^{-m})$ satisfy $0 \leq b_m - b_{m+d} \leq C^2\, 4^{-m}$ for all m, d, they converge, and G has curvature $\kappa = 2 \lim_m b_m$ at 0.*
2. *If $|G(t)| \leq c$ for $|t| \leq 1/(j+1)$, for a natural number j and a point $c < \frac{3}{2}$, then G has a curvature at 0. In particular this holds when G is continuous at 0 with a modulus: for every k some j has $|G(t)| \leq 1/(k+1)$ for $|t| \leq 1/(j+1)$.*
3. *If $G \geq 0$ on $[-1, 1]$ and G is bounded on $[-1, 1]$, then there is a point $c \geq 0$ with $G(t) = \cosh(ct) - 1$ for $|t| \leq 1$, and G has curvature c^2 at 0.*

In cases 1 and 2, $G = B_\kappa$ at every point. No omniscience principle and no choice principle is used.

Proof. 1. With $u = 2^{-m-1}$ duplication gives $b_m - b_{m+1} = 2 \cdot 4^m G(u)^2 \geq 0$, and $|G(u)| \leq Cu^2$ bounds the step by $C^2\, 4^{-m}/8$; summing, $0 \leq b_m - b_{m+d} \leq C^2\, 4^{-m}$. The limit of a sequence of points exists without choice when representatives of the terms are computed from finitely many given ones, and here they are. Put $H = 1 + G$; if $C \leq 1$, then $H(v) \geq 0$ on $[-1, 1]$, and duplication $H(2v) = 2H(v)^2 - 1$ gives $H(v) = \sqrt{(1 + H(2v))/2}$, so $H(2^{-m})$ is the mth half-angle iterate of $H(1)$, and a regular sequence for b_m is computed from one for $H(1)$ by the operations of Proposition 12.19 and the square root. A general C is reduced to $C \leq 1$ by the rescaled cost $t \mapsto G(t/2^N)$, which obeys the law, with $N \geq C$; 4^N times its curvature is a curvature of G.

Let L be the limit. The rate passes to the limit, $|b_i - L| \leq C^2\, 4^{-i}$. At a dyadic point $q = n/2^i \leq 1$, the law along multiples of 2^{-i} and $n^2 G(2^{-i}) = q^2 b_i$ give $|G(q) - Lq^2| \leq C^2 q^4 + C^2 q^2\, 4^{-i}$. For $0 \leq t \leq 1$ choose a dyadic q with $q \leq t \leq q + 2^{1-i}$ by the search of Theorem 21.42; the law gives a Lipschitz bound for G on $[0, 1]$, so $|G(t) - Lt^2| \leq C^2 t^4 + K 2^{-i}$ for every i, and $|G(t) - Lt^2| \leq C^2 t^4$. Evenness extends this to $|t| \leq 1$. Given k, the right side is at most $t^2/(k+1)$ once $C^2 t^2 \leq 1/(k+1)$: the curvature is $2L$.

2. Rescaling as in part 1 makes the bound hold on $[-1, 1]$. First let $c = \frac{1}{2}$. With $g_m = |G(t/2^m)|$ for $|t| \leq 1$, duplication $G(2v) = G(v)(4 + 2G(v))$ and $|4 + 2G(v)| \geq 3$ give $g_{m+1} \leq g_m/2$, and by induction $g_m \leq 2^{-m-1}$ and $4^m g_m \leq 1 - 2^{-m-1}$. So $|G(u)| \leq 4^{-m}$ whenever $|u| \leq 2^{-m}$. A search over the scales 2^{-i}, using cotransitivity, finds for each k either an i with $2^{-i-2} \leq |u| \leq 2^{-i}$, where $|G(u)| \leq 4^{-i} \leq 16u^2$, or $|u| \leq 2^{-k}$, where $|G(u)| \leq 4^{-k}$; as k is arbitrary and the bound is closed, $|G(u)| \leq 16u^2$ on $[-1, 1]$, and part 1 applies. For $c < \frac{3}{2}$ one halving brings the bound to $\frac{1}{2}$. Put $y = G(t)$ and $z = G(2t) = 2y^2 + 4y$, both of size at most c. If $y > \frac{1}{2}$, then $z - \frac{5}{2} = 2(y - \frac{1}{2})(y + \frac{5}{2}) > 0$, so $|z| > \frac{5}{2} > c$; if $y < -\frac{1}{2}$, then $y > -\frac{3}{2}$ and $z + \frac{3}{2} = 2(y + \frac{1}{2})(y + \frac{3}{2}) < 0$, so $|z| > \frac{3}{2} > c$. Both are impossible, so $|y| \leq \frac{1}{2}$. Continuity at 0 with a modulus gives the bound $\frac{1}{2}$ near 0 at once. Finally Lemma 21.40 gives $G = B_\kappa$.

3. Duplication spreads $G \geq 0$ from $[-1, 1]$ to the line. Put $y = 1 + G(1) \geq 1$ and $r = \sqrt{y^2 - 1}$, and let $c = \log(y + r)$; then $e^{-c} = y - r$, so $\cosh c - 1 = G(1)$ and $c \geq 0$. The rate c is computed, not chosen. As in Theorem 21.42, halving keeps the unique nonnegative root of $2x^2 + 4x$, and the law is a recursion along multiples, so G and $\cosh(c\,\cdot) - 1$ agree at every $k/2^n$; the bounded squeeze of Theorem 21.42(c), on the grid of 1, gives $G(t) = \cosh(ct) - 1$ on $[0, 1]$, and evenness gives $[-1, 0]$. Its curvature is c^2. □

The bound $\frac{3}{2}$ in part 2 is where the argument stops: $1 + G = -\frac{1}{2}$, a third of a turn of the oscillating branch, is a fixed point of the doubling, $2(-\frac{1}{2})^2 - 1 = -\frac{1}{2}$. Whether a bound c with $\frac{3}{2} \leq c < 2$ near 0 gives a curvature on $\mathbb{R}_\delta$ with no omniscience principle is open; at $c = 2$ it fails, since the wild cosine is bounded by 2. Part 3 needs its sign premise: the wild cosine is never positive and has no curvature.

21.10 The prices of the branches

Lemma 21.40 names the solution with curvature κ without deciding anything: it is B_κ. The classical statement names it differently, as $\cosh(ct) - 1$ or as $\cos(ct) - 1$, and that naming is

a decision about the sign of κ. This section prices the decisions a reader of the cost asks for. The forcing theorems are free; the prices sit in the decisions, as in Chapter 22.

Proposition 21.46 (The ratio law is the logarithmic law)**.** *For every function F on $\mathbb{R}_{\delta}$: F obeys the multiplicative law $F(xy) + F(xz) = 2F(x)F(y) + 2F(x) + 2F(y)$ for all positive x, y with $yz = 1$ exactly when $t \mapsto F(e^t)$ obeys the composition law. The proof is free.*

Proof. Forward: $x = e^s$, $y = e^t$ give $xy = e^{s+t}$, and the inverse of a positive point is unique, so $xz = e^{s-t}$. Backward: put $s = \log x$ and $t = \log y$ (Proposition 21.5). □

So every result of this chapter stated in logarithmic coordinates holds in the ratio coordinates in which Recognition Science states its law, with no calibration and no price.

Theorem 21.47 (The prices of naming the branch)**.** *Each of the following assertions is equivalent to the principle named, with no choice principle in either direction.*

1. Naming the branch. *Every cost G that obeys the law, has $G(0) = 0$ and some curvature κ at 0 is $\cosh(ct) - 1$ for some point c, or $\cos(ct) - 1$ for some point c: LLPO.*
2. The flat decision. *Every such cost is identically 0 or not: WLPO.*
3. The three branches. *Every such cost is $\cosh(ct) - 1$ with $c > 0$, or identically 0, or $\cos(ct) - 1$ with $c > 0$: LPO.*
4. The sign decision with a curvature. *Every such cost is ≥ 0 everywhere or ≤ 0 everywhere: LLPO. Without the curvature premise, the sign decision for every cost that obeys the law with $G(0) = 0$ implies LLPO.*
5. Comparing two costs. *For all points s, t, $\cosh s \leq \cosh t$ or $\cosh t \leq \cosh s$: LLPO.*
6. The costly-join decision. *For every positive point x, $J(x) > 0$ or $J(x) = 0$: LPO.*

Proof. LLPO is the dichotomy $r \geq 0 \vee r \leq 0$ for every point, WLPO the decision of $r = 0$, and LPO the trichotomy (Corollary 12.18).

1. Given LLPO, decide $\kappa \geq 0$ or $\kappa \leq 0$. In the first case $\kappa = c^2$ and $G = B_{c^2} = \cosh(ct) - 1$; in the second $G = \cos(ct) - 1$ (Lemma 21.40). Conversely, for a point r apply the assertion to B_r, which has curvature r: a hyperbolic branch has curvature $c^2 \geq 0$, an oscillating branch $-c^2 \leq 0$, and the curvature is unique; so $r \geq 0$ or $r \leq 0$.

2. The flat cost has curvature 0, so $B_r \equiv 0$ exactly when $r = 0$; this gives WLPO from the assertion, and WLPO decides $\kappa = 0$.

3. Given LPO, decide $\kappa > 0$, $\kappa = 0$ or $\kappa < 0$; a positive point has a positive square root, so the three cases give the three branches. Conversely the branch of B_r decides $r > 0$, $r = 0$ or $r < 0$, because the three curvatures c^2 with $c > 0$, 0, and $-c^2$ with $c > 0$ are positive, zero and negative.

4. A cost ≥ 0 everywhere has curvature ≥ 0: if $\kappa < 0$, then $G(t) \leq \kappa t^2/4 < 0$ for small $t \neq 0$. Likewise ≤ 0 everywhere gives $\kappa \leq 0$. So the sign decision for B_r gives $r \geq 0$ or $r \leq 0$, which is LLPO, and this uses only costs B_r, which obey the law with $B_r(0) = 0$; that proves the last

sentence. Conversely LLPO decides the sign of κ, and $B_\kappa \geq 0$ when $\kappa \geq 0$ and $B_\kappa \leq 0$ when $\kappa \leq 0$.

5. LLPO orders the line, which gives the dichotomy. Conversely let a be a point. If $\cosh(a+1) \leq \cosh(a-1)$, then $a \leq 0$: for $a > 0$, $|a-1| < a+1$, and cosh is even and strictly increasing on $[0, \infty)$. If $\cosh(a-1) \leq \cosh(a+1)$, then $a \geq 0$ by the same argument. So every point has a weak sign: LLPO.

6. $J(x) = (x-1)^2/(2x)$ with $2x > 0$, so $J(x) > 0$ exactly when $x - 1$ is apart from 0, and $J(x) = 0$ exactly when $x = 1$. For a point r take $x = 1 + |r|$: the assertion gives $|r| \# 0$ or $r = 0$, that is $r > 0$, $r < 0$ or $r = 0$, the trichotomy, which is LPO. Conversely LPO decides $x - 1 > 0$, $x - 1 = 0$ or $x - 1 < 0$. □

Conditions and scope.

Base and premises
The line; costs given as functions with the law and, where stated, a curvature at 0 given as a point.

Conclusion
Naming the branch, deciding the sign of a cost that has a curvature, or ordering two costs: exactly LLPO. Deciding flatness: exactly WLPO. Deciding the three branches or whether a ratio costs anything: exactly LPO.

Not claimed
The converse of the last sentence of part 4, a sign decision for costs with no curvature premise, which is open. Recognition Science's own calibration gives the curvature as the number 1; for it every decision above is free.

21.11 Exact values of the cost

The cost J has a small set of exact properties that the later chapters and the ledger of Section 21.5 use. On the line they are all free.

Proposition 21.48 (The shape of the cost)**.** *For all points s, t, every natural number k, and all points $t_1, \dots, t_n$:*

1. Averaging never costs more: $\cosh \frac{s+t}{2} - 1 \leq \frac{1}{2}\big((\cosh s - 1) + (\cosh t - 1)\big)$.
2. Closed loops: *if $t_1 + \cdots + t_n = 0$, then $\sum_i (\cosh t_i - 1) \geq 0$, and the sum is 0 only when every $t_i = 0$.*
3. Quadratic bounds: $\frac{t^2}{2} \leq \cosh t - 1 \leq \frac{t^2}{2} \cosh t$.
4. Repeated comparisons: $\cosh(kt) = T_k(\cosh t)$, *where $T_0 = 1$, $T_1(y) = y$ and $T_{n+2}(y) = 2yT_{n+1}(y) - T_n(y)$ are Chebyshev's polynomials.*

The proofs are free.

Proof. 1. The law at $a = (s+t)/2$, $b = (s-t)/2$ gives $G(s) + G(t) = 2G(a)G(b) + 2G(a) + 2G(b)$ for $G = \cosh - 1$, so $\frac{1}{2}(G(s) + G(t)) - G(a) = G(a)G(b) + G(b) \geq 0$.

2. Each term is nonnegative and at most the sum. A term $\cosh t - 1 = 0$ forces $t = 0$: $\cosh - 1$ is even and one-to-one on $[0, \infty)$, and equality is stable, so the sign of t need not be decided. The hypothesis $t_1 + \cdots + t_n = 0$ is not used: both conclusions hold for all points $t_1, \ldots, t_n$, and the closed loop is the case the ledger meets.

3. Both bounds come from the series: $\cosh t - 1 = \sum_{j \geq 1} t^{2j}/(2j)!$, whose first term is $t^2/2$, and $t^{2j}/(2j)! \leq \frac{t^2}{2} \cdot t^{2j-2}/(2j-2)!$.

4. The law at $((n+1)t, t)$ is $\cosh((n+2)t) = 2\cosh t \cosh((n+1)t) - \cosh(nt)$, the Chebyshev recursion, and two-step induction finishes. □

Part 4 says that a comparison repeated k times costs a fixed polynomial of the cost of one comparison, $1 + J(x^k) = T_k(1 + J(x))$. Part 2 says that a closed loop of comparisons costs nothing only when every step is trivial.

Proposition 21.49 (Rational costs and the inverse of the cost)**.**

1. *For natural numbers $p, q \geq 1$ and a point ℓ with $e^\ell q = p$, $2pq(\cosh \ell - 1) = (p-q)^2$: the cost of the rational ratio p/q is the ratio $(p-q)^2/(2pq)$.*
2. *For every point $y \geq 0$ there is exactly one point $t \geq 0$ with $\cosh t - 1 = y$, and $x = e^t$ solves $x^2 + 1 = 2(1+y)x$.*

The proofs are free.

Proof. 1. $J(p/q) = \frac{1}{2}(p/q + q/p) - 1 = (p-q)^2/(2pq)$, with $e^{-\ell} = q/p$.

2. Put $r = \sqrt{(1+y)^2 - 1}$ and $t = \log(1 + y + r)$; then $e^{-t} = 1 + y - r$, so $\cosh t = 1 + y$, and $t \geq 0$ because $1 + y + r \geq 1$. Uniqueness is injectivity of $\cosh - 1$ on $[0, \infty)$, and $e^{2t} + 1 = 2\cosh t\, e^t$. □

The golden ratio enters through the ledger of Section 21.5. Its rungs have exact costs, and every additive ladder approaches it at a rate fixed by φ alone.

Theorem 21.50 (Golden ladders)**.** *Let $F(n)$ be the Fibonacci numbers and ℓ a point with $e^\ell = \varphi$.*

1. Even rungs. $2J(\varphi^{2n}) + 2 = F(2n-1) + F(2n+1)$ *for $n \geq 1$, a Lucas number.*
2. Odd rungs. $2J(\varphi^{2n+1}) + 2 = \sqrt{5}\, F(2n+1)$ *for $n \geq 0$.*
3. The attractor. *For all points a, b, the ladder $s_0 = a$, $s_1 = b$, $s_{n+2} = s_{n+1} + s_n$ satisfies $|s_{n+1} - \varphi s_n| = |b - \varphi a|\,(\varphi - 1)^n$.*
4. Record growth. *The binary words of length n with no two adjacent ones number $F(n+2)$, and $|F(n+1)/F(n) - \varphi| \leq 1/F(n)^2$ for $n \geq 1$.*
5. The record ladder. *The levels $s_n = F(n+2)$ are positive and have a costly initial join: $s_2 = s_0 + s_1$ with join ratio $s_1/s_0 = 2$, of cost $J(2) = \frac{1}{4} > 0$.*

Here $J(\varphi^m)$ means $\cosh(m\ell) - 1$. The proofs are free.

Proof. Every root x of $x^2 = x + 1$ has $x^{m+1} = F(m+1)x + F(m)$, by induction. Apply this to φ and $\psi = 1 - \varphi$, with $e^{-\ell} = \varphi^{-1} = \varphi - 1 = -\psi$.

1. $2J(\varphi^{2n}) + 2 = \varphi^{2n} + \psi^{2n} = F(2n)(\varphi + \psi) + 2F(2n-1) = F(2n) + 2F(2n-1) = F(2n-1) + F(2n+1)$.

2. For odd powers $(-\psi)^{2n+1} = -\psi^{2n+1}$, so $2J(\varphi^{2n+1}) + 2 = \varphi^{2n+1} - \psi^{2n+1} = F(2n+1)(\varphi - \psi) = \sqrt{5}\, F(2n+1)$, using $\varphi - \psi = 2\varphi - 1 = \sqrt{5}$.

3. The error $e_n = s_{n+1} - \varphi s_n$ has $e_{n+1} = s_{n+1} + s_n - \varphi s_{n+1} = (1 - \varphi)e_n$, because $\varphi(1 - \varphi) = -1$; and $|1 - \varphi| = \varphi - 1$.

4. Let c_n count the admissible words of length n and d_n those that stay admissible after a leading one. A leading zero changes nothing, and a leading one forces the next letter to be zero, so $c_{n+1} = c_n + d_n$ and $d_{n+1} = c_n$; with $c_0 = 1$, $c_1 = 2$ this is $c_n = F(n+2)$. For the ratio, part 3 at $(a, b) = (0, 1)$ gives $|F(n+1) - \varphi F(n)| = r^n$ with $r = \varphi - 1$, and $r^n F(n) \le 1$ by two-step induction, since $r^2 + r = 1$: $r^{n+2}F(n+2) = r^2(r^n F(n)) + r(r^{n+1}F(n+1)) \le 1$. Divide by $F(n)$.

5. $F(4) = 3 = 1 + 2 = F(2) + F(3)$, the parts 1 and 2 are distinct, and $J(2) = \frac{1}{4}$. □

Part 4 counts the records of a two-state register in which no two adjacent cells are both set, and their number grows by the golden ratio: the golden ratio is read off records by counting, with an explicit error. Part 5 is the premise of the costly join (Section 22.6) met by the record count itself, with apart parts, so it costs nothing. Part 3 says that every additive ladder, whatever its first two levels, is drawn to the golden ratio geometrically, with ratio $\varphi - 1 \approx 0.618$ per rung.

21.12 The cost on ratio records

Before the line, a cost can be stated on ratio records themselves (Chapter 4): a map F from records (u, d), with u a signed count and d a positive count, to records. A record is a display; the records $(3, 3)$ and $(1, 1)$ display the same value. Recognition Science states five premises for such a cost. The questions are which of them follow from the others, and which further premises force $F = J$ on values.

Definition 21.51 (The five premises on records)**.** Write $q \equiv q'$ when two records have the same value, $\bar{q}$ for the reduced record of q, 1 for the record $(1, 1)$, 0 for $(0, 1)$, and q^{-1} for the record that inverts q (with value 0 at value 0). A cost F on records has:

(F1) *value-invariance* when $F(q) \equiv F(\bar{q})$ for every record q;

(F2) *the law* when $F(xy) + F(x/y) \equiv 2F(x)F(y) + 2F(x) + 2F(y)$ for all records x, y of nonzero value;

(F3) *reciprocity* when $F(q) \equiv F(q^{-1})$ for every q;

(F4) *literal unit-zero* when $F(1) = 0$ as records;

(F5) *the calibration at two* when $F(2) \equiv J(2) = \frac{1}{4}$.

Further, F is *bounded near balance* when some ratio M bounds the value of $F(q)$ for every q with value in $[\frac{1}{2}, 2]$; *positive* when $F(q)$ has nonnegative value whenever q has positive value; and has the *signed unit* when $F(-1) \equiv J(-1) = -2$.

Theorem 21.52 (The ledger of the five premises)**.** *Proved over the classical rationals:*

1. *(F1), (F2) and (F5) give $F(1) \equiv 0$ and $F(q) \equiv F(q^{-1})$ for every record q of nonzero value, and with (F1) at value 0 they give (F3); (F4) is not needed. So reciprocity is redundant, and so is the value of unit-zero.*
2. *Each of (F1), (F2), (F4) and (F5) is independent of the other four premises: for each there is a cost with the other four and not it.*
3. *(F1)–(F5) with a bound near balance do not give $F(q) \equiv J(q)$ at every nonzero value.*
4. *(F1)–(F5) with a bound near balance, positivity and the signed unit give $F(q) \equiv J(q)$ at every record of nonzero value.*
5. *None of the three premises added in part 4 can be dropped: for each there is a cost with (F1)–(F5) and the other two, not equal to J.*

Proof. 1. By (F1) the cost depends only on values, since records of equal value have the same reduced record. The law at $(2, 1)$ reads $F(2) + F(2) \equiv 2F(2)F(1) + 2F(2) + 2F(1)$; with $F(2) \equiv \frac{1}{4}$ this is $\frac{5}{2}F(1) \equiv 0$. The law at $(1, q)$ then reads $F(q) + F(q^{-1}) \equiv 2F(q)$ for $q \not\equiv 0$; at value 0 the inverse record has value 0, and (F1) suffices.

2. In parts 2 and 3 the formulas are written for records of nonzero value. At the records of value 0 each cost below is constant, given by its formula with 0^{-1} read as 0; the law asks nothing there, and a constant meets (F1) and (F3).

(F1): for a record with numerator value N and denominator d put $c_1 = q \cdot 9^{v_3(d)}$ and $c_2 = q^{-1} \cdot 9^{v_3(|N|)}$, with v_3 the exponent of 3, and let F return the literal 0 at the literal unit and a record of $(c_1 + c_2)/2 - 1$ elsewhere. Multiplying records multiplies c_1 and c_2, and inversion swaps them, so $G = F + 1$ solves $G(xy) + G(x/y) = 2G(x)G(y)$; $F(2) \equiv \frac{1}{4}$; but $F(3/3) \equiv 8$ while $F(1/1) = 0$. *(F2):* J with its values at 3 and $\frac{1}{3}$ set to 0 fails the law at $(3, 3)$, where $J(9) = \frac{32}{9}$. *(F4):* J returning a different record of value 0 at the unit. *(F5):* the zero cost.

3. Let $\chi(v) = v(-1)^{v_3(v)}$, a multiplicative map on nonzero rationals with $\chi(1/v) = 1/\chi(v)$ and $\chi(2) = 2$. The cost $F(v) = \frac{1}{2}(\chi(v) + \chi(v)^{-1}) - 1$ meets (F1)–(F5) and $F \leq 1$ on $[\frac{1}{2}, 2]$, but $F(3) = -\frac{8}{3}$ while $J(3) = \frac{2}{3}$.

4. By part 1, $G = 1 + F$ is a function of values on the group of nonzero rationals with $G(1) = 1$, and the law is d'Alembert's equation $G(xy) + G(x/y) = 2G(x)G(y)$ there. On an abelian group every such solution is $G = \frac{1}{2}(\chi + \chi^{-1})$ for a multiplicative map χ into the nonzero complex numbers [203]. For $q > 0$ positivity gives $G(q) \geq 1$, so $\chi(q)$ is a root of $x^2 - 2G(q)x + 1$, which has two positive roots; so $\chi(q) > 0$. The calibration gives $\chi(2) \in \{2, \frac{1}{2}\}$, and $1/\chi$ gives the same G, so we may take $\chi(2) = 2$. For $q \geq 1$ and each natural number b choose a with $2^a \leq q^b < 2^{a+1}$; the ratio $r = q^b/2^a$ lies in $[1, 2)$, so $J(\chi(r)) = F(r) \leq M$, and $\chi(r) = \chi(q)^b/2^a$ lies in a fixed interval $[1/C, C]$ with C computed from M. Hence $(\chi(q)/q)^b$ lies in $[1/(2C), 2C]$

for every b, and Bernoulli's inequality forces $\chi(q) = q$. For $0 < q < 1$ use $\chi(1/q) = 1/\chi(q)$. Finally $\chi(-1)^2 = \chi(1) = 1$, and the signed unit $G(-1) = -1$ gives $\chi(-1) = -1$, so $\chi(q) = q$ for $q < 0$ too, and $F = J$ on nonzero values.

5. *The bound:* $\chi_3(v) = v \cdot 9^{-v_3(v)}$ is positive on positive values and has $\chi_3(-1) = -1$; its cost exceeds $9^b/2 - 1$ at $3^b/(3^b + 1) \in [\frac{1}{2}, 1]$, and $F(6) = J(\frac{2}{3}) = \frac{1}{12}$ while $J(6) = \frac{25}{12}$. *Positivity:* the cost of part 3. *The signed unit:* $F = J \circ |\cdot|$, with $F(-1) = 0$. □

Conditions and scope.

Base and premises

Ratio records and their values, read over the classical rationals; the five premises of Definition 21.51.

Conclusion

Reciprocity and the value of unit-zero follow from value-invariance, the law and the calibration at two. Value-invariance, the law, the literal unit-zero and the calibration at two are each independent: in particular "a cost reads only the value of a ratio" is a premise in its own right, because the law and the calibration allow a cost that depends on how a ratio is written. On values, a bound near balance, positivity and the signed unit together force J, and each is needed.

Not claimed

A price on $\mathbb{R}_\delta$. On the line, Theorems 21.14 and 21.42 force J from the law with the calibration (Theorem 21.14), or with the bit and any one of the four conditions of Theorem 21.42, for free.

Physically read, a cost that obeys the composition law and charges a binary distinction $\frac{1}{4}$ already charges a ratio the same as its inverse, and charges the unit record $(1, 1)$ nothing. The law and the calibration alone allow a cost that charges the record $(3, 3)$, of the same value, 8 (Theorem 21.52, part 2). What such a cost does not do on its own is ignore how a ratio is written: that is the separate premise that a cost reads values.

Questions to work through

1. The cost $F(x) = \frac{1}{2}(x - 1)^2$ is calibrated at 1. Show that it violates the multiplicative law, and find which hypothesis of Theorem 21.13 fails for $C(t) = \frac{1}{2}(e^t - 1)^2$ when it is compared with $g = \cosh - 1$. (Theorem 21.14.)

2. Let $\lambda = \frac{2}{3}$, so that $\kappa = 3$ in Proposition 21.16. Compute both sides of the composition law for $\lambda(\cosh - 1)$ at $s = t = 1$ and show directly that they differ.

3. Let $\alpha = \arctan\frac{1}{5}$ and $\beta = \arctan\frac{1}{239}$. Using part 1 of Theorem 21.8 and the addition formulas, show that $\cos(4\alpha - \beta) = \sin(4\alpha - \beta)$, and conclude Machin's identity $\pi_\delta = 4(4\alpha - \beta)$. (Theorem 21.8.)

4. A finitely supported law with weights summing to 1 has a cost C that obeys the composition law and satisfies $C(2) = J(4) = \frac{9}{8}$. Find the law. (Theorem 21.22.)

5. Write out the first thirteen letters of the recognition history. Check the bound $-2/\varphi < N\varphi^{-3} - b_N < 2/\varphi^2$ for $N \le 13$, and the rule of part (e) for each of the thirteen postings. (Theorem 21.26.)

6. For a point c, show without Theorem 21.14 that the coded cost of the constant function c obeys the composition law at $s = t = 2$ and at $s = t = 3$ only when $c = 0$. (Theorem 21.29, part 2.)

Chapter 22

The forcing chain of Recognition Science, priced

Recognition Science derives physical structure as a chain of forced steps. Records of repeated distinction give arithmetic. Comparing two records gives the cost J of Chapter 21. A costly join of levels forces the golden ratio. A kept, linked pair forces three dimensions and eight binary states, and from there the chain reaches the constants of nature. Its central steps are uniqueness results: given their premises, nothing else is possible. This chapter asks a sharper question of every step. On the line $\mathbb{R}_\delta$, where every number is a construction from ratios, what does the step cost in logical strength? Does it hold outright, or does it need an omniscience principle, and which one?

Here is a small case. For $x > 0$ the cost of the ratio x is

$$J(x) = \frac{1}{2}\Big(x + \frac{1}{x}\Big) - 1 = \frac{(x-1)^2}{2x}.$$

One step of the chain, row T2.8 of Table 22.1, asserts that a ratio other than 1 costs something. If x is apart from 1, so that $|x - 1|$ exceeds some positive ratio, then $(x-1)^2 > 0$ and $2x > 0$, and $J(x) > 0$ with no principle at all. But "x is other than 1" can also be read as $x \neq 1$, the bare refutation of $x = 1$, and that reading is weaker. Take a binary sequence α and the point j_α of its positive jump presentation (Chapter 12), and put $x = 1 + j_\alpha$. Then $x \neq 1$ says only that α is not identically 0, while $J(x) > 0$ says that j_α is apart from 0, which happens exactly when α has a 1 at some index. So the uniform statement "every $x \neq 1$ has $J(x) > 0$" turns every refutation of "α is all zero" into an index where α is 1. That is Markov's principle, and Theorem 22.16 proves the converse: the statement is equivalent to it. The fact itself is free; the whole price lies in reading "other than 1" as a negation.

The same pattern runs through the whole chain, and it gives the chapter its main result. Column 2 of Table 22.1 restates the steps of Recognition Science's forcing chain in this book's own words, in two forms: the statements of its rungs, and one numbered sequence of steps, from T−2, "nothing is not something", to T41, the count of modes, with a bound on the period of a walk through the eight states. Every price in this chapter is a theorem about that restatement. Every step of the numbered sequence is free on the line (Theorem 22.45). Table 22.1 at the end of the chapter prices both forms in 130 rows: the first rung (segments T1–T4: the arithmetic of records, the cost J, the golden hierarchy, three dimensions and eight states), the 43 steps of the numbered sequence that are not already rows, the second rung (thirty-five groups of consequences of the first, one of them in two forms) and the chain of fifteen constants. Of the 130 rows, 119 are free. Ten have an exact price, one lies below a named principle, and none is open. All eleven priced rows belong to the rungs; none is a step of the numbered sequence. Every price sits in one of three places: in a decision form of a fact whose apartness or double-negation form is free; in a real-valued type of object that a step asks for, or in a countermodel to a dropped premise; or in a reading of a premise that is weaker than apartness. In each case a free reading of the same step is proved beside the priced one.

Several conventions hold throughout. We write MP for the binary form $\mathrm{MP}_{\mathrm{bin}}$ of Markov's principle (Definition 7.8); WLPO, LPO and LLPO are as there. A statement is *free* when it is proved with no omniscience principle and no choice principle. Maps are rules given as data, and from "every input has exactly one output" we do not form a map (Section 21.1); *the LPO function* and *the WLPO function* are the data forms of Definition 21.1: one map d from binary sequences to $\{0, 1\}$ with $d(\alpha) = 1$ exactly when α has a 1, respectively exactly when α is all 0. For x apart from 0, written $x \# 0$, $1/x$ is its inverse (Proposition 12.19, item 3); $x < y$, $x \le y$ and $x \neq y$ are as in Chapter 12, so $x \le y$ means $\neg(y < x)$ and $x = y$ means $\neg(x \# y)$. A ratio compared with a point stands for its point. The cost J is that of Definition 21.11, the golden ratio $\varphi = (1 + \sqrt{5})/2$ that of Definition 21.18, and log and π are those of Section 21.2.

The jump point carries most of the converse directions. For a binary sequence α, j_α is the point of the presentation (12.3) with every sign $+1$. It is never negative; it is 0 exactly when α is all 0 (Lemma 12.5); and it is positive exactly when α has a 1, as shown with Table 12.2. Two further facts are used without comment: a product apart from 0 has both factors apart from 0 (the product is invertible, so each factor is), and two points that are within $1/(k + 1)$ of each other for every k are equal (Lemma 12.52, item 2, in both directions).

22.1 Steps, native statements and prices

Over the classical real numbers every proposition about a number is true or false, and a step either follows from its premises or does not; it has no price. To price a step we read its statement on $\mathbb{R}_\delta$. Real numbers become points of the line; a quotient a/b becomes a times the inverse of a point b apart from 0; $x \neq y$ keeps its meaning, the negation of $x = y$; and a decision "A or B" stays a decision. We call the result the *native statement* of the step. It is the only form of the step whose price has content.

Definition 22.1 (The price of a step)**.** Let S be the native statement of a step. S is *free* when it is proved with no omniscience principle and no choice principle. S has *exact price* P, for a named principle P, when $S \Leftrightarrow P$ is proved with no choice principle in either direction. S lies *between* P and Q when $S \Rightarrow P$ and $Q \Rightarrow S$ are proved with no choice principle; it lies *below* Q when only $Q \Rightarrow S$ is proved. In the last two cases we say S is *bracketed*. A step is *open* when none of these is proved. "With no choice principle" concerns the proofs of these implications, not the principles they connect: a named principle in a price may itself include a choice principle.

The rows fall into seven segments. T1 is the arithmetic of records, eight steps that are this book's own beginning. T2 derives the cost J from an instrument that reads energies against a reference, in twelve steps. T3 derives golden scaling from a costly join, in ten. T4 obtains eight binary states from three dimensions and three dimensions from a kept linked pair, in six. These four segments form the first rung. The second rung collects thirty-five groups of consequences of the first rung: records and information, the geometry of the cost, composition and balancing, the divergence that the cost defines, the golden hierarchy, the three-bit carrier and persistent topology. One group, the first homology of the circle, has a relation form and a data form, and the table gives each its own row. Another segment runs from φ to the constants of nature in fifteen steps. The seventh segment is the numbered sequence, T−2 to T41 and the period bound. Two of its steps, the uniqueness of J and of φ, are rows of the first rung and of

the constants already; the other 43 have their own rows, and Section 22.10 reads them one by one. The table lists the numbered sequence after the first rung.

The first segment shows the contrast that runs through the chapter. Records compare for free; points of the line do not.

Proposition 22.2 (Records compare for free; points do not)**.** *Equality of counts, of signed counts and of ratio records is decidable, with no omniscience principle and no choice principle. The assertion that $x = y$ or $x \neq y$ for all points x, y of $\mathbb{R}_\delta$ is equivalent to WLPO.*

Proof. Two counts are compared by removing one successor from each until one of them runs out (Chapter 3). Signed counts are balanced pairs, and (a, b) and (c, d) name the same signed count when $a + d = b + c$, a comparison of counts; ratio records (u, d) and (v, e) with positive denominators are equal when $ue = vd$, a comparison of signed counts (Chapter 4). For points, $x = y$ says $x - y = \eta(0)$, so the second assertion is the first item of Corollary 12.18. □

The core of T1 is proved in Chapters 1 to 4: the orbit of repeated distinction is initial (Theorem 2.3); counts, signed counts and ratios obey the laws of the natural numbers, the integers and the rationals (Theorems 3.3, 3.10, 4.5 and 4.12); and the reading of this arithmetic does not depend on the carrier that realizes it (Theorems 2.7 and 3.14). The remaining steps code hereditarily finite sets by Ackermann's numbering, under which the arithmetic of records interprets the three axioms of adjunctive set theory with extensionality (extensionality, the empty set, and adjunction $x \cup \{y\}$), and compare two carriers of the records, the natural numbers and the lists over a one-element set, formula by formula. All eight steps are free.

22.2 The meter reads ratios

The derivation of the cost begins with an instrument. It reads a nonnegative quantity d, an energy, against a reference r. It pools: reading $d + e$ gives the reading of d plus the reading of e. It never reads below zero, and it reads its own reference as 1. On $\mathbb{R}_\delta$ a reference must be positive for "reads as 1" to make sense, and we build that into the meter. The theorem behind the meter is Cauchy's functional equation [171], which on the line needs monotonicity rather than continuity.

Lemma 22.3 (Monotone additive maps of the line are linear)**.** *Let $f : \mathbb{R}_\delta \to \mathbb{R}_\delta$ satisfy $f(x + y) = f(x) + f(y)$ for all x, y, and $f(x) \leq f(y)$ whenever $x \leq y$. Then $f(x) = f(1)\,x$ for every x. The proof is free.*

Proof. Additivity gives $f(0) = f(0) + f(0)$, so $f(0) = 0$, then $f(-x) = -f(x)$, and by induction $f(nx) = nf(x)$ for natural numbers n. For a ratio a/b with $b > 0$ this gives $bf(a/b) = f(a) = af(1)$, so $f(a/b) = (a/b)f(1)$. Since $0 \leq 1$, monotonicity gives $f(1) \geq f(0) = 0$.

Fix a point x and a natural number k. By Lemma 12.52, item 4, there are ratios $a < b$ with $a < x < b$ and $b - a \leq 1/(k + 1)$. Monotonicity gives $af(1) \leq f(x) \leq bf(1)$, and $f(1) \geq 0$ gives $af(1) \leq xf(1) \leq bf(1)$. So $f(x)$ and $f(1)x$ both lie in an interval of length $(b - a)f(1) \leq f(1)/(k + 1)$. Take a natural number $M \geq 1$ above $f(1)$ and apply this with $M(k + 1)$ in place of k: the two points are within $1/(k + 1)$, for every k, hence equal. □

Definition 22.4 (A meter)**.** A *meter* assigns to every positive point r (the reference) and every point $d \geq 0$ a point $\mathrm{read}(d, r)$, so that for each $r > 0$: $\mathrm{read}(d, r) \geq 0$; $\mathrm{read}(d + e, r) = \mathrm{read}(d, r) + \mathrm{read}(e, r)$ for $d, e \geq 0$; and $\mathrm{read}(r, r) = 1$. The *ratio meter* is $\mathrm{read}(d, r) = d/r$.

Theorem 22.5 (A meter reads the ratio to its reference)**.** *Every meter satisfies* $\mathrm{read}(d, r) = d/r$ *for every* $r > 0$ *and every* $d \geq 0$*, and the ratio meter is a meter. The proof is free.*

Proof. The ratio meter meets the three laws by the field laws of Proposition 12.19. Conversely fix $r > 0$ and write $\mathrm{read}(d)$ for $\mathrm{read}(d, r)$. Pooling gives $\mathrm{read}(0) = 0$. Put $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$, so that $x^+, x^- \geq 0$ and $x^+ - x^- = x$, and extend the reading to the whole line by

$$E(x) = \mathrm{read}(x^+) - \mathrm{read}(x^-).$$

E is additive. Indeed $(x + y)^+ + x^- + y^- = (x + y)^- + x^+ + y^+$, because $(x + y)^+ - (x + y)^- = x + y = (x^+ - x^-) + (y^+ - y^-)$; both sides are sums of nonnegative points, so reading them and pooling gives $E(x + y) = E(x) + E(y)$. E is monotone: if $x \leq y$ then $(y - x)^- = 0$, so $E(y) - E(x) = E(y - x) = \mathrm{read}((y - x)^+) \geq 0$. Lemma 22.3 gives $E(x) = E(1)x$. For $d \geq 0$ we have $d^+ = d$ and $d^- = 0$, so $\mathrm{read}(d) = E(d) = E(1)d$. Self-reference gives $E(1)r = 1$, so $E(1) = 1/r$. □

An instrument that pools energies and reads its own reference as 1 can only read the ratio to that reference. Row T2.1b of Table 22.1 states the meter differently: as one real-valued function of two real arguments, defined also at references that are not positive. That type of object has a price.

Theorem 22.6 (A total meter costs at least LPO)**.** *Call a map* $\mathrm{read} : \mathbb{R}_\delta \times \mathbb{R}_\delta \to \mathbb{R}_\delta$ *a* total meter *when, for every* $r > 0$*, its restriction to* $d \geq 0$ *satisfies the laws of Definition 22.4. Then:*

(a) a total meter exists if and only if some map $g : \mathbb{R}_\delta \to \mathbb{R}_\delta$ *satisfies* $r\,g(r) = 1$ *for every* $r > 0$*;*

(b) such a map g *gives LPO;*

(c) the LPO function gives such a map g*.*

So the existence of a total meter lies between LPO and the LPO function; Theorem 22.63 shows that it is exactly the LPO function. No choice principle is used.

Proof. (a) Given a total meter, put $g(r) = \mathrm{read}(1, r)$; for $r > 0$, Theorem 22.5 gives $g(r) = 1/r$. Given g, put $\mathrm{read}(d, r) = d\,g(r)$; for $r > 0$ this is the ratio meter.

(b) Let α be a binary sequence, $j = j_\alpha$ and $t = j\,g(j)$. Since $1/3 < 2/3$, cotransitivity gives $1/3 < t$ or $t < 2/3$. If $1/3 < t$, then $t \# 0$, so the factor j is apart from 0; it is not negative, so $j > 0$, and α has a 1. If $t < 2/3$, then $j > 0$ is impossible, since it would give $t = 1$; so α has no 1, and each $\alpha(n)$, being binary, is 0. That is LPO.

(c) Let D be the LPO function. For a regular sequence s let $p_s(n) = 1$ exactly when $s_n > 2/(n+1)$. The point named by s is positive exactly when p_s has a 1, and this belongs to the point, not to the sequence; so $D(p_s)$ depends only on the point r. Put $g(r) = 1/r$ if $D(p_s) = 1$, which then supplies $r > 0$, and $g(r) = 0$ otherwise. Two sequences naming r give the same bit and, in the first case, the same inverse; so g is a map, and $r\,g(r) = 1$ whenever $r > 0$. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; a total meter is one map on all pairs of points, with the meter laws at positive references.

Conclusion

A total meter gives LPO, and the LPO function gives one; the meter at positive references alone is free (Theorem 22.5).

Not claimed

The exact place of the total meter, which Theorem 22.63 settles: the LPO function. The derivation of J in this chapter reads only positive references, so it never needs the total type.

22.3 The apparatus and the native reporting theorem

The meter reads the energy that a comparison discards, against the energy of a reference. Which energies can an apparatus produce? A selection theorem answers: three premises about the balanced comparison and the empty operand force the energy law. The meter then forces the reported cost to be J, up to a scale. We state the apparatus on $\mathbb{R}_\delta$.

Definition 22.7 (The comparison apparatus)**.** The *complex numbers* $\mathbb{C}_\delta$ of the line are the pairs (x, y) of points, written $x + iy$, with the usual sum and product; $\mathrm{Re}(x + iy) = x$ and $|x + iy|^2 = x^2 + y^2$. A *signal* is a map f from $\{0, 1, \dots, 7\}$ to $\mathbb{C}_\delta$, and its *energy* is $\mathscr{E}(f) = \sum_{t<8} |f(t)|^2$. A *residual map* is a $\mathbb{C}_\delta$-linear map T of signals. Given two *operand pulses* u, v, the comparison of magnitudes $A, B \in \mathbb{R}_\delta$ at a phase z with $|z|^2 = 1$ is the signal $Au + zBv$, and its *discarded energy* is

$$\varepsilon(A, B, z) = \mathscr{E}\big(T(Au + zBv)\big).$$

Reference programs are built from records $\langle z\rangle$, one for each phase z with $|z|^2 = 1$, by sums and by multiplication with points. A program p *evaluates* to $p(A, B)$, with $\langle z\rangle(A, B) = \varepsilon(A, B, z)$, $(p + p')(A, B) = p(A, B) + p'(A, B)$ and $(c\,p)(A, B) = c\,p(A, B)$. Its *same weight* $\sigma(p)$ and *cross weight* $\chi(p)$ are additive and homogeneous in the same way, with $\sigma\langle z\rangle = 1/2$ and $\chi\langle z\rangle = -\mathrm{Re}(z)/2$.

Theorem 22.8 (The selection premises force the energy law)**.** *Let T, u, v and a reference program p satisfy*

$$\varepsilon(1, 1, 1) = 0, \qquad p(1, 0) = 0, \qquad p(1, 1) > 0.$$

Then $Tv = -Tu$, $\sigma(p) = 0$, the scale $k = \mathscr{E}(Tu)$ is positive, the gauge $\kappa = 4\chi(p)$ is positive, and for all points A, B

$$\varepsilon(A, B, 1) = k(A - B)^2, \qquad p(A, B) = \kappa\, k\, AB.$$

The proof is free.

Proof. A signal of energy 0 is 0: a sum of nonnegative points that is 0 has every term 0, and $x^2 + y^2 = 0$ gives $x^2 = y^2 = 0$, hence $x = y = 0$, since $x \# 0$ would make $x^2 > 0$. At $A = B = z = 1$ the comparison is $u + v$, so $T(u + v) = 0$, and linearity gives $Tv = -Tu$. Then $T(Au + zBv) = (A - zB)Tu$, and for $|z|^2 = 1$

$$\varepsilon(A, B, z) = |A - zB|^2 k = k\big(A^2 + B^2 - 2AB\,\mathrm{Re}(z)\big).$$

For a record this equals $2k\big(\frac{1}{2}(A^2 + B^2) + 2 \cdot (-\frac{1}{2}\mathrm{Re}(z))AB\big)$, and sums and multiples keep the form, so every program satisfies

$$p(A, B) = 2k\big(\sigma(p)(A^2 + B^2) + 2\chi(p)AB\big).$$

At $(1, 1)$ this is $2k(2\sigma(p) + 2\chi(p)) > 0$, so the factor $2k$ is apart from 0; since $k \geq 0$, $k > 0$. At $(1, 0)$ it is $2k\sigma(p) = 0$, and $2k$ is invertible, so $\sigma(p) = 0$. Then $p(1, 1) = 4k\chi(p)$, so $4\chi(p) = p(1, 1)/k > 0$, and $p(A, B) = 4\chi(p)kAB$. Finally $\varepsilon(A, B, 1) = k(A^2 + B^2 - 2AB) = k(A - B)^2$. □

No step argues by contradiction on the sign of k: k is a factor of a positive point, and that makes it apart from 0.

Remark 22.9 (A source witness)**.** Rows T2.3a and T2.3b of Table 22.1 restate a definite apparatus, the source witness. Its *pulse* with amplitudes a, b is the signal $(a, -a, 0, 0, -b, b, 0, 0)$; its residual map sends f to $t \mapsto (f(t) + f(t + 4))/2$, indices modulo 8; the operands are $u = \mathrm{pulse}(s, 0)$ and $v = \mathrm{pulse}(0, s)$ for an amplitude s with $|s|^2 > 0$ (the witness takes $s = 1/\sqrt{2}$); and the reference program is $(\kappa/2)(\langle i\rangle - \langle 1\rangle)$ with a gauge $\kappa > 0$. Then $Tu = (s, -s, 0, 0, s, -s, 0, 0)/2 = -Tv$, $k = |s|^2$, $\sigma = 0$, $\chi = \kappa/4$, $p(1, 0) = 0$ and $p(1, 1) = \kappa k$, so the three premises of Theorem 22.8 hold. The formation residual of row T2.3b, built with φ, acts on these operands as $1 + \varphi^{-2}$ times this residual, on $\mathbb{R}_\delta$ as over the classical complex numbers. Reading the classical apparatus of the two rows as the one above is an identification, made in classical mathematics; it is not part of the native statement.

Definition 22.10 (Reports at ratios)**.** Let $k, \kappa > 0$. The *apparatus law* at counts is $E(a, b) = k(a - b)^2$ and $R(a, b) = \kappa kab$ for natural numbers a, b. A map $F : \mathbb{R}_\delta \to \mathbb{R}_\delta$ *reports* a meter at ratios when $F(a/b) = \mathrm{read}(E(a, b), R(a, b))$ for all natural numbers $a, b > 0$. F is *continuous at positive points* when for every $x > 0$ and every k there is a j with $|F(y) - F(x)| \leq 1/(k + 1)$ for every $y > 0$ within $1/(j + 1)$ of x.

Theorem 22.11 (The native reporting theorem)**.** *Let a meter, the apparatus law with $k, \kappa > 0$, and a map F that reports the meter at ratios be given. Then $F(a/b) = (2/\kappa)\,J(a/b)$ for all natural numbers $a, b > 0$. If moreover F is continuous at positive points, then $F(x) = (2/\kappa)\,J(x)$ for every $x > 0$; for $\kappa = 2$, $F = J$ on the positive points. The proof is free.*

Proof. For $a, b > 0$ the reference $R(a, b) = \kappa kab$ is positive, so Theorem 22.5 gives

$$F\Big(\frac{a}{b}\Big) = \frac{k(a - b)^2}{\kappa kab} = \frac{2}{\kappa} \cdot \frac{(a - b)^2}{2ab} = \frac{2}{\kappa}\, J\Big(\frac{a}{b}\Big),$$

because $J(a/b) = (a/b + b/a)/2 - 1 = (a^2 + b^2 - 2ab)/(2ab)$.

Now let F be continuous at positive points and fix $x > 0$ and k. For $y > 0$,

$$J(y) - J(x) = \frac{(y - x)(xy - 1)}{2xy},$$

so for $y \geq x/2$ we get $|J(y) - J(x)| \leq C|y - x|$ with $C = (1 + 2/x^2)/2$: J is continuous at x with an explicit modulus. Take j for F at x with tolerance $1/(2k + 2)$, and by Lemma 12.52, item 4, a ratio q within $\min(1/(j + 1),\ x/2,\ \kappa/(4C(k + 1)))$ of x; it is positive because $q \geq x/2$. Then

$$\left|F(x) - \frac{2}{\kappa}J(x)\right| \leq |F(x) - F(q)| + \frac{2}{\kappa}|J(q) - J(x)| \leq \frac{1}{2k + 2} + \frac{1}{2k + 2}.$$

This holds for every k, so $F(x) = (2/\kappa)J(x)$. The moduli are used inside the proof of an equation, so no choice of moduli is made. □

Conditions and scope.

Base and premises

A meter (Definition 22.4); the apparatus law at counts; literal reports at every positive ratio a/b; continuity of F at positive points.

Conclusion

$F = (2/\kappa)J$ on the positive points, with no omniscience and no choice. With Theorem 22.8, every apparatus meeting the three selection premises has the apparatus law with $k = \mathscr{E}(Tu)$ and $\kappa = 4\chi(p)$, so the theorem applies to it.

Identification

a and b are the counted operands of one comparison, the meter reads discarded energy against the reference, and F is the reported dimensionless cost. That a physical instrument obeys the meter laws is a premise, not a result.

The balanced premise says that the report of the balanced comparison is $F(1) = 0$. With literal reports this gives the first premise of Theorem 22.8: $F(1) = \mathrm{read}(\varepsilon(1, 1, 1), p(1, 1)) = \varepsilon(1, 1, 1)/p(1, 1)$, and $p(1, 1) > 0$. The composition law with its calibration forces J by a second route (Theorem 21.14); both routes end at the same function.

Continuity is one way to pass from ratios to all positive points. Two further premises of the chain, reciprocity and monotonicity on $[1, \infty)$ (row T2.2b), do the same work without it.

Theorem 22.12 (Reciprocity and order replace continuity). *Let c be a point and $F : \mathbb{R}_\delta \to \mathbb{R}_\delta$ satisfy $F(a/b) = c\,J(a/b)$ for all natural numbers $a, b > 0$; $F(1/x) = F(x)$ for every $x > 0$; and $F(x) \leq F(y)$ whenever $1 \leq x \leq y$. Then $F(x) = c\,J(x)$ for every $x > 0$. The proof is free.*

Proof. First let $x \geq 1$, and fix k. Take ratios $l < x < u$ with $u - l$ as small as we please (Lemma 12.52, item 4). Whether $l < 1$ is a comparison of ratios; if it holds, replace l by 1, which keeps $l \leq x$ because $x \geq 1$. Now $1 \leq l \leq x \leq u$, both l and u are positive ratios, and the order premise gives $c\,J(l) = F(l) \leq F(x) \leq F(u) = c\,J(u)$. Hence

$$-|c|\,|J(l) - J(x)| \leq F(x) - c\,J(x) \leq |c|\,|J(u) - J(x)|,$$

and the explicit modulus of J at x from the proof of Theorem 22.11 makes both bounds smaller than $1/(k+1)$ once $u - l$ is small enough. So $F(x) = c\,J(x)$.

For general $x > 0$ the dichotomy $x \geq 1$ or $x \leq 1$ costs LLPO (Corollary 12.18), but we need it only under a double negation. An equation on the line is the negation of an apartness, so it follows from its own double negation. And $\neg\neg(x \geq 1 \vee x \leq 1)$ holds outright: if neither held, then $\neg\neg(x < 1)$ and $\neg\neg(1 < x)$, hence $\neg\neg(x < 1 \wedge 1 < x)$, which is absurd. In the case $x \geq 1$ the first part applies; in the case $x \leq 1$ we have $1/x \geq 1$, so $F(x) = F(1/x) = c\,J(1/x) = c\,J(x)$, since $J(1/x) = J(x)$. Hence $\neg\neg(F(x) = c\,J(x))$, and so $F(x) = c\,J(x)$. □

Can the passage from ratios to points be dropped altogether? Not for free. To alter a report at one irrational point, a map must recognize that point, and that needs a decision of equality on the line as data.

Lemma 22.13 (An equality test is the WLPO function)**.** *A map $d : \mathbb{R}_\delta \times \mathbb{R}_\delta \to \{0,1\}$ with $d(x,y) = 1$ exactly when $x = y$ exists if and only if the WLPO function exists. No choice principle is used.*

Proof. Given d, the map $\alpha \mapsto d(j_\alpha, 0)$ is the WLPO function, because $j_\alpha = 0$ exactly when α is all 0. Conversely let D be the WLPO function. For regular sequences s and t naming x and y, let u be the regular sequence of $x - y$ and $\beta(n) = 1$ exactly when $|u_n| > 2/(n+1)$, a comparison of ratios. By Bishop's zero condition (Lemma 12.15), β is all 0 exactly when $x = y$. So $D(\beta)$ depends only on the points x and y, and $d(x,y) = D(\beta)$ is a map with the required property. □

Theorem 22.14 (Ratio reports alone do not force J, given the WLPO function)**.** *Suppose the WLPO function exists. Then it is not the case that every map F on the positive points with $F(a/b) = J(a/b)$ for all natural numbers $a, b > 0$ equals J at every positive point. No choice principle is used.*

Proof. Let d be the equality test of Lemma 22.13, let $\sqrt{2}$ be the nonnegative square root of 2 (Proposition 12.19, item 5), and put $F(x) = J(x) + d(x, \sqrt{2})$, the bit read as the point 0 or 1. A ratio is never $\sqrt{2}$: $(a/b)^2 = 2$ with $b > 0$ would give $a^2 = 2b^2$ in the natural numbers, so a is even, $a = 2a'$, and $b^2 = 2a'^2$ is a smaller solution, which is impossible by descent. So $d(a/b, \sqrt{2}) = 0$ and $F(a/b) = J(a/b)$ at every positive ratio, while $F(\sqrt{2}) = J(\sqrt{2}) + 1 \neq J(\sqrt{2})$. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ and the WLPO function, a single map deciding "all zero" for every binary sequence.

Conclusion

Reports at ratios do not by themselves force J; the step needs a further premise. With continuity (Theorem 22.11) or with reciprocity and order (Theorem 22.12) the step is free.

Not claimed

A lower bound. Showing that some principle is needed to alter a report would take a model of the line in which every map is continuous; none is given here. WLPO as a proposition does not define F: a map needs the decision as data.

22.4 What deciding the consequences of J costs

Once J is forced, the chain reads off its consequences. Each holds on the line at no cost when it is stated with apartness. The decision forms are another matter, and the opening example belongs here.

Proposition 22.15 (The consequences of J are free)**.** *For every $x > 0$: $J(x) = (x-1)^2/(2x)$; $J(x) \geq 0$; $J(x) = 0$ exactly when $x = 1$; $J(1/x) = J(x)$; $J(x) > 0$ exactly when $x \# 1$; and $1 \leq x \leq y$ implies $J(x) \leq J(y)$. In particular $J(1) = 0$. The proofs are free.*

Proof. $\frac{1}{2}(x + 1/x) - 1 = (x^2 + 1 - 2x)/(2x)$. A square is nonnegative and $2x > 0$, so $J(x) \geq 0$. If $J(x) = 0$ then $(x-1)^2 = 0$, and $x - 1 \# 0$ would make the square positive; so $\neg(x - 1 \# 0)$, which is $x = 1$. The converse is immediate, and $J(1/x) = J(x)$ because the definition is symmetric in x and $1/x$. Since $2x$ is positive and invertible, $J(x) > 0$ exactly when $(x-1)^2 > 0$, and a square is positive exactly when its base is apart from 0. Finally

$$J(y) - J(x) = \frac{(y - x)(xy - 1)}{2xy},$$

and for $1 \leq x \leq y$ both factors $y - x$ and $xy - 1$ are nonnegative, so the difference is nonnegative. □

Theorem 22.16 (A ratio other than 1 has positive cost: exactly MP)**.** *On $\mathbb{R}_\delta$ each of the following assertions, made for every $x > 0$, is equivalent to the principle named, with no choice principle in either direction:*

(i) $x \neq 1$ *implies* $J(x) > 0$*: MP;*

(ii) $J(x) \neq 0$ *implies* $J(x) > 0$*: MP;*

(iii) $J(x) = 0$ *or* $J(x) \neq 0$*: WLPO;*

(iv) $J(x) = 0$ *or* $J(x) > 0$*: LPO.*

Proof. By Proposition 22.15, $J(x) = 0$ exactly when $x = 1$, so $J(x) \neq 0$ exactly when $x \neq 1$, and (i) and (ii) are the same assertion.

(ii) gives MP. Let α be a binary sequence that is not all 0, and put $x = 1 + j_\alpha > 0$. Then $j_\alpha \neq 0$, since $j_\alpha = 0$ would make α all 0; so $x \neq 1$ and $J(x) \neq 0$. By (ii), $J(x) > 0$, so $x \# 1$ by Proposition 22.15, that is, $j_\alpha \# 0$. As j_α is never negative, $j_\alpha > 0$, and α has a 1.

MP gives (ii). Let $J(x) \neq 0$. If $\neg(J(x) > 0)$, then $J(x) \leq 0$, and with $J(x) \geq 0$ this gives $J(x) = 0$; so $\neg\neg(J(x) > 0)$, and MP in the form $\neg\neg(y > 0) \to y > 0$ (Corollary 12.18) gives $J(x) > 0$.

For (iii) and (iv), let r be a point and $x = 1 + r^2$. Then

$$J(1 + r^2) = \frac{(1 + r^2)^2 + 1 - 2(1 + r^2)}{2(1 + r^2)} = \frac{r^4}{2(1 + r^2)}.$$

So $J(x) = 0$ exactly when $r = 0$, and $J(x) > 0$ exactly when $r \# 0$. Assertion (iii) therefore decides $r = 0$ or $r \neq 0$ for every r, which is WLPO by Corollary 12.18; conversely WLPO decides whether the point $J(x)$ is 0. Assertion (iv) decides $r = 0$ or $r > 0$ or $r < 0$ for every r, which is LPO; conversely LPO decides $J(x) > 0$, $J(x) = 0$ or $J(x) < 0$, and the last is excluded by $J(x) \geq 0$. □

Conditions and scope.

Base and premises
$\mathbb{R}_\delta$ and J on the positive points.

Conclusion
The step "$J(x) > 0$ for $x \neq 1$" (row T2.8), read on the line in its own form, is exactly Markov's principle; its apartness form "$J(x) > 0$ for $x \# 1$" is free. The decision "zero or not" is WLPO and "zero or positive" is LPO.

Not claimed
Any price for a single x: for one given point the facts of Proposition 22.15 hold outright. The price belongs to the uniform assertion over all positive points.

The last step of the segment, T2.9, is the cost-law route: on the line, a comparison cost that obeys the composition law and the unit calibration is J (Theorem 21.14), with no omniscience and no choice. The reporting route and the cost-law route meet at the same function. Of the twelve rows of T2, nine are free, two have an exact price, and one is bracketed (Table 22.1).

22.5 The golden hierarchy

The second segment of the first rung turns one costly join into the golden ratio. A carrier of states moves by a step map, and each state has a positive reading. The steps ask that the readings be presented additively, that the step preserve positive readings, and that the first new level be a join of earlier levels with a positive cost. These conditions leave one scaling: every step multiplies the reading by φ.

Definition 22.17 (Orbits, additive readings and costly joins)**.** An *orbit with positive readings* is a set S with a map $T : S \to S$ and a map $r : S \to \mathbb{R}_\delta$ with $r(s) > 0$ for every s. An *additive reading* of it consists of an abelian group G, a group homomorphism $A : G \to G$, a map $\ell : S \to G$ and an additive map $v : G \to \mathbb{R}_\delta$ with $v(\ell(s)) = r(s)$ and $\ell(Ts) = A(\ell(s))$ for every s. For a sequence of positive levels $s_0, s_1, s_2, \ldots$ and natural numbers $a \leq b$, the *join ratio* is s_{b-a}/s_0. The levels have a *costly initial join* when some $a \leq b < 2$ satisfy $s_2 = s_a + s_b$ and $J(s_{b-a}/s_0) > 0$. A *realized hierarchy* of the orbit is a state s_* whose levels $L_k = r(T^k s_*)$ satisfy $L_1/L_0 > 1$, $L_{k+2}/L_{k+1} = L_{k+1}/L_k$ for every k, and $L_2 = L_1 + L_0$.

The theorem uses J only through two of its properties: $J(1) = 0$, and $J(x) > 0$ exactly when $x \# 1$.

Lemma 22.18 (A costly initial join posts additively)**.** *If positive levels have a costly initial join, then $s_2 = s_1 + s_0$. The proof is free.*

Proof. The pairs $a \le b < 2$ are $(0,0)$, $(0,1)$ and $(1,1)$, and which one is given is decided on the natural numbers. If $a = b$, the join ratio is $s_0/s_0 = 1$ and $J(1) = 0$, which is not positive. So $(a,b) = (0,1)$ and $s_2 = s_0 + s_1$. □

Lemma 22.19 (Positive comparisons fix one scalar)**.** *Let $L, M : G \to \mathbb{R}_\delta$ be additive maps on an abelian group, let $e \in G$ have $L(e) > 0$, and suppose $L(x) > 0$ implies $M(x) > 0$ for every $x \in G$. Then $M(x)/M(e) = L(x)/L(e)$ for every x, so $M = \big(M(e)/L(e)\big)L$. The proof is free.*

Proof. $M(e) > 0$ by the hypothesis at e. Suppose $M(x)/M(e) < L(x)/L(e)$. By Lemma 12.52, item 1, there is a ratio p/q with $q > 0$ strictly between them. Then $L(qx - pe) = qL(x) - pL(e) > 0$, while $M(qx - pe) = qM(x) - pM(e) < 0$, against the hypothesis. So $\neg\big(M(x)/M(e) < L(x)/L(e)\big)$, which is $L(x)/L(e) \le M(x)/M(e)$. The same argument at $-x$ gives the reverse inequality, and the two give equality (Proposition 12.19, item 2). □

A classical proof of this lemma argues by contradiction to obtain a strict inequality. On the line that is unnecessary: refuting a strict inequality is all that $\le$ asks.

Theorem 22.20 (Golden scaling is free)**.** *Let an orbit with positive readings have an additive reading (G, A, ℓ, v) whose step preserves positive values: $v(x) > 0$ implies $v(Ax) > 0$ for every $x \in G$. Let $s_* \in S$. Then:*

(a) some $\rho > 0$ satisfies $r(Ts) = \rho\, r(s)$ for every $s \in S$;

(b) if the levels $r(T^k s_*)$ have a costly initial join, then $r(Ts) = \varphi\, r(s)$ for every $s \in S$, and these levels form a realized hierarchy;

(c) every realized hierarchy has $L_1/L_0 = \varphi$, $L_n = \varphi^n L_0$ and $L_0/L_n = \varphi^{-n}$ for every n.

No omniscience principle and no choice principle is used.

Proof. (a) Apply Lemma 22.19 to $L = v$ and $M = v \circ A$, which is additive, with $e = \ell(s_*)$; $v(e) = r(s_*) > 0$, and the step hypothesis is the positivity hypothesis of the lemma. With $\rho = v(Ae)/v(e) > 0$,

$$r(Ts) = v(\ell(Ts)) = v(A\ell(s)) = \rho\, v(\ell(s)) = \rho\, r(s).$$

(b) By (a), $r(T^k s_*) = \rho^k r(s_*)$. Lemma 22.18 gives $\rho^2 r(s_*) = \rho r(s_*) + r(s_*)$, and $r(s_*)$ is invertible, so $\rho^2 = \rho + 1$. The positive root is unique: $\rho^2 - \rho - 1 = (\rho - \varphi)(\rho + \varphi - 1)$, and $\rho + \varphi - 1 > 0$ is invertible, so $\rho = \varphi$. The levels $\varphi^k r(s_*)$ have ratio $\varphi > 1$ throughout, and $L_2 = L_1 + L_0$ by Lemma 22.18.

(c) Self-similarity makes every ratio L_{k+1}/L_k equal to $\lambda = L_1/L_0$, so $L_n = \lambda^n L_0$ by induction. Then $L_2 = L_1 + L_0$ gives $\lambda^2 = \lambda + 1$, and $\lambda > 1 > 0$ gives $\lambda = \varphi$ as in (b). The weights follow: $L_0/L_n = \varphi^{-n}$. □

Conditions and scope.

Base and premises
An orbit with positive readings on $\mathbb{R}_\delta$, an additive reading whose step preserves positive values, and a costly initial join of the levels of one state.

Conclusion
Every step multiplies readings by φ; levels are φ^n times the base and weights are φ^{-n}. Free.

Not claimed
That a costly join is supplied. The supply is the next step, and its price depends on how "costly" is read (Theorem 22.24).

The hypotheses are not empty. The model below has every one of them, with φ entering as an input of the model.

Proposition 22.21 (A model of golden scaling)**.** *Let S be the natural numbers with $T(n) = n+1$ and $r(n) = \varphi^n$; let $G = \mathbb{Z} \times \mathbb{Z}$, $A(a,b) = (a+b, a)$, $v(a,b) = a\varphi + b$ and $\ell(n) = A^n(0,1)$. This is an orbit with positive readings and an additive reading whose step preserves positive values, and the levels of 0 have a costly initial join. The proof is free.*

Proof. A is a homomorphism and v is additive. $v(A(a,b)) = (a+b)\varphi + a = \varphi(a\varphi + b)$ because $\varphi^2 = \varphi + 1$; so $v \circ A = \varphi v$ preserves positivity, and by induction $v(\ell(n)) = \varphi^n v(0,1) = \varphi^n$. The levels are $1, \varphi, \varphi^2$ with $\varphi^2 = \varphi + 1$, and the join ratio φ is apart from 1, so $J(\varphi) > 0$ by Proposition 22.15. □

Over the classical real numbers excluded middle holds for every statement about a number, so these steps have no price there; the native statements above are the ones whose price means something. The six steps T3.1–T3.6 are free.

22.6 Three readings of the first costly join

Golden scaling needs a costly initial join. Step T3.7 supplies it from the fact that the join is a recognition: its two parts differ, so the join costs something. There are three ways to read "its two parts differ", and each has its own price. The parts may be *apart*: the join ratio ρ satisfies $\rho \# 1$. They may be merely *distinct*: $\rho \neq 1$. Or their difference may be *witnessed by an integer pairing*, which is the reading of step T3.7 in Table 22.1.

Definition 22.22 (Pair kinematics and the deformation-erasure principle)**.** A *pair kinematics* is a set of configurations with an equivalence relation, read "deforms into", and two configurations, the *split* and the *pair*. A *pairing observable* is a map P from configurations to $\mathbb{Z}$ that is invariant under deformation and has $P(\text{split}) = 0$. The *deformation-erasure principle* (DEP) holds for a kinematics when some pairing observable has $P(\text{pair}) \neq 0$. The *rigid kinematics* of a point x of the line has the points of $\mathbb{R}_\delta$ as configurations, equality as deformation, split 1 and pair x; DEP holds for it exactly when some map $P : \mathbb{R}_\delta \to \mathbb{Z}$ has $P(1) = 0$ and $P(x) \neq 0$. Finally, *strong extensionality for integer-valued maps*, $\mathrm{SE}_{\mathbb{Z}}$, is the principle that every map $P : \mathbb{R}_\delta \to \mathbb{Z}$ is strongly extensional: $P(x) \neq P(y)$ implies $x \# y$.

The converse directions below rest on one device, a join whose parts differ by a prescribed point.

Lemma 22.23 (The tilted join). *For a point t put $s_0 = 1 + |t|$, $s_1 = 1 + |t| + t$ and $s_k = s_0 + s_1$ for $k \geq 2$. These levels are positive, $s_2 = s_0 + s_1$, and the join ratio $\rho = s_1/s_0$ satisfies $(\rho - 1)(1 + |t|) = t$. Hence $\rho \neq 1$ exactly when $t \neq 0$, and $\rho \# 1$ exactly when $t \# 0$. The proof is free.*

Proof. $s_1 \geq 1$ because $|t| \geq -t$, and $s_0 \geq 1$. Then $(\rho - 1)s_0 = s_1 - s_0 = t$. Since $1 + |t|$ is positive and invertible, $\rho - 1$ is 0 exactly when t is, and apart from 0 exactly when t is. □

Theorem 22.24 (Three readings of the first costly join). *Consider positive levels $s_0, s_1, \ldots$ and natural numbers $a \leq b < 2$ with $s_2 = s_a + s_b$, and let ρ be the join ratio s_{b-a}/s_0.*

(a) If $\rho \# 1$, the levels have a costly initial join. Conversely, levels have a costly initial join exactly when $s_2 = s_0 + s_1$ and $s_1/s_0 \# 1$. Free.

(b) The assertion "$\rho \neq 1$ gives a costly initial join", for all such levels and indices, is equivalent to MP.

(c) The assertion "DEP for the rigid kinematics of ρ gives a costly initial join", for all such levels and indices, is equivalent to $\mathrm{SE}_{\mathbb{Z}}$.

(d) MP implies $\mathrm{SE}_{\mathbb{Z}}$.

No choice principle is used.

Proof. (a) By Proposition 22.15, $J(\rho) > 0$ exactly when $\rho \# 1$, so $\rho \# 1$ with $s_2 = s_a + s_b$ is a costly initial join for the same a, b. Conversely, a costly initial join has $(a, b) = (0, 1)$ by Lemma 22.18, and then $J(s_1/s_0) > 0$ gives $s_1/s_0 \# 1$.

(b) Assume MP. If $\rho \neq 1$, then $\rho - 1 \neq 0$, and MP in the form $x \neq 0 \to x \# 0$ (Table 12.2) gives $\rho \# 1$; (a) gives the costly join. Conversely assume the assertion and let $r \neq 0$. The tilted join at $t = r$ has $\rho \neq 1$ by Lemma 22.23, so it has a costly initial join, so $\rho \# 1$ by (a), so $r \# 0$. That is MP in the same form.

(c) Assume $\mathrm{SE}_{\mathbb{Z}}$ and DEP for the rigid kinematics of ρ: some P has $P(1) = 0$ and $P(\rho) \neq 0$. Strong extensionality gives $\rho \# 1$, and (a) gives the costly join. Conversely assume the assertion, and let $P : \mathbb{R}_\delta \to \mathbb{Z}$ and points x, y satisfy $P(x) \neq P(y)$. Put $t = x - y$ and take the tilted join, with ratio ρ. The map

$$Q(c) = P\big(y + (c - 1)(1 + |t|)\big) - P(y)$$

has $Q(1) = 0$ and $Q(\rho) = P(y+t) - P(y) = P(x) - P(y) \neq 0$, so DEP holds for the rigid kinematics of ρ. The assertion gives a costly initial join, so $\rho \# 1$ by (a), so $t \# 0$ by Lemma 22.23: $x \# y$.

(d) If $P(x) \neq P(y)$ then $x \neq y$, because P is a map; so $x - y \neq 0$, and MP gives $x - y \# 0$. □

How far below MP does $\mathrm{SE}_{\mathbb{Z}}$ sit? With the WLPO function the two coincide, and the same function settles the rigid form of the step.

Proposition 22.25 (Equality as data closes the gap)**.** *Assume the WLPO function. Then for every point $x \neq 1$, DEP holds for the rigid kinematics of x; and* $\mathrm{SE}_{\mathbb{Z}}$ *implies MP, so the assertion of Theorem 22.24(c) is equivalent to MP. No choice principle is used.*

Proof. Let d be the equality test of Lemma 22.13, and put $P(c) = 0$ if $d(c, 1) = 1$ and $P(c) = 1$ otherwise. This is a map to $\mathbb{Z}$ with $P(1) = 0$, and $P(x) = 1$ for every $x \neq 1$: DEP for the rigid kinematics of x. Now let $r \neq 0$. Then $1 + r \neq 1$, so $P(1 + r) = 1 \neq 0 = P(1)$, and $\mathrm{SE}_{\mathbb{Z}}$ gives $(1 + r) \# 1$, that is, $r \# 0$. That is MP. □

Conditions and scope.

Base and premises
Positive levels on $\mathbb{R}_\delta$ with $s_2 = s_a + s_b$; the three readings of "the parts differ".

Conclusion
Apart parts: free. Distinct parts: exactly MP. Parts separated by an integer pairing (the route of step T3.7): exactly $\mathrm{SE}_{\mathbb{Z}}$, which MP implies, and which equals MP in the presence of the WLPO function.

Not claimed
That $\mathrm{SE}_{\mathbb{Z}}$ implies MP: this converse is open. It is expected to fail, since in models where every map from the line to $\mathbb{Z}$ is constant, strong extensionality holds trivially while MP need not; no such model is constructed here. The rigid form "$x \neq 1$ gives DEP" costs exactly the WLPO function (Theorem 22.64).

A classical proof of step T3.7 picks the pairing "1 at the pair, 0 elsewhere" by deciding, for each configuration, whether it equals the pair. On the line that decision is an equality test, and an equality test is the WLPO function (Lemma 22.13). In physical words: if the two parts of the first join are known to be apart, the join costs something for free; if they are only known not to coincide, extracting the positive cost is Markov's principle; if their difference is witnessed by an integer pairing, the price is the strong extensionality of integer readings of the line.

22.7 Eight states, and where three dimensions come from

The last segment of the first rung concerns a pair of loops that the world keeps: a configuration that cannot be deformed into the split one. DEP expresses that some integer pairing tells them apart. Three of its steps are logic, one is arithmetic, and two are topology.

Proposition 22.26 (What a kept pair gives)**.** *For every pair kinematics, DEP implies that the pair does not deform into the split. The* winding kinematics*, with configurations $\mathbb{Z}$, equality as deformation, split* 0 *and pair* 2*, satisfies DEP. The* erasable *kinematics, with configurations $\mathbb{Z}$, in which every configuration deforms into every other, split* 0 *and pair* 2*, does not. The proofs are free.*

Proof. If the pair deformed into the split, every pairing observable would give $P(\text{pair}) = P(\text{split}) = 0$. In the winding kinematics the identity map is a pairing observable, with value $2 \neq 0$ at the pair. In the erasable kinematics every configuration deforms into the split, so every pairing observable is 0 everywhere. □

Proposition 22.27 (Eight states exactly in dimension three)**.** *For a natural number D, $2^D = 8$ if and only if $D = 3$. The binary words of length n are 2^n in number and pairwise distinct, and every binary word of length 3 is one of* 000, 001, 010, 011, 100, 101, 110, 111. *The proofs are free.*

Proof. $2^0, 2^1, 2^2, 2^3 = 1, 2, 4, 8$, and $2^D \geq 16$ for $D \geq 4$. The words of length $n+1$ are the words of length n followed by 0 and by 1, so their number doubles at each length, and two words are equal exactly when they agree letter by letter, a finite check. □

This is arithmetic given $D = 3$; it is not the selection of the dimension. The selection is the content of steps T4.4 and T4.5, and it is the subject of Chapter 23. On the native spheres of $\mathbb{R}_\delta$, with continuous maps given with their moduli and chains presented by generators and relations, a natively embedded circle can be linked, that is, some loop in its complement bounds no native surface there, exactly when the dimension is 3. The proof uses no omniscience principle and no choice principle. Chapter 23 also treats the circle's first homology, the last group of the second rung: it is $\mathbb{Z}$ as a relation, with no principle, and its data form costs exactly the WLPO function (Theorem 22.65). Given that theorem, every native spatial realization in dimension D that satisfies DEP has $2^D = 8$; reading the kept pair of segment T4 as a linked circle in this topology is an identification, which the chain takes as a premise and which is not derived here. All six steps of T4 are free.

22.8 The second rung

The second rung collects what the first rung certifies, in thirty-five groups: records and information, the geometry of the cost, composition and balancing, the divergence that the cost defines, the golden hierarchy, the three-bit carrier and persistent topology. On the line, thirty-one groups hold with no omniscience, three have an exact price in one decision form, and the last, the first homology of the circle, is free as a relation and costs exactly the WLPO function as a function (Chapter 23 and Theorem 22.65). We prove the headline identities and the three prices first, and then the remaining identities and counts (Propositions 22.34, 22.35 and 22.36); Table 22.1 lists every group with the result that settles it.

Proposition 22.28 (Records and capacity)**.** *Let k, n, m, M be natural numbers. If an injective map sends M messages to the words of length n over k letters, then $M \leq k^n$. If an injective map sends the words of length m over $k \geq 1$ letters to M cells, then $k^m \leq M$. For $k \geq 2$ no injective map sends the words of length M over k letters to M cells: a store cannot hold its own history. The proofs are free.*

Proof. An injective map from a finite set of size N to one of size M gives $N \leq M$, by induction on M: remove the image of one element and its preimage. There are k^n words of length n, which gives the first two statements. For the third, the second gives $k^M \leq M$, but $k^M \geq 2^M > M$. □

Proposition 22.29 (The geometry of the cost)**.** *Let a, b, c, x, y, z be positive points.*

(a) $J(a/b) = (a-b)^2/(2ab)$, and $J(ca/(cb)) = J(a/b)$.

(b) If $c' \geq 0$ and $s \geq 0$ satisfy $s^2 = c'(c'+2)$, then $J(x) \leq c'$ exactly when $1+c'-s \leq x \leq 1+c'+s$.

(c) $J(xy) + J(x/y) = 2J(x)J(y) + 2J(x) + 2J(y)$. With $S(x) = (x - 1/x)/2$, $S(x)^2 = J(x)\big(J(x)+2\big)$ and $J(xy) = J(x)J(y) + J(x) + J(y) + S(x)S(y)$.

(d) If $m = tx + (1-t)y > 0$, then $tJ(x) + (1-t)J(y) - J(m) = t(1-t)(x-y)^2/(2xym)$.

(e) If $s > 0$ and $s^2 = ab$, then $J(a/z) + J(b/z) - J(a/s) - J(b/s) = (a+b)(z-s)^2/(2abz)$, so two parties meet most cheaply at the geometric mean.

(f) If $s \geq 0$ and $s^2 = ab$, then $s \leq (a+b)/2$, with equality exactly when $a = b$.

The proofs are free.

Proof. (a) $J(a/b) = (a/b + b/a)/2 - 1 = (a^2 + b^2 - 2ab)/(2ab)$, and $ca/(cb) = a/b$.

(b) Multiplying by $2x > 0$, $J(x) \leq c'$ becomes $(x-1)^2 \leq 2c'x$, that is, $\big(x - (1+c')\big)^2 \leq (1+c')^2 - 1 = s^2$. For nonnegative p and s, $p^2 \leq s^2$ holds exactly when $p \leq s$, since $p^2 - s^2 = (p-s)(p+s)$ and $p + s \geq 0$; applied to $p = |x - 1 - c'|$ this is the stated interval.

(c) $\big(J(xy)+1\big) + \big(J(x/y)+1\big) = \frac{1}{2}\big(xy + \frac{1}{xy} + \frac{x}{y} + \frac{y}{x}\big) = \frac{1}{2}(x + \frac{1}{x})(y + \frac{1}{y}) = 2\big(J(x)+1\big)\big(J(y)+1\big)$, which expands to the composition law. Next, $S(x)^2 = \big(\frac{1}{2}(x + \frac{1}{x})\big)^2 - 1 = (J(x)+1)^2 - 1$. Finally $(J(x)+1)(J(y)+1) + S(x)S(y) = \frac{1}{4}\big[(x + \frac{1}{x})(y + \frac{1}{y}) + (x - \frac{1}{x})(y - \frac{1}{y})\big] = \frac{1}{2}(xy + \frac{1}{xy}) = J(xy) + 1$.

(d) The left side is $\frac{1}{2}\big(t/x + (1-t)/y - 1/m\big)$, and $m\big(ty + (1-t)x\big) - xy = t(1-t)(x-y)^2$ after expanding.

(e) The left side equals $\frac{1}{2}(a+b)\big(1/z + z/(ab) - 2/s\big)$, because $s/(ab) = 1/s$; over the denominator abz the bracket is $(ab + z^2 - 2zs)/(abz) = (z-s)^2/(abz)$.

(f) $\big((a+b)/2\big)^2 - s^2 = (a-b)^2/4 \geq 0$, and the comparison of squares in (b) applies; equality holds exactly when $(a-b)^2 = 0$. □

The cost defines a divergence between distributions. It is half of Pearson's χ^2 [173], and it bounds total variation.

Proposition 22.30 (The divergence of the cost)**.** *Let $q_0, \dots, q_{n-1}$ be positive points and $p_0, \dots, p_{n-1}$ points, and put*

$$D(p\|q) = \frac{1}{2}\sum_{i<n} \frac{(p_i - q_i)^2}{q_i}, \qquad \mathrm{TV}(p,q) = \frac{1}{2}\sum_{i<n} |p_i - q_i|.$$

If every p_i is positive, then $\sum_i p_i J(p_i/q_i) = D(p\|q)$. $D(p\|q) = 0$ exactly when $p_i = q_i$ for every i. If $\sum_i q_i = 1$, then $2\,\mathrm{TV}(p,q)^2 \leq D(p\|q)$, with equality for $q = (\frac{1}{2}, \frac{1}{2})$ and $p = (\frac{1}{2} + e, \frac{1}{2} - e)$, for every point e. The proofs are free.

Proof. By Proposition 22.29(a), $p\,J(p/q) = (p-q)^2/(2q)$. A sum of nonnegative points is 0 exactly when each term is, and $(p_i-q_i)^2/(2q_i) = 0$ exactly when $p_i = q_i$. For the bound, the weighted Cauchy–Schwarz inequality $\big(\sum_i q_i w_i\big)^2 \le \big(\sum_i q_i\big)\big(\sum_i q_i w_i^2\big)$ for $q_i \ge 0$ follows from $\sum_{i,j} q_i q_j (w_i - w_j)^2 \ge 0$. With $w_i = |p_i - q_i|/q_i$ and $\sum_i q_i = 1$ it gives $\big(\sum_i |p_i - q_i|\big)^2 \le 2D(p\|q)$, that is, $4\,\mathrm{TV}^2 \le 2D$. In the example both sides equal $2e^2$. □

The first of the three prices concerns the reciprocal ambiguity. A cost determines a ratio only up to inversion: $J(x) = J(1/x)$.

Theorem 22.31 (The reciprocal ambiguity costs exactly LLPO)**.** *For positive points x, y,*

$$J(x) - J(y) = \frac{(x-y)(xy-1)}{2xy}.$$

Hence, with no principle: $J(x) = J(y)$ implies $\neg\neg(x = y \vee xy = 1)$; $x \# y$ and $xy \# 1$ imply $J(x) \# J(y)$; and $x = y$ or $xy = 1$ implies $J(x) = J(y)$. The assertion that $J(x) = J(y)$ implies $x = y$ or $xy = 1$, for all positive x, y, is equivalent to LLPO, with no choice principle in either direction.

Proof. The identity: $J(x) - J(y) = \frac{1}{2}(x-y) + \frac{1}{2}(1/x - 1/y) = \frac{1}{2}(x-y)(1 - 1/(xy))$. Write $w = (x-y)(xy-1)$, so $J(x) = J(y)$ exactly when $w = 0$. Let $w = 0$ and suppose that neither $x - y = 0$ nor $xy - 1 = 0$. If $x - y \# 0$, then $x - y$ is invertible and $xy - 1 = 0$, which is excluded; so $\neg(x - y \# 0)$, which is $x - y = 0$, also excluded. Hence $\neg\neg(x = y \vee xy = 1)$. A product of points apart from 0 is apart from 0, which gives the second statement, and the third holds because $J(1/x) = J(x)$.

Assume LLPO. By Table 12.2 it gives the domain law: a product equal to 0 has a factor equal to 0. Applied to $w = 0$ it gives $x = y$ or $xy = 1$.

Conversely assume the assertion; we prove the domain law, which gives LLPO by the same table. Let $uv = 0$, and put $x = \exp((u-v)/2)$ and $y = \exp(-(u+v)/2)$, using the exponential of Proposition 12.88. Both are positive, $x/y = \exp(u)$ and $xy = \exp(-v)$, so

$$w = (x-y)(xy-1) = y\big(\exp(u) - 1\big)\big(\exp(-v) - 1\big).$$

First, $\exp(h) \# 1$ implies $h \# 0$. Indeed, let $|\exp(h) - 1| > \delta > 0$ and $\mu = \min(\delta/2, 1)$; by cotransitivity $|h| > \mu/2$ or $|h| < \mu$, and in the second case Proposition 12.88, item 2, gives $|\exp(h) - 1| \le |h| + |h|^2 \le 2|h| < \delta$, which is excluded. Second, $\exp(h) = 1$ implies $h = 0$: $h > 0$ gives $\exp(h) \ge 1 + h > 1$, and $h < 0$ gives $\exp(h) = 1/\exp(-h) < 1$, so $h \# 0$ is impossible. Now if the product $(\exp(u) - 1)(\exp(-v) - 1)$ were apart from 0, both factors would be, so $u \# 0$ and $v \# 0$, and uv would be apart from 0, against $uv = 0$. So the product is 0, $w = 0$ and $J(x) = J(y)$. The assertion gives $x = y$, so $\exp(u) = 1$ and $u = 0$; or $xy = 1$, so $\exp(-v) = 1$ and $v = 0$. That is the domain law. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$, J on positive points, and the exponential of Proposition 12.88.

Conclusion

Deciding which branch of the reciprocal ambiguity holds is exactly LLPO. Its double-negated form, its apartness form and its converse are free.

Not claimed

Any price for the identity or for a single pair: for given x and y the identity holds outright.

The second price sits in the data-processing inequality. Pass a distribution through a channel and the divergence cannot grow; for divergences of this kind the inequality goes back to Csiszár [172]. The proof divides by the output probabilities, so it needs to know which outputs have probability 0.

Theorem 22.32 (Omitting empty bins costs exactly LPO)**.** *A* channel *from n inputs to m outputs is a family of points $K(j \mid i) \geq 0$ with $\sum_{j<m} K(j \mid i) = 1$ for each $i < n$. For $q_i > 0$ and points p_i put $P_j = \sum_i K(j \mid i)p_i$, $Q_j = \sum_i K(j \mid i)q_i$ and $T_j = \sum_i K(j \mid i)(p_i - q_i)^2/q_i$.*

(a) With no principle: $(P_j - Q_j)^2 \leq Q_j T_j$ for every j; $Q_j = 0$ implies $P_j = 0$; if every Q_j is positive, then $D(P\|Q) \leq D(p\|q)$; and for any set of outputs given by a list of yes and no on which every Q_j is positive, the sum of $(P_j - Q_j)^2/(2Q_j)$ over that set is at most $D(p\|q)$.

(b) The assertion that every output of every channel, for every positive input distribution q with $\sum_i q_i = 1$, satisfies $Q_j > 0$ or $Q_j = 0$ is equivalent to LPO, with no choice principle in either direction.

Proof. (a) $P_j - Q_j = \sum_i K(j \mid i)q_i \cdot (p_i - q_i)/q_i$, and the weighted Cauchy–Schwarz inequality with weights $K(j \mid i)q_i \geq 0$ gives $(P_j - Q_j)^2 \leq Q_j T_j$. If $Q_j = 0$ this gives $(P_j - Q_j)^2 \leq 0$, so $P_j = Q_j = 0$. If $Q_j > 0$, then $(P_j - Q_j)^2/Q_j \leq T_j$, and $\sum_j T_j = \sum_i \big(\sum_j K(j \mid i)\big)(p_i - q_i)^2/q_i = 2D(p\|q)$, because every column of the channel sums to 1. Summing over all outputs, or over the given set, gives the two inequalities, since every $T_j \geq 0$.

(b) LPO gives $Q_j > 0$, $Q_j = 0$ or $Q_j < 0$, and $Q_j \geq 0$ excludes the last. Conversely let r be a point and take one input, $q_0 = 1$, and two outputs with $K(0 \mid 0) = r^2/(1+r^2)$ and $K(1 \mid 0) = 1/(1+r^2)$. Then $Q_0 = r^2/(1 + r^2)$. If $Q_0 > 0$ then $r^2 > 0$, so $r \# 0$; if $Q_0 = 0$ then $r^2 = 0$, so $r = 0$. That decides $r > 0$, $r = 0$ or $r < 0$ for every r, which is LPO (Corollary 12.18). □

The inequality itself is free, bin by bin and over any known support; only the literal recipe "drop the bins of probability zero" is priced. The third price concerns the golden hierarchy under a budget. Levels k apart have ratio φ^k; a cost budget c admits the gaps with $J(\varphi^k) \leq c$.

Theorem 22.33 (The largest admissible gap as a value costs exactly WLPO)**.**

(a) Let $c \geq 0$ and $s \geq 0$ with $s^2 = c(c + 2)$. Every natural number k with $J(\varphi^k) \leq c$ satisfies $k \log \varphi \leq \log(1 + c + s)$. Free.

(b) The assertion that for every point $c \geq 0$ some natural number K satisfies $J(\varphi^K) \leq c$, while every natural number k with $J(\varphi^k) \leq c$ has $k \leq K$, is equivalent to WLPO, with no choice principle in either direction.

Proof. (a) By Proposition 22.29(b), $\varphi^k \le 1+c+s$. The logarithm is increasing and $\log \varphi^k = k \log \varphi$.

(b) First, $J(\varphi) = (\varphi + \varphi^{-1})/2 - 1 = \varphi - \frac{3}{2} > 0$, since $\varphi^{-1} = \varphi - 1$ and $\varphi > 1.6$ (Proposition 22.37). Assume the assertion and let x be a point. Apply it to $c = J(\varphi)/(1 + x^2) \ge 0$, obtaining K; whether $K = 0$ is decided on the natural numbers. If $K = 0$, then $x \ne 0$: for $x = 0$ the budget is $J(\varphi)$, the gap 1 is admissible, and $1 \le K$ would follow. If $K \ge 1$, then $J(\varphi) \le J(\varphi^K) \le c$ by Proposition 22.15, so $J(\varphi)x^2/(1 + x^2) \le 0$, hence $x^2 \le 0$ and $x = 0$. So $x = 0$ or $x \ne 0$: WLPO.

Conversely assume WLPO. For a point u, $u \le c$ holds exactly when $\max(u - c, 0) = 0$, so WLPO decides it. For $k \ge 1$ we have $J(\varphi^k) \ge \varphi^k/2 - 1$ and, by Bernoulli's inequality, $\varphi^k \ge 1 + k(\varphi - 1) \ge 1 + 3k/5$; so $J(\varphi^k) > c$ as soon as $k > (10c + 5)/3$. Take a natural number B above $(10c + 5)/3$ (Proposition 12.19, item 4). Decide admissibility for each $k \le B$; the gap 0 is admissible because $J(1) = 0$; the largest admissible $k \le B$ is the required K. □

The remaining seventeen groups are identities of the cost and of the golden weights, limits of partial sums, and finite counts on the three-bit carrier. All are free. We prove them in three propositions, with $G(t) = \cosh t - 1$ as in Definition 21.11, so that $J(e^t) = G(t)$.

Proposition 22.34 (The cost in logarithmic form)**.** *Let t, u be points and a, b, c, x, y positive points.*

(a) (R2-05) $J(x) = G(\log x)$, and for every t the partial sums $\sum_{k<N} t^{2k+2}/(2k + 2)!$ converge to $G(t)$.

(b) (R2-06) With $h = y - x$,

$$J(y) = J(x) + \frac{1 - x^{-2}}{2}\, h + \frac{h^2}{2x^3} - \frac{h^3}{2x^3 y};$$

at $x = 1$ this reads $J(y) = \frac{1}{2}(y - 1)^2 - (y - 1)^3/(2y)$. If $0 < \theta < 1$ and $x \# y$, then $\theta J(x) + (1-\theta)J(y) > J(\theta x + (1-\theta)y)$. In logarithmic form, $G(t + u) + G(t - u) - 2G(t) = 2\big(G(t) + 1\big)G(u) \ge u^2$, and this second difference is positive when $u \# 0$.

(c) (R2-09) $d(a, b) = |\log a - \log b|$ is a metric on the positive points: $d(a, b) \ge 0$, $d(a, b) = d(b, a)$, $d(a, b) = 0$ exactly when $a = b$, and $d(a, c) \le d(a, b) + d(b, c)$. Moreover $J(a/b) = G\big(d(a, b)\big)$, while J itself fails the triangle inequality: $J(2) + J(2) < J(4)$.

(d) (R2-10) $t^2/2 \le G(t)$. If $|u| \le q < 1$, then

$$\frac{u^2}{2(1 + q)} \le J(1 + u) \le \frac{u^2}{2(1 - q)}.$$

(e) (R2-13) Put $C_n(x) = J(x^n) + 1 = \frac{1}{2}(x^n + x^{-n})$. Then $C_0(x) = 1$ and $C_{n+2}(x) = 2C_1(x)C_{n+1}(x) - C_n(x)$, so $C_n(x) = T_n\big(C_1(x)\big)$ for the Chebyshev polynomials $T_0 = 1$, $T_1(w) = w$, $T_{n+2}(w) = 2wT_{n+1}(w) - T_n(w)$.

(f) (R2-14) Let $x_0, \dots, x_{n-1}$ be positive, and let $a_i \ge 0$ and c_i satisfy $G(a_i) = c_i$ and $J(x_i) \le c_i$ for every $i < n$. Then $J(x_0 x_1 \cdots x_{n-1}) \le G(a_0 + a_1 + \cdots + a_{n-1})$.

(g) (R2-16) Let $x_0, \dots, x_{n-1}$ and r be positive with $x_0 x_1 \cdots x_{n-1} = r^n$. Then $nJ(r) \le \sum_{i<n} J(x_i)$, with equality exactly when $x_i = r$ for every $i < n$.

The proofs are free.

Proof. (a) By Proposition 21.5, item 1, and Lemma 21.3, item 1, $x = e^{\log x}$ and $x^{-1} = e^{-\log x}$, so $J(x) = G(\log x)$. Let E_N be the partial sums of the exponential (Proposition 12.88). In $\frac{1}{2}\big(E_{2N+1}(t) + E_{2N+1}(-t)\big)$ the odd powers cancel, and what remains is $1 + \sum_{k<N} t^{2k+2}/(2k+2)!$. Both partial sums converge, so these sums converge to $\frac{1}{2}(e^t + e^{-t}) - 1 = G(t)$.

(b) By the identity $J(y) - J(x) = h(xy-1)/(2xy)$ of the proof of Theorem 22.11, the formula follows from $(1 - x^{-2}) + h/x^3 - h^2/(x^3 y) = 1 - 1/(xy)$. Multiplied by $x^3 y$ this reads $x^3 y - xy + hy - h^2 = x^3 y - x^2$, which holds because $hy - h^2 = hx = xy - x^2$. At $x = 1$ the first-order term vanishes. For convexity, Proposition 22.29(d) gives the gap $\theta(1-\theta)(x-y)^2/(2xy\bar{x})$ with $\bar{x} = \theta x + (1-\theta)y > 0$, a product of positive points when $x \# y$. In logarithmic form, $\cosh(t+u) + \cosh(t-u) = 2\cosh t \cosh u$ (proof of Proposition 21.12, item 1), so the second difference is $2\cosh t(\cosh u - 1) = 2\big(G(t)+1\big)G(u)$. Here $G(t) + 1 \geq 1$, since $G \geq 0$, and $G(u) \geq u^2/2$ by (d); so the second difference is at least u^2, which is positive when $u \# 0$.

(c) $d \geq 0$ and $d(a,b) = d(b,a)$ are properties of $|\cdot|$, and so is the triangle inequality, applied to the points $\log a$, $\log b$, $\log c$. $d(a,b) = 0$ exactly when $\log a = \log b$, which holds exactly when $a = b$, since exp and log invert each other (Proposition 21.5, items 1 and 2). By (a) and Proposition 21.5, item 3, $J(a/b) = G(z)$ with $z = \log a - \log b$. Now $G(-z) = G(z)$, and $G(|z|) = G(z)$: an equation follows from its double negation, $\neg\neg(z \geq 0 \vee z \leq 0)$ holds outright, by the argument in the proof of Theorem 22.12 with 0 in place of 1, and in either case $|z|$ is z or $-z$. Finally $J(2) + J(2) = \frac{1}{2} < \frac{9}{8} = J(4)$.

(d) The terms $t^{2k+2}/(2k+2)!$ are nonnegative, so the partial sums of (a) increase, and each is at most their limit $G(t)$ (Lemma 12.52, item 2); the one with $N = 1$ is $t^2/2$. Next, $J(1+u) = u^2/\big(2(1+u)\big)$ by Proposition 22.15, and $0 < 1 - q \leq 1 + u \leq 1 + q$ gives $1/(1+q) \leq 1/(1+u) \leq 1/(1-q)$.

(e) $C_0 = 1$, and $C_1 C_{n+1} = \frac{1}{4}(x + x^{-1})(x^{n+1} + x^{-n-1}) = \frac{1}{4}(x^{n+2} + x^{-n-2} + x^n + x^{-n}) = \frac{1}{2}(C_{n+2} + C_n)$. Induction on n, two steps at a time, gives $C_n = T_n(C_1)$.

(f) Put $t_i = \log x_i$. By (a) and Proposition 21.5, item 3, $J(x_0 \cdots x_{n-1}) = G(t_0 + \cdots + t_{n-1})$ and $J(x_i) = G(t_i)$. On $[0, \infty)$ the map G is nondecreasing and strictly increasing: for $0 \leq p \leq q$ we have $1 \leq e^p \leq e^q$ (Lemma 21.3, item 4), and $G(p) = J(e^p) \leq J(e^q) = G(q)$ by Proposition 22.15; for $0 \leq p < q$ we have $1 \leq e^p < e^q$ (item 5), and $J(e^q) - J(e^p) = (e^q - e^p)(e^p e^q - 1)/(2e^p e^q) > 0$. Hence $G(|t_i|) = G(t_i) \leq G(a_i)$ gives $|t_i| \leq a_i$: if $a_i < |t_i|$, then $G(a_i) < G(|t_i|)$. So $|t_0 + \cdots + t_{n-1}| \leq a_0 + \cdots + a_{n-1}$, and $G(t_0 + \cdots + t_{n-1})$, which is G of the absolute value of the sum, is at most $G(a_0 + \cdots + a_{n-1})$.

(g) Put $t_i = \log x_i$ and $m = \log r$, so that $\sum_{i<n} t_i = nm$ (Proposition 21.5, item 3). Let $g(v) = e^v - 1 - v$, which is nonnegative by Lemma 21.3, item 2. For all points t and m, expanding the definitions,

$$G(t) = G(m) + (t-m)\sinh m + \tfrac{1}{2}\big(e^m g(t-m) + e^{-m} g(m-t)\big).$$

Summing over $i < n$, the middle terms add to $(\sum_i t_i - nm)\sinh m = 0$, so $\sum_i J(x_i) = nJ(r) + \frac{1}{2}\sum_i R_i$ with $R_i = e^m g(t_i - m) + e^{-m} g(m - t_i) \geq 0$. This gives the inequality. Equality holds exactly when $\sum_i R_i = 0$, that is, when every $R_i = 0$, since a sum of nonnegative points is 0 exactly when each term is. As $e^{\pm m} > 0$, $R_i = 0$ exactly when $g(t_i - m) = g(m - t_i) = 0$, and then

$2G(t_i - m) = g(t_i - m) + g(m - t_i) = 0$, so $J(e^{t_i - m}) = 0$, $e^{t_i - m} = 1$ and $t_i = m$ (Proposition 22.15, Lemma 21.3, item 5); then $x_i = e^{t_i} = r$. Conversely $x_i = r$ gives $R_i = 0$. □

The golden hierarchy prices its gaps with J and weighs its levels by powers of φ^{-1}. An infinite sum below is the limit of its partial sums over $n < N$.

Proposition 22.35 (Golden gaps, golden weights and the golden distribution)**.** *Let $r = \varphi - 1$, so that $r\varphi = \varphi^2 - \varphi = 1$ and $r + r^2 = r\varphi = 1$.*

(a) (R2-24) For every natural number k, $J(\varphi^k) = G(k \log \varphi)$; and $J(\varphi) = \varphi - \frac{3}{2}$, $J(\varphi^2) = \frac{1}{2}$.

(b) (R2-25) The golden weights r^n refine, $r^n = r^{n+1} + r^{n+2}$, and $\sum_{n<N} r^n = \varphi^2(1 - r^N)$; so $\sum_{n \geq 0} r^n = \varphi^2$, and $\sum_{n \geq 0} r^{m+n} = \varphi^2 r^m$ for every m.

(c) (R2-26) The weights $p_n = r^{n+2}$ are positive, with $p_n \varphi^2 = r^n$. Their total is $\sum_n p_n = 1$, their tails are $\sum_n p_{m+n} = r^m$, their mean is $\sum_n n p_n = \varphi$, their second moment is $\sum_n n^2 p_n = \varphi^3 + \varphi^2$, and their variance is $\sum_n (n - \varphi)^2 p_n = \varphi^3$, which is also the limit of $\sum_{n<N} n^2 p_n - \left(\sum_{n<N} n p_n\right)^2$. Each weight has $-\log p_n = (n+2)\log\varphi$, and the entropy is $\sum_n p_n(-\log p_n) = (\varphi + 2)\log\varphi$.

The proofs are free.

Proof. (a) By Proposition 22.34(a) and Proposition 21.5, item 3, $J(\varphi^k) = G(\log \varphi^k) = G(k \log \varphi)$. Since $\varphi^{-1} = r = \varphi - 1$, $J(\varphi) = \frac{1}{2}(2\varphi - 1) - 1 = \varphi - \frac{3}{2}$; and $\varphi^2 + \varphi^{-2} = (\varphi + 1) + (1 - r) = 3$, so $J(\varphi^2) = \frac{1}{2}$.

(b) $r^n - r^{n+1} - r^{n+2} = r^n(1 - r - r^2) = 0$. For the finite sum, both sides are 0 at $N = 0$, and the step from N to $N + 1$ adds r^N on the left and $\varphi^2(r^N - r^{N+1}) = \varphi^2 r^N r^2 = r^N$ on the right. Multiplying by r^m gives $\sum_{n<N} r^{m+n} = \varphi^2 r^m (1 - r^N)$.

The limits here and in (c) follow from the closed forms once r^N, Nr^N and $N^2 r^N$ tend to 0, because sums and products of convergent sequences converge to the sums and products of the limits. The binomial expansion of $(1 + r)^N$ has nonnegative terms, so $(1 + r)^N \geq \binom{N}{3} r^3$, and $r^N(1 + r)^N = 1$. For $N \geq 4$, using $(N - 1)(N - 2) \geq N^2/4$ and $r^3 > \frac{1}{5}$ (as $r = \varphi - 1 > \frac{3}{5}$ by Proposition 22.37(c)),

$$0 \leq N^2 r^N \leq \frac{N^2}{\binom{N}{3} r^3} = \frac{6N}{(N-1)(N-2)\, r^3} \leq \frac{24}{N r^3} < \frac{120}{N}.$$

So for $N \geq 120(k + 1)$ the points $r^N \leq Nr^N \leq N^2 r^N$ are within $1/(k + 1)$ of 0.

(c) $p_n > 0$ as a power of the positive point r, and $p_n \varphi^2 = r^n (r\varphi)^2 = r^n$. From (b), $\sum_{n<N} p_n = r^2\varphi^2(1 - r^N) = 1 - r^N$ and $\sum_{n<N} p_{m+n} = r^m(1 - r^N)$. By induction on N,

$$\sum_{n<N} n p_n = \varphi - (N + \varphi) r^N, \qquad \sum_{n<N} n^2 p_n = \varphi^3 + \varphi^2 - (N^2 + 2\varphi N + 3\varphi + 2) r^N.$$

At $N = 0$ both are 0, since $\varphi^2 = \varphi + 1$ and $\varphi^3 = 2\varphi + 1$ give $\varphi^3 + \varphi^2 = 3\varphi + 2$. The steps need $(N + \varphi) - (N + 1 + \varphi) r = Nr^2$ and $(N^2 + 2\varphi N + 3\varphi + 2) - \big((N + 1)^2 + 2\varphi(N + 1) + 3\varphi + 2\big) r = N^2 r^2$,

and both follow from $\varphi r = 1$, $\varphi - r = 1$ and $1 - r = r^2$. Expanding $(n-\varphi)^2 = n^2 - 2\varphi n + \varphi^2$ and using $\varphi^2 = \varphi + 1$,

$$\sum_{n<N} (n-\varphi)^2 p_n = \varphi^3 - (N^2 + 2\varphi + 1) r^N.$$

The limits are φ, $\varphi^3 + \varphi^2$ and φ^3, and the second moment less the squared mean tends to $(\varphi^3 + \varphi^2) - \varphi^2 = \varphi^3$. Finally $\log r = -\log \varphi$, because $r\varphi = 1$ and $\log 1 = 0$ (Proposition 21.5, item 3); so $-\log p_n = (n+2)\log\varphi$, and

$$\sum_{n<N} p_n (n+2)\log\varphi = \log\varphi\Big(\sum_{n<N} n p_n + 2\sum_{n<N} p_n\Big) = \log\varphi\big(\varphi + 2 - (N+\varphi+2) r^N\big),$$

which tends to $(\varphi+2)\log\varphi$. □

The last groups concern the three-bit carrier, the cube $\{0,1\}^3$ of Proposition 22.27, and the eight-tick signals of Definition 22.7.

Proposition 22.36 (The three-bit carrier and its clock)**.** *Let $B = \{0,1\}^3$. Call two words* adjacent *when they differ in exactly one coordinate, let $h(x,y)$ be the number of coordinates in which x and y differ, and call a sequence $x_0, \dots, x_n$ of words in which consecutive words are adjacent a* walk *of length n from x_0 to x_n.*

(a) (R2-28) The cube has $\binom{3}{k} 2^{3-k}$ faces of dimension k, each obtained by freeing k coordinates and fixing the others: 8 *vertices,* 12 *edges,* 6 *squares and* 1 *cell.*

(b) (R2-29) For all $x, y \in B$ some walk of length $h(x,y)$ runs from x to y, and every walk from x to y has length at least $h(x,y)$. Every walk from a word to itself has even length.

(c) (R2-30) If a sequence of M slots shows every word of B, then $M \geq 8$. The Gray cycle 000, 100, 110, 010, 011, 111, 101, 001 *shows each word once and changes one coordinate at each step, the step from the last word back to the first included.*

(d) (R2-31) The states of n independent three-bit records, the maps from $\{0, \dots, n-1\}$ to B, correspond one to one with the numbers $0, \dots, 8^n - 1$; and an injective encoding of these states by binary words of length k has $3n \leq k$.

(e) (R2-32) For $A \subseteq \{0,1,2\}$ put $\chi_A(x) = (-1)^{\sum_{i \in A} x_i}$. Then $\sum_{x \in B} \chi_A(x)\chi_{A'}(x)$ is 8 if $A = A'$ and 0 otherwise. The Laplacian $(Lf)(x) = \sum_{i<3}\big(f(x) - f(x^{(i)})\big)$, where $x^{(i)}$ is x with coordinate i changed, satisfies $L\chi_A = 2|A|\chi_A$. Every map $f : B \to \mathbb{R}_\delta$ satisfies $f = \sum_A \hat{f}(A)\chi_A$ with $\hat{f}(A) = \frac{1}{8}\sum_x f(x)\chi_A(x)$, and if $f = \sum_A a_A \chi_A$, then $a_A = \hat{f}(A)$ for every A.

(f) (R2-33) Let $s \geq 0$ with $s^2 = 2$, and $\omega = \frac{s}{2}(1 + \mathrm{i})$ in $\mathbb{C}_\delta$. Then $\omega^2 = \mathrm{i}$, $\omega^8 = 1$, $|\omega|^2 = 1$, and the imaginary part of $\omega^7 - 1$ is apart from 0. For signals f, with indices read modulo 8, put $(Sf)(k) = f(k-1)$ and $(Cf)(k) = \omega^k f(k)$ for $0 \leq k \leq 7$. Then $CS = \omega\, SC$; S^8 and C^8 are the identity; S and C keep the energy $\mathscr{E}$; and for the constant signal $\mathbf{1}$ the imaginary part of $(CS\mathbf{1})(0) - (SC\mathbf{1})(0)$ is apart from 0, so CS and SC differ.

(g) (R2-34) Take a pair kinematics (Definition 22.22) in which the pair deforms into the split, and a map P from its configurations to $\mathbb{Z}$ with $P(\mathrm{split}) = 0$. Then either P is not invariant under deformation, or $P(\mathrm{pair}) = 0$. The same assertion for every pair kinematics and every map P into $\mathbb{R}_\delta$ is equivalent to WLPO, with no choice principle in either direction.

Parts (a) to (f), and the first assertion of (g), are free.

Proof. (a) A face of dimension k is given by its k free coordinates, chosen in $\binom{3}{k}$ ways, and the values of the other $3-k$, chosen in 2^{3-k} ways; different choices give different faces.

(b) Changing the coordinates in which x and y differ, one at a time, is a walk of length $h(x,y)$. For adjacent x' and x'' the numbers $h(x',y)$ and $h(x'',y)$ differ by exactly one, so along a walk of length n from x the distance to y drops by at most n, and reaching y, where it is 0, needs $n \geq h(x,y)$. Each step changes the number of coordinates equal to 1 by one, so its parity alternates, and a walk that returns to its start has even length.

(c) For each word the first slot that shows it is found by a bounded search, and different words have different first slots. This injective map from the eight words to $\{0,\dots,M-1\}$ gives $M \geq 8$, by the count in the proof of Proposition 22.28. The Gray cycle is checked step by step.

(d) Read a word x as the digit $x_0 + 2x_1 + 4x_2 \in \{0,\dots,7\}$. The map sending a state w to $\sum_{i<n} 8^i \cdot (\text{digit of } w(i))$ is a bijection onto $\{0,\dots,8^n-1\}$, with inverse the reading of base-8 digits. An injective encoding by binary words of length k gives $8^n \leq 2^k$ by Proposition 22.28; as $8^n = 2^{3n}$ and $k < 3n$ would give $2^k < 2^{3n}$, we get $3n \leq k$.

(e) $\chi_A \chi_{A'} = \chi_U$, where U is the set of coordinates in exactly one of A and A'. If U is empty the sum is 8. Otherwise take $i \in U$: changing coordinate i pairs the words of B and changes the sign of χ_U, so the sum is 0. Changing coordinate i changes the sign of χ_A exactly when $i \in A$, so each $i \in A$ contributes $2\chi_A(x)$ to $(L\chi_A)(x)$ and each other i contributes 0. For the expansion, with $x \oplus y$ the coordinatewise sum modulo 2,

$$\sum_A \chi_A(x)\chi_A(y) = \sum_A \chi_A(x \oplus y) = \prod_{i<3} \bigl(1 + (-1)^{x_i + y_i}\bigr),$$

which is 8 when $y = x$ and 0 otherwise; so $\sum_A \hat{f}(A)\chi_A(x) = \frac{1}{8}\sum_y f(y) \sum_A \chi_A(y)\chi_A(x) = f(x)$. If $f = \sum_A a_A \chi_A$, then $\hat{f}(A') = \frac{1}{8}\sum_A a_A \sum_x \chi_A(x)\chi_{A'}(x) = a_{A'}$ by orthogonality. Only the ring operations of $\mathbb{R}_\delta$ and $8 \cdot \frac{1}{8} = 1$ are used.

(f) $\omega^2 = \frac{s^2}{4}(1+\mathrm{i})^2 = \frac{2}{4} \cdot 2\mathrm{i} = \mathrm{i}$, so $\omega^8 = \mathrm{i}^4 = 1$, and $|\omega|^2 = 2 \cdot \frac{s^2}{4} = 1$. Hence $\omega^7 = \omega^{-1} = \bar{\omega} = \frac{s}{2}(1-\mathrm{i})$, and the imaginary part of $\omega^7 - 1$ is $-s/2$; it is apart from 0, because $s^2 = 2$ makes s apart from 0 and $s \geq 0$ makes it positive. For $1 \leq k \leq 7$, $(\omega SCf)(k) = \omega \cdot \omega^{k-1} f(k-1) = (CSf)(k)$, and at $k = 0$, $(\omega SCf)(0) = \omega \cdot \omega^7 f(7) = f(7) = (CSf)(0)$. S^8 shifts by 8, and C^8 multiplies by $\omega^{8k} = 1$. $|\omega^k f(k)|^2 = |\omega|^{2k}|f(k)|^2 = |f(k)|^2$, and S permutes the entries, so both keep $\mathscr{E}$. Finally $(CS\mathbf{1})(0) = 1$ and $(SC\mathbf{1})(0) = (C\mathbf{1})(7) = \omega^7$, and the imaginary part of $1 - \omega^7$ is $s/2$.

(g) $P(\text{pair})$ is an integer, so $P(\text{pair}) = 0$ or not is decided. In the second case P is not invariant, since invariance would give $P(\text{pair}) = P(\text{split}) = 0$. For maps into $\mathbb{R}_\delta$, WLPO decides $P(\text{pair}) = 0$ or $P(\text{pair}) \neq 0$ (Corollary 12.18), and the same argument applies. Conversely let v be a point. Take the kinematics with two configurations, the split and the pair, in which every configuration deforms into every other, and put $P(\text{split}) = 0$ and $P(\text{pair}) = v$. Then P is invariant exactly when $v = 0$, so the assertion gives $v \neq 0$ or $v = 0$, which is WLPO by Corollary 12.18. □

With Proposition 22.26, part (g) is the second-rung group R2-34: an invariant integer pairing that is nonzero at the pair forbids the deformation, and for integer values the decision is free.

The 48 symmetries of the cube in R2-28 are the signed permutations of the coordinates, by Proposition 24.44(d).

22.9 The constants

From φ the chain reaches the constants of nature in fifteen steps. On the line, fourteen of them are free, and the fifteenth is a decision. We give each with arithmetic a reader can repeat by hand: every decimal bound below becomes a comparison of integers once the denominators are cleared. Three kinds of statement must be kept apart here. A *derivation* is a theorem about points of the line. An *identification* reads such a point as a physical quantity. A *calibration* fixes units or inserts a measured value. The section ends by sorting the steps into these kinds.

Proposition 22.37 (The golden ratio)**.**

(a) $\varphi^2 = \varphi + 1$ *and* $1 < \varphi < 2$.

(b) For natural numbers a *and* $b > 0$, $b\varphi - a \# 0$*:* φ *is apart from every ratio.*

(c) $1.61 < \varphi < 1.62$.

(d) With Fibonacci numbers $F_0 = 0$, $F_1 = 1$, $F_{n+2} = F_{n+1} + F_n$, $\varphi^{n+2} = \varphi^{n+1} + \varphi^n$ *and* $\varphi^{n+1} = F_{n+1}\varphi + F_n$ *for every* n*. In particular* $\varphi^5 = 5\varphi + 3$*, and* $10.7 < \varphi^5 < 11.3$.

The proofs are free.

Proof. (a) With $\sqrt{5} \geq 0$ and $(\sqrt{5})^2 = 5$, $\varphi^2 = (6+2\sqrt{5})/4 = \varphi+1$. Since $(\sqrt{5}-1)(\sqrt{5}+1) = 4 > 0$ and $\sqrt{5}+1 > 0$, $\sqrt{5} > 1$ and $\varphi > 1$; since $(3-\sqrt{5})(3+\sqrt{5}) = 4$, $\sqrt{5} < 3$ and $\varphi < 2$.

(b) From $\varphi^2 = \varphi + 1$, $(b\varphi - a)\big(b(1-\varphi) - a\big) = a^2 - ab - b^2$. The integer $a^2 - ab - b^2$ is not 0 when $b > 0$: a solution of $a^2 = ab + b^2$ has $b < a < 2b$, and then $(b, a-b)$ is a solution with a smaller positive second entry, since $b(a-b) + (a-b)^2 = a^2 - ab = b^2$; descent excludes this. A nonzero integer is apart from 0, so the product is, and so is its factor $b\varphi - a$.

(c) For natural numbers p and $q > 0$,

$$(q\varphi - p)\big(q(\varphi - 1) + p\big) = q^2 + pq - p^2,$$

and the second factor is positive. So $p/q < \varphi$ when $p^2 < q^2 + pq$, and $\varphi < p/q$ when $q^2 + pq < p^2$. Now $161^2 = 25921 < 26100 = 100^2 + 100 \cdot 161$ and $162^2 = 26244 > 26200 = 100^2 + 100 \cdot 162$.

(d) Multiply $\varphi^2 = \varphi + 1$ by φ^n, and induct with the Fibonacci recurrence. Then $\varphi^5 = F_5\varphi + F_4 = 5\varphi + 3$, and (c) gives $11.05 < \varphi^5 < 11.1$. □

We use the comparison in (c) repeatedly. For instance, $1618033^2 = 2\,618\,030\,789\,089 < 2\,618\,033\,000\,000$ and $1618034^2 = 2\,618\,034\,025\,156 > 2\,618\,034\,000\,000$ give $1.618033 < \varphi < 1.618034$.

Proposition 22.38 (The locked constants and the action quantum)**.** *Put $r = \varphi - 1$. Then $r\varphi = 1$ and $1 - r = r^2$. The locked constants $\alpha_{\mathrm{lock}} = (1 - 1/\varphi)/2$ and $C_{\mathrm{lock}} = (\varphi - 1)^5$ satisfy $\alpha_{\mathrm{lock}} = (\varphi - 1)^2/2$, and every point w with $\varphi^5 w = 1$ equals C_{lock}. With the native time unit $\tau_0 = 1$, the action quantum $\hbar = C_{\mathrm{lock}}\tau_0$ satisfies*

$$\hbar = (\varphi - 1)^5 = \varphi^{-5} = 5\varphi - 8, \qquad 0.088 < \hbar < 0.093.$$

The proofs are free.

Proof. $r\varphi = \varphi^2 - \varphi = 1$ and $r^2 + r = \varphi^2 - \varphi = 1$. So $1 - 1/\varphi = 1 - r = r^2$. The inverse of φ^5 is unique, and $\varphi^5 r^5 = 1$; also $(5\varphi + 3)(5\varphi - 8) = 25(\varphi^2 - \varphi) - 24 = 1$, so $r^5 = 5\varphi - 8$. The bounds say $1.6176 < \varphi < 1.6186$, and the comparison of Proposition 22.37(c) gives them: $16176^2 = 261\,662\,976 < 261\,760\,000 = 10000^2 + 10000 \cdot 16176$ and $16186^2 = 261\,986\,596 > 261\,860\,000$. □

Proposition 22.39 (The cube's curvature and the recognition length)**.** *The cube has 8 vertices, 12 edges and 6 faces, so its Euler characteristic is $\chi = 2$. Its faces meet at the angle $\pi/2$; at each vertex three faces meet, leaving the angular deficit $2\pi - 3 \cdot \pi/2 = \pi/2$. Then:*

(a) the eight deficits total $4\pi = 2\pi\chi$;

(b) the curvature coefficient $8 \cdot (\pi/2)/(2\pi)$ equals $\chi = 2$, so the curvature cost $\lambda \mapsto (\text{coefficient}) \cdot \lambda^2$ is $2\lambda^2$;

(c) exactly one point $\lambda > 0$ satisfies $2\lambda^2 = 1$.

The proofs are free.

Proof. The counts are a listing of $\{0,1\}^3$ and its faces. (a) $8 \cdot \pi/2 = 4\pi$ and $2\pi \cdot 2 = 4\pi$. (b) $4\pi/(2\pi) = 2$. (c) The point $\lambda_0 = \sqrt{1/2}$ (Proposition 12.19, item 5) is a positive root, and $\lambda_0\sqrt{2} = 1$ because both sides are nonnegative with square 1. If $\lambda, \mu > 0$ are roots, then $(\lambda - \mu)(\lambda + \mu) = 0$ and $\lambda + \mu > 0$ is invertible, so $\lambda = \mu$. □

Part (a) is Descartes' theorem on the total angular defect, which for the cube is a count. In (b) the coefficient is derived; that the curvature cost of step S6 has the quadratic form $\lambda \mapsto \text{coefficient} \cdot \lambda^2$ is a modelling input of that step, an identification. The constant π is that of Section 21.2, the half-period of the sine (Theorem 21.8); it is $\pi = 4(\arctan\frac{1}{2} + \arctan\frac{1}{3})$, an identity due to Euler, with Gregory's series $\arctan(1/m) = \sum_{j\ge 0}(-1)^j/\big((2j+1)m^{2j+1}\big)$.

The next constants need decimal brackets for exp and log. For $x \ge 0$ put $S_N(x) = \sum_{m<N} x^m/m!$. Every $S_N(x)$ is at most $\exp(x)$, since the partial sums increase to it (Lemma 12.52, item 2). For $0 \le x \le 1$ and $N \ge 1$, $\exp(x) \le S_N(x) + 2x^N/N!$, because $x^{m+1}/(m+1)! \le \frac{1}{2}x^m/m!$ for $m \ge 1$, so the tail from N on is at most $(x^N/N!)(1 + \frac{1}{2} + \frac{1}{4} + \cdots)$. Since log is the inverse of the increasing map exp on positive points (Definition 21.4), $\exp(x) < a$ gives $x < \log a$, and $a < \exp(x)$ gives $\log a < x$.

Proposition 22.40 (Newton's constant, the Einstein coupling and the Boltzmann scale)**.**

(a) $2.7725887 < 4\log 2 < 2.7725888$.

(b) In the native units of the chain, with recognition length $\lambda_{\mathrm{rec}} = 1$, $c = 1$ *and* $\hbar\varphi^5 = 1$, *the Clausius form* $G = \lambda_{\mathrm{rec}}^2 c^3/(4\log 2\cdot\hbar)$ *of Newton's constant equals* $\varphi^5/(4\log 2)$. *The Einstein coupling* $\kappa = 8\pi G/c^4$ *equals* $2\pi\varphi^5/\log 2$, *and* $\kappa\hbar = 2\pi/\log 2$.

(c) The Boltzmann scale $k_R = \log\varphi$ *satisfies* $0.47 < k_R < 0.49$*; more precisely* $0.4812 < k_R < 0.4813$.

The proofs are free.

Proof. (a) Let $x_- = 0.6931471805$ and $x_+ = 0.6931471806$. Exact rational arithmetic gives

$$S_{12}(x_-) + \frac{2x_-^{12}}{12!} = 1.99999999990\ldots < 2, \qquad S_{12}(x_+) = 2.00000000005\ldots > 2.$$

So $x_- < \log 2 < x_+$, and $2.772588722 < 4\log 2 < 2.7725887224$.

(b) $G = 1/(4\log 2\cdot\hbar) = \varphi^5/(4\log 2)$ because $1/\hbar = \varphi^5$. Then $\kappa = 8\pi G = 2\pi\varphi^5/\log 2$ and $\kappa\hbar = 2\pi/\log 2$. With (a) and $11.05 < \varphi^5 < 11.1$, $3.98 < G < 4.01$.

(c) With $1.618033 < \varphi < 1.618034$,

$$S_6(0.4812) + \frac{2\cdot 0.4812^6}{6!} = 1.6180308\ldots < \varphi, \qquad S_6(0.4813) = 1.6181581\ldots > \varphi,$$

so $0.4812 < \log\varphi < 0.4813$. □

The chain does not derive the fine-structure constant. Step S14 says that the coupling is a boundary datum, which no closure condition blind to the coupling can pin, and Proposition 22.43(a) below proves it on the line. What the chain forces is the kinematic part of α^{-1}, $4\pi\varphi^5 - 5\log\varphi$.

Theorem 22.41 (The forced content of α^{-1})**.** *The point* $\alpha_{\mathrm{kin}}^{-1} = 4\pi\varphi^5 - 5\log\varphi$ *satisfies* $136.947 < \alpha_{\mathrm{kin}}^{-1} < 136.959$*; in particular it lies below* 137.030. *The proof is free.*

Proof. Gregory's series is alternating with terms decreasing in size, so its value lies between any two consecutive partial sums. With A_J and B_J the sums of the first J terms for $m = 2$ and $m = 3$, this gives $4(A_{10} + B_8) < \pi < 4(A_{11} + B_7)$, that is,

$$3.1415925 < 3.14159257797\ldots < \pi < 3.14159268738\ldots < 3.1415927.$$

By Proposition 22.37(d), $11.090165 < \varphi^5 < 11.09017$, so $139.36311 < 4\pi\varphi^5 < 139.36319$; and $2.406 < 5\log\varphi < 2.4065$ by Proposition 22.40(c). Hence $136.95661 < \alpha_{\mathrm{kin}}^{-1} < 136.95719$. □

Step S13 keeps a witness value $\sigma\exp(-f/\sigma)$, built on the seed $\sigma = 4\pi\cdot 11 = 44\pi$ with a gap term f through an exponential; the photon normalization that fixes it is one of the calibrations listed at the end of this section. We treat it as a construction. Reading it as the inverse fine-structure constant is an identification that no theorem here uses; by step S14 the coupling itself is a boundary datum. Its band and the uniqueness of its exponential form are both free.

Theorem 22.42 (The band and the exponential form of the witness α_W^{-1})**.** *Let $\sigma = 44\pi$.*

(a) For every $s \geq 0$ with $s^2 = 2$, put $w_8 = \big(348 + 210s - (204 + 130s)\varphi\big)/7$, $f = w_8 \log\varphi$ and $\alpha_W^{-1} = \sigma\exp(-f/\sigma)$. Then $137.030 < \alpha_W^{-1} < 137.039$.

(b) Let $g : \mathbb{R}_\delta \to \mathbb{R}_\delta$ have positive values and $g(0) = \sigma$, and suppose that $\log\circ g$ has uniform derivative $-1/\sigma$ on every bounded interval: for all natural numbers T and k there is a j such that

$$\left|\log g(t) - \log g(u) + \frac{t-u}{\sigma}\right| \leq \frac{|t-u|}{k+1}$$

whenever $|t|, |u| \leq T$ and $|t-u| \leq 1/(j+1)$. Then $g(x) = \sigma\exp(-x/\sigma)$ for every x.

The proofs are free.

Proof. (a) Since $1414213^2 = 1\,999\,998\,409\,369 < 2 \cdot 10^{12} < 2\,000\,001\,237\,796 = 1414214^2$, $1.414213 < s < 1.414214$. On these ranges w_8 decreases in s (its coefficient $(210 - 130\varphi)/7$ is negative, as $\varphi > 21/13$) and in φ, so $2.4905686 < w_8 < 2.4906241$, and with Proposition 22.40(c), $1.19846 < f < 1.19874$. The bracket for π gives $138.23007 < \sigma < 138.23008$. For $0 \leq y \leq 1$ the series of $\exp(-y)$ alternates with decreasing terms, so $1 - y \leq \exp(-y) \leq 1 - y + y^2/2$. With $y = f/\sigma$,

$$\sigma - f \leq \alpha_W^{-1} \leq \sigma - f + \frac{f^2}{2\sigma},$$

and the brackets give $137.03133 < \alpha_W^{-1} < 137.03682$.

(b) Put $h = \log\circ g$ and fix x, a natural number $T \geq |x|$ and k; take j from the hypothesis. With $N = (T+1)(j+1)$ the points $t_i = ix/N$, $i \leq N$, lie in $[-T, T]$ and are spaced by $|x|/N \leq 1/(j+1)$. Telescoping,

$$\left|h(x) - h(0) + \frac{x}{\sigma}\right| \leq \sum_{i<N}\left|h(t_{i+1}) - h(t_i) + \frac{t_{i+1} - t_i}{\sigma}\right| \leq \frac{|x|}{k+1}.$$

This holds for every k, so $h(x) = \log\sigma - x/\sigma$. Applying exp, which inverts log and turns sums into products, $g(x) = \sigma\exp(-x/\sigma)$. □

The witness, as a function of the gap, meets the hypothesis of (b), so the statement is not empty, and on the line the uniqueness of its exponential form is a theorem. Reading α_W^{-1} as the inverse fine-structure constant is an identification, and no part of the theorem; the forced part α_{kin}^{-1} lies below the band of (a).

Proposition 22.43 (What the constants cannot fix)**.**

(a) Let $A > 0$. No point t satisfies $\kappa A = t$ for every $\kappa > 0$. If a property P of points does not depend on its argument, that is, $P(\kappa) \Leftrightarrow P(\kappa')$ for all κ, κ', and holds for some κ, then no point t satisfies $\kappa A = t$ for every $\kappa > 0$ with $P(\kappa)$.

(b) *The exponents of length, time and mass in $c^a\hbar^b G^d$ are $a+2b+3d$, $-a-b-2d$ and $b-d$. The determinant of this linear map is -2, so the monomial is dimensionless only for $a=b=d=0$.*

(c) *For positive points $\hbar_{\mathrm{SI}}$ and c_{SI}, the map $G \mapsto \pi\hbar_{\mathrm{SI}}G/c_{\mathrm{SI}}^5$ sends positive points to positive points and is injective.*

The proofs are free.

Proof. (a) $\kappa = 1$ and $\kappa = 2$ would give $A = t = 2A$, so $A = 0$. A property that holds for some κ and does not depend on its argument holds at 1 and at 2. (b) c, $\hbar$ and G have dimensions LT^{-1}, ML^2T^{-1} and $L^3M^{-1}T^{-2}$. The matrix with rows $(1,2,3)$, $(-1,-1,-2)$ and $(0,1,-1)$ has determinant $1\cdot 3 - 2\cdot 1 + 3\cdot(-1) = -2$, and a linear map of nonzero determinant has only the zero solution. (c) $\pi\hbar_{\mathrm{SI}}/c_{\mathrm{SI}}^5$ is positive, hence invertible. □

Read physically, (a) says that no closure condition blind to the coupling can pin the value of α: it is a boundary datum. Parts (b) and (c) say that values in SI units need an anchor, and that the map from an input G to the squared SI tick is a calibration, not a prediction.

The one priced constant is a test. The units check of step S11 tests four approximate equalities $|a-b| \le 10^{-6}$ as yes-or-no questions.

Theorem 22.44 (The units test costs exactly WLPO and fails at its own values)**.**

(a) *The assertion that $|a-b| \le t$ or $\neg(|a-b| \le t)$ for all points a, b, t is equivalent to WLPO. A map f from triples of points to $\{0,1\}$ with $f(a,b,t) = 1$ exactly when $|a-b| \le t$ exists if and only if the WLPO function exists. No choice principle is used.*

(b) *Let units τ_0, ℓ_0, c satisfy $c\tau_0 = \ell_0$ with τ_0 and ℓ_0 apart from 0, and put $\tau_{\mathrm{rec}} = 2\pi\tau_0/(8\log\varphi)$ and $\lambda_{\mathrm{kin}} = 2\pi\ell_0/(8\log\varphi)$. Then $\tau_{\mathrm{rec}}/\tau_0 = \lambda_{\mathrm{kin}}/\ell_0 = \pi/(4\log\varphi)$ and $c\,\tau_{\mathrm{rec}} = \lambda_{\mathrm{kin}}$. Exactly one $K \ge 0$ has $K^2 = \varphi$, and $1 < K < 3/2$. Moreover $\pi/(4\log\varphi) > 3/2$. Free.*

(c) *The units check of step S11 takes base units $\tau_0 = \ell_0 = c = 1$, scaled units $\tau_0 = \ell_0 = 2$, $c = 1$, and tolerance 10^{-6}. It tests the base clock ratio against the base length ratio, the base clock ratio against 1, the scaled clock ratio against the scaled length ratio, and the base clock ratio against the scaled clock ratio, and it passes when all four tests pass. For every map f as in (a), the check fails: the second test returns 0.*

Proof. (a) $|r-0| \le 0$ holds exactly when $r = 0$, so the decision at $(r,0,0)$ is $r = 0$ or $r \ne 0$, which is WLPO (Corollary 12.18); and $f(j_\alpha, 0, 0)$ is the WLPO function. Conversely, $|a-b| \le t$ holds exactly when $\max(|a-b|, t) = t$, an equality, which WLPO decides and the equality test of Lemma 22.13 decides as data.

(b) The ratios are $2\pi/(8\log\varphi)$ in both cases, and $c\tau_{\mathrm{rec}} = 2\pi\ell_0/(8\log\varphi)$. The root K exists and is unique by Proposition 12.19, item 5; $(1+K)(K-1) = \varphi - 1 > 0$ gives $K > 1$, and $(3-2K)(3+2K) = 9 - 4\varphi > 0$, as $\varphi < 2$, gives $K < 3/2$. Next, $(31+48\varphi)(79-48\varphi) = 2449 - 2304(\varphi^2 - \varphi) = 145 > 0$, so $48\varphi < 79$, and $79/48 = 1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{48} = S_4(\frac{1}{2}) \le \exp(\frac{1}{2})$; so $\log\varphi < \frac{1}{2}$. With $\pi > 3$ and $\log\varphi > 0$, $\pi/(4\log\varphi) > 3/2$.

(c) By (b), the clock and length ratios are equal in each pack, and halving both τ_{rec} and τ_0 leaves the ratio unchanged; so the first, third and fourth tests compare equal points and return 1. The second compares $\pi/(4\log\varphi)$ with 1, and their difference exceeds $1/2 > 10^{-6}$, so it returns 0. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; the check's packs and tolerance as step S11 states them.

Conclusion

The test as a total yes-or-no decision is exactly WLPO, and as one map it is exactly the WLPO function. At the values of step S11 the check fails for every faithful test, with no principle: one of its comparisons is refuted outright. The clock ratio $\pi/(4\log\varphi)$, about 1.632, is compared with 1 rather than with K, and it differs from both.

Not claimed

Anything about the physical constants themselves: the packs are unit conventions, and the theorem is about the test.

Derived, identified and calibrated

Of the fifteen steps of the constants, fourteen are free and one, the test, costs exactly WLPO. The steps divide as follows.

Derived

Theorems about points of the line: the properties and bounds of φ and its powers; α_{lock}, C_{lock} and $\hbar = \varphi^{-5}$ with its bounds; the cube's total deficit 4π and the coefficient $\chi = 2$; the unique positive root of $2\lambda^2 = 1$; $G = \varphi^5/(4\log 2)$, $\kappa = 2\pi\varphi^5/\log 2$ and $k_R = \log\varphi$ with their bounds; the identities of the units test; $\alpha_{\mathrm{kin}}^{-1}$ with its bounds; the band and the uniqueness of the exponential form; and the two boundary statements.

Identified

Readings, named and not priced: the quadratic form of the curvature cost; G as the Clausius coefficient with one bit per face; k_R in units of temperature; the gap formula α_W^{-1} read as the inverse fine-structure constant, an identification that no theorem here uses (by step S14 the coupling is a boundary datum); and the line's log and π read as the classical logarithm and circle constant.

Calibrated

Six calibrations, named and not priced: the native units (tick, voxel and c equal to 1, with λ_{rec} the voxel); the unit of bit cost; the photon normalization; the measured values of c, $\hbar$ and G in SI units; the anchor of the SI bridge (seconds per tick, metres per voxel, joules per unit of cost); and the measured value of α^{-1} with its uncertainty.

Selecting the complete table of constants physically is not one of the fifteen steps, and this chapter does not price it.

22.10 The numbered chain, step by step

The chain also comes as one numbered sequence. It begins with T−2, "nothing is not something", and T−1, the floor below logic; it passes through the Boolean floor of logic and cost (T0–T4), the uniqueness of J and of φ (T5, T6), the count of eight states and three dimensions (T7, T8), and a run of structural steps (T9–T15); and it ends with the constants, the masses on the golden ladder, a series of formulas compared with measured values, and three counts (T16–T41). A separate statement bounds the period of a walk through the eight states. What T7 states, as restated in Table 22.1, is the count $2^3 = 8$, the number of states of three bits; it contains no process and no period. A period enters only through the separate bound: a periodic walk that reaches all eight states has period at least 8. We write Cn for step Tn of this sequence, and P8 for the period bound, to keep these steps apart from the segments T1–T4 of the first rung, which number other statements. Each step is a short list of statements.

We read the sequence on the line as in Section 22.1, statement by statement. A real number becomes a point of $\mathbb{R}_\delta$; φ, π, log, exp and square roots are those of the line; a power φ^t with a point t as exponent is $e^{t \log \varphi}$, which at an integer t is the ordinary power; a measured number written into a statement, such as 0.2229, is the same ratio, $2229/10000$; and a quotient x/y is x times the inverse of a positive point y. Statements about counts, finite lists and finite sets are read as they stand. The native statement keeps every statement of the restated step and adds none; step C11 is restated for every assignment of virtues to its fourteen names.

Many steps are a definition followed by arithmetic, and the native statement then proves exactly that arithmetic. Where a step compares a formula with a measured value, the theorem is the comparison of the ratios and points involved. Whether the formula is the physical quantity it is named after is an identification, and whether the measured value is right is an observation; neither enters the price.

Theorem 22.45 (The numbered chain is free)**.** *Read on $\mathbb{R}_\delta$ as above, every step C−2, C−1, C0, ..., C41 of the numbered chain, and the period bound P8, holds with no omniscience principle and no choice principle.*

Proof. C−2, C−1 and C0–C4 are Proposition 22.47. C5, the uniqueness of J, is Theorem 21.14 with Theorem 22.12; C6, φ as the unique positive root of $\lambda^2 = \lambda + 1$, is Proposition 22.37(a) with the uniqueness argument in the proof of Theorem 22.20(b). The counting steps C7, C12, C14, C27, C28, C35, C39, C40, C41, the arithmetic of C8 and the counts of C31 are Proposition 22.48; the rest of C8 and the bound P8 are Proposition 22.49, and the topological form of C8 and the linking statement of C9 are step T4.4, proved free in Chapter 23. The symmetry count of C12 is Proposition 24.44(d) (row R2-28). C9, C15, C25, C29, C30, C31 and C37 are Proposition 22.50, and C36 is Proposition 22.15. C10, C11 and C13 are Proposition 22.51. C16–C18, C20–C24, C26, C32–C34 and C38 are Theorem 22.52, and C19 is Theorem 22.53. Each of these results is proved below or earlier with no omniscience principle and no choice principle. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; each step of the numbered chain as restated in Table 22.1, read statement by statement; the measured numbers written into its statements, taken as the same ratios.

Conclusion

Every step of the numbered chain, T−2 to T41, and the period bound are free. None of the eleven priced rows of Table 22.1 is a step of this sequence.

Not claimed

That the formulas are the physical quantities they are named after, or that the measured values are correct. For C11, that any particular fourteen maps are virtues: the restated step covers every assignment of virtues to the fourteen names, and its proof uses nothing beyond the law that defines a virtue. For C8, more than the integer encoding of linking in the step; its topological form is step T4.4.

What makes the native proofs free is the same in every step. They compute with counts, signed counts and ratios, which compare for free (Proposition 22.2); they use the field laws and the order of the line in its apartness form; they bound φ, π, $\log 2$ and $\log \varphi$ by the integer comparisons of Section 22.9; and they never decide whether two points are equal, or which sign a point has, unless it is known to be apart from 0. A classical proof of such a step may split on excluded middle or count a finite set through a choice principle: for C−1 it may argue by cases on whether P holds; a Gray cycle in every dimension, or a time at which a periodic walk reaches each corner, may be obtained by a choice; and a finite set may be counted through an enumeration chosen for it. The native proofs need none of this: $P \neq \neg P$ needs no case split, the Gray cycle is written down, each corner's first visit is found by a bounded search, and the counts are made on explicit lists.

Definition 22.46 (The Boolean floor)**.** The *floor* has two states, the *empty* state e and the *marked* state m. The *join* of two states is m when either is m, and e otherwise. A state is *consistent* when it is e, and two states are *independent* when at least one of them is e. A *cost* on the floor assigns to each state Γ a point $C(\Gamma) \geq 0$ of $\mathbb{R}_\delta$, with $C(\Gamma) = 0$ exactly when Γ is consistent, and $C(\mathrm{join}(\Gamma_1, \Gamma_2)) = C(\Gamma_1) + C(\Gamma_2)$ for independent Γ_1, Γ_2. The *floor cost* is $C(e) = 0$, $C(m) = 1$.

Proposition 22.47 (The floor of the chain is free)**.**

(a) (C−2) A set with no element is not equal to a set with one element.

(b) (C−1) No proposition equals its negation, and two different propositions exist. A set with an element has two different elements exactly when some property holds of one of its elements and fails of another; the floor has both.

(c) (C0) The floor cost is a cost. It is positive exactly on the inconsistent state, and every cost that agrees with it on a set S of states agrees with it on the join of any two independent states of S.

(d) (C1) An inconsistent state has positive cost, a state of cost 0 is consistent, and $C(m) > 0$.

(e) (C2) Every state is e or m, and $e \neq m$; a state of cost 0 is e, and a state of positive cost is m.

(f) (C3) $C(e) = 0$; $\mathrm{join}(e, \Gamma) = \Gamma$ and $C(\mathrm{join}(e, \Gamma)) = C(\Gamma)$ for every state; costs add over independent joins.

(g) *(C4) The floor is a two-point floor normalized by* $C(e) = 0$ *and* $C(m) = 1$*: two different states, every state one of them, consistent exactly when empty. It carries a recognition structure, its states with the relation of equality, and a recognition, the ordered pair* (e, e)*.*

The proofs are free.

Proof. (a) An equation between the two sets would carry the element of the second to an element of the first.

(b) Suppose $P = \neg P$. If P held, then $\neg P$ would hold too, which is absurd; so $\neg P$. But $\neg P$ is P, so P holds, a contradiction. The argument makes no case split on P. The propositions "true" and "false" differ, because the first holds and the second does not. If $x \neq y$, the property "equals x" holds of x and fails of y; conversely an element with a property and an element without it are different. On the floor, $e \neq m$.

(c)–(g) Every statement concerns the two states, and each is checked state by state. This is a case split on a two-element set, not on a proposition. The values are $C(e) = 0$ and $C(m) = 1$, and $0 < 1$; so $C \geq 0$, $C = 0$ exactly at e, and $C > 0$ exactly at m. The independent pairs are (e, e), (e, m) and (m, e), with joins e, m, m and costs $0 = 0 + 0$, $1 = 0 + 1$ and $1 = 1 + 0$, and $\mathrm{join}(e, \Gamma) = \Gamma$ for both states. If a cost C' agrees with C on S, and $\Gamma_1, \Gamma_2 \in S$ are independent, additivity of both gives $C(\mathrm{join}(\Gamma_1, \Gamma_2)) = C(\Gamma_1) + C(\Gamma_2) = C'(\Gamma_1) + C'(\Gamma_2) = C'(\mathrm{join}(\Gamma_1, \Gamma_2))$. The relation of equality and the pair (e, e) are given outright. □

The counting steps are identities between natural numbers, integers and ratios. They hold for the reason records compare for free: both sides are computed.

Proposition 22.48 (Every count and ratio identity of the chain is free)**.** *Each of the following holds, with no omniscience principle and no choice principle.*

(a) *(C7, C8)* $2^3 = 8$*. For a natural number* D*, each of* $D - 2 = 1$ *(in the integers) and* $2^D = 8$ *gives* $D = 3$*.*

(b) *(C12) The numbers recorded in step C12 for its three gauge layers form the triple* $(3, 2, 1)$*;* $3 + 2 + 1 = 6 = 2 \cdot 3$*, the number of faces of the cube* $\{0, 1\}^3$*; and the recorded numbers are not* 4*,* 3 *and* 2*.*

(c) *(C14, C27, C28) The cube* $\{0, 1\}^3$ *has* 3 *pairs of opposite faces, one for each axis, and* $3 \neq 2$*,* $3 \neq 4$*.*

(d) *(C31)* $56 = 7 \cdot 8$*, and* $50 \leq 56 \leq 70$*.*

(e) *(C35)* $\frac{23}{4} - \frac{31}{4} = -2$*,* $-2 - \frac{10}{4} = -\frac{9}{2}$ *and* $-\frac{9}{2} - \frac{22}{4} = -10$*.*

(f) *(C39) List the pairs* (n, l) *of natural numbers with* $1 \leq n \leq 7$*,* $l < \min(4, n)$ *and* $n + l \leq 8$ *in order of increasing* $n + l$*, and of increasing* n *when* $n + l$ *ties, and give* (n, l) *the capacity* $2(2l + 1)$*. There are* 19 *pairs; their running totals are* 2*,* 4*,* 10*,* 12*,* 18*,* 20*,* 30*,* 36*,* 38*,* 48*,* 54*,* 56*,* 70*,* 80*,* 86*,* 88*,* 102*,* 112*,* 118*; and the number* 1 *falls in the first pair,* $(1, 0)$*, which has* $l = 0$*.*

(g) *(C40) A set listed as four different letters U, C, A, G has exactly four elements;* $4^3 = 64$*; and* $3 \cdot 8 - 4 = 20$.

(h) *(C41)* $2^{3+1} = 16$, $3 + 2 = 5$ *and* $16 - 5 = 11$.

Proof. Each item compares counts, signed counts or ratios given by explicit expressions, and by Proposition 22.2 such comparisons are decided by computing both sides. We record the computations. (a) $2^3 = 8$; $D - 2 = 1$ gives $D = 3$ in the integers; and $2^0, 2^1, 2^2 = 1, 2, 4$ while $2^D \geq 16$ for $D \geq 4$. (b) The record is $(3, 2, 1)$ term by term, $3 + 2 + 1 = 6$, and the cube has two faces for each of its three axes, one with the coordinate 0 and one with 1; $3 \neq 4$, $2 \neq 3$, $1 \neq 2$. (c) The faces with a given coordinate fixed are the two opposite faces of that axis, so there are 3 pairs. (d) $7 \cdot 8 = 56$. (e) In quarters: $23 - 31 = -8$, $-8 - 10 = -18$ and $-18 - 22 = -40$. (f) The pairs in order are $(1, 0)$, $(2, 0)$, $(2, 1)$, $(3, 0)$, $(3, 1)$, $(4, 0)$, $(3, 2)$, $(4, 1)$, $(5, 0)$, $(4, 2)$, $(5, 1)$, $(6, 0)$, $(4, 3)$, $(5, 2)$, $(6, 1)$, $(7, 0)$, $(5, 3)$, $(6, 2)$, $(7, 1)$, with capacities $2, 2, 6, 2, 6, 2, 10, 6, 2, 10, 6, 2, 14, 10, 6, 2, 14, 10, 6$; summing gives the totals, and the first total, 2, is at least 1. (g) Every element is one of the four letters and no two letters are equal; $4^3 = 64$ and $24 - 4 = 20$. (h) $2^4 = 16$. □

The items prove the arithmetic of the steps as restated in Table 22.1. The steps define the number of generations, and the sizes of the two mixing matrices, as the number of face pairs of the cube; the three gauge numbers are values assigned to the layers, not consequences of a group; the quark residues are quarter-rung values, and in the restatement the integer rungs are the core convention and the quarter ladder is a hypothesis; the pairs (n, l) are subshells and 118 the last element; and the 16 modes split as step C41 assigns them. Reading these numbers as generations, gauge groups, quarks, elements, codons or modes is an identification; the proposition proves the arithmetic.

Proposition 22.49 (Dimension three and the period bound)**.**

(a) *(C8) Call a natural number D* compatible *when $D - 2 = 1$, $2^D = 8$, 2^D divides* 360*, the 2^D points of $\{0, 1\}^D$ can be listed in a cycle in which each point differs from the next, and the last from the first, in exactly one coordinate, and three structural conditions of step C8, stated in a form every dimension meets. Exactly one natural number is compatible, namely* 3.

(b) *(P8) Let w be a sequence of points of $\{0, 1\}^3$ and $p > 0$ a natural number with $w(n+p) = w(n)$ for every n. If every point of $\{0, 1\}^3$ is a value of w, then $p \geq 8$.*

The proofs are free.

Proof. (a) For $D = 3$: $3 - 2 = 1$, $2^3 = 8$, $360 = 45 \cdot 8$, and the cycle

$$000,\ 100,\ 110,\ 010,\ 011,\ 111,\ 101,\ 001$$

lists the eight points once each, with each point, and the last against the first, differing from the next in one coordinate. A compatible D has $D - 2 = 1$, so $D = 3$ by Proposition 22.48(a).

(b) By induction on q, $w(r+qp) = w(r)$, so every value of w is one of $w(0), \ldots, w(p-1)$. Equality of points of $\{0, 1\}^3$ is decided coordinate by coordinate, so for each point x the least $r < p$ with

$w(r) = x$ is found by checking $r = 0, 1, \dots, p-1$ in turn. Different points have different first visits, so this is an injective map from the eight points to $\{0, \dots, p-1\}$, and $p \geq 8$ by the count in the proof of Proposition 22.28. □

The golden ladder carries the masses of the chain. For an integer n we write φ^n for the ordinary power when $n \geq 0$ and for $(\varphi - 1)^{-n}$ when $n < 0$; since $\varphi(\varphi - 1) = 1$, $\varphi^{a+b} = \varphi^a \varphi^b$ for all integers, and $\varphi^n = e^{n \log \varphi}$. The ladder is $m(r) = \varphi^{-5} \varphi^r$ for integers r.

Proposition 22.50 (The golden ladder is free)**.**

(a) (C15) $\varphi > 1$ *and* $\varphi^2 = \varphi + 1$*. For every* $y > 0$*, integer* r *and point* g*,* $y\varphi^{r-8+g} > 0$*; the ladder is this law at* $y = \varphi^{-5}$ *and* $g = 8$*.*

(b) (C29) Every $m(r)$ *is positive, and* $r < s$ *gives* $m(r) < m(s)$*; so* $r_1 < r_2 < r_3$ *gives* $m(r_1) < m(r_2) < m(r_3)$*.*

(c) (C25) $m_e = m(2)$ *is positive, and every* $m_p = m(r_p)$ *satisfies* $m_p / m_e = \varphi^{r_p - 2}$*.*

(d) (C30) The valence mass $3m(4)$ *and the binding mass* $m(14)$*, at the binding rung* 14 *of the step, are positive,* $m(14) > 40 \cdot 3m(4)$*, and their sum is positive.*

(e) (C31) $1 < 56/\varphi^8 < 6/5$*.*

(f) (C37) $\eta = \varphi^{-44}$ *is positive, and* $\eta \varphi^{44} = 1$*.*

(g) (C9) For positive points $c_1, \dots, c_N$ *and a set* S *of indices,* $\sum_{i \in S} J(c_i) = 0$ *exactly when* $c_i = 1$ *for every* $i \in S$*.*

The proofs are free.

Proof. (a) Proposition 22.37(a). Exponentials are positive (Section 21.2), so $\varphi^{r-8+g} = e^{(r-8+g)\log\varphi} > 0$, and so is its product with y. At $y = \varphi^{-5}$, $g = 8$ the law gives $\varphi^{-5}\varphi^r = m(r)$.

(b) Powers of the positive point φ are positive. For $k \geq 1$, $\varphi^k > 1$ by induction, since $\varphi^{k+1} - \varphi^k = \varphi^k(\varphi - 1) > 0$. Then $m(s) - m(r) = m(r)(\varphi^{s-r} - 1) > 0$ for $r < s$.

(c) $m(r_p)/m(2) = \varphi^{r_p - 5}\varphi^{5-2} = \varphi^{r_p - 2}$.

(d) $m(14) = \varphi^{10} m(4)$, and $\varphi^{10} = 55\varphi + 34$ by Proposition 22.37(d). Since $86^2 = 7396 < 7755 = 55^2 + 55 \cdot 86$, the comparison in the proof of Proposition 22.37(c) gives $86/55 < \varphi$, so $\varphi^{10} > 86 + 34 = 120$ and $m(14) > 120\, m(4)$.

(e) $\varphi^8 = 21\varphi + 13$. From $\varphi < 2$, $\varphi^8 < 55 < 56$. From $101^2 = 10201 < 10332 = 63^2 + 63 \cdot 101$, $\varphi > 101/63$, so $\varphi^8 > 101/3 + 13 = 140/3$, that is, $6\varphi^8 > 280 = 5 \cdot 56$.

(f) $\varphi^{-44}\varphi^{44} = \varphi^0 = 1$.

(g) Each $J(c_i) \geq 0$ (Proposition 22.15). If $a, b \geq 0$ and $a + b = 0$, then $0 \leq a \leq a + b = 0$, so $a = 0$ and $b = 0$; by induction on S, a sum of nonnegative points is 0 exactly when each term is. And $J(c) = 0$ exactly when $c = 1$. □

Three steps remain that are neither counts nor numbers.

Proposition 22.51 (Conservation, virtues and the half-period phase)**.**

(a) (C10) An eight-tick signal *is a map from* $\{0,\dots,7\}$ *to the complex numbers* $\mathbb{C}_\delta$ *of the line. Step C10 defines the total charge Z of a signal to be* 0*, so every map of signals conserves it. If a map R of a set and an integer-valued z satisfy* $z(R(s)) = z(s)$ *for every s in a subset P that R maps into itself, then* $z(R^n(s)) = z(s)$ *for every* $s \in P$ *and every n.*

(b) (C11) A moral state *carries a finite ledger of directed bonds with positive multipliers, an agent, a tick and a positive energy. Its* skew *is the sum of the logarithms of the multipliers of the agent's outgoing bonds minus that of its incoming bonds, and a finite list of states is* admissible *when its skews sum to* 0*. A* virtue *is a map of lists of states that sends admissible lists to admissible lists and, on lists whose ticks lie within* 8 *of each other, gives ticks at most* 8 *after every old tick. For every assignment of virtues to the fourteen named virtues of step C11, each of them and every composition of them sends admissible lists to admissible lists. The identity map is a virtue, and the empty list is admissible.*

(c) (C13) $\cos\pi = -1$, $\sin\pi = 0$, $\cos 0 = 1$ *and* $\sin 0 = 0$*. So the eight-tick phase* $\cos(k\pi/4) + \mathrm{i}\sin(k\pi/4)$ *is* -1 *at* $k = 4$ *and* 1 *at* $k = 0$*.*

The proofs are free.

Proof. (a) Both sides are 0. For the second statement, R^n maps P into P, and $z(R^{n+1}(s)) = z(R(R^n(s))) = z(R^n(s))$; induct on n.

(b) Each virtue sends admissible lists to admissible lists by definition. A composition applies the virtues of a list of names in turn; induct on the list. The identity keeps the list and its ticks, and the empty sum is 0.

(c) Theorem 21.8 gives $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$, and the addition formulas for the cosine and the sine (Section 21.2) give $\cos\pi = \cos^2(\pi/2) - \sin^2(\pi/2) = -1$ and $\sin\pi = 2\sin(\pi/2)\cos(\pi/2) = 0$. At 0 the two series give $\cos 0 = 1$ and $\sin 0 = 0$. □

Item (a) is as thin as it reads: step C10, as restated, defines the total charge as 0, and the theorem proves no more. Item (b) is the law that defines a virtue, carried through compositions; the step needs nothing else.

The constants and predictions of the sequence come next. Several of these steps contain measured values; we name them where they enter, and $\alpha_m = 0.0072973525643$ denotes the measured fine-structure constant used in the steps, as a ratio. The count $W = 17$ is the number of wallpaper groups [174], a classical count that enters as a given number.

Theorem 22.52 (The constants and comparisons of the numbered chain)**.** *With no omniscience principle and no choice principle:*

(a) (C16) In native units $\tau_0 = \ell_0 = 1$ *put* $c = \ell_0/\tau_0$, $\hbar = \varphi^{-5}\tau_0$, $G = \varphi^5/(4\log 2)$ *and* $\ell_P = \sqrt{\hbar G/c^3}$*. Then* $c = 1$*;* $\hbar$ *is the integer power* φ^{-5}*, and* $G \cdot 4\log 2$ *the integer power* φ^5*;* $G\hbar = 1/(4\log 2)$*; and* $\ell_P = \sqrt{1/(4\log 2)}$*.*

(b) *(C17) The seed, the cube's total angular deficit* 4π *times* $11 = 12-1$ *(the twelve edges less the one active edge of a tick), is* 44π*; the seam numerator and denominator are* $6 \cdot 17 + 1 = 103$ *and* $6 \cdot 17 = 102$*; and* $137.030 < \alpha_W^{-1} < 137.039$*, with* α_W^{-1} *as in Theorem 22.42(a) at* $s = \sqrt{2}$*.*

(c) *(C18) With* $B = -2 \cdot 11 = -22$*,* $R_0 = 2^2 \cdot 17 - (8-2) = 62$ *and coherence exponent* $2^3 - 3 = 5$*, the electron's structural mass on its rung* $r_e = 2$*,* $2^B \varphi^{-5} \varphi^{R_0} \varphi^{r_e - 8}$*, equals* $2^{-22}\varphi^{51}$ *and is positive.*

(d) *(C20, C24)* $s_W = (3-\varphi)/6$ *satisfies* $0 < s_W < 1/2$ *and* $|s_W - 0.2229| < 0.01$*, and* $0.76 < (3+\varphi)/6 < 0.78$*. With the measured masses of step C24,* $m_H = 125.25$*,* $m_W = 80.3692$ *and* $m_Z = 91.1876$*:* $|m_H/m_W - \varphi| < 0.1$ *and* $m_W/m_Z < 1$*.*

(e) *(C21)* $2/17 = 2/W$*, and* $|2/17 - 0.1179| < 0.0009$*.*

(f) *(C22)* $|\varphi^{-3} - \frac{3}{2}\alpha_m - 0.22500| < 0.00067$*,* $|\frac{1}{24} - 0.04182| < 0.00085$ *and* $|\frac{1}{2}\alpha_m - 0.00369| < 0.00011$*.*

(g) *(C23, C26, C34)* $\Omega_\Lambda = \frac{11}{16} - \alpha_m/\pi$ *satisfies* $0.6851 < \Omega_\Lambda < 0.6852$*; so* $0 < \Omega_\Lambda < 1$*,* $0 < 1 - \Omega_\Lambda < 1$ *and* $|\Omega_\Lambda - 0.6847| < 0.0073$*. Also* $\frac{13}{12} = \frac{12+1}{12}$*, and* $|67.4 \cdot \frac{13}{12} - 73.04|/73.04 < 0.0005$*.*

(h) *(C32)* $b_0 = 11/(2\pi)$ *with* $11 = 3 \cdot 2^2 - 1$*; and with* $\hbar = \varphi^{-5} = e^{-5\log\varphi}$ *and* $\ell_0 = 1$*, every point* r *satisfies* $\hbar/(\ell_0 \varphi^r) = (\hbar/\ell_0)\,\varphi^{-r}$*.*

(i) *(C33)* $\kappa = 2\pi\varphi^5/\log 2$ *is positive, and* $96.9 < \kappa < 102.5$*.*

(j) *(C38)* $\frac{3}{5} = \frac{3}{3+2}$*, and* $a_C = \frac{3}{5}\,\alpha_m \cdot 197.327/1.2$ *satisfies* $0.7 < a_C < 0.75$ *and* $|a_C - 0.714| < 0.036$*.*

Proof. We use the brackets $1.618033 < \varphi < 1.618034$, $3.1415925 < \pi < 3.1415927$, $2.772588722 < 4\log 2 < 2.7725887224$ and $0.4812 < \log\varphi < 0.4813$ of Section 22.9; every other number below is a ratio.

(a) $c = 1/1$. $G \cdot 4\log 2 = \varphi^5$ because $4\log 2 > 0$ is invertible, and $G\hbar = \varphi^5\varphi^{-5}/(4\log 2)$. Then $\hbar G/c^3 = 1/(4\log 2)$, and ℓ_P is its nonnegative square root.

(b) By Proposition 22.39(a) the deficits total 4π, and $4\pi \cdot 11 = 44\pi$; $6 \cdot 17 = 102$. The band is Theorem 22.42(a).

(c) $2^{-22}\varphi^{-5}\varphi^{62}\varphi^{-6} = 2^{-22}\varphi^{51}$, a product of positive points.

(d) $0.2303276 < s_W < 0.2303279$, so $0 < s_W < 1/2$ and $0.0074 < s_W - 0.2229 < 0.0075$; and $0.7696721 < (3+\varphi)/6 < 0.7696724$. The ratio $m_H/m_W = 125.25/80.3692$ lies between 1.5584328 and 1.5584329, so $0.0596 < \varphi - m_H/m_W < 0.0597$; and $80.3692 < 91.1876$.

(e) $0.1179 - 2/17 = 0.0043/17 < 0.0003$.

(f) $\varphi^3 = 2\varphi + 1$ and $(2\varphi+1)(2\varphi-3) = 4(\varphi^2 - \varphi) - 3 = 1$, so $\varphi^{-3} = 2\varphi - 3$, between 0.236066 and 0.236068. With $\frac{3}{2}\alpha_m = 0.01094602884645$, the first difference lies between 0.000119 and 0.000123. Next $\frac{1}{24} - 0.04182 = -0.000153\ldots$ and $\frac{1}{2}\alpha_m - 0.00369 = -0.0000413\ldots$, both ratios.

(g) The bracket for π gives $0.0023228194 < \alpha_m/\pi < 0.0023228196$, so $0.6851771804 < \Omega_\Lambda < 0.6851771806$, and the three bounds follow; $0 < \Omega_\Lambda - 0.6847 < 0.00048$. Finally $67.4 \cdot 13/12 = 73.01666\ldots$, and $73.04 - 73.01666\ldots = 7/300 < 0.0005 \cdot 73.04 = 0.03652$.

(h) $3 \cdot 4 - 1 = 11$. Since $e^a e^{-a} = 1$, $\hbar/(\ell_0 e^{r \log \varphi}) = (\hbar/\ell_0) e^{-r \log \varphi}$.

(i) $\varphi^5 = 5\varphi + 3$, so $\kappa = 2\pi(5\varphi + 3)/\log 2$, a product and quotient of positive points; the brackets give $100.5292 < \kappa < 100.5293$.

(j) $a_C = 0.7199823447\ldots$, a ratio, and $a_C - 0.714 < 0.006$. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ with its φ, π, log and exp; native units; the measured values written into the steps (α_m, 0.2229, 0.1179 ± 0.0009, the three mixing entries with their errors, the boson masses in GeV, the Hubble rates 67.4 and 73.04, 0.6847 ± 0.0073, $\hbar c = 197.327$ MeV fm, $r_0 = 1.2$ fm), taken as ratios.

Conclusion

Each comparison as stated, with no omniscience and no choice.

Identification

Reading $(3 - \varphi)/6$ as the weak mixing angle, $2/17$ as the strong coupling, the three expressions of (f) as mixing-matrix entries, Ω_Λ as the dark-energy fraction, b_0 and the scale law as the running of the strong coupling, κ as Einstein's coupling, and a_C as the Coulomb coefficient of nuclear binding.

Not claimed

The identifications themselves, and the measured values.

Step C19 builds the muon and tau masses from the electron's structural mass by residues on the golden ladder. The ingredients are a gap term $\mathrm{gap}(Z) = \log(1 + Z/\varphi)/\log\varphi$ at the charged-lepton index $Z = 1332$; a shift $2W + (W + E)/(4E_p) + \alpha^2 + E\alpha^3$ with $W = 17$, $E = 12$ edges and $E_p = 11$ passive edges; a step $E_p + 1/(4\pi) - \alpha^2$ from electron to muon; and a step $6 - \frac{1}{2}(2W + 3)\alpha$ from muon to tau, with 6 the cube's faces. Here $\alpha = 1/\alpha_W^{-1}$ is the construction value of C17, the witness value of Theorem 22.42(a), not a derived coupling.

Theorem 22.53 (The lepton window). *Let $\alpha = 1/\alpha_W^{-1}$, with α_W^{-1} as in Theorem 22.42(a) at $s = \sqrt{2}$, and put*

$$\rho_\mu = \mathrm{gap}(1332) - \Big(34 + \frac{29}{44} + \alpha^2 + 12\alpha^3\Big) + \Big(11 + \frac{1}{4\pi} - \alpha^2\Big), \qquad \rho_\tau = \rho_\mu + 6 - \frac{37}{2}\alpha,$$

$m_\mu = 2^{-22}\varphi^{51}\varphi^{\rho_\mu}$ and $m_\tau = 2^{-22}\varphi^{51}\varphi^{\rho_\tau}$, with $\varphi^t = e^{t\log\varphi}$. Then, with $z = \frac{15}{44} + \frac{1}{4\pi} - 2\alpha^2 - 12\alpha^3$,

$$2^{22} m_\mu = (\varphi^{27} + 1332\varphi^{26})\, e^{z\log\varphi}, \qquad 2^{22} m_\tau = (\varphi^{33} + 1332\varphi^{32})\, e^{(z - 37\alpha/2)\log\varphi},$$

and

$$105 < m_\mu < 107, \qquad 1768 < m_\tau < 1792.$$

The proof is free.

Proof. Since $e^{\log y} = y$, $\varphi^{\mathrm{gap}(1332)} = 1 + 1332/\varphi$. The constant terms of ρ_μ are $-34 - \frac{29}{44} + 11 = -24 + \frac{15}{44}$, so $\rho_\mu = \mathrm{gap}(1332) - 24 + z$ and

$$2^{22} m_\mu = \varphi^{51}\Big(1 + \frac{1332}{\varphi}\Big)\varphi^{-24} e^{z\log\varphi} = (\varphi^{27} + 1332\varphi^{26})\, e^{z\log\varphi}.$$

Likewise $\rho_\tau = \mathrm{gap}(1332) - 18 + z - \frac{37}{2}\alpha$, which gives the second identity. By Proposition 22.37(d), with $F_{25}, \dots, F_{33} = 75025$, 121393, 196418, ..., 2178309, 3524578,

$$\varphi^{27} + 1332\varphi^{26} = 161891894\,\varphi + 100054693,$$
$$\varphi^{33} + 1332\varphi^{32} = 2905032166\,\varphi + 1795408617,$$

and $1.618033 < \varphi < 1.618034$ puts the first between 362001119 and 362001282 and the second between 6495846527 and 6495849433.

From $137.030 < 1/\alpha < 137.039$, $0.007297 < \alpha < 0.007298$, so $0 < 2\alpha^2 + 12\alpha^3 < 0.000112$ and $0.1349945 < \frac{37}{2}\alpha < 0.135013$. With $0.0795774 < 1/(4\pi) < 0.0795775$ and $0.3409090 < \frac{15}{44} < 0.3409091$,

$$0.4203744 < z < 0.4204866, \qquad 0.2853614 < z - \tfrac{37}{2}\alpha < 0.2854921.$$

Multiplying by $0.4812 < \log\varphi < 0.4813$, $y = z\log\varphi$ lies in $(0.20228, 0.20239)$ and $y' = (z - \frac{37}{2}\alpha)\log\varphi$ in $(0.13731, 0.13741)$. For $0 \le x \le 1$, $1 + x + x^2/2 + x^3/6 \le e^x \le 1 + x + x^2/2 + x^3/3$ (Section 22.9), which gives $1.2241 < e^y < 1.2257$ and $1.1471 < e^{y'} < 1.1478$. With $2^{22} = 4194304$,

$$105.64 < \frac{362001119 \cdot 1.2241}{4194304} < m_\mu < \frac{362001282 \cdot 1.2257}{4194304} < 105.79,$$

$$1776.5 < \frac{6495846527 \cdot 1.1471}{4194304} < m_\tau < \frac{6495849433 \cdot 1.1478}{4194304} < 1777.7.$$

□

Reading m_μ and m_τ as the muon and tau masses in MeV, and α as the fine-structure constant, are identifications; the theorem uses neither, and by step S14 the coupling itself is a boundary datum. The theorem shows that the two formulas of step C19, evaluated on the line, land in the windows the step states for them, with no principle.

22.11 Where the prices sit

Ten steps have an exact price, each beside a free reading of the same step.

1. A ratio other than 1 has positive cost (T2.8): MP, by Theorem 22.16. Free beside it: $J(x) > 0$ for $x \# 1$.

2. The route of step T3.7 to the first costly join, an integer pairing that separates its parts: strong extensionality for integer-valued maps, $\mathrm{SE}_{\mathbb{Z}}$, by Theorem 22.24(c). MP implies it, and with the WLPO function it is MP (Proposition 22.25); whether it implies MP outright is open. Free beside it: the join of apart parts.

3. The costly join of merely distinct parts (T3.7b): MP, by Theorem 22.24(b).

4. The reciprocal ambiguity of equal costs (R2-08): LLPO, by Theorem 22.31. Free beside it: the double-negated, apartness and converse forms.

5. Omitting the output bins of probability zero from the data-processing inequality (R2-20): LPO, by Theorem 22.32. Free beside it: the inequality bin by bin and over any known support.

6. The largest admissible level gap as a value (R2-27): WLPO, by Theorem 22.33. Free beside it: the bound $k \log \varphi \leq \log(1 + c + s)$.

7. The constants' approximate-equality test $|a - b| \leq 10^{-6}$ as a yes-or-no decision (S11): WLPO, and as one map the WLPO function, by Theorem 22.44. Free beside it: the identities of the test; and at the values of step S11 the check returns false.

8. The total real-valued meter type (T2.1b): the LPO function, by Theorem 22.63. Free beside it: the meter at positive references.

9. The rigid form of DEP for every point other than 1 (T3.7c): the WLPO function, by Theorem 22.64. Free beside it: the join of apart parts.

10. The first homology of the circle in data form (R2-35d): the WLPO function, by Theorem 22.65. Free beside it: the relation form R2-35.

One step is bracketed, with its exact place open: dropping continuity and the order premises (T2.6) lies below the WLPO function (Theorem 22.14); no lower bound is proved.

Every price sits in one of three places. Five are decision forms of facts whose apartness or double-negation form is free (T2.8, R2-08, R2-20, R2-27, S11). Two concern real-valued object types or a countermodel to a dropped premise (T2.1b, T2.6). Four are readings of a premise weaker than apartness (T3.7, T3.7b, T3.7c and the data form of R2-35). None of them is the mathematical content of a step: every forced structure of the chain holds on the line with no omniscience and no choice. All eleven belong to the rungs: the first rung's segments T2 and T3, the second rung, and the constants. None is a step of the numbered chain, which is free from T−2 to T41 and through the period bound (Theorem 22.45). Table 22.1 gives all 130 rows: 119 free, 10 with an exact price, 1 bracketed, none open.

Table 22.1. The forcing chain of Recognition Science, restated and priced on $\mathbb{R}_\delta$, in 130 rows. Column 2 restates each step in this book's own words; the prices are theorems about these statements. The rows run in the order of the chain: the first rung (T1–T4), the numbered chain (C−2 to C41 and the period bound P8, where Cn is step Tn of that sequence; Section 22.10), the second rung (R2-01 to R2-35, the last in relation and data form) and the constants (S1–S15). "Free" means proved with no omniscience principle and no choice principle; a named principle is an exact price, an equivalence proved in both directions with no choice; "between" and "below" are brackets (Definition 22.1). The last column names the result of this book that settles the step, or the elementary argument that does. Counts: 119 free, 10 exact prices, 1 bracketed, 0 open.

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
T1: the arithmetic of records (8 steps, 8 free)			
T1.1	Records of repeated distinction form an initial structure: the orbit maps uniquely into every structure with a point and a self-map.	free	Thm 2.3
T1.2	Counts, signed counts and ratios obey the laws of $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$.	free	Thms 3.3, 4.5, 4.12
T1.3	The arithmetic of records supports a fragment of Heyting arithmetic and, by Ackermann codes, a model of adjunctive set theory with extensionality, in which $\{\emptyset\} \neq \emptyset$.	free	Ackermann coding
T1.4	Hereditarily finite codes, and the number two, are derived from records, not posited.	free	Ackermann coding
T1.5	The arithmetic does not depend on the carrier that holds the records.	free	Thm 2.7
T1.6	Two independent carriers, the natural numbers and the lists over a one-element set, give the same arithmetic.	free	Thms 2.7, 3.14
T1.7	The hereditarily finite model is definable in the arithmetic of records: every code is the value of a closed numeral.	free	Ackermann coding
T1.8	A formula is valid on the natural numbers exactly when it is valid on the lists over a one-element set.	free	transport by isomorphism
T2: the cost (12 steps, 9 free)			
T2.1	A meter that pools, is nonnegative and reads its reference as 1 reads d/r.	free	Thm 22.5
T2.1b	The meter is one real-valued map of two real arguments.	the LPO function	Thms 22.6, 22.63
T2.2	A balanced zero, an empty operand and a positive reference force opposite operand images, energy $k(A-B)^2$ with $k>0$ and reference κkAB.	free	Thm 22.8
T2.2b	Reports at ratios, reciprocity and order on $[1,\infty)$ force $(\kappa/2)F = J$ without continuity.	free	Thm 22.12

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
T2.3a	A source witness, the eight-slot pulse with its half-sum residual, meets the selection premises.	free; identification classical	Rem. 22.9
T2.3b	The formation residual acts on the operands as $1+\varphi^{-2}$ times the source residual.	free; identification classical	Rem. 22.9
T2.4	Literal reports at ratios a/b are $(2/\kappa)J(a/b)$.	free	Thm 22.11
T2.5	Continuity at positive points extends this to every $x>0$; at $\kappa=2$, $F=J$.	free	Thm 22.11
T2.6	Continuity cannot be dropped: a report can be altered at an irrational point.	below the WLPO function	Thm 22.14
T2.7	$J(x)=(x-1)^2/(2x)$; $J\geq 0$; $J(x)=0$ iff $x=1$; $J(1/x)=J(x)$; $J(x)>0$ iff $x\#1$; J increases on $[1,\infty)$.	free	Prop. 22.15
T2.8	$x\neq 1$ gives $J(x)>0$.	MP; the decisions of $J=0$ cost WLPO and LPO	Thm 22.16
T2.9	The composition law with its calibration forces J.	free	Thm 21.14
T3: the golden hierarchy (10 steps, 7 free)			
T3.1	A costly initial join posts additively: $s_2=s_1+s_0$.	free	Lem. 22.18
T3.2	Positive comparisons fix one scalar between two additive readings.	free	Lem. 22.19
T3.3	One rule holds along the orbit: $r(Ts)=\rho r(s)$ with $\rho>0$.	free	Thm 22.20(a)
T3.4	The step factor satisfies $\lambda^2=\lambda+1$, so $\lambda=\varphi$.	free	Thm 22.20(b)
T3.5	The realized hierarchy exists.	free	Thm 22.20(b), Prop. 22.21
T3.6	Levels are φ^n times the base, and weights are φ^{-n}.	free	Thm 22.20(c)
T3.7	The join is a recognition, so it costs something; the route: an integer pairing separates its parts.	$\mathrm{SE}_{\mathbb{Z}}$, strong extensionality (implied by MP; MP given the WLPO function)	Thm 22.24(c), Prop. 22.25
T3.7a	The same step with apart parts, $\rho\#1$.	free	Thm 22.24(a)
T3.7b	The same step with distinct parts, $\rho\neq 1$.	MP	Thm 22.24(b)
T3.7c	Every point $x\neq 1$ carries DEP on its rigid kinematics.	the WLPO function	Prop. 22.25, Thm 22.64
T4: eight states and three dimensions (6 steps, 6 free)			
T4.1	DEP separates the pair from the split.	free	Prop. 22.26
T4.2	The winding kinematics keeps the pair.	free	Prop. 22.26
T4.3	An erasable world keeps no pair.	free	Prop. 22.26

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
T4.4	Detecting nontrivial linking forces $D = 3$.	free	Ch. 23
T4.5	Linking is detected at $D = 3$: a loop linking the equatorial circle of the 3-sphere bounds nothing in its complement.	free	Ch. 23
T4.6	Given $D = 3$, $2^D = 8$: exactly eight binary states.	free	Prop. 22.27
The numbered chain, T−2 to T41 and the period bound (43 rows, 43 free)			
C−2	A set with no element is not a set with one element.	free	Prop. 22.47(a)
C−1	No proposition equals its negation; two different propositions exist; a set has two different elements iff some property holds of one element and fails of another.	free	Prop. 22.47(b)
C0	The floor cost $C(e) = 0$, $C(m) = 1$ is a cost: zero exactly on the consistent state, positive exactly on the inconsistent one, additive over independent joins, fixed on them by its values on the parts.	free	Prop. 22.47(c)
C1	An inconsistent state has positive cost; a state of cost 0 is consistent; $C(m) > 0$.	free	Prop. 22.47(d)
C2	The floor has exactly two states; cost 0 selects e, positive cost selects m.	free	Prop. 22.47(e)
C3	The empty state is neutral for join and for cost; independent joins add costs.	free	Prop. 22.47(f)
C4	The floor is the two-point floor normalized by 0 and 1; it carries a recognition structure and a recognition.	free	Prop. 22.47(g)
C5 (J is unique) is rows T2.9 and T2.2b; C6 (φ is the unique positive root) is rows S1 and T3.4.			
C7	$2^3 = 8$, the number of states of three bits (a count, not a period); $2^D = 8$ iff $D = 3$.	free	Props. 22.48(a), 22.27
C8	The linking condition $D - 2 = 1$ gives $D = 3$; $2^D = 8$ gives $D = 3$; exactly one D meets the compatibility conditions. The topological form is T4.4.	free	Props. 22.48(a), 22.49(a)
C9	A loop of positive entries has $\sum_i J(c_i) = 0$ iff every entry is 1; linking forces $D = 3$ (T4.4).	free	Prop. 22.50(g); Ch. 23
C10	The total charge Z of an eight-tick signal is defined as 0, so every evolution conserves it; what each step conserves, n steps conserve.	free	Prop. 22.51(a)
C11	For every assignment of virtues (maps keeping total skew 0) to the fourteen names, each of them and every composition keeps total skew 0.	free	Prop. 22.51(b)

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
C12	The cube's 48 symmetries (R2-28); the recorded gauge numbers are $(3, 2, 1)$, sum to the cube's $6 = 2 \cdot 3$ faces, and are not $(4, 3, 2)$.	free	Props. 22.48(b), 24.44(d)
C13	The eight-tick phase is -1 at $k = 4$ and 1 at $k = 0$: $\cos \pi = -1$, $\sin \pi = 0$.	free	Prop. 22.51(c)
C14	The cube has 3 pairs of opposite faces, not 2 or 4.	free	Prop. 22.48(c)
C15	$\varphi > 1$, $\varphi^2 = \varphi + 1$, and $y\varphi^{r-8+g} > 0$ for every $y > 0$.	free	Prop. 22.50(a)
C16	With $\tau_0 = \ell_0 = 1$: $c = 1$, $\hbar = \varphi^{-5}$, $G \cdot 4 \log 2 = \varphi^5$, $G\hbar = 1/(4 \log 2)$, $\ell_P = \sqrt{1/(4 \log 2)}$.	free	Thm 22.52(a)
C17	The seed $4\pi \cdot 11 = 44\pi$; the seam $103/102$ from $6 \cdot 17$; $137.030 < \alpha_W^{-1} < 137.039$.	free	Thm 22.52(b)
C18	The electron sits on rung 2, with structural mass $2^{-22}\varphi^{51} > 0$.	free	Thm 22.52(c)
C19	The muon and tau formulas give $105 < m_\mu < 107$ and $1768 < m_\tau < 1792$.	free	Thm 22.53
C20	$\lvert(3 - \varphi)/6 - 0.2229\rvert < 0.01$.	free	Thm 22.52(d)
C21	$2/17 = 2/W$ with $W = 17$, and $\lvert 2/17 - 0.1179\rvert < 0.0009$.	free	Thm 22.52(e)
C22	$\lvert\varphi^{-3} - \frac{3}{2}\alpha_m - 0.225\rvert < 0.00067$, $\lvert\frac{1}{24} - 0.04182\rvert < 0.00085$, $\lvert\frac{1}{2}\alpha_m - 0.00369\rvert < 0.00011$.	free	Thm 22.52(f)
C23	$\Omega_\Lambda = \frac{11}{16} - \alpha_m/\pi$ lies in $(0, 1)$.	free	Thm 22.52(g)
C24	$0 < \frac{3-\varphi}{6} < \frac{1}{2}$, $0.76 < \frac{3+\varphi}{6} < 0.78$; with the measured masses, $\lvert m_H/m_W - \varphi\rvert < 0.1$ and $m_W < m_Z$.	free	Thm 22.52(d)
C25	$m_e = m(2) > 0$, and $m(r_p)/m_e = \varphi^{r_p - 2}$.	free	Prop. 22.50(c)
C26	$\frac{13}{12} = \frac{12+1}{12}$; $67.4 \cdot \frac{13}{12}$ is within 0.05% of 73.04; $\lvert\Omega_\Lambda - 0.6847\rvert < 0.0073$.	free	Thm 22.52(g)
C27	The number of generations, defined as the cube's face pairs, is 3 and not 4.	free	Prop. 22.48(c)
C28	Both mixing matrices have the size the step defines, the face-pair count 3.	free	Prop. 22.48(c)
C29	$r < s$ gives $m(r) < m(s)$, and every $m(r)$ is positive.	free	Prop. 22.50(b)
C30	The valence mass $3m(4)$ and binding mass $m(14)$ are positive, and $m(14) > 40 \cdot 3m(4)$.	free	Prop. 22.50(d)
C31	$1 < 56/\varphi^8 < 1.2$; $56 = 7 \cdot 8$; $50 \le 56 \le 70$.	free	Props. 22.50(e), 22.48(d)
C32	$b_0 = 11/(2\pi)$ with $11 = 3 \cdot 2^2 - 1$; $\hbar/(\ell_0\varphi^r) = (\hbar/\ell_0)\varphi^{-r}$ for every point r.	free	Thm 22.52(h)
C33	$\kappa = 2\pi\varphi^5/\log 2 > 0$, and $96.9 < \kappa < 102.5$.	free	Thm 22.52(i)
C34	$0 < 1 - \Omega_\Lambda < 1$, and $\lvert\Omega_\Lambda - 0.6847\rvert < 0.0073$.	free	Thm 22.52(g)

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
C35	The quark residues $\frac{23}{4} - \frac{31}{4} = -2$, then $-\frac{9}{2}$ and -10; integer rungs are the core convention, the quarter ladder a hypothesis.	free	Prop. 22.48(e)
C36	$J(1) = 0$; $J \geq 0$; $J(x) = 0$ iff $x = 1$; $J(x) = (x-1)^2/(2x)$.	free	Prop. 22.15
C37	$\eta = \varphi^{-44}$ is positive.	free	Prop. 22.50(f)
C38	$\frac{3}{5} = \frac{3}{3+2}$; $a_C = \frac{3}{5}\alpha_m \cdot 197.327/1.2$ has $0.7 < a_C < 0.75$ and $\lvert a_C - 0.714\rvert < 0.036$.	free	Thm 22.52(j)
C39	Subshells ordered by $n + l$, then n, with capacities $2(2l+1)$: 19 running totals, the last 118; element 1 is in an s subshell.	free	Prop. 22.48(f)
C40	Four bases; $4^3 = 64$; $3 \cdot 8 - 4 = 20$.	free	Prop. 22.48(g)
C41	$2^{3+1} = 16$ modes, $3 + 2 = 5$ active, $16 - 5 = 11$ passive.	free	Prop. 22.48(h)
P8	A periodic walk through all eight corners of the cube has period at least 8.	free	Prop. 22.49(b)
Second rung (36 rows for 35 groups, 32 free)			
R2-01	Lossless storage of M messages in words of length n over k letters needs $M \leq k^n$.	free	Prop. 22.28
R2-02	Erasure is final: what processing merges stays merged, and alternatives told apart later were told apart before.	free	composition of maps
R2-03	Keeping histories needs capacity; a store over $k \geq 2$ symbols cannot hold its own history.	free	Prop. 22.28
R2-04	The relative cost of a pair a, b is $J(a/b)$, unchanged by a common scale.	free	Prop. 22.29(a)
R2-05	In log form $J(e^t) = \cosh t - 1 = \sum_{k\geq 1} t^{2k}/(2k)!$.	free	Prop. 22.34(a)
R2-06	J has its unique minimum 0 at 1 and is strictly convex, also in log form; Taylor's formula with exact cubic remainder.	free	Props. 22.15, 22.34(b)
R2-07	A budget c confines x to $[1 + c - s, 1 + c + s]$, $s = \sqrt{c(c+2)}$.	free	Prop. 22.29(b)
R2-08	$J(x) = J(y)$ gives $x = y$ or $xy = 1$.	LLPO; the $\neg\neg$, apartness and converse forms free	Thm 22.31
R2-09	$\lvert \log x - \log y\rvert$ is a metric; J fails the triangle inequality.	free	Prop. 22.34(c)
R2-10	J is quadratic near 1, with explicit two-sided bounds.	free	Prop. 22.34(d)
R2-11	$J(xy) + J(x/y) = 2J(x)J(y) + 2J(x) + 2J(y)$.	free	Prop. 22.29(c)
R2-12	$S = (x - 1/x)/2$ has $S^2 = J(J+2)$; composition keeps orientation.	free	Prop. 22.29(c)
R2-13	The costs of powers follow the Chebyshev recurrence.	free	Prop. 22.34(e)

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
R2-14	Budgets compose: a product of budgeted steps stays within the summed log budget.	free	Prop. 22.34(f)
R2-15	Mixing has a nonnegative Jensen gap (midpoint formula).	free	Prop. 22.29(d)
R2-16	Equal log steps minimize total cost at a fixed product: $nJ(r) \le \sum_i J(x_i)$ when $\prod_i x_i = r^n$.	free	Prop. 22.34(g)
R2-17	Two parties meet most cheaply at the geometric mean.	free	Prop. 22.29(e)
R2-18	The AM–GM discrepancy is a cost.	free	Prop. 22.29(f)
R2-19	$\sum_i p_i J(p_i/q_i) = \frac{1}{2}\sum_i (p_i - q_i)^2/q_i$, zero iff $p = q$.	free	Prop. 22.30
R2-20	Data processing does not increase the divergence, omitting the output bins of probability zero.	LPO for omitting the bins; the inequality free	Thm 22.32
R2-21	$2\,\mathrm{TV}^2 \le D$, sharply.	free	Prop. 22.30
R2-22	$\varphi^{n+2} = \varphi^{n+1} + \varphi^n$ and $\varphi^{n+1} = F_{n+1}\varphi + F_n$.	free	Prop. 22.37(d)
R2-23	φ is not a ratio of counts: $b\varphi - a \# 0$.	free	Prop. 22.37(b)
R2-24	A gap of k levels costs $J(\varphi^k) = \cosh(k \log \varphi) - 1$; $J(\varphi) = \varphi - \frac{3}{2}$, $J(\varphi^2) = \frac{1}{2}$.	free	Prop. 22.35(a)
R2-25	The golden weights φ^{-n} refine and sum to φ^2.	free	Prop. 22.35(b)
R2-26	The normalized golden distribution: total, tails, mean φ, variance φ^3, entropy $(\varphi + 2)\log\varphi$.	free	Prop. 22.35(c)
R2-27	A budget bounds the gap, $k \log \varphi \le \log(1 + c + s)$; the largest admissible gap as a value.	WLPO for the value; the bound free	Thm 22.33
R2-28	The cube has 8 vertices, 12 edges, 6 square faces and 1 cell; its 48 symmetries are the signed permutations.	free	Props. 22.36(a), 24.44(d)
R2-29	Graph distance on the cube is Hamming distance; closed walks have even length.	free	Prop. 22.36(b)
R2-30	A pass through all states needs eight slots; the Gray cycle changes one bit per step.	free	Prop. 22.36(c)
R2-31	n product records have 8^n states and need $3n$ bits.	free	Prop. 22.36(d)
R2-32	Parity modes diagonalize the cube's Laplacian; signals expand uniquely in them.	free	Prop. 22.36(e)
R2-33	The eight-tick clock C and shift S satisfy the Weyl relation with $\omega^8 = 1$, $\omega^2 = i$.	free	Prop. 22.36(f)
R2-34	An invariant obstruction separates kept from erased pairs.	free; its decision form for maps into $\mathbb{R}_\delta$ is WLPO	Props. 22.26, 22.36(g)
R2-35	H_1 of the circle is $\mathbb{Z}$ as a relation: every cycle has one integer winding, homologous cycles share it, winding 0 bounds, and a cycle winds once.	free	Ch. 23

continued on the next page

Table 22.1, continued

Step	Statement (restated)	Price on $\mathbb{R}_\delta$	Settled by
R2-35d	The same with the winding one function on all chains of the circle.	the WLPO function	Ch. 23, Thm 22.65
The constants (15 steps, 14 free)			
S1	$\varphi^2 = \varphi + 1$, $1 < \varphi < 2$, and φ is irrational.	free	Prop. 22.37(a, b)
S2	$\varphi^n = F_n\varphi + F_{n-1}$ and decimal bounds such as $1.61 < \varphi < 1.62$ and $10.7 < \varphi^5 < 11.3$.	free	Prop. 22.37(c, d)
S3	The locked constants $\alpha_{\text{lock}} = (1 - 1/\varphi)/2$ and $C_{\text{lock}} = \varphi^{-5}$.	free	Prop. 22.38
S4	$\hbar = \varphi^{-5}$ with $\tau_0 = 1$, and $0.088 < \hbar < 0.093$.	free	Prop. 22.38
S5	The cube's eight angular deficits total 4π.	free	Prop. 22.39(a)
S6	The curvature cost is $2\lambda^2$, the coefficient being $\chi(S^2) = 2$.	free; the quadratic form an identification	Prop. 22.39(b)
S7	$\lambda_0 = 1/\sqrt{2}$ is the unique positive root of $2\lambda^2 = 1$.	free	Prop. 22.39(c)
S8	$G = \varphi^5/(4\log 2)$, with $2.7725887 < 4\log 2 < 2.7725888$.	free	Prop. 22.40(a, b)
S9	$\kappa = 8\pi G = 2\pi\varphi^5/\log 2$ and $\kappa\hbar = 2\pi/\log 2$.	free	Prop. 22.40(b)
S10	$k_R = \log\varphi$, with $0.47 < k_R < 0.49$.	free	Prop. 22.40(c)
S11	$K = \sqrt{\varphi}$, and the units test $\lvert a - b\rvert \le 10^{-6}$ as a yes-or-no check.	WLPO (as one map, the WLPO function); the identities free	Thm 22.44
S12	$4\pi\varphi^5 - 5\log\varphi$ lies in $(136.947, 136.959)$.	free	Thm 22.41
S13	$44\pi\exp(-f/(44\pi))$ lies in $(137.030, 137.039)$, and the exponential form is unique.	free; the reading as α^{-1} an identification	Thm 22.42
S14	α is a boundary datum: no closure blind to the coupling pins it.	free	Prop. 22.43(a)
S15	No nontrivial dimensionless monomial $c^a\hbar^b G^d$; SI values need an anchor.	free	Prop. 22.43(b, c)

22.12 The verdict with no cited premise

Recognition Science gathers its first rung into one theorem, its verdict: from a list of premises about the world, the cost of a comparison is J, the levels of a recognition hierarchy are powers of φ, space has three dimensions and the recognition period is eight. The proof of that theorem used one result of transcendental number theory without proving it, the six exponentials theorem, in a form adapted to the cost. This section proves that result, so the verdict holds from its premises alone. The verdict is a theorem about the classical real numbers; it is proved in classical mathematics, and its price on $\mathbb{R}_\delta$ is not determined here. The last part of the section says how far its transcendence step reaches on the line.

Definition 22.54 (The premises and conclusions of the verdict)**.** In classical mathematics, let F be a real function on the positive reals, κ a real number, D a natural number, and $s_0, s_1, \dots$ the levels of a recognition hierarchy, read by $s_n = r(T^n b)$ from a base state b, a self-map T and a reading r of the states: positive reals with $s_1/s_0 > 1$, equal adjacent ratios $s_{n+2}/s_{n+1} = s_{n+1}/s_n$, and the additive posting $s_2 = s_1 + s_0$. For $c > 0$ write $J_c(x) = J(x^c)$. The *premises* are:

(V1) $\kappa > 0$;

(V2) $F(xy) + F(x/y) = \kappa F(x)F(y) + 2F(x) + 2F(y)$ for all $x, y > 0$;

(V3) $F(x) \neq F(y)$ for some $x, y > 0$;

(V4) $t \mapsto F(e^t) + 1$ is nondecreasing on $[0, \infty)$;

(V5) *(the cost is native to the carrier)* for every $c > 0$ with $\frac{\kappa}{2}F = J_c$ on the positive reals, J_c maps positive rationals to rationals;

(V6) *(its character is an automorphism)* for every such c, every positive rational is r^c for some positive rational r;

(V7) the hierarchy above exists;

(V8) *(a completed recognition is kept)* some spatial realization in dimension D satisfies the deformation-erasure principle of Definition 22.22; a spatial realization is a pair kinematics together with the identification that a deformation-invariant integer pairing nonzero on the pair exists exactly when some embedded circle in the sphere S^D has a complement with nonzero first homology.

The *conclusions* are: $\frac{\kappa}{2}F(x) = J(x)$ for every $x > 0$; $s_n = \varphi^n s_0$; φ is the only positive root of $r^2 = r + 1$; $D = 3$; $2^D = 8$; and $s_0/s_n = \varphi^{-n}$.

Premises (V2)–(V4) are the cost laws of Section 21.3 with an amplitude κ, (V5) and (V6) are the two qualitative calibration premises of the carrier, (V7) is the hierarchy of Section 22.6, and (V8) is the linking premise of Chapter 23. The cited result entered between (V5), (V6) and the conclusion $c = 1$.

Theorem 22.55 (The six exponentials theorem)**.** *In classical mathematics:*

1. *If d is real and 2^d, 3^d and 5^d are rational, then d is rational.*
2. Trace form. *If c is real and $n^c + n^{-c}$ is rational for $n = 2, 3, 4, 5$, then c is rational.*

Both parts are cases of the six exponentials theorem of Lang and Ramachandra [204, 205]: if x_1, x_2 and y_1, y_2, y_3 are each linearly independent over the rationals, one of the six numbers $e^{x_i y_j}$ is transcendental; here $x = (1, d)$, or $(1, c)$, and $y = (\log 2, \log 3, \log 5)$, and in part 2 each n^c is algebraic, a root of $X^2 - tX + 1$ with t rational. That two bases would suffice is the open four exponentials problem [206], so, with present knowledge, a proof must use all three. Because every number here is real, Rolle's theorem replaces both the Schwarz lemma and the algebraic zero estimate of the textbook proof. The argument uses four tools.

Lemma 22.56 (The four tools)**.**

1. Siegel's lemma. *Let a be an $M \times N$ matrix of integers of size at most A, and B a natural number with $(2NAB+1)^M < (B+1)^N$. Then some nonzero integer vector v with entries of size at most B has $av = 0$. If $N \geq 2M$ and $N \geq 1$, the size $B = 2NA + 1$ always works. The proof is free.*

2. Extrapolation. *In classical mathematics, if f is a smooth function on the line that vanishes at N distinct points of $[0, R]$, and $|f^{(N)}| \leq M$ on $[0, R]$, then $|f(x)| \leq MR^N/N!$ for $0 \leq x \leq R$.*

3. Zero estimate. *In classical mathematics, an exponential sum $\sum_{i \in S} a_i e^{\lambda_i t}$ with distinct real frequencies λ_i that vanishes at $|S|$ distinct real points has every $a_i = 0$.*

4. Growth. *For every $c > 0$ and every sequence f with $0 \leq f(n) \leq n^a G^{n^5}$ for all $n \geq 1$, for some fixed a and $G \geq 1$, some $n \geq 1$ has $f(n)\,(9n^5 c)^{n^6} < (n^6)!$.*

Proof. 1. Consider the $(B+1)^N$ vectors with entries in $\{0, \dots, B\}$. Each entry of av lies in an interval of at most $2NAB + 1$ integers, so av takes at most $(2NAB+1)^M$ values, and two vectors $v \neq v'$ have $av = av'$; their difference works. The count is a finite comparison of explicit lists. For $N \geq 2M$ and $B = 2NA + 1$: $2NAB + 1 \leq B^2 + 1 < (B+1)^2$, so $(2NAB+1)^M < (B+1)^{2M} \leq (B+1)^N$.

2. Let $z_1 < \cdots < z_N$ be the zeros and x a point of $[0, R]$ not among them; choose λ with $g = f - \lambda \prod_j (t - z_j)$ zero at x. Then g has $N+1$ zeros in $[0, R]$, and N applications of Rolle's theorem give ξ with $g^{(N)}(\xi) = 0$, that is $f^{(N)}(\xi) = \lambda N!$. So $|f(x)| = |\lambda| \prod |x - z_j| \leq MR^N/N!$.

3. Induction on $|S|$. Multiply by $e^{-\lambda_0 t}$ and differentiate: the derivative is an exponential sum with $|S| - 1$ terms and distinct frequencies $\lambda_i - \lambda_0$, coefficients $a_i(\lambda_i - \lambda_0)$, and by Rolle it vanishes at $|S| - 1$ distinct points; so $a_i = 0$ for $i \neq 0$, and then $a_0 = 0$.

4. $\log f(n) + n^6 \log(9n^5 c) \leq a \log n + n^5 \log G + n^6 \log(9c) + 5n^6 \log n$, while $\log(n^6)! \geq n^6(6 \log n - 1)$. The difference grows like $n^6 \log n$. □

Proof of Theorem 22.55. Part 1. Suppose d is irrational and $p^d = r_p$ is rational for $p = 2, 3, 5$, with height H (the largest numerator or denominator). Choose n by Lemma 22.56, item 4, with $c = (1 + |d|) \log 30$ and the explicit bounds below; put $K = 2n^3$, $S = n^2$ and $S' = 2n^2$.

Siegel step. Look for integers a_k, $k = (k_1, k_2)$ with $k_1, k_2 < K$, not all 0, such that

$$\Phi(x) = \sum_k a_k \, e^{(k_1 + k_2 d)x}$$

vanishes at the n^6 points $x = \log m$, $m = 2^{s_0} 3^{s_1} 5^{s_2}$, $s_j < S$. At such a point $\Phi(\log m) = \sum_k a_k m^{k_1} (m^d)^{k_2}$ with $m^d = r_2^{s_0} r_3^{s_1} r_5^{s_2}$ rational; clearing the denominators gives n^6 integer equations in $K^2 = 4n^6$ unknowns, with entries of size at most $A = 30^{2n^5} H^{6n^5}$, and Siegel's lemma, with $B = 2 \cdot 4n^6 A + 1$, gives a solution with $|a_k| \leq 8n^6 \, 30^{2n^5} H^{6n^5} + 1$.

Extrapolation. Φ has $N = n^6$ zeros in $[0, R]$, $R = S' \log 30$, and $|\Phi^{(N)}| \leq K^2 \max |a_k| \, \Lambda^N e^{\Lambda R}$ with $\Lambda = K(1 + |d|)$. By the extrapolation tool, $|\Phi(x)| \leq K^2 \max |a_k| \, \Lambda^N e^{\Lambda R} R^N / N!$ on $[0, R]$.

Liouville step. At a point $\log m$ of the larger grid $s_j < S'$, Φ is a rational with denominator at most H^{12n^5}, so it is 0 or at least H^{-12n^5} in size. The growth tool makes the extrapolation bound smaller, so Φ vanishes at all $8n^6$ points of the larger grid.

Zero estimate. The frequencies $k_1 + k_2 d$ are $4n^6$ distinct reals, because d is irrational, and the $8n^6$ points are distinct, by unique factorization. The zero estimate gives every $a_k = 0$, a contradiction.

Part 2. Now $u_p = p^c$ satisfies $u_p + u_p^{-1} = t_p$ rational, so the values $\Phi(\log m)$ lie in the field generated by u_2, u_3, u_5, not in the rationals. Two repairs, with no field theory, restore the argument. Scale $U_p = \mathrm{den}(t_p) u_p$, a root of an integer quadratic $U^2 = TU - Q$; every monomial $\prod_p U_p^{l_p}$ has explicit integer coordinates in the eight products $\prod_p U_p^{e_p}$, $e \in \{0,1\}^3$, from the recursion $U^l = A_l U + B_l$, $A_{l+1} = TA_l + B_l$, $B_{l+1} = -QA_l$. In the Siegel step all eight coordinates are set to 0 at each of the n^6 points: $8n^6$ equations in $9n^6$ unknowns, with $K = 3n^3$. Here there are fewer than twice as many unknowns as equations, so Siegel's lemma is used with its general condition. For M equations and N unknowns with $8N = 9M$, and entries of size at most $A \geq 1$, the size $B = (2NA)^8$ works: $2NAB + 1 = (2NA)^9 + 1$, and with $x = 2NA \geq 1$,

$$(x^9 + 1)^8 = x^{72}(1 + x^{-9})^8 < x^{72}(1 + x^{-8})^9 = (x^8 + 1)^9,$$

so $(2NAB + 1)^M < (B + 1)^N$. The entries here are of size at most $n^a G^{n^5}$ for fixed a and G, so B is too, after enlarging a and G, and the growth tool applies as in part 1. In the Liouville step a nonzero value w with integer coordinates satisfies a nonzero integer polynomial of degree at most 8, found by Siegel's lemma on the coordinates of $1, w, \ldots, w^8$ (8 equations, 9 unknowns, the same size $B = (2NA)^8$ with $N = 9$); its lowest nonzero coefficient gives $1 \leq |w| \cdot 8B \max(1, |w|)^8$, a lower bound for $|w|$. No conjugate and no norm is used, so the cases in which some u_p is rational, or the square roots are dependent, need no separate treatment. The extrapolation runs on $[0, 3n^2 \log 30]$; the growth tool makes Φ vanish at all $27n^6$ points with $s_j < 3n^2$; and the zero estimate, with $9n^6$ distinct frequencies, kills every coefficient. Rationality at $n = 4$ is not used. □

The verdict needs one more step from the exponent. For $c > 0$, (V5) says that $J(q^c)$ is rational at every positive rational q, and (V6) that every positive rational is a rational cth power. The two premises leave different sets of exponents.

Proposition 22.57 (The exponent census)**.** *In classical mathematics, for $c > 0$:*

1. *J_c maps positive rationals to rationals exactly when c is a positive integer;*
2. *every positive rational is a rational cth power exactly when $c = 1/k$ for a positive integer k;*
3. *so both hold exactly when $c = 1$. Only part 1 uses the trace form of Theorem 22.55; part 2 uses its part 1, and the failure of part 1 of this proposition at $c = 1/k$ for $k \geq 2$ uses neither.*

Proof. We use one arithmetic step. *If a/b is a positive rational in lowest terms and $t = 2^{a/b} + 2^{-a/b}$ is rational, then $b = 1$.* Indeed $u = 2^{a/b}$ is a root of $X^2 - tX + 1$, so u lies in a field of degree at most 2 over the rationals, where the 2-adic valuation takes values in $\frac{1}{2}\mathbb{Z}$; it takes the value a/b

at u, so $b \le 2$. If $b = 2$, then a is odd and $t = 2^{a/2}(1 + 2^{-a})$ is a rational multiple of $\sqrt{2}$, which is irrational.

1. Integer exponents give rational costs. Conversely $J(n^c)$ rational for $n = 2, \dots, 5$ makes $n^c + n^{-c}$ rational, so $c = a/b$ is rational by the trace form, and $J(2^c)$ rational makes $2^c + 2^{-c}$ rational, so $b = 1$.

2. If $c = 1/k$, then $r = q^k$ gives $r^c = q$. Conversely, with $q = 2, 3, 5$ the premise makes $2^{1/c}$, $3^{1/c}$ and $5^{1/c}$ rational, so $1/c = a/b$ is rational by part 1 of Theorem 22.55; then $2^{a/b}$ and its inverse are rational, so their sum is, and $b = 1$: $c = 1/a$.

3. A positive integer of the form $1/k$ is 1. For $k \ge 2$, $J(2^{1/k})$ is irrational by the arithmetic step, since $2J(2^{1/k}) + 2 = 2^{1/k} + 2^{-1/k}$. □

Read physically, an exponent $c = k$ reads the kth power of a ratio: a coarser reading whose costs stay records. An exponent $c = 1/k$ reads a root: a finer reading that reaches every ratio, but whose costs leave the carrier. Only the unit reading does both.

Theorem 22.58 (The verdict of Recognition Science, with no cited premise)**.** *In classical mathematics, the premises (V1)–(V8) give the conclusions of Definition 22.54. No result of transcendental number theory is cited; the proof uses the textbook facts on measurable solutions of d'Alembert's equation [98] and Alexander duality.*

Proof. The cost. Put $F' = \frac{\kappa}{2}F$. By (V2), F' obeys the composition law (21.1). A solution of the law is either constant -1 or has $F'(1) = 0$ (Theorem 21.29, part 1, read classically), and (V3) excludes the constant; the law at $x = 1$ then gives $F'(x^{-1}) = F'(x)$. In logarithmic coordinates $H(t) = F'(e^t) + 1$ solves d'Alembert's equation and is even. By (V4) it is nondecreasing on $[0, \infty)$, hence measurable; measurable solutions of d'Alembert's equation are continuous, and the continuous ones with $H(0) = 1$ are $\cosh(c_0 t)$ and $\cos(c_0 t)$ [98], of which only the first is nondecreasing on $[0, \infty)$ for $c_0 \neq 0$. (V3) excludes $H \equiv 1$, and evenness lets us take $c = |c_0| > 0$. So $F' = J_c$. By (V5) and Proposition 22.57, part 1, c is a positive integer; by (V6) and part 2 it is $1/k$; so $c = 1$ and $\frac{\kappa}{2}F = J$.

The hierarchy. The adjacent ratios are all equal to $r = s_1/s_0 > 1$, and $s_2 = s_1 + s_0$ gives $r^2 = r + 1$, so $r = \varphi$, the only positive root; hence $s_n = \varphi^n s_0$ and $s_0/s_n = \varphi^{-n}$.

The dimension. By (V8) some embedded circle in S^D has a complement with nonzero first homology. For $D \le 1$ no circle has a complement that carries a loop: none embeds in S^0, and one fills S^1. For $D \ge 2$, Alexander duality identifies the first homology of the complement of an embedded circle with the reduced cohomology of the circle in degree $D - 2$, which is nonzero only for $D = 3$. So $D = 3$, and $2^D = 8$. Chapter 23 proves the corresponding statement on the line, natively and for free (Theorem 23.8). □

Conditions and scope.

Base and premises

Classical real numbers; the premises (V1)–(V8).

Conclusion

J at unit amplitude, the golden levels, $D = 3$ and the period 8, with no cited premise:

the transcendence step is Theorem 22.55, and the proof uses only the textbook facts on d'Alembert's equation [98] and Alexander duality.

Identification
The premises themselves: that physical costs compose by (V2), that the cost is native to the carrier and its character an automorphism, that the hierarchy exists, and that a completed recognition is kept. The theorem derives the conclusions from them; it does not supply them.

Not claimed
A price on $\mathbb{R}_\delta$; the price of the transcendence step on the line is treated in Proposition 22.61.

The census also settles which single premise selects the exponent.

Proposition 22.59 (The selection ledger)**.** *In classical mathematics, for $c > 0$:*

1. *$J(2^c) = \frac{1}{4}$ exactly when $c = 1$; and for each $x > 0$ with $x \neq 1$, $J(x^c) = J(x)$ exactly when $c = 1$;*
2. *the second derivative of $t \mapsto J_c(e^t)$ at 0 is 1 exactly when $c = 1$;*
3. *$J_c \leq J_m$ at every $x \neq 1$ for every positive integer m exactly when $c \leq 1$;*
4. *if J_c maps positive rationals to rationals, then (V6) holds exactly when $c = 1$, and J_c is least among the integer exponents exactly when $c = 1$.*

Proof. 1. J is strictly increasing on $[1, \infty)$ and $J(y) = J(y^{-1})$, so $J(x^c) = J(x)$ with $x \neq 1$ forces $x^c = x$ or $x^c = x^{-1}$, and $c > 0$ leaves $c = 1$; at $x = 2$, $J(2) = \frac{1}{4}$. 2. $J_c(e^t) = \cosh(ct) - 1$ has second derivative c^2 at 0. 3. J grows with $|\log x|$, so $J(x^c) \leq J(x^m)$ for all $m \geq 1$ exactly when $c \leq 1$. 4. By the census c is a positive integer; with (V6) it is also $1/k$; and least among integers with c an integer means $c \leq 1$. □

Part 1 is Recognition Science's bit at the classical level: the value of one binary distinction selects the exponent. Proposition 21.44 proves the same selection on $\mathbb{R}_\delta$ for free.

Why the ledger is $\mathbb{Z}[\varphi]$

The ledger of Section 21.5 lives in $\mathbb{Z}[\varphi]$. The reason is a universal property.

Proposition 22.60 (The free ring with a join scale)**.**

1. *No rational q has $q^2 = q + 1$.*
2. *For every commutative ring A and every $r \in A$ with $r^2 = r + 1$ there is exactly one ring map $\mathbb{Z}[\varphi] \to A$ that sends φ to r.*
3. *Every ring map from $\mathbb{Z}[\varphi]$ to a commutative ring of characteristic zero is injective.*

Parts 2 and 3 are free; part 1 is proved over the classical rationals.

Proof. 1. $(2q-1)^2 = 5$, and 5 is not the square of a rational, by descent as for 2. 2. Send $a + b\varphi$ to $a + br$; this respects the product because $r^2 = r + 1$, and a ring map is determined by the image of φ. 3. Let $f(a + b\varphi) = 0$. With the conjugate $\bar{x} = (a + b) - b\varphi$, $x\bar{x} = a^2 + ab - b^2$ is an integer, and $f(x)f(\bar{x}) = f(x\bar{x}) = a^2 + ab - b^2$, which is 0 in characteristic zero only if the integer is 0. By part 1 (in integers: $a^2 + ab - b^2 = 0$ forces $a = b = 0$), $x = 0$. □

A hierarchy with a costly initial join has ratio r with $r^2 = r + 1$ (Section 22.6). By part 2 the ring that r generates over the integers is the image of $\mathbb{Z}[\varphi]$, by part 3 that image is a faithful copy, and by part 1 no smaller ring will do: the rationals carry no join scale. So the level ratios of every such hierarchy are φ^n read faithfully, and $\mathbb{Z}[\varphi]$ is forced, not chosen.

The transcendence step on the line

Theorem 22.55 is classical. Its statement can be read on $\mathbb{R}_\delta$, and the question is what it costs there.

Proposition 22.61 (The six exponentials step on the line)**.** *Let ℓ_2, ℓ_3, ℓ_5 be points with $e^{\ell_p} = p$, and d a point such that $e^{d\ell_p}$ is a ratio for $p = 2, 3, 5$. Consider the conclusion*

$$d \text{ is an integer.} \tag{22.1}$$

1. *In classical mathematics,* (22.1) *holds.*
2. *With no choice principle: if $\neg\neg$*(22.1) *holds under these hypotheses, then* (22.1) *holds.*
3. *With no choice principle: the step reduces to its irrational core, the statement that under these hypotheses d cannot differ from every ratio. A ratio exponent a/b with $2^{a/b}$ a ratio is an integer by arithmetic alone.*

Proof. 1. Read through a bijective reading, the hypotheses are those of Theorem 22.55, part 1, so d is rational, and 2^d rational makes it an integer by the 2-adic argument of Proposition 22.57. 2. "d is an integer" is equivalent, with no choice, to "d is within $\frac{1}{3}$ of an integer z and $d = z$"; the integer z is found from one stage of a regular sequence naming d, and the equation is stable. 3. Under the core, $\neg\neg(d = a/b$ for some ratio a/b in lowest terms); and if $d = a/b$, the ratio $r = e^{d\ell_2}$ has $r^b = 2^a$, so $b \mid a$ by unique factorization and d is an integer. So $\neg\neg$(22.1), and part 2 finishes. □

Whether the irrational core holds on $\mathbb{R}_\delta$ with no omniscience principle and no choice principle is open. A free proof would replace each tool of Lemma 22.56 by a form proved on the line. Siegel's lemma is already free. Rolle's theorem on the line costs LLPO (Theorem 12.87), so a free extrapolation and a free zero estimate must avoid it. The next results do that for the extrapolation, and for the zero estimate in two cases.

Proposition 22.62 (Exponential sums on the line)**.** *On $\mathbb{R}_\delta$, with no omniscience principle and no choice principle:*

1. Newton's form. *In every commutative ring, for nodes $z_0, z_1, \dots$, every x and all N, k,*

$$x^k = \sum_{l<N} h_{k-l}(z_0, \dots, z_l) \prod_{i<l} (x - z_i) + \prod_{i<N} (x - z_i)\, h_{k-N}(z_0, \dots, z_{N-1}, x),$$

   with h_m the complete homogeneous symmetric polynomial of degree m ($h_m = 0$ for $m < 0$); the same holds for every polynomial by linearity.

2. The size of h. *If $0 \le z_i \le R$, then $0 \le h_m(z_0, \dots, z_n) \le \binom{n+m}{m} R^m$.*

3. Extrapolation. *Let $f(x) = \sum_{k<n} a_k e^{\lambda_k x}$ vanish at N increasing nodes in $[0, R]$. Then for $0 \le x \le R$,*

$$|f(x)| \le \Big(\sum_{k<n} |a_k|\, |\lambda_k|^N e^{|\lambda_k| R} \Big) \frac{R^N}{N!}.$$

4. Two terms. *If $\lambda - \mu$ is apart from 0, $z_1 < z_2$, and $ae^{\lambda z} + be^{\mu z}$ vanishes at z_1 and z_2, then $a = b = 0$.*

5. Frequencies in progression. *If d is apart from 0, $z_0 < \cdots < z_{n-1}$, and $\sum_{k<n} a_k e^{(\lambda_0 + kd)z}$ vanishes at every z_j, then every $a_k = 0$.*

Proof. 1. Induction on N with the exchange rule $h_m(z_0, \dots, z_{N-1}, x) = h_m(z_0, \dots, z_N) + (x - z_N) h_{m-1}(z_0, \dots, z_N, x)$.

2. $h_m(z_0, \dots, z_n)$ is a sum of $\binom{n+m}{m}$ monomials of degree m.

3. Expand each exponential as its power series and apply Newton's form to every monomial. The Newton coefficients of f at the nodes vanish, level by level, because f vanishes at the nodes; the remainder term is $\prod (x - z_i) \sum_k a_k \sum_j \lambda_k^j h_{j-N}(\dots)/j!$, and part 2 with $\binom{j}{N}/j! = 1/(N!(j-N)!)$ bounds it by the right side. The limit over the partial sums is taken with explicit moduli; no point is chosen.

4. Cramer's rule: the determinant $e^{\lambda z_1 + \mu z_2} - e^{\mu z_1 + \lambda z_2}$ is apart from 0, because the exponential is strictly increasing and $(\lambda - \mu)(z_2 - z_1)$ is apart from 0.

5. With $y = e^{dz}$ the sum is $e^{\lambda_0 z} P(y)$ for the polynomial $P(Y) = \sum_k a_k Y^k$; the points $y_j = e^{dz_j}$ are n points pairwise apart, and a polynomial of degree below n vanishing at them is 0, by Newton's form with the divided differences, which divide only by differences of nodes. □

The general zero estimate, for n frequencies with no arithmetic structure, and its case on the two-parameter lattice $s + td$ that the six exponentials need, remain open on the line: classically they follow from Rolle's theorem, and what is missing is a lower bound, free on the line, for the generalized Vandermonde determinant $\det(e^{\lambda_k z_j})$.

22.13 The brackets of the chain, closed

Three steps of the forcing chain that the earlier sections place only between two bounds have exact prices: the total meter costs exactly the LPO function (Theorem 22.6 brackets it), the rigid form of the deformation-erasure principle costs exactly the WLPO function (Proposition 22.25 bounds it above), and so does the winding number of the circle as data (Theorem 23.40 brackets it). A fourth step, dropping continuity and the order premises (T2.6), stays below the WLPO function (Section 22.11). The last two closures use one device: a map into a discrete set that is not constant on the line already decides as much as the WLPO function. The first reads a two-valued point as a bit.

Theorem 22.63 (The total inverse is the LPO function)**.** *A map $g : \mathbb{R}_\delta \to \mathbb{R}_\delta$ with $r\,g(r) = 1$ for every $r > 0$ exists if and only if the LPO function exists. Hence a total meter (Theorem 22.6) exists exactly when the LPO function does. No choice principle is used.*

Proof. Let g be given and α a binary sequence, with jump point $j_\alpha \geq 0$; put $t_\alpha = j_\alpha\, g(j_\alpha)$. If α has a 1, then $j_\alpha > 0$ and $t_\alpha = 1$; if α is all 0, then $j_\alpha = 0$ and $t_\alpha = 0$. Since $\neg\neg(\alpha$ has a 1 or is all 0) and equality is stable, $\neg\neg(t_\alpha = 0 \vee t_\alpha = 1)$, and the bit of Lemma 28.2, item 3, read at stage 7 of a regular sequence for t_α, is a map d on binary sequences: if $s_7 > \frac{1}{2}$ then $t_\alpha \neq 0$, so $t_\alpha = 1$; otherwise $t_\alpha \neq 1$, so $t_\alpha = 0$. If $d(\alpha) = 1$, then $j_\alpha\, g(j_\alpha) = 1$, so j_α is invertible and positive, and α has a 1; if α has a 1, then $t_\alpha = 1$ and $d(\alpha) = 1$. So d is the LPO function. The converse is Theorem 22.6(c), and (a) of that theorem carries the price to the meter. □

Theorem 22.64 (Integer readings of the line)**.**

1. *A map $P : \mathbb{R}_\delta \to \mathbb{Z}$ with $P(0) \neq P(1)$ gives the WLPO function.*
2. *The rigid form of the deformation-erasure principle, "for every point $x \neq 1$, DEP holds for the rigid kinematics of x", holds exactly when the WLPO function exists.*
3. *A map $F : [0,1] \to \mathbb{Z}$ with $F(0) = 0$ and $F(1) \neq 0$ exists exactly when the WLPO function exists.*

No choice principle is used.

Proof. 1. *Bisection on representatives.* Starting from $[a_0, b_0] = [0,1]$, given by explicit regular sequences, let m be the midpoint, computed from the sequences of a_n, b_n; keep $[m, b_n]$ if $P(m) = P(0)$ and $[a_n, m]$ otherwise, an integer test. Then $P(a_n) = P(0)$, $P(b_n) \neq P(0)$, $b_n - a_n = 2^{-n}$, and the intervals nest. The left ends form a sequence with an explicit modulus, given by explicit regular sequences, so their diagonal names a point c with $|c - a_n|, |c - b_n| \leq 2^{-n}$, and no representative is chosen. One integer test picks the side that disagrees with c: $y_n = b_n$ if $P(c) = P(0)$, else $y_n = a_n$; then $P(y_n) \neq P(c)$ and $|y_n - c| \leq 2^{-n}$.

Gluing by a stop time. For a binary sequence α let $\sigma(m) = m$ until α first takes the value 1, and frozen from then on. The sequence $y_{\sigma(m+1)}$ again has an explicit modulus; let r_α be its limit. If α is all 0, then $r_\alpha = c$; if $\alpha(n) = 1$, then $r_\alpha = y_{\sigma(n)}$ and $P(r_\alpha) \neq P(c)$. So $d(\alpha) = 1$ exactly when $P(r_\alpha) = P(c)$ defines the WLPO function: if $P(r_\alpha) = P(c)$, then α has no 1, so it is all 0; and if α

is all 0, then $r_\alpha = c$. Every point is the class of an explicit representative and every test is an equality of integers.

2. Given the WLPO function, Proposition 22.25 gives the rigid form. Conversely, DEP for the rigid kinematics of 2 gives a map P to $\mathbb{Z}$ with $P(1) = 0 \neq P(2)$, and $t \mapsto P(1+t)$ separates 0 from 1; part 1 applies.

3. Given F, the map $P(t) = F(\min(1, \max(0, t)))$ separates 0 from 1; part 1 applies. Conversely the WLPO function decides $s \leq \frac{1}{2}$ as data, and $F(s) = 0$ if $s \leq \frac{1}{2}$, $F(s) = 1$ otherwise, is such a map (Theorem 23.40, part 2). □

Part 1 says that the line admits no integer reading that is not constant, short of the WLPO function: a single integer-valued map that tells 0 from 1 already decides, as data, whether every binary sequence is identically zero. This is the device behind the circle.

Theorem 22.65 (The winding number as data costs exactly the WLPO function)**.** *Each of the following exists exactly when the WLPO function exists, with no choice principle in either direction: the data form of $H_1(S^1) = \mathbb{Z}$, a quarter-turn colouring of the circle, an antipodal two-colouring of the circle, and a map $F : [0,1] \to \mathbb{Z}$ with $F(0) = 0$ and $F(1) \neq 0$ (Definition 23.39). Moreover a quarter-turn colouring gives the data form through the WLPO function, and the WLPO function gives the data form without unique choice of path labels.*

Proof. Theorem 23.40, part 1, gives the chain data form ⇒ quarter-turn colouring ⇒ antipodal two-colouring ⇒ the non-constant unit principle, and Theorem 22.64, part 3, closes it at the WLPO function. It remains to go back up from the WLPO function.

A quarter-turn colouring from the WLPO function. The WLPO function decides $a = 0$ and $\neg(a > 0)$ as bits for every point a (Lemma 28.2). For $a \neq 0$ exactly one of $\neg(a > 0)$ and $\neg(-a > 0)$ holds: both would make a not apart from 0, hence 0; neither would give $\neg\neg(a > 0)$ and $\neg\neg(-a > 0)$, whose sum is 0, which is impossible. Colour $p = (x, y)$ by its quadrant: if $y = 0$, by the sign test of x (0 or 2); else if $x = 0$, by the sign test of y (1 or 3); else by both sign tests. The quarter turn $(x, y) \mapsto (-y, x)$ adds 1 modulo 4 in every case, using $x \neq 0$ when $y = 0$, from the circle equation.

The data form from the WLPO function. Theorem 23.40, part 2, needs one integer label $n(\gamma)$ for each path, with $4n(\gamma)$ the swept angle plus $\theta(\gamma(0)) - \theta(\gamma(1))$; it took the label by unique choice, because the swept angle exists only as a proposition. The WLPO function computes it. With it, $\leq$ on the line is decided as data, and "every dyadic point of cell k of a grid has chart coordinate at least $\frac{1}{8}$ in chart j" is one bit. A grid on which every cell passes exists, by the modulus of the path; the least such grid is found by a terminating search, and each cell's chart is the first that passes. Continuity carries the dyadic test to the whole cell, so the grid is admissible and the swept angle is the sum of the chart angles; the label is found by a decided search over the integers. The rest is the proof of Theorem 23.40, part 2.

The data form from a quarter-turn colouring. The colouring gives the WLPO function through the chain above, and the WLPO function gives the data form. □

Conditions and scope.

Base and premises
The line, the native circle and the reading of maps of Section 21.1.

Conclusion
The total meter: exactly the LPO function. The rigid form of DEP for every point other than 1: exactly the WLPO function. The data form of $H_1(S^1) = \mathbb{Z}$, the quarter-turn colouring, the antipodal two-colouring and the non-constant unit principle: each exactly the WLPO function. Unique choice of path labels is not needed.

Not claimed
Whether $\mathrm{SE}_{\mathbb{Z}}$ implies MP, which remains open (Theorem 22.24); and the exact price of dropping continuity and the order premises (T2.6), which stays below the WLPO function. Both readings of Theorem 23.40, part 3, stand: BU_2 and continuity on Cantor space each refute all four data forms.

With these closures, Table 22.1 reads: of 130 rows, 119 free, 10 with an exact price and 1 bracketed, none open. Read physically, one integer label on all records of motion around a circle, read as one function, is exactly as strong as deciding, as data, whether a binary record is blank; and so is a single integer reading of the line that tells two points apart.

22.14 The moral code, priced

Morality is the physics of relations between participants. It runs on the record, the relations and the cost J of this chapter: a relation out of balance carries the cost $J(x)$, and the moral code reads who wrote on each relation and what each end received [198]. On true books the code admits an act exactly when the act has consent, and it permits exactly the least costly acts with consent. This section reads that code on the line, at the scope of [198], and prices it. With the code's hypothesis in its free form (H3), every conclusion of its main theorem is free. A price appears only where a moral verdict is run as a decision on exact amounts over every conceivable record: deciding consent costs WLPO, and reading the writes off every record costs the WLPO function. Three steps of the usual proofs cost LLPO, MP and $\mathrm{MP}^{\vee}$: a decided alternative, a strict inequality from "not equal", and a finite De Morgan step. Each priced step has a free form beside it, and the code's theorems use only the free forms.

Amounts are points of $\mathbb{R}_\delta$ and ticks are natural numbers. Comparisons are those of this chapter: $a \geq 0$ means $\neg(a < 0)$, and $a \neq 0$ is the negation of $a = 0$. Where [198] reads a multiplier x, the line reads its log-flow t, with $x = e^t$; so $J(x) = \cosh t - 1$ (Definition 21.11), and $x = 1$ exactly when $t = 0$.

Definition 22.66 (The record of relations, on the line)**.** A *world* W has a finite set B of relations. A relation b has a source end and a destination end, held by participants $s(b)$ and $d(b)$, and each end holds a log-amount, $h_b(s)$ and $h_b(d)$. Its log-flow is $t_b = h_b(s) - h_b(d)$, its multiplier $x_b = e^{t_b}$ and its cost $J(x_b) = \cosh t_b - 1$; it is *at balance* when $x_b = 1$. Each end keeps an entry for b on its books, and the books are *true* when every entry is the multiplier. Each relation has a reserve θ_b, held by neither end. The *real cost* of W is $C(W) = \sum_{b \in B} J(x_b)$. The participants'

writes P are a set of pairs (participant, relation); b is *movable* when it is live and $(s(b), b)$ and $(d(b), b)$ both lie in P. The relations have *distinct ends* when $s(b) \neq d(b)$ for every b.

An *exchange* on b of amount a posts a at the source end and $-a$ at the destination end; one of the two ends emits it at a tick, and it is absorbed at a later tick, at least one tick after its emission. A *reserve post* on b adds an amount to θ_b; the source end, the destination end, both or neither emit it. An *act* α is a finite list of exchanges together with a finite list of reserve posts; doing nothing is the empty act. It reaches the world in which every holding has changed by the net of the posts at its end, so that t_b changes by twice the net posted at the source end, with true books, nothing ended, and every reserve changed by the net of its posts.

In α, an end k of b *absorbs* on b the sum of the posts at its end of the exchanges on b that the other end emits, and at b's reserve the net of the reserve posts on b that k did not emit. It *consents* on b when it emits on b, by an exchange or a reserve post, and both amounts it absorbs are ≥ 0. The act *has consent* when every end whose holding on a relation it changes, and each end of every relation whose reserve it changes, consents there.

A *recorded world* is a world with its reserves together with a record: a finite list of exchanges and reserve posts, each with its relation, amount and emitters. The recorded world R_α of α is the world it reaches, its reserves after it, and its own posts. *Writes* ω assign to each recorded world a set of pairs (participant, relation). They *meet* (W) on a set $\mathscr{R}$ of recorded worlds when, for every $R \in \mathscr{R}$, every relation b of W and each end of b, the pair (participant at that end, b) lies in $\omega(R)$ exactly when that end consents on b as the record of R shows.

Definition 22.67 (The code)**.** Let W have true books and every relation live, let $\mathscr{A}$ be a set of acts, and let ω be writes. An act $\alpha \in \mathscr{A}$ is *admissible* when both ends of every relation whose multiplier or reserve α changes write on that relation in $\omega(R_\alpha)$. Let W_α be the world α reaches and Φ the record's motion through one interval. An admissible act β *overtakes* α when $\sum_{n \leq N} C(\Phi^n W_\beta) < \sum_{n \leq N} C(\Phi^n W_\alpha)$ for every large enough N. The code *permits* the admissible acts of $\mathscr{A}$ that no admissible act of $\mathscr{A}$ overtakes, and *refuses* every other act of $\mathscr{A}$. The *balanced record* puts every movable relation at balance and keeps every other multiplier. The *balancing act* emits two exchanges on every movable relation b and nothing else: with $g = t_b$, $g^+ = \max(g, 0)$ and $g^- = \max(-g, 0)$, the source end emits one of amount $-g^+/2$ and the destination end one of amount $g^-/2$.

The code of [198, Definition 3.1] admits an act when four clauses hold: it moves only live relations; every new entry is written truthfully; it changes a relation's entry or its multiplier only when both ends of that relation write on it; and every live relation it ends is written on by one of its own ends. Admission in Definition 22.67 is these clauses on such worlds: an act of exchanges moves only live relations, ends nothing and writes true entries, so the clauses ask only that both ends of a changed relation write on it, and the code asks the same of every changed reserve. Where the record holds still between acts, Φ is the identity, so each act's world is the same at every checkpoint and one act overtakes another exactly when its world costs less; the code then permits α exactly when α is admissible and $C(W_\alpha) \leq C(W_\beta)$ for every admissible $\beta \in \mathscr{A}$ [198, Proposition 5.2]. On the line this reduction is free, because "no admissible act reaches a smaller cost" is the relation $\leq$ itself. In the balancing act each end gives half of what it holds above the other, and the end that holds less gives 0. It tests nothing: g^+ and g^- are defined at every point.

Theorem 22.68 (The moral code on the record is free)**.** *Let W have true books and every relation live, let the record the code reads hold still between acts, let P be participants' writes, $\mathscr{A}$ a set of acts and ω writes, and let G be a price on positive points with $G(x) \geq 0$, and $G(x) = 0$ exactly when $x = 1$. Let*

(H1) the code's menu be the recorded worlds of the acts of $\mathscr{A}$;

(H2) every exchange of every act of $\mathscr{A}$ be on a relation of W;

(H3) ω meet (W) on the recorded worlds of the acts of $\mathscr{A}$, with $\omega(R_\alpha) \subseteq P$ for every $\alpha \in \mathscr{A}$;

(H4) the balancing act be an act of $\mathscr{A}$.

Then, with no omniscience principle and no choice principle:

(a) An act of $\mathscr{A}$ is admissible exactly when it has consent.

(b) The code permits an act of $\mathscr{A}$ exactly when it has consent and reaches the balanced record; equivalently, when it has consent and its price $\sum_b G(x'_b)$, at the multipliers x'_b it reaches, is at most the price of every act of $\mathscr{A}$ with consent.

(c) The code permits the balancing act.

(d) No net taking is admitted or permitted: an act of $\mathscr{A}$ in which an end whose holding on b, or b's reserve, the act changes absorbs an amount < 0 on b or at b's reserve is neither admissible nor permitted.

(e) Love takes two: if the code permits an act of $\mathscr{A}$ that brings a relation with $x_b \neq 1$ to balance, both ends of that relation consent there.

(f) On every run of exchanges, let the code read at each tick the world that the posts made by then carry, let every exchange be absorbed within L ticks of its emission, and let $t \leq s$ be ticks that each come at least L ticks after every emission up to it. If no exchange is emitted after t and at or before s, the code reads the same world at s as at t. Under the motion that holds still, take k relations and a menu of single acts *on them: a set of admissible acts, each of which sets the log-flows of the k relations once, after which the motion carries each relation on its own and nothing else acts on them. If some act of the menu brings all k relations to balance, the code permits exactly the acts of the menu that reach balance.*

(g) Every exchange posts equal and opposite amounts and is absorbed at least one tick after its emission.

The same holds when ω meets (W) on every recorded world and $\omega(R_\alpha) \subseteq P$ for every $\alpha \in \mathscr{A}$, as in [198, Theorem 9.10], since that gives (H3) at once.

Proof. (a) By (H2) the act moves only live relations; it ends nothing and writes true entries. An exchange posts a at the source end and $-a$ at the destination end, so the act changes t_b by twice the net n_b it posts at the source end, and $2n_b = 0$ exactly when $n_b = 0$. So it changes x_b exactly when it changes the holdings at both ends, and admission asks that both ends of

every relation whose holdings or reserve change write on it in $\omega(R_\alpha)$. By (H3) that is that both ends consent there: consent. Every condition here is an implication from "$\neq 0$" to a finite membership, and nothing is decided.

(b) By (a) and (H3), an admissible act changes x_b only when both ends of b write on it inside P, that is only on a movable relation, every relation being live. So it keeps every other relation, and its world costs at least the balanced record's cost $\sum_{b \text{ not movable}} J(x_b)$. If it costs at most that, the costs $J(x'_b) \geq 0$ of its movable relations sum to at most 0, so each is 0 and each $x'_b = 1$ (Proposition 22.15). The balancing act attains the bound: each end emits on each movable relation; the source end absorbs $g^-/2 \geq 0$ and the destination end $g^+/2 \geq 0$; and the source end's net post is $(g^- - g^+)/2 = -g/2$, so t_b goes to 0. It has consent, so it is admissible by (a). With the reduction of Definition 22.67, the code permits exactly the acts with consent that reach the balanced record. For G the same bound holds term by term.

(c) follows from (b). (d) The end that absorbs an amount < 0 does not consent, so by (a) the act is not admissible, and the code permits only admissible acts. (e) A permitted act has consent by (a), and bringing $x_b \neq 1$ to 1 changes both holdings of b.

(f) At such a tick every exchange emitted by then has been absorbed, so the posts carry the record of completed exchanges, which no exchange changes between t and s. Under the motion that holds still, each act's cost on the k relations, $\sum_i J(x_i)$ at the multipliers it sets, is the same at every checkpoint, so β overtakes α exactly when β's cost is smaller. An act of the menu that brings the k relations to balance costs 0 and is admissible, so an act permitted on the menu is not overtaken by it and costs at most 0; each term $J(x_i) \geq 0$ is then 0, and every $x_i = 1$. Conversely an act of the menu that reaches balance costs 0, which no act undercuts, and it is admissible, so the code permits it. This argument is direct and free.

(g) is the definition of an exchange. The last sentence holds because (W) on every recorded world is (W) on the recorded worlds of the acts of $\mathscr{A}$, and $\omega(R_\alpha) \subseteq P$ is assumed. □

For comparison, here are the conclusions of [198, Theorem 9.10] in full. Its hypotheses are those of Theorem 22.68, with (H3) read as (W) on every recorded world and the writes on the recorded world of every act of $\mathscr{A}$ inside P, and with a price G that is zero exactly at balance. It concludes:

(a) an act of $\mathscr{A}$ is admissible exactly when it has consent;

(b) the code permits an act of $\mathscr{A}$ exactly when it has consent and leaves the balanced record, every movable relation at balance and the rest where they were; equivalently, when it has consent and its price $\sum_b G(x'_b)$ is least among the acts of $\mathscr{A}$ with consent;

(c) the code permits the balancing act;

(d) no net taking in $\mathscr{A}$ is admitted or permitted, on a relation or at its reserve: an act of $\mathscr{A}$ in which an end whose holding it changes, or an end of a relation whose reserve it changes, absorbs a net loss there is neither admissible nor permitted; nor is an act of $\mathscr{A}$ with a taking that its taker does not give back, that is, whose taker's other exchanges post less than the taking's amount at the victim's end on net, when it changes the victim's holding;

(e) love takes two: if the code permits an act of $\mathscr{A}$ that brings a relation from off balance to balance, both ends of that relation consent there, each emitting on it and neither absorbing a net loss on it or at its reserve;

(f) on every run of exchanges, with the timing of (f) above, the code reads the same world at s as at t; the motion that holds still keeps balance and repairs nothing, so balance leads under it, and the code permits, on relations each written by its participant and carried by this motion on its own, exactly the act to balance on every menu of single acts that offers it;

(g) every exchange posts equal and opposite amounts, absorbed at least one tick after its emission.

Theorem 22.68 proves (a), (b), (c), (e) and (g) as they stand, for $G \geq 0$, and the first sentence of (d). The second sentence of (d) is a case of the first: the victim's end absorbs on the relation the net of the taker's exchanges, which is the taker's other posts there less the taking's amount, an amount < 0. In (f), the reading of the world and the code law are Theorem 22.68(f), proved directly; "balance leads" is free when it is read with "apart" and costs MP when it is read with "$\neq$" (Theorem 22.72(vi)).

Corollary 22.69 (The record fixes the emitters, and the duty to help)**.** *Free:*

(a) Two acts with one recorded world have the same emitters: at every end of every relation, one emits there, by an exchange or a reserve post, exactly when the other does.

(b) Under the hypotheses of Theorem 22.68, with doing nothing an act of $\mathscr{A}$, let every movable relation be at balance. Then the code permits doing nothing and the balancing act; it permits exactly the acts of $\mathscr{A}$ with consent that keep every multiplier; and no act of $\mathscr{A}$ without consent is admissible or permitted.

(c) Under the hypotheses of Theorem 22.68, with doing nothing an act of $\mathscr{A}$, the code permits doing nothing exactly when every movable relation is at balance, and refuses it exactly when not every movable relation is at balance.

(d) Let W have true books, every relation live and distinct ends, and let P be any participants' writes. Take $\mathscr{A}$ to be doing nothing and the balancing act, and ω the emitter writes*: the pairs (participant at an end, b) for the ends whose record shows an emission on b. Then (H1)–(H4) hold. So on every such world, with the record holding still between acts, the code refuses doing nothing exactly when not every movable relation is at balance.*

Proof. (a) The record lists every post with its emitters. (b) The balanced record now keeps every multiplier; doing nothing has consent, and the balancing act posts only amounts 0 and has consent; apply Theorem 22.68(a) and (b). (c) Doing nothing has consent and keeps every multiplier, so by Theorem 22.68(b) it is permitted exactly when W is the balanced record; the second half is the negation of the first. (d) Doing nothing's record shows no emission, so its writes are empty. In the balancing act's record the emitting ends are the two ends of the movable relations, and each absorbs no loss, so with distinct ends the emitter writes are exactly the consenting ends and lie in P. Whether a finite record shows an emission is a finite search. □

Part (c) is the duty to help: where a movable relation is off balance, doing nothing is forbidden. Part (d) states the duty on every world with distinct ends, with writes that cost nothing to read. Its last form, [198, Corollary 9.14], "some movable relation is off balance", costs $\mathrm{MP}^{\vee}$ when (W) holds on the menu only (Theorem 22.72(vii)); with (W) on every recorded world it is free, because that hypothesis already decides whether each movable relation is at balance. That hypothesis is not free to supply: writes meeting (W) on every recorded world exist only with the WLPO function (Theorem 22.72(iii)).

The code's two clauses produce consent. Read the record of an act α at an end k of a relation b: $u = (u_1, u_2)$ is what k sent, the net of its own posts on b at its holding and at the reserve; $v = (v_1, v_2)$ is what it received, the net of the posts on b it did not emit, at the same two places; and its change is $\Delta = u + v$. A *book* of k on b is a pair ψ that meets the three cases of [198, Definition 9.5]: $\psi = (0, 0)$ if $\Delta_1 = 0$ and $\Delta_2 = 0$; $\psi = u + v$ if $\Delta_1 \neq 0$ or $\Delta_2 \neq 0$, and $v_1 \geq 0$ and $v_2 \geq 0$; $\psi = u$ if $\Delta_1 \neq 0$ or $\Delta_2 \neq 0$, and $v_1 < 0$ or $v_2 < 0$. The book is *true* when $\psi = \Delta$. The truth clause asks that the books of both ends of a relation be true; the both-ends clause, with the emitters as writes, asks that both ends of a relation that moved emit on it.

Proposition 22.70 (Consent from the two clauses, and consent forced)**.** *Free:*

(a) Two books of one end on one relation are equal, and a book is true exactly when, if $\Delta_1 \neq 0$ or $\Delta_2 \neq 0$, then $v_1 \geq 0$ and $v_2 \geq 0$.

(b) Let the relations of W have distinct ends, and take a book of each end on each relation on the record of α. Then α has consent exactly when the truth clause and the both-ends clause hold on every relation of its recorded world. Under the hypotheses of Theorem 22.68, an act of $\mathscr{A}$ is admissible exactly when the two clauses hold there.

(c) The third case cannot be replaced by the true book: on one relation, let the source end emit an exchange of amount 1, a taking of 1, and the destination end an exchange of amount 0. With Δ as every book and the emitters as writes both clauses hold, yet the destination end absorbs -1 and the act lacks consent.

(d) On one relation, call a history of exchanges a net taking *when some end whose holding it changes absorbs an amount < 0, and an* unrequested gift *when some end whose holding it changes emits nothing. A history has consent exactly when it is neither. So consent is the greatest condition on histories that refuses every net taking and every unrequested gift, and a condition on histories is consent exactly when it refuses those and passes every other history.*

Proof. (a) Equality of points is stable under double negation, and under double negation one of the three cases applies, so two books coincide. If the relation did not move, $\psi = (0, 0) = \Delta$. If it moved and $v \geq 0$, then $\psi = u + v = \Delta$. If it moved and some $v_i < 0$, then $\psi = u$, and $\psi = \Delta$ would give $v = 0$; so a true book excludes $v_i < 0$, which is $v_i \geq 0$. (b) By (a), at an end of a moved relation the book is true exactly when the end received no net loss, and the both-ends clause asks that both ends emitted: together, that both ends consent. A relation moves exactly when the holding of either end or its reserve changes, and then both holdings or the reserve change; so the clauses on every relation are consent. The second sentence is this with Theorem 22.68(a). (c) Both ends emitted, and the true book is true by definition. (d)

Absorbing no net loss is by definition the negation of absorbing an amount < 0, and whether an end emitted on the relation is a finite search, so its double negation gives it. □

The price at the multiplier [198, Section 9.6] reaches the same verdicts from the shape of J. A taking moves a relation from e^s to e^t while the silent end's entry stays at e^s; charge it the absolute slope of the cost at s, $\sinh|s|$, times the distance moved.

Proposition 22.71 (The price at the multiplier)**.** *Free:*

(a) For all points s and t, $\cosh s - 1 \le \cosh t - 1 + \sinh|s| \cdot |s - t|$, strictly when t is apart from s, and equality holds only if $t = s$. With $r = e^s$ and $y = e^t$ this reads $J(r) \le J(y) + \sinh|\log r| \cdot |\log(r/y)|$.

(b) Let B be a finite set of relations with log-flows t_b, $\mathscr{A}$ a set of acts with log-flows $t_b(\alpha)$ after each act, "silent" any rule on acts and relations, and $M \subseteq B$, such that every act that changes a relation outside M is silent on it, and some act changes no relation with a silent end and puts every relation of M at balance and keeps the rest. Let $c_b(\alpha)$ be any values meeting the cases $c_b(\alpha) = \sinh|t_b| \cdot |t_b - t_b(\alpha)|$ if $t_b(\alpha) \neq t_b$ and b is silent in α, and $c_b(\alpha) = 0$ otherwise, and let $p(\alpha) = \sum_b \big(\cosh t_b(\alpha) - 1 + c_b(\alpha)\big)$. Then $p(\alpha) \le p(\beta)$ for every $\beta \in \mathscr{A}$ exactly when α changes no relation with a silent end, puts M at balance and keeps the rest.

(c) Read Theorem 22.68 on worlds: the menu is the worlds the acts of $\mathscr{A}$ reach, the writes on each are the ends that consent in the act reaching it, and b is silent in α when an end of b does not consent there. An act of $\mathscr{A}$ whose reserve posts are consented has least charged price exactly when the code permits it; so does every act with no reserve post. On one relation with the source end holding 1 and the destination end 0, no participant writing on it, and $\mathscr{A}$ doing nothing and a taking of 1 from the reserve by neither end, every hypothesis holds, the taking has least charged price for every charge meeting the cases, and the code refuses it and permits doing nothing.

Proof. (a) Put $d = t - s$. Then

$$\cosh t - \cosh s - d\sinh s = \tfrac{1}{2}e^s\big(e^d - 1 - d\big) + \tfrac{1}{2}e^{-s}\big(e^{-d} - 1 + d\big),$$

a sum of two terms ≥ 0, both positive when d is apart from 0; and $d\sinh s \ge -\sinh|s| \cdot |d|$. Equality makes d apart from 0 impossible, so $d = 0$. (b) Compare term by term with the balanced price. A relation outside M that α changes is silent, so by (a) it contributes at least its old cost, and exactly that only if it is unchanged; a relation of M contributes $\cosh t_b(\alpha) - 1 + c_b(\alpha) \ge 0$, which is 0 exactly when it is at balance with no charge. The act of the hypothesis attains the bound. (c) On these worlds a relation outside the movable ones cannot change with both ends consenting, so (b) applies with the balancing act; an act has consent exactly when it changes no relation with a silent end and its reserve posts are consented, and Theorem 22.68(b) finishes. In the example the taking changes no multiplier and carries no charge, while no end consents at the reserve. □

Consent at the reserve refuses what the charge cannot see: the charge reads multipliers, and a reserve moves none.

Here are the prices. Each priced step is stated as a statement about every case, and each is beside a free form proved above.

Theorem 22.72 (The priced steps of the moral code)**.** *Each of the following assertions is equivalent to the principle named, with no choice principle in either direction.*

(i) For every world and every act, the act has consent or it does not: WLPO. For every act and relation, the relation is silent in the act or it is not: WLPO. Free form: consent stated, not decided (Theorem 22.68).

(ii) For every recorded world, relation and end, some book meets the cases: WLPO. Free form: books on the act's own record (Proposition 22.70(b)).

(iii) For a world with at least one relation, whose relations all have distinct ends: writes meeting (W) on every recorded world exist, exactly when the WLPO function exists. Free form: (W) on the menu's recorded worlds, which the emitter writes meet with no decision (Corollary 22.69(d)).

(iv) A rule f from points to $\{0,1\}$ naming an end that holds at least as much, $f(g)=1$ giving $g\geq 0$ and $f(g)=0$ giving $g\leq 0$, exists exactly when the WLPO function exists; so does an exact test, $f(g)=1$ exactly when $g\geq 0$. Free form: the balancing act, which tests nothing.

(v) For every history of exchanges and reserve posts on one relation and each end, let U say that the end's holding and the reserve are unchanged and N that both amounts the end absorbs are ≥ 0. Then $(\neg U\to N)$ gives $U\vee N$: LLPO. Free form: the implication $\neg U\to N$.

(vi) For all points s,t with $t\neq s$, $\cosh s-1<\cosh t-1+\sinh|s|\cdot|s-t|$: MP. So is balance leading under the motion that holds still [198, Lemma 5.11(a)]: for every $t\neq 0$, $\sum_{n\leq N}J(e^0)<\sum_{n\leq N}J(e^t)$ for every large enough N. Free forms: Proposition 22.71(a), and the code law of Theorem 22.68(f), proved directly.

(vii) For every world and participants' writes, if not every movable relation is at balance, then some movable relation has $x_b\neq 1$: $\mathrm{MP}^{\vee}$. Free form: Corollary 22.69(c) and (d).

Proof. (i) WLPO decides $a=0$ and $a\geq 0$ for every point (Table 12.2); consent is a finite conjunction, over the relations and their two ends, of implications between such statements and a finite search for an emission, so it is decided. Conversely, on one relation let the source end emit an exchange of amount 0 and the destination end one of amount z. Both ends emit; the source end absorbs z and the destination end 0. So the source end consents, and the act has consent, exactly when $z\geq 0$: $z<0$ would make $z\neq 0$ and the absorbed amount negative. Deciding consent, or silence, decides $z\geq 0$ for every z, which is WLPO (Table 12.2).

(ii) Given WLPO the tests of the cases are decided and a book exists. Conversely, let the source end emit an exchange of amount $1+|x|$ and the destination end one of amount -1. The source end's holding changes by $|x|$ and it absorbs -1, so a book is $(0,0)$ if $x=0$ and $(1+|x|,0)$ if $x\neq 0$. Its first entry p satisfies $0<p$ or $p<1$; the first excludes $x=0$, the second excludes $x\neq 0$, and $x=0$ is stable. So $x=0$ or $x\neq 0$.

(iii) Let ω meet (W) everywhere and let b be a relation. For a point z, let R_z be W with zero reserves and the record in which the source end of b emits an exchange of amount 0 and the destination end one of amount z. As in (i), the source end consents on b in R_z exactly when $z\geq 0$. Membership in the finite set $\omega(R_z)$ is a bit, so $z\mapsto[(s(b),b)\in\omega(R_z)]$ is a map from the

line to $\{0,1\}$ with value 1 at 0 and 0 at -1; by Theorem 22.64, part 1, applied at $u \mapsto u-1$, it gives the WLPO function. Conversely the WLPO function decides $\leq$ as data (Theorem 28.3, part 2); list, relation by relation, each end whose record passes the test. With distinct ends each pair (participant, relation) names one end, so these writes meet (W) on every recorded world.

(iv) $f(1) = 0$ would give $1 \leq 0$, so $f(1) = 1$, and likewise $f(-1) = 0$; so f is a map with two values, and Theorem 22.64, part 1, gives the WLPO function. Conversely the WLPO function decides $g \geq 0$ as data (Theorem 28.3, part 2), an exact test.

(v) Let the source end emit an exchange of amount $-z$ and the destination end one of amount $-z^-$. The source end's holding changes by $-z^+$ and it absorbs $-z^-$, so U is $z \leq 0$ and N is $z \geq 0$. And $\neg U \to N$ holds: $z^- > 0$ would give $z < 0$ and $z^+ = 0$. So the assertion gives $z \leq 0$ or $z \geq 0$ for every z, which is LLPO (Corollary 12.18). Conversely, put $a = |\Delta_1| + |\Delta_2|$ and $l = v_1^- + v_2^-$, both ≥ 0, from the change Δ and the absorbed amounts v. LLPO gives $a \leq l$ or $l \leq a$. If $a \leq l$ and $a > 0$, then $\neg U$, so N, so $l = 0$, which is impossible; so $a = 0$, and U. If $l \leq a$ and $l > 0$, then $a > 0$ and the same argument gives $l = 0$; so $l = 0$, and N.

(vi) At $s = 0$ the assertion is: $t \neq 0$ gives $J(e^t) > 0$, which is Theorem 22.16(i); so it gives MP. Conversely MP makes $t \neq s$ apart (Table 12.2), and Proposition 22.71(a) applies. Under the motion that holds still the two sums are 0 and $(N+1)J(e^t)$, so balance leads exactly when every $t \neq 0$ has $J(e^t) > 0$: MP again.

(vii) For a point $r \neq 0$, take two relations from one participant to another, both ends writing on both, at log-flows r^+ and r^-, so $r = r^+ - r^-$. Both at balance would give $r = 0$, so not both are; the assertion names one with $x_b \neq 1$. If it is the first, $r^+ \neq 0$, and $r < 0$ would give $r^+ = 0$, so $r \geq 0$; if the second, $r \leq 0$. So every nonzero point has a weak sign, which is $\mathrm{MP}^\vee$ (Theorem 14.19). Conversely, let the movable relations have log-flows $t_1, \dots, t_m$, not all 0, and argue by induction on m; for $m = 1$ there is nothing to prove. Fix regular sequences for them; let $e(2k) = 1$ when stage k of t_1 witnesses $t_1 \# 0$ and $e(2k+1) = 1$ when stage k of some t_i with $i \geq 2$ witnesses $t_i \# 0$, and let f mark the first event of e. Then f has at most one 1, and it is not all 0, since then every t_i would be 0. By $\mathrm{MP}^\vee$ its odd entries are all 0, and then $t_1 = 0$ would leave no event at all, so $t_1 \neq 0$; or its even entries are all 0, and then $t_2, \dots, t_m$ are not all 0, and the induction applies. □

The first theorem of [198] fixes the judgment itself. A *case* (S, o, c) is a set S of acts, an act o, doing nothing, and a real price $c(s)$ for every act; a *judgment form* $\mathcal{F}$ returns the set $\mathcal{F}(S,o,c)$ of acts it permits, and the least form is $\mathcal{L}(S,o,c) = \{s \in S : c(s) \leq c(t) \text{ for every } t \in S\}$. The joint case of (S,o,c) and (T,o',d) is $(S \times T, (o,o'), c \oplus d)$, with $(c \oplus d)(a,b) = c(a) + d(b)$.

Theorem 22.73 (The judgment is unique; quoted from [198, Theorem 6.2], not proved here)**.** *In classical mathematics, let a judgment form $\mathcal{F}$ meet: (J1) if $o \in S$ then $\mathcal{F}(S,o,c) \subseteq S$; (J2) if $s, t \in S$ differ from o, $c(s) = c(t)$ and $s \in \mathcal{F}(S,o,c)$, then $t \in \mathcal{F}(S,o,c)$; (J3) $\mathcal{F}$ of a joint case is the product of the two judgments; (J4) if $\mathcal{F}$ permits o, then $c(o) \leq c(t)$ for every $t \in S$; (J5) $\mathcal{F}$ permits some act on every situation the record allows. Then on every situation σ the record allows, $\mathcal{F}(A_\sigma, o_\sigma, c_\sigma) = \mathcal{L}(A_\sigma, o_\sigma, c_\sigma)$: $\mathcal{F}$ permits exactly the admissible acts of least real cost, the acts the code permits where the record holds still between acts, on true books.*

Here a situation the record allows, its admissible acts A_σ and the prices c_σ, the real costs of the worlds the acts reach, are those of [198, Section 6.1]: a world whose menu is every act executable with some available energy, every act's writes inside some participants' writes P, with the balancing act of P on the menu. The theorem is quoted from [198], where it is proved in classical mathematics; this book does not prove it, and its price on the line is not determined here.

Conditions and scope.

Base and premises

The line, the cost J, and the worlds, acts, records and code of Definitions 22.66 and 22.67, at the scope of [198]: true books, every relation live, and the record the code reads holding still between acts where a theorem says so. The identification (W) of the code's writes with the ends that consent is a hypothesis of Theorem 22.68, as in [198, Theorem 9.10]; on the menu of Corollary 22.69(d) the emitter writes meet it.

Conclusion

With the code's hypothesis in its free form (H3), every conclusion of [198, Theorem 9.10], listed in full after Theorem 22.68, with [198, Proposition 9.2, Theorem 9.6(b), Proposition 9.7(d), Corollaries 9.11, 9.13 and 9.14, Theorem 9.12 and Theorem 9.16] and consent forced as the greatest condition [198, Section 8.4], is free in the forms stated here. Deciding consent or silence, and a total book rule, cost exactly WLPO; on a world with at least one relation whose relations all have distinct ends, writes meeting (W) on every recorded world cost exactly the WLPO function, and so does any rule that names the end holding more; "no change or no net loss" costs LLPO; the strict inequality from "$\neq$", and balance leading read with "$\neq$", cost MP; and the finite De Morgan step in the duty to help costs $\mathrm{MP}^\vee$.

Not claimed

A proof or a price on the line for Theorem 22.73, which is quoted, classical and unpriced; any statement of the moral code beyond [198]; and the physical premises that [198, Section 11] lists as steps not yet derived, among them that the code's writes are the emitters and that its records are the books of the book rule, which enter here as stated hypotheses.

Read physically, a world that keeps its record carries, with its holdings, the posts that reached it and who emitted each one. That record tells apart two acts that reach the same holdings with different emitters, and every verdict of the code follows from it with no decision. What costs something is a rule written as a function over every conceivable record: it has to decide whether an absorbed amount is a loss, a weak sign on the line. The code never needs that rule beyond the records its own acts produce.

Questions to work through

1. (Theorem 22.16.) Give a second proof that assertion (ii) implies MP, without the jump point: apply (ii) at $x = 1 + r^2$ and use the row of Table 12.2 that prices $x \neq 0 \to x < 0 \vee x > 0$.

2. (Theorem 22.20 and Proposition 22.21.) In the model, compute $\ell(n) = A^n(0,1)$ in terms of Fibonacci numbers, check $\nu(\ell(n)) = \varphi^n$ directly from Proposition 22.37(d), and verify each hypothesis of Theorem 22.20(b) for the state $s_* = 0$.

3. (Theorem 22.31.) Show, with no principle, that for $x, y \geq 1$ the equation $J(x) = J(y)$ implies $x = y$. Explain why LLPO enters the general statement only through the reciprocal branch $xy = 1$.

4. (Theorem 22.44.) Using the brackets for π and $\log \varphi$ in Section 22.9, show that the clock ratio $\pi/(4 \log \varphi)$ lies between 1.631 and 1.633, and that it differs from $K = \sqrt{\varphi}$ by more than $1/10$.

5. (Theorem 22.42(b).) Let $h : \mathbb{R}_\delta \to \mathbb{R}_\delta$ have uniform derivative 0 on every bounded interval. Show by the telescoping argument that h is constant, and deduce from (b) that a function g satisfying its hypotheses has $g(x)\, g(-x) = \sigma^2$ for every x.

6. (Proposition 22.49(b).) Let $D \geq 1$. Show, with no principle, that a sequence of points of $\{0,1\}^D$ with period $p > 0$ that reaches every point has $p \geq 2^D$, and write down one such sequence with $p = 2^D$ in which consecutive points differ in exactly one coordinate.

Chapter 23

Why three dimensions

Why should a world of distinctions have three spatial dimensions? Recognition Science answers through records that keep a pair. Two linked loops of string cannot be pulled apart without cutting one of them: the link is a record that no continuous deformation erases. Recognition Science identifies such a kept pair with a closed curve linked in space. That identification is a premise of its forcing chain, and from it the dimension of space must be one in which a closed curve can be linked. That this dimension is three is the topological fact on which the step rests; Chapter 22 lists the step in its forcing chain. This chapter proves that fact on $\mathbb{R}_\delta$, and asks what it costs.

Classically the fact goes back to Alexander: a circle embedded in the D-sphere has a complement whose first homology group is $\mathbb{Z}$ when $D = 3$ and zero for every other $D \geq 2$ [191], [194, Proposition 2B.1]. The D-sphere is D-dimensional space closed up by one point at infinity, so $D = 3$ is ordinary space. The classical proofs bisect an arc again and again, keeping a half in whose complement some cycle fails to bound, and finish with a compactness argument; along the way they decide freely whether a point lies on the curve or off it. None of these moves is free on the line. Bisection decides at every stage which half to keep; the compactness argument has the prices that Chapter 13 computes; and deciding membership in a closed set is an omniscience principle.

Here is the smallest case, and it shows how the decisions disappear. In the 2-sphere S^2 of points (x_0, x_1, x_2) with $x_0^2 + x_1^2 + x_2^2 = 1$, the plane closed up by one point, take the equatorial circle C where $x_2 = 0$. Is C linked: does some loop that keeps away from C fail to bound a surface that also keeps away from C? Let γ be a loop that keeps a distance $r > 0$ from C. Its third coordinate never comes near zero: $|\gamma_2(t)| \geq c = \min(r/2, 1/4)$ at every time t, since a point with a small third coordinate lies close to the point of C obtained by rescaling its first two coordinates. Whether γ runs above or below C looks like a sign decision, and signs of points are not free (Corollary 12.18). Here none is needed. At $t = 0$ the number $\gamma_2(0)$ is apart from 0, so it is positive or negative, and that much is free. The sign then cannot change. The loop is uniformly continuous, so over every time step shorter than its modulus at c the third coordinate moves by less than c, while passing from at least c to at most $-c$ takes a move of $2c$; finitely many such steps cover $[0, 1]$. Say the loop runs above, so $\gamma_2(t) \geq c$ for every t. The map

$$H(s, p) = \Big(s\, p_0,\ s\, p_1,\ \sqrt{1 - s^2\big(p_0^2 + p_1^2\big)} \Big), \qquad 0 \leq s \leq 1,$$

slides each point of the upper hemisphere along its meridian to the pole $(0, 0, 1)$: $H(1, p) = p$ there, $H(0, p)$ is the pole, and the third coordinate only grows, so nothing approaches C. The square of pairs (s, t), cut along its diagonal, gives two triangles $(s, t) \mapsto H(s, \gamma(t))$. Their boundaries, with the boundary of the constant triangle at the pole, add up to γ, because the two remaining edges are the paths of the same point $\gamma(0) = \gamma(1)$ and cancel. Every loop of the complement bounds. The equatorial circle of S^2 is not linked, and nothing was decided.

What made the case free is a sign that cannot flip. The complement of C reads into two points, above and below, which form the 0-sphere $S^0 = \{-1, 1\}$, and a uniformly continuous path in S^0 is constant. That needs no decision either: two points of S^0 closer than 1 are equal. The general argument lowers the dimension of the sphere step by step (Section 23.4). For $D = 2$ the descent ends at exactly this fact about S^0. For $D \geq 4$ it ends at a sphere of positive dimension, in which any two points are joined by a path. For $D = 3$ it ends at the circle, whose winding number stops it; there an explicit loop around the equatorial circle of S^3 sweeps four quarter turns and bounds nothing (Section 23.7).

Theorem 23.8 states the result: for every natural number D, some natively embedded circle in the native D-sphere is linked if and only if $D = 3$, with no omniscience principle and no choice principle. The definitions are those of singular homology, built from the line (Section 23.1). The proof replaces bisection by explicit moduli, compactness by finitely many comparisons and Lebesgue numbers computed from moduli, and the winding function by a relation between chains and values. The one price in the chapter belongs to a different question: whether the winding number of the circle can be given as one integer-valued function on all chains of paths. Section 23.10 relates that function to colourings of the circle and to the WLPO function, and Theorem 22.65 shows that it costs exactly the WLPO function. It is refuted by the Borsuk–Ulam property for two-colourings of the circle, and the WLPO function by the continuity of Boolean functions on Cantor space; both properties hold in Brouwer's intuitionistic mathematics and fail classically.

Two conventions hold throughout. A ratio used as a point stands for its point, as in Chapter 13. For points of the line, $x \leq y$ means $\neg(y < x)$, as in Proposition 12.19.

23.1 Spheres, simplices and chains built from the line

Everything in this section is built from $\mathbb{R}_\delta$ and its ordered field structure, which costs nothing (Proposition 12.19): sums, products, max, min, $|x|$, nonnegative square roots, inverses of points apart from 0, and cotransitivity of $<$. Equality of points is equality on the line. Deciding it costs WLPO (Corollary 12.18), and no definition below asks for it.

Definition 23.1 (Native spheres)**.** Let D be a natural number. The *native D-sphere* S^D is the set of tuples $x = (x_0, \dots, x_D)$ of points of $\mathbb{R}_\delta$ with $x_0^2 + \cdots + x_D^2 = 1$, with the sup distance $d(x, y) = \max_i |x_i - y_i|$. We write $\langle x, y\rangle = \sum_i x_i y_i$ and $|x|^2 = \langle x, x\rangle$, so that $d(x, y)^2 \leq |x - y|^2 \leq (D{+}1)\, d(x, y)^2$. The *native circle* is S^1. The *poles* of S^D are $N = (1, 0, \dots, 0)$ and $S = (-1, 0, \dots, 0)$; for $D \geq 1$ its *equator* is the set of points with $x_0 = 0$, a copy of S^{D-1}. A *removed set* is given as the image of a map $K : X \to S^D$ from any set X. A point y is *r-apart* from K when $d(y, K(x)) \geq r$ for every x, and *apart* from K when it is r-apart for some $r > 0$.

Definition 23.2 (Uniform continuity, apartness and embeddings)**.** A map g between sets with distances d_A and d_B is *uniformly continuous* when there is a map ω from the line to the line, a *modulus*, such that $\varepsilon > 0$ implies $\omega(\varepsilon) > 0$ and $d_A(a, b) < \omega(\varepsilon)$ implies $d_B(g(a), g(b)) < \varepsilon$. A map g into S^D is *uniformly apart* from a removed set K when a single $r > 0$ has $d(g(a), K(x)) \geq r$ for every a and every x. A *native embedding* of the circle is a uniformly continuous $f : S^1 \to S^D$ with a second modulus ω' such that $d(f(a), f(b)) < \omega'(\varepsilon)$ implies $d(a, b) < \varepsilon$.

The definitions assert that the moduli and the radius r exist. A map does not carry them as data; this matters in Section 23.10.

Definition 23.3 (Model simplices and native singular simplices)**.** For $k \geq 1$ the *model k-simplex* Δ^k is the set of tuples $(y_1, \dots, y_k)$ of points of the line with $0 \leq y_1 \leq y_2 \leq \cdots \leq y_k \leq 1$, with the sup distance; Δ^0 is a single point, and $\Delta^1 = [0, 1]$. Its *faces* $d_0, \dots, d_k : \Delta^{k-1} \to \Delta^k$ insert one coordinate: d_0 appends 1 at the end, d_k puts 0 in front, and for $0 < i < k$ the face d_i repeats the coordinate y_{k-i}. Thus $d_0 = 1$ and $d_1 = 0$ on Δ^1, and

$$d_0(s) = (s, 1), \qquad d_1(s) = (s, s), \qquad d_2(s) = (0, s)$$

on Δ^2, whose vertices are $(0, 0)$, $(0, 1)$ and $(1, 1)$. A *native singular k-simplex* of $S^D \setminus K$ is a uniformly continuous map $\sigma : \Delta^k \to S^D$ that is uniformly apart from K; a 0-simplex is a point apart from K. When X is empty nothing is removed, and we speak of simplices of S^D.

Each face map preserves sup distances, so every face $\sigma \circ d_i$ of a simplex of $S^D \setminus K$ is again one, with the same modulus and radius.

Definition 23.4 (Native chains, cycles and boundaries)**.** Let M be a set. A *combination* of members of M is a finite list of terms (n, σ) with $n \in \mathbb{Z}$ and $\sigma \in M$. Two combinations are *equivalent*, $a \sim b$, when one is reached from the other by finitely many moves, each made anywhere in the list and in either direction: exchange two neighbouring terms; merge two neighbouring terms (n, σ) and (m, σ) that carry the same member into $(n + m, \sigma)$; delete a term $(0, \sigma)$. The classes form the *free abelian group* on M, added by concatenation; $-a$ is a with every coefficient negated, and na is defined from these. A *native k-chain* of $S^D \setminus K$ is a combination of native singular k-simplices of $S^D \setminus K$. For $k \geq 1$ its *boundary* is computed term by term:

$$\partial(n, \sigma) = (n, \sigma \circ d_0),\ (-n, \sigma \circ d_1),\ \dots,\ \big((-1)^k n, \sigma \circ d_k\big).$$

A k-chain z with $k \geq 1$ is a *cycle* when $\partial z \sim \emptyset$, the empty list. It *bounds* in $S^D \setminus K$ when $z \sim \partial L$ for a $(k + 1)$-chain L of $S^D \setminus K$. We write $H_k(S^D \setminus K) = 0$ for the statement that every k-cycle of $S^D \setminus K$ bounds there, and $\tilde{H}_0(S^D \setminus K) = 0$ for the statement that every 0-chain of $S^D \setminus K$ whose coefficients sum to 0 is the boundary of a 1-chain of $S^D \setminus K$.

Two terms carry the same member when the two maps are equal, that is, equal at every point; a merge needs a proof of that equality, never a decision. This is why chains are lists. The usual description, a formal sum with the coefficients of each simplex collected, would have to decide which simplices coincide, and maps into the line have no such decision. The boundary of an equivalent chain is equivalent, because each move on a chain becomes a few moves on its boundary. The presentation does not collapse: every move keeps the sum of the coefficients, so $(1, \sigma)$ is never equivalent to the empty list. Statements such as $H_k(S^D \setminus K) = 0$ speak about cycles and boundaries only; we never form a quotient group.

Definition 23.5 (Linking)**.** Let f be a native embedding of the circle into S^D, and write $S^D \setminus f$ for S^D with the image of f removed. A *loop* of the complement is a native 1-simplex γ of $S^D \setminus f$ with $\gamma(0) = \gamma(1)$. The embedding f is *linked* when some loop of its complement does not bound there: no 2-chain L of $S^D \setminus f$ has $\partial L \sim (1, \gamma)$. The dimension D *detects linking* when some native embedding of the circle into S^D is linked.

Remark 23.6 (The notions are not empty)**.** For every $D \geq 1$ the *equatorial circle* $e(x, y) = (x, y, 0, \dots, 0)$ is a native embedding of S^1 into S^D, with both moduli the identity, since it preserves distances. (It is not the equator $x_0 = 0$ of Definition 23.1.) For $D \geq 2$ the point $(0, 0, 1, 0, \dots, 0)$ is 1-apart from it, so its complement has points, and constant loops there. Every constant loop bounds: the constant 2-simplex has three equal faces, and the boundary $c - c + c$ is equivalent to c. So "does not bound" is a genuine condition.

Remark 23.7 (The classical reading)**.** Classically each native notion agrees with its usual counterpart. A continuous map on a compact model simplex is uniformly continuous (Heine–Cantor); a continuous injective map on the compact circle has a uniformly continuous inverse on its image; a compact curve and a point off it have positive distance, so "uniformly apart from the curve" describes simplices of the open complement; every singular 1-cycle is homologous to a sum of loops, so the first homology group is nonzero exactly when some single loop does not bound [194, Section 2.A]; the list presentation is the free abelian group; and the ordered model simplices are the standard simplices of singular homology [193], [194, Section 2.1]. These agreements are classical theorems, and we do not prove them natively. The theorem below concerns the native notions, and none of them mentions the number three.

23.2 The theorem: a circle links only in the 3-sphere

Theorem 23.8 (Three dimensions are forced, with no omniscience and no choice)**.** *For every natural number D, some native embedding of the circle into the native D-sphere has a loop in its complement that bounds no native 2-chain of the complement if and only if $D = 3$. The proof uses no omniscience principle and no choice principle.*

Conditions and scope.

Base and premises
The line $\mathbb{R}_\delta$ with its ordered field structure, and the native notions of Definitions 23.1–23.5. Nothing else: no omniscience principle, no countable or unique choice, no decision of equality or order on the line.

Conclusion
Three is the only dimension in which some natively embedded circle is linked. In S^3 the equatorial circle is linked (Theorem 23.34); in every other S^D every loop of the complement of every natively embedded circle bounds there.

Identification
Recognition Science reads a kept pair of records, a pairing that no deformation erases, as a linked curve in space, and reads the dimension of space as D. These readings belong to the forcing chain (Chapter 22); the theorem does not supply them. The classical meaning of the native notions (Remark 23.7) is likewise a reading, not a native theorem.

Not claimed

That every natively embedded circle in S^3 is linked: the theorem exhibits one, the equatorial circle. The homology groups as quotient groups. An integer-valued winding function on chains, which costs exactly the WLPO function (Section 23.10 and Theorem 22.65).

The proof has two halves that meet at the third homology of the sphere. The first half turns a linked circle into a cycle of the sphere that does not bound. Cut the circle into two arcs that meet in two points. The complement of an arc has no first homology (Lemma 23.20). A Mayer–Vietoris argument then shows that a linked circle needs a 2-cycle of the sphere minus the two meeting points that bounds nothing there, and a second one turns that into a 3-cycle of the sphere that bounds nothing (Proposition 23.23). The second half shows that every 3-cycle of S^D bounds when $D = 2$ or $D \geq 4$, by a descent through lower spheres, each step lowering the degree of homology or the dimension of the sphere:

$$\begin{aligned} H_3(S^D) \Leftarrow H_2(S^D \setminus \{N, S\}) \Leftarrow H_2(S^{D-1}) \Leftarrow H_1(S^{D-1} \setminus \{N, S\}) \\ \Leftarrow H_1(S^{D-2}) \Leftarrow \tilde{H}_0(S^{D-2} \setminus \{N, S\}) \Leftarrow \tilde{H}_0(S^{D-3}), \end{aligned} \tag{23.1}$$

where each arrow reads "if the right side is 0, so is the left". The steps alternate. A Mayer–Vietoris step covers the sphere by the complements of its two poles, each of which is acyclic (Lemmas 23.12 and 23.15). A retraction step slides the complement of the poles onto the equator, a sphere one dimension lower (Lemma 23.16). Dimensions 0 and 1 are settled directly (Proposition 23.24), and in S^3 an explicit loop is linked (Theorem 23.34).

23.3 Tools that work in every dimension

Six tools carry the proof. None depends on D, and none uses a decision. The first is algebra of the free abelian group. The next two are the subdivision and prism operators of singular homology [194, Section 2.1], built on the model simplices. The last three concern the sphere: a sphere with one point removed is acyclic; a point far from the two places where two arcs meet is far from one of the arcs; and a Mayer–Vietoris gluing.

Lemma 23.9 (Combinations that agree lie in the intersection)**.** *Let M be a set and A, B properties of its members.*

1. *If a combination a of members with property A is equivalent, as a combination of members of M, to a combination b of members with property B, then a is equivalent to a combination c of members with both properties.*
2. *If two combinations of members with property A are equivalent as combinations of members of M, they are equivalent through moves whose lists contain only members with property A.*

Proof. Label the terms of a combination by natural numbers. A labelling is *good* when terms with equal labels carry the same member and, for each label, the coefficients of the terms carrying it sum to 0; a combination is *grouped* when it has a good labelling. A grouped combination is equivalent to the empty list: exchanges bring the terms of one label together, merges collapse them into one term with coefficient 0, and a deletion removes it.

Each move preserves grouping, in both directions, whatever surrounds it. An exchange permutes the labels. Splitting $(n+m,\sigma)$ into (n,σ) and (m,σ) gives both parts the label of the whole. When (n,σ) and (m,σ) with labels i and j are merged, the merged term keeps i if $i=j$; otherwise every term labelled j is relabelled i, which keeps the labelling good, since all terms of both labels carry σ and two sums 0 add to 0. Whether $i=j$ is a comparison of natural numbers. Deleting a term with coefficient 0 changes no sum, and an inserted one gets a label not used before. Now $-b$ followed by b is grouped: give the ith terms of $-b$ and of b the label i. If $a\sim b$, the moves that turn b into a, made after the prefix $-b$, preserve grouping, so $-b$ followed by a is grouped.

For part 1, take a good labelling of $-b$ followed by a. Whether the label of a term of a also occurs on a term of b is decided by comparing finitely many natural numbers. Let c be the terms of a whose labels occur in b, in their order, and r the other terms of a. Each term of c carries the same member as a term of b, so its member has both properties. Each label of r occurs only in r, so r is grouped and $r\sim\emptyset$. Exchanges give $a\sim c$ followed by r, hence $a\sim c$.

For part 2, let a and b have property A. A good labelling of $-b$ followed by a is good when the members are read as members with property A, because two such members that are equal as members of M are equal. So $-b$ followed by a is grouped among members with property A, and so is b followed by $-b$. Hence, within such members, a is equivalent to b followed by $-b$ followed by a (insert the grouped list $b,-b$ in front), and that is equivalent to b (remove the grouped list $-b,a$). □

Part 1 is the algebraic heart of every Mayer–Vietoris argument below: a chain written once with simplices that avoid one set and once with simplices that avoid another can be written with simplices that avoid both. Part 2 lets us compute equivalence among all maps into the sphere, whatever property the simplices of a particular complement carry.

Lemma 23.10 (Subdivision)**.** *For $k=1,2,3,4$ there are operators Sd taking each map $\sigma:\Delta^k\to S^D$ to a k-chain, and for $k=1,2,3$ operators T taking it to a $(k+1)$-chain, extended to chains term by term, such that:*

1. *$\partial\,\mathrm{Sd}\,c\sim\mathrm{Sd}\,\partial c$ for $k\geq 2$;*
2. *$\partial Tc\sim c-\mathrm{Sd}\,c$ for $k=1$, and $\partial Tc+T\partial c\sim c-\mathrm{Sd}\,c$ for $k=2,3$;*
3. *every simplex occurring in $\mathrm{Sd}\,\sigma$ or $T\sigma$ is σ composed with a map between model simplices, so its values are values of σ, and it is uniformly continuous when σ is;*
4. *for every uniformly continuous σ and every $\varepsilon>0$ some m makes every simplex τ of $\mathrm{Sd}^m\sigma$ small: $d(\tau(x),\tau(y))<\varepsilon$ for all x,y.*

The operators respect equivalence of chains.

Proof. The construction is the cone construction of barycentric subdivision [194, Section 2.1], carried out with ratio weights. A linear simplex in a model is a list of vertices of the model; it names the affine map that sends the vertices of the standard simplex to them. Each linear simplex gets an interior point b: the convex combination of its vertices with weights $(c,1-c)$, $(c,c,1-2c)$ and $(c,c,c,1-3c)$ in degrees 1, 2 and 3, where $c=1/4$, and weights $(c,c,c,c,1-4c)$

with $c = 1/5$ in degree 4. The cone $b \cdot [w_0, \dots, w_j] = [b, w_0, \dots, w_j]$ satisfies $\partial(b \cdot u) = u - b \cdot \partial u$. Linear subdivision is $S[w_0] = [w_0]$ and $S\tau = b(\tau) \cdot S(\partial\tau)$, and the homotopy is $T\tau = b(\tau) \cdot (\tau - T\partial\tau)$ with $T = 0$ on points. Then $\partial S = S\partial$ follows from the cone formula and $\partial\partial = 0$ alone, by cancelling identical vertex lists; and

$$\partial T\tau = \tau - T\partial\tau - b \cdot \partial(\tau - T\partial\tau) = \tau - T\partial\tau - b \cdot S\partial\tau = \tau - T\partial\tau - S\tau,$$

using $\partial T\partial\tau = \partial\tau - S\partial\tau$ one degree lower. For a map σ, $\mathrm{Sd}\,\sigma$ and $T\sigma$ compose σ with the affine maps of the linear subdivision of the identity simplex; faces commute with these maps because the face maps are affine, so the identities transfer. With these weights every affine identity used is a ring identity among ratios. Affine maps between models are Lipschitz for the sup distances, with constant 3 (constant 4 in degree 4), so composition keeps uniform continuity, with a modulus divided by that constant. One subdivision shrinks every coordinate spread of the vertices by the factor $3/4$ ($4/5$ in degree 4). After m steps the vertex coordinates of every piece spread at most $(3/4)^m$ ($(4/5)^m$ in degree 4), and so do the points of its affine image, which are convex combinations of the vertices. Bernoulli's inequality gives $(3/4)^m(3+m) \le 3$ and $(4/5)^m(4+m) \le 4$, so by the Archimedean property of the line (Proposition 12.19) some m makes the spread smaller than the modulus of σ at ε; two values of a piece of $\mathrm{Sd}^m\sigma$ are values of σ at two such points, which gives item 4. Each operator acts term by term, so it respects the three moves. The number m is found inside the proof of a statement; nothing is chosen. □

Lemma 23.11 (Prisms). *For $k = 1, 2, 3$, let $F : [0,1] \times \Delta^k \to S^D$, write $F_t = F(t, \cdot)$, and let ∂F be the alternating list of the face homotopies $F \circ (\mathrm{id} \times d_i)$; the prism of a homotopy of a point is its path. There is a $(k+1)$-chain PF, the* prism *of F, with*

$$\partial PF + P(\partial F) \sim F_1 - F_0.$$

Every simplex of PF takes only values of F, and it is uniformly continuous when F is uniformly continuous for the distance $\max(|t - t'|, d(y, y'))$.

Proof. A point of Δ^{k+1} is $(y_1, \dots, y_{k+1})$. For $j = 0, \dots, k$ the jth piece reads the time from the coordinate y_{k+1-j} and the point of Δ^k from the other k coordinates, which remain in order: $\pi_j(y) = F\big(y_{k+1-j}, (y_1, \dots, \widehat{y_{k+1-j}}, \dots, y_{k+1})\big)$; the prism is $PF = \sum_j (-1)^j \pi_j$. For $k = 1$ this is $\pi_0(a,b) = F(b,a)$ and $\pi_1(a,b) = F(a,b)$, and the faces of Definition 23.3 give

$$\partial(\pi_0 - \pi_1) = \big(F_1 - \delta + F(\cdot, 0)\big) - \big(F(\cdot, 1) - \delta + F_0\big) = F_1 - F_0 - \big(F(\cdot, 1) - F(\cdot, 0)\big),$$

where $\delta(s) = F(s,s)$ is the diagonal and $F(\cdot, 0)$, $F(\cdot, 1)$ are the paths of the two ends of the 1-simplex; the last bracket is $P(\partial F)$. In every degree each face of each piece is a face of F_1 or of F_0 (the top and bottom), a diagonal shared by two neighbouring pieces with opposite signs, or a piece of the prism of a face homotopy with the sign of that face; collecting them gives the law. Coordinate maps are 1-Lipschitz for sup distances, so every piece is uniformly continuous with the modulus of F. □

Lemma 23.12 (A sphere with one point removed is acyclic)**.** *For every D and every $r > 0$ there is $\rho > 0$ such that, for every point p of S^D:*

1. *every uniformly continuous loop that is r-apart from p is the boundary of a 2-chain of S^D whose simplices are ρ-apart from p;*
2. *every 2-cycle of S^D whose simplices are r-apart from p is the boundary of a 3-chain of S^D whose simplices are ρ-apart from p;*
3. *for $k = 1, 2, 3$, every k-cycle of S^D whose simplices are apart from p is the boundary of a $(k+1)$-chain of S^D.*

Proof. Push points away from p along great circles. For x in S^D put $c = \langle x, p\rangle$ and $m = 1 - c$, so that $|x - p|^2 = 2m$. For $t \in [0, 1]$ put

$$A = t^2(1 + c), \qquad Q = A + m, \qquad k = A - 2tc - m, \qquad H(t, x) = \frac{2t\,x + k\,p}{Q}.$$

No square root is needed. Since $k = (A - m) - 2tc$,

$$|2tx + kp|^2 = 4t^2 + 4tkc + k^2 = 4t^2(1 - c^2) + (A - m)^2 = 4Am + (A - m)^2 = Q^2,$$

so $H(t, x)$ lies on S^D whenever $Q > 0$. At $t = 1$, $k = 0$ and $Q = 2$, so $H(1, x) = x$; at $t = 0$, $H(0, x) = -p$ when $m > 0$. Moreover $\langle H, p\rangle = (2tc + k)/Q = (A - m)/Q$, so

$$|H(t, x) - p|^2 = 2 - \frac{2(A - m)}{A + m} = \frac{4m}{Q} \geq 2m = |x - p|^2,$$

because $Q \leq (1 + c) + m = 2$: the homotopy never moves a point towards p.

To have one homotopy on the whole sphere, slow it down near p. Put $e = \min(r^2, 1)$, $\phi(x) = \min(1, 2m/e)$ and $G(s, x) = H(1 - s\phi(x), x)$ for $s \in [0, 1]$. Suppose $Q < e/4$ at some (s, x). Then $m \leq Q < e/4$, so $\phi < 1/2$, the time $t = 1 - s\phi$ exceeds $1/2$, and $c = 1 - m > 3/4$; hence $Q \geq A > \frac{1}{4} \cdot \frac{7}{4} \geq e/4$, a contradiction. So $Q \geq e/4$ everywhere, G uses the inverse of a point at least $e/4$, and G is uniformly continuous on $[0, 1] \times S^D$. We have $G(0, x) = x$. If x is r-apart from p then $2m \geq d(x, p)^2 \geq r^2 \geq e$, so $\phi(x) = 1$ and $G(1, x) = -p$; and $|G(s, x) - p|^2 \geq r^2$ along the way, so $G(s, x)$ is ρ-apart from p with $\rho = r/(D + 1)$.

For item 1, let γ be the loop and $F(s, t) = G(s, \gamma(t))$. The prism law gives $\partial PF \sim F_1 - F_0 - P(\partial F)$. Here $F_0 = \gamma$, F_1 is the constant path c_{-p} at $-p$, and $P(\partial F)$ is the path of $\gamma(1)$ minus the path of $\gamma(0)$, which cancel. The constant 2-simplex at $-p$ has boundary $c_{-p} - c_{-p} + c_{-p} \sim c_{-p}$. Hence $\gamma \sim \partial(\text{constant} - PF)$, and every simplex here is ρ-apart from p.

For item 2, let z be the 2-cycle and apply the same homotopy to each of its simplices; summing the prism laws, $\partial Pz + P(\partial z) \sim G_1 \circ z - z$. The boundary of one 2-simplex has coefficient sum $1 - 1 + 1 = 1$, so the coefficient sum of z equals that of $\partial z \sim \emptyset$, which is 0; thus $G_1 \circ z$, a combination of copies of one constant simplex, merges to the empty list. The prism operator acts term by term and respects equivalence, so $P(\partial z) \sim P(\emptyset) = \emptyset$. Hence $z \sim \partial(-Pz)$.

For item 3, take for r the least of the radii of the simplices of z. The same computation gives $z \sim G_1 \circ z - \partial Pz$, and $G_1 \circ z$ is the coefficient sum of z times the constant k-simplex at $-p$. For $k = 2$ that sum is 0, as above. For odd k the constant k-simplex is a boundary: the constant $(k + 1)$-simplex has $k + 2$ equal faces with alternating signs, which merge to one copy. □

The next two tools work with two removed sets at once. The Lebesgue number of a cover is classically produced by compactness; here it is computed from the moduli of the embedding.

Lemma 23.13 (Lebesgue numbers for two arcs). *Let $f : S^1 \to S^D$ be a native embedding, and let R_1, R_2 be properties of points of the circle, the two pieces, with meeting points $j(x)$, $x \in X$, that have both properties. Assume the pieces are separated near the meeting points (for every $\varepsilon > 0$ some $\eta > 0$ has: if a has R_1, b has R_2 and $d(a,b) < \eta$, then $d(a, j(x)) < \varepsilon$ for some x) and have finite nets (for every $\eta > 0$ each piece has a finite list of its points within η of each of its points). Then for every $r > 0$ there is $\delta > 0$ such that every point y of S^D that is r-apart from all the $f(j(x))$ is δ-apart from $f(R_1)$ or δ-apart from $f(R_2)$.*

Proof. Let ω and ω' be the moduli of f. Put $\varepsilon = \omega(r/2)$, let η be given by the separation at ε, and put $m = \min(r/2, \omega'(\eta))$, $b = m/2$ and $\delta = b/4$. Take nets of both pieces of fineness $\omega(\delta)$. For each net point c, cotransitivity of $<$ at $b/2 < b$ gives $b/2 < d(y, f(c))$ or $d(y, f(c)) < b$. Going through the two finite lists, each piece has either all its net points $b/2$-far from y, or some net point b-near. If all net points of a piece are $b/2$-far, then every point a of the piece is within $\omega(\delta)$ of a net point c, so $d(f(a), f(c)) < \delta$ and $d(y, f(a)) > b/2 - \delta = \delta$: the point y is δ-apart from that piece. If both pieces have b-near net points c_1 and c_2, then $d(f(c_1), f(c_2)) < 2b \le \omega'(\eta)$, so $d(c_1, c_2) < \eta$, so $d(c_1, j(x)) < \varepsilon$ for some x, so $d(f(c_1), f(j(x))) < r/2$, and $d(y, f(j(x))) < b + r/2 < r$, which is impossible. Only finitely many comparisons were made, each by cotransitivity. □

Lemma 23.14 (Mayer–Vietoris gluing). *Let K_1, K_2 and K_{12} be removed sets in S^D, every point of K_{12} a point of K_1 and of K_2. Assume the Lebesgue condition (for every $r > 0$ some $\delta > 0$ makes every point r-apart from K_{12} δ-apart from K_1 or from K_2) and $H_2(S^D \setminus K_{12}) = 0$. Let γ be a uniformly continuous loop apart from K_1, K_2 and K_{12} that bounds in $S^D \setminus K_1$ and bounds in $S^D \setminus K_2$. Then γ is the boundary of a 2-chain whose simplices are apart from K_1 and from K_2.*

Proof. This is the classical argument [194, Section 2.2] with every step made explicit. Let $\gamma \sim \partial a$ with a apart from K_1 and $\gamma \sim \partial b$ with b apart from K_2. Then $a - b$ is a 2-cycle whose simplices are apart from K_{12}, so $a - b \sim \partial e$ for a 3-chain e apart from K_{12}, with some radius r. Let δ be the Lebesgue number for r. Subdivide e by Lemma 23.10, m times, until every simplex τ of $\mathrm{Sd}^m e$ has its values within $\delta/2$ of one another. The first vertex of τ is a value of e, hence r-apart from K_{12}, hence δ-apart from K_1 or from K_2; then every value of τ is $\delta/2$-apart from the same set. Sorting the terms by this outcome gives $\mathrm{Sd}^m e = e_1 + e_2$, with e_1 apart from K_1 and e_2 apart from K_2. Since $\partial \mathrm{Sd}^m e \sim \mathrm{Sd}^m \partial e \sim \mathrm{Sd}^m a - \mathrm{Sd}^m b$,

$$\mathrm{Sd}^m a - \partial e_1 \sim \mathrm{Sd}^m b + \partial e_2,$$

the left side apart from K_1 and the right side apart from K_2, because subdivision keeps values. Lemma 23.9 writes both as one chain c apart from K_1 and from K_2, and $\partial c \sim \mathrm{Sd}^m \partial a \sim \mathrm{Sd}^m \gamma$. Finally $\gamma - \mathrm{Sd}^m \gamma \sim \partial \sum_{i<m} T(\mathrm{Sd}^i \gamma)$, by Lemma 23.10, and these simplices take values of γ. So $\gamma \sim \partial\big(c + \sum_{i<m} T(\mathrm{Sd}^i \gamma)\big)$. Part 2 of Lemma 23.9 moves between chains of the complements and chains of all maps. □

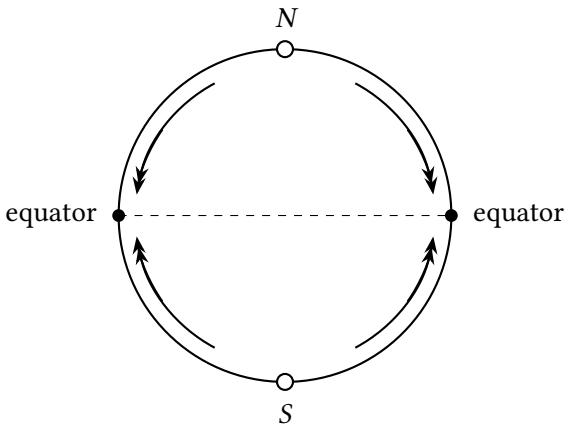


Figure 23.1. A section of S^D through its poles. The complements of N and of S (open circles) cover the sphere. The points of their intersection slide along meridians onto the equator (arrows), which is a sphere one dimension lower; near the poles the slide is slowed down, and points a given distance from both poles reach the equator.

Table 23.1. The three endings of the descent (23.1). Each row lists the last statement the descent needs and why it holds or fails.

Dimension	Last statement needed	Status
$D = 2$	$H_1(S^0) = 0$	Holds: uniformly continuous paths in the two points of S^0 are constant (Lemma 23.17).
$D \geq 4$	$\tilde{H}_0(S^{D-3}) = 0$	Holds: S^{D-3} has positive dimension, and any two of its points are joined by a path (Lemma 23.18).
$D = 3$	$H_1(S^1) = 0$	Fails: every cycle of the circle carries a winding number, and a cycle of winding number 1 bounds nothing (Theorem 23.38), so the descent stops. Separately, $H_3(S^3)$ does not vanish (Corollary 23.37, from the linked equatorial circle).

23.4 The suspension ladder

The sphere S^D is covered by the complements of its two poles, $S^D \setminus \{N\}$ and $S^D \setminus \{S\}$. Each is acyclic (Lemma 23.12). Their intersection, the sphere without its poles, slides along meridians onto the equator, a copy of S^{D-1} (Figure 23.1). Classically these two facts make S^D the suspension of S^{D-1}, and the Mayer–Vietoris sequence gives $H_k(S^D) \cong H_{k-1}(S^{D-1})$ [194, Section 2.2]. We prove the two implications that the ladder (23.1) uses. A Mayer–Vietoris step lowers the degree; a retraction step lowers the dimension; one of each lowers the degree and the dimension together. The descent can end in three ways, which Table 23.1 lists.

Lemma 23.15 (Mayer–Vietoris for two point complements)**.** *Let a and b be points of S^D with $d(a,b) > 0$, and let $k \in \{1,2,3\}$. If $\tilde{H}_0(S^D \setminus \{a,b\}) = 0$ when $k = 1$, or $H_{k-1}(S^D \setminus \{a,b\}) = 0$ when $k = 2,3$, then $H_k(S^D) = 0$.*

Proof. Every point y is $d(a,b)/4$-apart from a or from b: from $d(a,y) + d(y,b) \geq d(a,b) > d(a,b)/2$, cotransitivity gives $d(a,y) > d(a,b)/4$ or $d(y,b) > d(a,b)/4$. Let z be a k-cycle of S^D.

1. By Lemma 23.10, one number m of subdivisions makes every simplex of $\mathrm{Sd}^m z$ smaller than $d(a,b)/8$; there are finitely many simplices, and one m serves all.

2. The first vertex of each simplex is $d(a,b)/4$-apart from a or from b, and then the whole simplex is $d(a,b)/8$-apart from the same point. Sorting gives $\mathrm{Sd}^m z = x_a + x_b$, with x_a apart from a and x_b apart from b.

3. Since $\mathrm{Sd}^m z$ is a cycle, $\partial x_a \sim -\partial x_b$. By Lemma 23.9, $\partial x_a \sim w$ for a $(k-1)$-chain w apart from both points. It is a cycle, since $\partial\partial = 0$; for $k = 1$ its coefficients sum to 0, as those of every boundary of a 1-chain do. By hypothesis $w \sim \partial e$ with e apart from both points.

4. Then $x_a - e$ is a k-cycle apart from a and $x_b + e$ a k-cycle apart from b. By Lemma 23.12, item 3, they bound in S^D: $x_a - e \sim \partial F_a$ and $x_b + e \sim \partial F_b$.

5. By Lemma 23.10, $z - \mathrm{Sd}^m z \sim \partial T'$, where T' telescopes the subdivision homotopies of $z, \mathrm{Sd}\, z, \dots, \mathrm{Sd}^{m-1} z$; the terms $T\partial$ vanish because each $\mathrm{Sd}^i z$ is a cycle.

So $z \sim \partial(F_a + F_b + T')$, a boundary of uniformly continuous simplices. Steps 3 and 4 use only that ∂ is linear, and are done once for all degrees. □

Lemma 23.16 (The complement of the poles slides onto the equator)**.** *Let D be a natural number. If $\tilde{H}_0(S^D) = 0$, then $\tilde{H}_0(S^{D+1} \setminus \{N, S\}) = 0$. If $H_1(S^D) = 0$, then $H_1(S^{D+1} \setminus \{N, S\}) = 0$. If $H_2(S^D) = 0$, then $H_2(S^{D+1} \setminus \{N, S\}) = 0$.*

Proof. Write a point of S^{D+1} as $x = (c, x')$ with $c = x_0$, and put $u = |x'|^2 = 1 - c^2$. For a scale κ with $0 < \kappa \le 1$ put $\phi(x) = \min(1, u/\kappa)$ and, for $t \in [0,1]$,

$$H(t,x) = \frac{\big((1 - t\phi)\, c,\ x'\big)}{\sqrt{n}}, \qquad n = (1 - t\phi)^2 c^2 + u.$$

Suppose $n < \kappa/8$. Then $u \le n < \kappa/8$, so $\phi < 1/8$ and $1 - t\phi > 7/8$, while $c^2 = 1 - u > 7/8$; hence $n \ge (1 - t\phi)^2 c^2 > (7/8)^3 > 1/8 \ge \kappa/8$, a contradiction. So $n \ge \kappa/8$ everywhere. The square root of the line (Proposition 12.19) then has an inverse bounded by $\sqrt{8/\kappa}$, so H is defined on all of $[0,1] \times S^{D+1}$, lands on the sphere, and is uniformly continuous. Four facts about it are short computations.

1. $H(0,x) = (c, x')/\sqrt{c^2 + u} = x$.

2. Where $u \ge \kappa$, $\phi = 1$ and $H(1,x) = (0, x')/\sqrt{u}$ lies on the equator.

3. Along H the quantity u does not decrease: the last coordinates of $H(t,x)$ have squared length $u/n \ge u$, because $n \le c^2 + u = 1$.

4. A point with $u \ge \kappa$ is $\kappa/2$-apart from both poles: its first coordinate has square $1 - u \le 1 - \kappa \le (1 - \kappa/2)^2$, so $|x_0 \mp 1| \ge \kappa/2$. Conversely, a point r-apart from both poles, with $r \le 1$, has $u \ge r^2$. For if $u < r^2$, every $|x_i|$ with $i \ge 1$ is below r and $c^2 > 1 - r^2 > 0$, so c is apart from 0 and has a sign; if $c > 0$ then $|c - 1| \le 1 - c^2 = u < r$, so $d(x, N) < r$, and if $c < 0$ likewise $d(x, S) < r$; both are impossible.

Let z be a k-cycle of $S^{D+1}\setminus\{N,S\}$, $k=1,2$, its simplices r-apart from both poles, where we may take $r\le 1$; put $\kappa=r^2$. Every value of z has $u\ge\kappa$. Summing the prism laws of Lemma 23.11 over the terms of z, with $F=H\circ(\mathrm{id}\times\sigma)$ for each simplex σ,

$$\partial Pz + P(\partial z) \sim H_1\circ z - z,$$

and $P(\partial z)\sim\emptyset$ because $\partial z\sim\emptyset$ and the prism acts term by term. By items 3 and 4 every simplex of Pz is $\kappa/2$-apart from both poles, and by item 2 the cycle $H_1\circ z$ lies on the equator. Dropping the first coordinate maps the equator onto S^D, preserving distances and commuting with faces, so it carries $H_1\circ z$ to a k-cycle z' of S^D (Lemma 23.9, part 2, transports the equivalences). By hypothesis $z'\sim\partial L$ in S^D. Putting 0 in front again carries L into the equator of S^{D+1}, at distance 1 from both poles, with boundary equivalent to $H_1\circ z$. Hence $z\sim\partial(L-Pz)$ in $S^{D+1}\setminus\{N,S\}$.

In degree 0 the prism of a point x is the path $t\mapsto H(t,x)$. For a 0-chain z of points apart from the poles with coefficient sum 0, these paths give $\partial Pz\sim H_1\circ z-z$; the dropped 0-chain has the same coefficient sum 0, bounds in S^D by hypothesis, and the argument closes as before. □

Physically read, Lemma 23.16 says that a loop or a membrane that keeps away from two antipodal events can be slid, without approaching them, onto the sphere of points midway between them.

The ladder needs two foundations, one for each way it can end.

Lemma 23.17 (On the 0-sphere every uniformly continuous simplex is constant)**.** *Every uniformly continuous map $g:[0,1]\to S^0$ is constant: $g(t)=g(0)$ for every t. Hence every 1-chain of S^0 is a boundary; in particular $H_1(S^0)=0$.*

Proof. The 0-sphere consists of the points v of the line with $v^2=1$. Two of them closer than 1 are equal. Indeed $(v-1)(v+1)=0$ and the two factors differ by 2, so by cotransitivity one of them exceeds $1/2$ in absolute value; it has an inverse, so the other factor is 0, and $v=1$ or $v=-1$. The same holds for w, and if the signs differed, $|v-w|$ would be 2. Now let ω be a modulus of g, put $h=\omega(1)/2$, and take N with $Nh\ge 1$. The points $s_n=\min(t,nh)$, $n=0,\dots,N$, run from $s_0=0$ to $s_N=t$ in steps of at most $h<\omega(1)$, so neighbouring values of g are closer than 1 and therefore equal; hence $g(t)=g(0)$. For a 1-chain, replace each term (n,g) by n times the constant 2-simplex at $g(0)$: its boundary is $n(c-c+c)\sim(n,c)=(n,g)$. □

Lemma 23.18 (Spheres of positive dimension are connected)**.** *For every $n\ge 1$, $\tilde H_0(S^n)=0$: every 0-chain of S^n whose coefficients sum to 0 is the boundary of a 1-chain of S^n.*

Proof. The poles are 2 apart, so every point x is $1/2$-apart from N or from S (the first line of the proof of Lemma 23.15). If x is apart from N, the homotopy of Lemma 23.12 for $p=N$ is a uniformly continuous path from x to $-N=S$; if x is apart from S, the same construction gives a path from x to N. By induction on the length of a 0-chain z, there are integers α,β whose sum is the coefficient sum of z, and a 1-chain L with $\partial L\sim\alpha N+\beta S-z$. If the coefficient sum is 0 then $\beta=-\alpha$, and $\alpha N-\alpha S$ is the boundary of α times a path from $e_1=(0,1,0,\dots,0)$ to N minus α times a path from e_1 to S; the point e_1 exists because $n\ge 1$, and it is 1-apart from both poles. So z is a boundary. □

Proposition 23.19 (The ladder)**.** *$H_3(S^D) = 0$ for $D = 2$ and for every $D \geq 4$. On the way, $H_1(S^m) = 0$ for every $m \geq 2$, and $H_2(S^m) = 0$ for $m = 1$ and every $m \geq 3$. No omniscience principle and no choice principle is used.*

Proof. Write MV for Lemma 23.15, applied to the poles, which are 2 apart, and R for Lemma 23.16. For $D = 2$:

$$H_1(S^0) = 0 \overset{\mathrm{R}}{\Longrightarrow} H_1(S^1 \setminus \{N, S\}) = 0 \overset{\mathrm{MV}}{\Longrightarrow} H_2(S^1) = 0$$
$$\overset{\mathrm{R}}{\Longrightarrow} H_2(S^2 \setminus \{N, S\}) = 0 \overset{\mathrm{MV}}{\Longrightarrow} H_3(S^2) = 0,$$

starting from Lemma 23.17. For $D = m + 4$, starting from Lemma 23.18 for S^{m+1}:

$$\tilde{H}_0(S^{m+1}) = 0 \overset{\mathrm{R}}{\Longrightarrow} \tilde{H}_0(S^{m+2} \setminus \{N, S\}) = 0 \overset{\mathrm{MV}}{\Longrightarrow} H_1(S^{m+2}) = 0 \overset{\mathrm{R}}{\Longrightarrow} H_1(S^{m+3} \setminus \{N, S\}) = 0$$
$$\overset{\mathrm{MV}}{\Longrightarrow} H_2(S^{m+3}) = 0 \overset{\mathrm{R}}{\Longrightarrow} H_2(S^{m+4} \setminus \{N, S\}) = 0 \overset{\mathrm{MV}}{\Longrightarrow} H_3(S^{m+4}) = 0.$$

The intermediate statements are the byproducts. □

The ladder has three endings (Table 23.1). For $D \geq 4$ it reaches the reduced zeroth homology of a sphere of positive dimension, which is a statement about paths: any two points can be joined. For $D = 2$ it stops one rung earlier, at the first homology of the 0-sphere, and holds there for the reason of the opening example: a uniformly continuous path cannot jump between the two points. For $D = 3$ it needs $H_1(S^1) = 0$, the first homology of the circle, and that is false: the circle carries a winding number. The descent breaks exactly where a cycle can wind, and this is why three is the dimension that links.

23.5 From a linked circle to the third homology of the sphere

Cut the circle into its right half $R = \{x_0 \geq 0\}$ and its left half $L = \{x_0 \leq 0\}$, which meet at the two *junctions* $(0, 1)$ and $(0, -1)$. For a native embedding f, the images $f(R)$ and $f(L)$ are removed sets (given by f on the points of each half), and so is the pair $J_f = \{f(0, 1), f(0, -1)\}$. The circle is the union of its halves only up to a double negation: "$x_0 \geq 0$ or $x_0 \leq 0$" for every point costs LLPO (Corollary 12.18). The argument never uses the union itself.

Lemma 23.20 (The complement of an arc is acyclic)**.** *For every D and every native embedding $f : S^1 \to S^D$, every uniformly continuous loop apart from $f(R)$ bounds in $S^D \setminus f(R)$, and every uniformly continuous loop apart from $f(L)$ bounds in $S^D \setminus f(L)$.*

Proof. The classical proof [194, Proposition 2B.1] bisects the arc until it finds a piece whose complement carries a non-bounding cycle, and the choice of the bad half at each stage is a decision. We glue instead of bisecting. The right half is parametrized by

$$w(u) = \Big(\frac{1 - u^2}{1 + u^2},\ \frac{2u}{1 + u^2}\Big), \qquad -1 \leq u \leq 1,$$

which is onto, with inverse $u = x_1/(1 + x_0)$; it is 4-Lipschitz, and $|u - v| \leq 2\, d(w(u), w(v))$. Let γ be a loop r-apart from $f(R)$, and let ρ be given by Lemma 23.12, item 1, for r: for every point

q of the arc, γ bounds a 2-chain ρ-apart from q. By uniform continuity of $f \circ w$ there is n such that, with $u_k = -1 + 2k/(n+1)$, each image piece $f(w[u_k, u_{k+1}])$ lies within $\rho/2$ of $f(w(u_k))$. Then γ bounds in the complement of each piece: the 2-chain that avoids $f(w(u_k))$ by ρ avoids the whole piece by $\rho/2$. Now glue. Suppose γ bounds in the complement of $f(w[u_0, u_{k+1}])$. Apply Lemma 23.14 with $K_1 = f(w[u_0, u_{k+1}])$, $K_2 = f(w[u_{k+1}, u_{k+2}])$ and K_{12} the single point $f(w(u_{k+1}))$. Its hypotheses hold: H_2 of the sphere minus a point vanishes by Lemma 23.12, item 2; the two pieces have finite nets from a grid in u, and they are separated near u_{k+1}, since points $w(u)$ and $w(v)$ with $u \le u_{k+1} \le v$ satisfy $|u - u_{k+1}| \le |u - v| \le 2\,d(w(u), w(v))$; so Lemma 23.13 supplies the Lebesgue condition. Lemma 23.14 gives a 2-chain with boundary equivalent to γ whose simplices are r_1-apart from the first piece and r_2-apart from the second. They are $\min(r_1, r_2)$-apart from the union $f(w[u_0, u_{k+2}])$. For a parameter u in $[u_0, u_{k+2}]$, $\neg\neg(u \le u_{k+1} \vee u \ge u_{k+1})$ holds with no principle, and the bound $d(\sigma(a), f(w(u))) \ge \min(r_1, r_2)$ is a negation, since $\le$ is the negation of $<$; a double negation suffices to prove a negation. Hence γ bounds in the complement of $f(w[u_0, u_{k+2}])$. After n gluings it bounds in the complement of the whole right half. The left half is the right half of f composed with the reflection $(x_0, x_1) \mapsto (-x_0, x_1)$. □

Lemma 23.21 (Gluing the two halves)**.** *Let $f : S^1 \to S^D$ be a native embedding with $H_2(S^D \setminus J_f) = 0$. Then every loop of the complement of f that bounds in $S^D \setminus f(R)$ and bounds in $S^D \setminus f(L)$ bounds in $S^D \setminus f$.*

Proof. Apply Lemma 23.14 with $K_1 = f(R)$, $K_2 = f(L)$ and $K_{12} = J_f$; the junctions lie in both halves. The halves are separated near the junctions: if $a \in R$, $b \in L$ and $d(a, b)$ is small, then $0 \le a_0 \le a_0 - b_0 \le d(a, b)$, so $a_1^2 = 1 - a_0^2$ is near 1, and a is near $(0, 1)$ or near $(0, -1)$; which one is read off by cotransitivity, since $|a_1| > 1/2$ makes a_1 apart from 0. Both halves have finite nets, from a grid in the parameter u and finitely many comparisons. So Lemma 23.13 gives the Lebesgue condition, and Lemma 23.14 gives a 2-chain c with boundary equivalent to the loop, its simplices r_1-apart from $f(R)$ and r_2-apart from $f(L)$. They are $\min(r_1, r_2)$-apart from the whole image of f: for a point x of the circle, $\neg\neg(x_0 \ge 0 \vee x_0 \le 0)$ holds with no principle, and the bound $d(\sigma(a), f(x)) \ge \min(r_1, r_2)$ is a negation, since $\le$ is the negation of $<$; a double negation suffices to prove a negation. So c is a 2-chain of $S^D \setminus f$. □

Lemma 23.22 (Two point complements from the third homology of the sphere)**.** *If $H_3(S^D) = 0$, then $H_2(S^D \setminus \{a, b\}) = 0$ for all points a, b of S^D with $d(a, b) > 0$.*

Proof. This is Mayer–Vietoris one degree up, for the complements of a and of b. Let z be a 2-cycle apart from a and from b. By Lemma 23.12, item 2, $z \sim \partial c_a$ with c_a apart from a and $z \sim \partial c_b$ with c_b apart from b. Then $c_a - c_b$ is a 3-cycle of S^D, so $c_a - c_b \sim \partial e$ by hypothesis. Subdivide e until its simplices are smaller than $d(a, b)/8$; as in the proof of Lemma 23.15, each is then apart from a or from b, and $\mathrm{Sd}^m e = e_a + e_b$. Now $\mathrm{Sd}^m c_a - \partial e_a \sim \mathrm{Sd}^m c_b + \partial e_b$, the left side apart from a and the right side apart from b, so by Lemma 23.9 both equal a chain w apart from both points, with $\partial w \sim \mathrm{Sd}^m \partial c_a \sim \mathrm{Sd}^m z$, using $\partial\partial = 0$ on 4-chains. Finally $z - \mathrm{Sd}^m z \sim \partial T'$ with $T' = \sum_{i<m} T(\mathrm{Sd}^i z)$, since z is a cycle, and T' takes values of z. So $z \sim \partial(w + T')$, a boundary apart from a and b. □

Proposition 23.23 (A linked circle needs a cycle that does not bound)**.** *Let D be a natural number. If $H_2(S^D \setminus \{a,b\}) = 0$ for all points a, b of S^D with $d(a,b) > 0$, then for every native embedding f of the circle into S^D every loop of the complement of f bounds there; so no such embedding is linked. In particular, if $H_3(S^D) = 0$, then D does not detect linking. No omniscience principle and no choice principle is used.*

Proof. Let f be a native embedding and γ a loop of its complement. It is apart from the image of f, hence from $f(R)$ and from $f(L)$, so it bounds in both complements (Lemma 23.20). The junctions are apart: with ω' the inverse modulus of f, $d(f(0,1), f(0,-1)) < \omega'(2)$ would give $d((0,1),(0,-1)) < 2$, but that distance is 2; so $d(f(0,1), f(0,-1)) \geq \omega'(2) > 0$. By hypothesis $H_2(S^D \setminus J_f) = 0$, and Lemma 23.21 shows that γ bounds in $S^D \setminus f$. So every loop of the complement bounds. The last sentence follows with Lemma 23.22. □

23.6 Dimensions zero and one

In the two lowest dimensions there is nothing to link, for two different reasons: the 0-sphere has no room for a circle, and a circle in the circle leaves no room beside itself.

Proposition 23.24 (No circle links in S^0 or S^1)**.** *No map $f : S^1 \to S^0$ is a native embedding. For every native embedding $f : S^1 \to S^1$, no point of S^1 is apart from the image of f. Hence neither $D = 0$ nor $D = 1$ detects linking. No omniscience principle and no choice principle is used.*

Proof. $D = 0$. Let $f : S^1 \to S^0$ be uniformly continuous with an inverse modulus ω'. The arc $u \mapsto z(u) = w(u)^2$, $u \in [-1,0]$, the square of the parametrization of the proof of Lemma 23.20 taken as a complex number, runs along the lower half of the circle from $(-1,0)$ to $(1,0)$, and it is 32-Lipschitz. So $f \circ z$ is a uniformly continuous map from an interval into S^0, hence constant, by the argument of Lemma 23.17: $f(1,0) = f(-1,0)$. Then $d(f(1,0), f(-1,0)) = 0 < \omega'(1)$, so $d((1,0),(-1,0)) < 1$; but that distance is 2.

$D = 1$. Let f be a native embedding of the circle in itself, with moduli ω and ω', and suppose a point p is r-apart from its image; we derive a contradiction. For q on the circle put $C = \langle p, q\rangle$, $S = p_0 q_1 - p_1 q_0$ and $E = 1 - C$. On the image of f, $E \geq r^2/2$, so the stereographic coordinate $\sigma(q) = S/E$ from p is defined there; it is Lipschitz with constant $6/\kappa^2$, $\kappa = r^2/2$, and the rational map that inverts it is 8-Lipschitz: $d(q,q') \leq 8|\sigma(q) - \sigma(q')|$. Put $H(x) = \sigma(f(x)) - \sigma(f(-x))$. Then $H(-x) = -H(x)$, and H is uniformly continuous. Let $\varepsilon = \omega'(1)/16$, and walk along the right half of the circle from $x^{(0)} = (0,-1)$ to $x^{(N)} = (0,1) = -x^{(0)}$ in steps over which H moves by less than ε. At each step cotransitivity gives $|H(x^{(i)})| > \varepsilon$ or $|H(x^{(i)})| < 2\varepsilon$. The second case is impossible: it gives $d(f(x), f(-x)) < 16\varepsilon = \omega'(1)$, hence $d(x,-x) < 1$, while $d(x,-x) = 2\max(|x_0|,|x_1|) \geq 1$ on the circle. So every $|H(x^{(i)})| > \varepsilon$, and $H(x^{(i)})$ has a sign. A step of less than ε cannot pass from above ε to below $-\varepsilon$, so all the signs agree; but $H(x^{(N)}) = -H(x^{(0)})$. Only finitely many comparisons were made. So the complement of f has no points, and a fortiori no loops. □

23.7 The 3-sphere links: a loop that sweeps four quarter turns

In S^3, inside the space of (x_0, x_1, x_2, x_3), take the equatorial circle $e(x, y) = (x, y, 0, 0)$. A loop that runs once around the circle of points $(0, 0, u, v)$ should link it, the way two rings of a chain link. To prove that such a loop bounds no 2-chain of the complement we need an invariant of chains that vanishes on every boundary and not on the loop. Classically it is the integral of the angle form around the origin of the plane of the last two coordinates, the linking cocycle. Here it is the angle swept around that origin, read through four overlapping charts by rational functions, with no arctangent and no decision. It comes as a relation between chains and values on the line, since a function would need data that a simplex does not carry. Figure 23.2 shows the charts and the loop.

Definition 23.25 (Charts and the diamond angle)**.** Let $\rho(x, y) = (y, -x)$ be the clockwise quarter turn of the plane $\mathbb{R}^2_\delta$. For a natural number j, *chart j* is the set of points p whose image $\rho^j(p) = (x', y')$ has $x' > 0$: chart 0 is $x > 0$, chart 1 is $y > 0$, chart 2 is $x < 0$ and chart 3 is $y < 0$. On chart j the *diamond angle* is

$$\theta_j(p) = j + \frac{y'}{x' + |y'|}, \qquad (x', y') = \rho^j(p).$$

The denominator is at least $x' > 0$, so it has an inverse. The angle counts quarter turns, anticlockwise: at $(1, 0)$, $(0, 1)$, $(-1, 0)$ and $(0, -1)$ it takes the values 0, 1, 2 and 3 in the charts that contain them. We write $|p| = \max(|x|, |y|)$ for the sup norm of a point of the plane.

Lemma 23.26 (Charts agree up to whole turns)**.** *On the overlap of charts j and $j + 1$, $\theta_{j+1} = \theta_j$. Chart $j + 4$ is chart j, with $\theta_{j+4} = \theta_j + 4$. Charts j and $j + 2$ do not meet. Hence, if p and q both lie in chart j and in chart k, with $j, k < 4$, then $\theta_j(q) - \theta_j(p) = \theta_k(q) - \theta_k(p)$.*

Proof. Since $\theta_j(p) = j + \theta_0(\rho^j p)$, it suffices to compare charts 0 and 1, where $x > 0$ and $y > 0$: $\theta_0 = y/(x + y)$, and $\theta_1 = 1 + (-x)/(y + x) = y/(x + y)$. Next, ρ^4 is the identity, and $\rho^{j+2}(p) = -\rho^j(p)$, whose first coordinate cannot be positive together with that of $\rho^j(p)$. For the last statement, k is j, $j + 1$ or $j + 3$ modulo 4, and on the overlap, which contains p and q, θ_k differs from θ_j by a constant (0 or ± 4). □

Definition 23.27 (The swept angle of a path)**.** Let $g : [0, 1] \to \mathbb{R}^2_\delta$ be uniformly continuous with $|g(t)| \geq c$ for all t, for some $c > 0$. An *admissible grid* for g is a number N with charts $J(0), \dots, J(N) < 4$ such that g maps the cell $[k/(N + 1), (k + 1)/(N + 1)]$ into chart $J(k)$ for each $k \leq N$. A point v of the line is a *swept angle* of g when, for some admissible grid,

$$v = \sum_{k=0}^{N} \Big(\theta_{J(k)}\big(g(\tfrac{k+1}{N+1})\big) - \theta_{J(k)}\big(g(\tfrac{k}{N+1})\big) \Big).$$

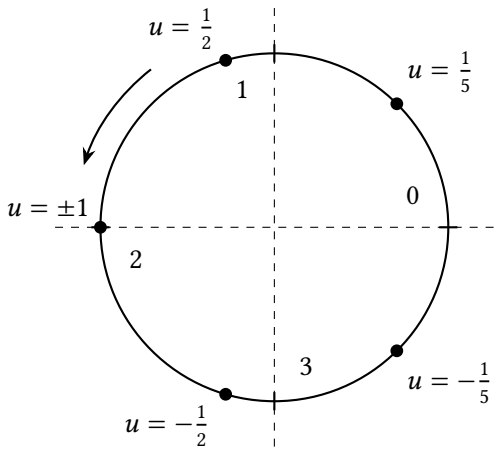


Figure 23.2. The plane of the last two coordinates of S^3. The dashed axes bound the four charts: chart 0 lies right of the vertical axis, chart 1 above the horizontal axis, chart 2 to the left and chart 3 below; the numbers inside the circle are the diamond angles of the four axis points. The loop of Lemma 23.33 runs once around the circle, anticlockwise from $u = -1$ (arrow); the dots mark the ends of the five pieces of Table 23.2.

Lemma 23.28 (A path that avoids the origin sweeps one angle)**.** *Every g as in Definition 23.27 has an admissible grid, and any two swept angles of g, on any two admissible grids, are equal.*

Proof. Let ω be a modulus of g and take N with $1/(N+1) < \omega(c/2)$. At the left end of each cell, $|g| \geq c > c/2$, so by cotransitivity one coordinate exceeds $c/2$ in absolute value, and a second use of cotransitivity gives its sign; that names a chart in which the relevant coordinate exceeds $c/2$. Over the cell, g moves by less than $c/2$, so the coordinate stays positive and the cell lies in the chart. These are finitely many comparisons. For uniqueness, refine both grids to the common grid of $(N+1)(N'+1)$ cells. Inside one coarse cell, the coarse difference is the sum of the fine differences read in the same chart, by telescoping. On each fine cell both readings use charts that contain both end points, so they agree by Lemma 23.26. □

The swept angle is a unique value, but it is a relation between g and v: the grid is computed from the modulus, and g asserts its modulus without handing it over. We never need a map from paths to values.

Lemma 23.29 (The projection stays away from the origin)**.** *Let $\pi(p) = (p_2, p_3)$ for p in S^3. If p is r-apart from the equatorial circle e, then $|\pi(p)| \geq \min(r/2, 1/4)$.*

Proof. Suppose $|p_2|$ and $|p_3|$ are both below $\min(r/2, 1/4)$. Then $s = p_0^2 + p_1^2 = 1 - p_2^2 - p_3^2 > 7/8$ and $1 - s < r/4$. The point $q = (p_0, p_1, 0, 0)/\sqrt{s}$ lies on the equatorial circle, and $|p_i - q_i| \leq (1/\sqrt{s} - 1) \leq (1 - s)/s < 2r/7$ for $i = 0, 1$, while $|p_i - q_i| < r/2$ for $i = 2, 3$. So $d(p, q) < r$, a contradiction. □

Lemma 23.30 (Boundaries sweep no angle). *Let $\tau : \Delta^2 \to \mathbb{R}^2_\delta$ be uniformly continuous with $|\tau| \geq c > 0$, and let v_0, v_1, v_2 be swept angles of its faces $\tau \circ d_0$, $\tau \circ d_1$, $\tau \circ d_2$. Then $v_0 - v_1 + v_2 = 0$.*

Proof. No Stokes theorem is needed; a ladder of shrunken copies of the triangle suffices. The boundary of Δ^2 runs from $(0,0)$ to $(0,1)$ along d_2, to $(1,1)$ along d_0, and back along d_1 reversed. Place M grid points on each edge, spaced $1/M$ apart, and for a level $\lambda = a/M$ scale them by λ towards $(0,0)$. Their images under τ, taken around the boundary, form a closed polygon P_a, read by adding chart differences of consecutive points. Choose M with $2/M$ below the modulus of τ at $c/2$. Then the four points $\tau(\lambda q)$, $\tau(\lambda q')$, $\tau(\lambda' q)$, $\tau(\lambda' q')$ for neighbouring grid points q, q' and neighbouring levels λ, λ' lie within $c/2$ of the first, hence in one chart, chosen as in Lemma 23.28. The reading of P_{a+1} minus that of P_a is a sum, over the small quadrilaterals, of chart differences around each; the rungs between levels are shared by neighbouring quadrilaterals and agree by Lemma 23.26, so the sum telescopes to 0. The polygon P_0 is constant, with reading 0. The polygon P_M is the boundary, and its reading is $v_2 + v_0 - v_1$, by Lemma 23.28 applied to each face. Every distance estimate here is a computation with ratios whose denominators divide M^2. □

Definition 23.31 (The linking value of a chain). For a native 1-simplex σ of $S^3 \setminus e$, the path $\pi \circ \sigma$ satisfies the hypothesis of Definition 23.27, by Lemma 23.29. A 1-chain $(n_1, \sigma_1), \dots, (n_m, \sigma_m)$ of $S^3 \setminus e$ has *linking value* v when $v = \sum_i n_i w_i$ for swept angles w_i of the paths $\pi \circ \sigma_i$.

Lemma 23.32 (The linking value is a cocycle). *Every 1-chain of $S^3 \setminus e$ has a linking value, and only one. Equivalent chains have equal linking values, and the boundary of every 2-chain of $S^3 \setminus e$ has linking value 0.*

Proof. Existence and uniqueness follow term by term from Lemma 23.28, by induction on the length of the chain. An exchange permutes a sum, a merge of two terms with the same simplex uses one swept angle for both, by uniqueness, and a deleted term contributes 0. For a 2-simplex τ of the complement, $\pi \circ \tau$ stays away from the origin by Lemma 23.29, and the faces of τ contribute $v_0 - v_1 + v_2 = 0$ by Lemma 23.30. □

Lemma 23.33 (The loop sweeps four quarter turns). *Put*

$$z(u) = \Big(\frac{(1-u^2)^2 - 4u^2}{(1+u^2)^2},\ \frac{4u(1-u^2)}{(1+u^2)^2} \Big), \qquad \gamma(t) = \big(0, 0, z(2t-1)\big), \quad t \in [0,1].$$

Then γ is a uniformly continuous loop of S^3, $1/2$-apart from the equatorial circle, and $\pi \circ \gamma$ has swept angle 4.

Proof. The point $z(u)$ is the square, as a complex number, of the point $w(u)$ of the proof of Lemma 23.20; so it lies on the circle, runs once around it as u goes from -1 to 1, and $z(-1) = z(1) = (-1, 0)$, which makes γ a loop. Its first two coordinates vanish, so $d(\gamma(t), e(x,y)) \geq |z(2t-1)| \geq 1/2$, since two numbers below $1/2$ in absolute value cannot have squares summing to 1. The map w is 4-Lipschitz, z is 32-Lipschitz, and $u = 2t - 1$ doubles distances, so $\varepsilon \mapsto \varepsilon/64$ is a modulus of γ.

Table 23.2. The angle swept by the loop of Lemma 23.33, piece by piece. The point $z(u)$ is given at the ends of each piece, with its diamond angle in the chart of that piece.

Piece in u	Chart	z at the ends	θ at the ends	Swept
$[-1,-\frac{1}{2}]$	2	$(-1,0)$, $(-\frac{7}{25},-\frac{24}{25})$	$2,\ 2+\frac{24}{31}$	$\frac{24}{31}$
$[-\frac{1}{2},-\frac{1}{5}]$	3	$(-\frac{7}{25},-\frac{24}{25})$, $(\frac{119}{169},-\frac{120}{169})$	$3-\frac{7}{31},\ 3+\frac{119}{239}$	$\frac{7}{31}+\frac{119}{239}$
$[-\frac{1}{5},\frac{1}{5}]$	0	$(\frac{119}{169},\mp\frac{120}{169})$	$-\frac{120}{239},\ \frac{120}{239}$	$\frac{240}{239}$
$[\frac{1}{5},\frac{1}{2}]$	1	$(\frac{119}{169},\frac{120}{169})$, $(-\frac{7}{25},\frac{24}{25})$	$1-\frac{119}{239},\ 1+\frac{7}{31}$	$\frac{119}{239}+\frac{7}{31}$
$[\frac{1}{2},1]$	2	$(-\frac{7}{25},\frac{24}{25})$, $(-1,0)$	$2-\frac{24}{31},\ 2$	$\frac{24}{31}$

Take the grid of 20 cells, so that each cell is an interval of length $1/10$ in u, with charts $2,2,2,2,2,3,3,3,0,0,0,0,1,1,1,2,2,2,2,2$. The charts are constant on the five pieces of Table 23.2, and on each piece the required sign is a polynomial inequality with an explicit positive lower bound. Write $z=(z_0,z_1)$. Then $z_0<0$ where $|u|\geq 1/2$, since there $4u^2-(1-u^2)^2\geq 7/16$; $z_1<0$ for $-1/2\leq u\leq -1/5$ and $z_1>0$ for $1/5\leq u\leq 1/2$, since there $4|u|(1-u^2)\geq 3/5$; and $z_0>0$ where $|u|\leq 1/5$, since there $(1-u^2)^2-4u^2\geq (24/25)^2-4/25>0$. Within a piece the cell differences telescope to the difference of the end values, which Table 23.2 lists. The total is

$$\frac{24}{31}+\Big(\frac{7}{31}+\frac{119}{239}\Big)+\frac{240}{239}+\Big(\frac{119}{239}+\frac{7}{31}\Big)+\frac{24}{31}=\frac{62}{31}+\frac{478}{239}=2+2=4.$$

□

Theorem 23.34 (The equatorial circle of the 3-sphere is linked)**.** *The loop γ of Lemma 23.33 bounds no 2-chain of $S^3\setminus e$. Hence the equatorial circle is a linked native embedding, and $D=3$ detects linking. No omniscience principle and no choice principle is used.*

Conditions and scope.

Base and premises

The native notions of Section 23.1 for $D=3$; the diamond-angle charts of Definition 23.25, which use only rational operations, absolute values and inverses of positive points.

Conclusion

One explicit loop of the complement of the equatorial circle of S^3 bounds no native 2-chain there. The invariant that shows it is a value relation on chains, unique on each chain, kept by equivalence and zero on boundaries.

Identification

Reading the swept angle as the linking number of two closed curves, and the two curves as a kept pair of records, is the physical reading of the theorem, not part of it.

Not claimed

That every native embedding of the circle into S^3 is linked. A linking function from chains to the line.

Proof. The equatorial circle is a native embedding (Remark 23.6), and γ is a loop of its complement (Lemma 23.33). If $(1, \gamma) \sim \partial L$ for a 2-chain L of $S^3 \setminus e$, then by Lemma 23.32 the linking value of $(1, \gamma)$, which is 4 by Lemma 23.33, equals that of ∂L, which is 0. But $4 \neq 0$. □

Read physically, the swept angle of a closed path in the plane normal to the curve's plane is its linking number with the curve, counted in quarter turns, and Lemma 23.30 says that a membrane kept away from the curve sweeps no angle around its own boundary. The loop sweeps four quarter turns, one full turn, so no membrane of the complement fills it.

23.8 Room in four dimensions: pulling a loop off the equatorial circle

The ladder proves that nothing is linked in S^4 without ever moving a loop, and Theorem 23.8 uses only the ladder. It is worth seeing the room directly, for the same equatorial circle; the proposition below does that, and nothing later depends on it. In S^4 the points away from the equatorial circle $e(x, y) = (x, y, 0, 0, 0)$ have three normal coordinates instead of two, and a loop in the normal directions can be slid off a direction it misses. Finding a missed direction without a decision is the interesting step. One tool turns homotopies into chains.

Lemma 23.35 (A homotopy through loops carries bounding)**.** *Let f be a native embedding into S^D and $H : [0,1] \times [0,1] \to S^D$ a uniformly continuous map that is uniformly apart from f, with $H(s,0) = H(s,1)$ for every s. If the loop $H(0,\cdot)$ bounds in $S^D \setminus f$, so does the loop $\gamma = H(1,\cdot)$.*

Proof. Cut the square of pairs (s,t) along its diagonal into two triangles and read H on each as a singular 2-simplex, A and B. Their faces are γ, $H(0,\cdot)$, the diagonal, counted once in each with opposite signs, and the two vertical sides $H(\cdot,0)$ and $H(\cdot,1)$, which are equal. So $\partial(B - A) \sim \gamma - H(0,\cdot)$, and γ bounds whenever $H(0,\cdot)$ does. This is the construction of the opening example. □

Proposition 23.36 (The equatorial circle of S^2 and of S^4 is not linked)**.** *For $D = 2$ and for $D = 4$, every loop of the complement of the equatorial circle in S^D bounds there. The proof uses explicit homotopies and no omniscience principle or choice principle.*

Proof. For $D = 2$ this is the opening example: the sign of the third coordinate is constant along the loop, and the hemisphere homotopy with Lemma 23.35 finishes the proof.

For $D = 4$ let γ be a loop r-apart from the equatorial circle, and write the normal part of a point p as $n(p) = (p_2, p_3, p_4)$. As in Lemma 23.29, $|n(\gamma(t))| \geq \zeta = \min(r/2, 1/4)/2$ for every t. For a vector v of $\mathbb{R}^3_\delta$ with $|v|$ bounded below, let $E(v) = (0, 0, v/|v|)$, a point of the normal 2-sphere; it is at least $1/\sqrt{3}$ from the equatorial circle. Three homotopies, each through loops that stay apart from the equatorial circle, carry γ to a constant loop, and three uses of Lemma 23.35 show that γ bounds.

Retraction. $R(s,p) = \big(sp_0,\ sp_1,\ \sqrt{1 - s^2(p_0^2 + p_1^2)}\, n(p)/|n(p)|\big)$. It lies on S^4, $R(1,p) = p$ because $|n(p)|^2 = 1 - p_0^2 - p_1^2$, and $R(0,p) = E(n(p))$. The normal part of $R(s,p)$ has length $\sqrt{1 - s^2(p_0^2 + p_1^2)} \geq |n(p)|$, so the homotopy stays apart from the equatorial circle.

Straightening. A uniformly continuous loop in the normal 2-sphere may pass through every one of its points; classically, Peano's curves fill whole squares [195]. So no missed direction need exist yet. Take M so large that γ moves by less than $\zeta/4$ over every time step $1/(M+1)$, and let P be the closed polygon through the samples $x_j = n(\gamma(j/(M+1)))$, $j = 0, \dots, M+1$, run at constant speed in t. Then $|P(t) - n(\gamma(t))| < \zeta/2$, so the segment from $n(\gamma(t))$ to $P(t)$ stays at least $\zeta/2$ from 0, and $E(s\,n(\gamma(t)) + (1-s)P(t))$ is a homotopy from $E \circ P$ to $E \circ n \circ \gamma$.

Contraction. The candidates are the points $q_k = (1, k, k^2)$ of the moment curve, for the integers k with $0 \le k \le 2M+2$. All bounds from here on are in the sup norm, in which $|q_k| \le Q = 1 + (2M+2)^2$ and $\zeta \le |x_j| \le 1$, the samples being normal parts of points of S^4; in this norm $|a \times b| \le 2|a||b|$ and $|\det(a,b,c)| \le 6|a||b||c|$. For a candidate q put $K(s,t) = E\big(sP(t) - (1-s)q\big)$: it runs from $E \circ P$ at $s = 1$ to the constant loop $E(-q)$ at $s = 0$, and it is defined as long as $sP(t) - (1-s)q$ stays away from 0, that is, as long as the polygon stays away from the ray through q. Put $e = \zeta/(16(1+2Q))$, and take the cross-product bound $\xi = \zeta e/2$, the chord bound $\eta = \xi e$, the determinant bound $\eta' = \eta e$ and $\mu = \eta' e$, all at most 1. Cotransitivity puts every chord, from x_j to x_{j+1}, into one of two cases: it is *spread*, with $|x_j \times x_{j+1}| > \eta$, or *nearly degenerate*, with $|x_j \times x_{j+1}| < 2\eta$. For each chord and each candidate q it decides again: q is *good* when $|\det(x_j, x_{j+1}, q)| > \eta'$ for a spread chord, or $|x_j \times q| > \xi$ for a nearly degenerate one, and *bad* when these are below $2\eta'$, respectively 2ξ. For a spread chord, $k \mapsto \det(x_j, x_{j+1}, q_k) = q_k \cdot (x_j \times x_{j+1})$ is a quadratic polynomial $f(k) = w_0 + w_1 k + w_2 k^2$ whose coefficients are the coordinates of $x_j \times x_{j+1}$. If $|f(k)| < \beta$ at three integers $k_1 < k_2 < k_3 \le A = 2M+2$, its coefficients are below $4\beta(1+2Q)$. Indeed, the divided difference $\big(f(k_2) - f(k_1)\big)/(k_2 - k_1) = w_1 + w_2(k_1 + k_2)$ is below 2β in absolute value, and so is the one for k_2, k_3; their difference divided by $k_3 - k_1 \ge 2$ is w_2, so $|w_2| < 2\beta$. Then $|w_1| < 2\beta + 2\beta(k_1 + k_2) < 2\beta(1 + 2A)$, and $|w_0| \le |f(k_1)| + |w_1|k_1 + |w_2|k_1^2 < \beta + 2\beta A + 6\beta A^2$; each of these is at most $4\beta(1+2Q) = 4\beta(3 + 2A^2)$. Three bad candidates would give $|x_j \times x_{j+1}| < 8\eta'(1+2Q) \le \eta$, against the spread, so at most two candidates are bad. For a nearly degenerate chord write $x_j = (y_0, y_1, y_2)$; the third coordinate of $x_j \times q_k$ is $ky_0 - y_1$, so if $|x_j \times q_k|$ and $|x_j \times q_{k'}|$ are below β for $k < k'$, then $|y_0|(k' - k) < 2\beta$, so $|y_0| < 2\beta$; the third coordinate then gives $|y_1| < \beta + 2\beta A$, and the second, $y_2 - y_0 k^2$, gives $|y_2| < \beta + 2\beta A^2$, so $|x_j| < 2\beta(1+Q)$; and two bad candidates would give $|x_j| < 4\xi(1+Q) < \zeta$. So at most one candidate is bad. Each of the $M+1$ chords rules out at most two of the $2M+3$ candidates, so some candidate is good for every chord; this count is over finitely many natural numbers.

With that candidate q, $|sP(t) - (1-s)q| \ge \mu$ for all s and t. This bound is a negation, so it suffices to prove it for t in the interval of one chord, which t is up to double negation. There $P(t) = (1-\lambda)x_j + \lambda x_{j+1}$ with $0 \le \lambda \le 1$, and $|P(t)| \ge \zeta/2$ by the straightening step. Suppose that $W = sP(t) - cq$, with $c = 1 - s$, has $|W| < \mu$. Then $s\zeta/2 < \mu + cQ$, that is, $c(\zeta/2 + Q) > \zeta/2 - \mu$. If the chord is spread, $P(t)$ lies in the plane of x_j and x_{j+1}, so $c|\det(x_j, x_{j+1}, q)| = |\det(x_j, x_{j+1}, W)| \le 6|W|$, and $c < 6\mu/\eta' = 6e$. If it is nearly degenerate, $x_j \times W = s\lambda(x_j \times x_{j+1}) - c(x_j \times q)$, so $c\xi < 2|W| + 2\eta < 4\eta$ and $c < 4e$. Either way $\zeta/2 < \mu + 6e(\zeta/2 + Q) \le e(4 + 6Q) \le \zeta/2$, which is absurd. So K is a homotopy through loops apart from the equatorial circle, ending at a constant loop, which bounds (Remark 23.6). □

The proof is a picture of why four dimensions leave room. Two normal coordinates around a circle in S^3 give a circle of directions, and a loop can wind around it; three normal coordinates in S^4 give a 2-sphere of directions, and a loop drawn as finitely many chords misses some direction, which can be found by counting.

23.9 Why three

All the pieces are in place.

Proof of Theorem 23.8. If $D = 3$, the equatorial circle is linked (Theorem 23.34). For $D \neq 3$ we prove a positive statement, stronger than the theorem's: for every native embedding f of the circle into S^D, every loop of the complement of f bounds a native 2-chain of the complement.

If $D = 2$ or $D \geq 4$, then $H_3(S^D) = 0$ (Proposition 23.19), and Lemma 23.22 gives $H_2(S^D \setminus \{a, b\}) = 0$ for all points a, b with $d(a, b) > 0$. The proof of Proposition 23.23 then composes three positive steps: a loop of the complement bounds in the complement of each half (Lemma 23.20); the two junctions of f are apart; and a loop that bounds in the complements of both halves bounds in the complement of f (Lemma 23.21, whose hypothesis is the two-point statement just obtained). If $D = 0$, no map of the circle into S^0 is a native embedding (Proposition 23.24), so there is nothing to check. If $D = 1$, the starting point of a loop of the complement would be a point of S^1 apart from the image of f, which Proposition 23.24 excludes; so the complement has no loop, and again there is nothing to check.

In each case with $D \neq 3$, a loop that bounds no 2-chain of the complement would contradict the positive statement, so no embedding is linked. Every case is a natural number D compared with 0, 1, 2 and 3, which is decided. Every lemma used was proved with no omniscience principle and no choice principle. □

The proof also settles which native spheres have vanishing third homology.

Corollary 23.37 (The third homology of the native spheres)**.** *For every $D \geq 2$, $H_3(S^D) = 0$ if and only if $D \neq 3$. For $D = 2$ and every $D \geq 4$, $H_2(S^D \setminus \{a, b\}) = 0$ for all points with $d(a, b) > 0$; for $D = 3$ this fails. No omniscience principle and no choice principle is used.*

Proof. For $D \neq 3$ this is Proposition 23.19 with Lemma 23.22. If $H_3(S^3) = 0$, or if every two-point complement of S^3 had vanishing H_2, Proposition 23.23 would show that no circle is linked in S^3, against Theorem 23.34. □

The descent (23.1) for $D = 3$ asks for $H_2(S^2) = 0$, and one rung lower for $H_1(S^1) = 0$. The corollary says that it must fail somewhere, and the next theorem says where: the first homology of the circle is $\mathbb{Z}$. Its cycles are read by the same swept angle, now for paths on the circle itself, whose points have $|p| \geq 1/\sqrt{2}$.

Theorem 23.38 (The first homology of the circle is $\mathbb{Z}$, as a relation)**.** *For 1-chains of S^1, say that a chain c has* label $n \in \mathbb{Z}$ *when c is a cycle with swept angle $4n$ (the sum, with coefficients, of the swept angles of its paths). Then every cycle has a label; no chain has two labels; equivalent chains have the same labels; labels add under concatenation; every boundary has label 0; a cycle with label 0 is a boundary; and some cycle has label 1. Moreover the swept angle of every cycle is $4k$ for an integer k, and the cycle is homologous to k copies of one fixed loop. No omniscience principle and no choice principle is used.*

Conditions and scope.

Base and premises

Native 1- and 2-chains of the circle (Definition 23.4 with nothing removed) and the swept angle of Definition 23.27.

Conclusion

The labels have the properties listed, which are those of the classical isomorphism $H_1(S^1) \cong \mathbb{Z}$ restricted to cycles; the label is a relation between cycles and integers, with no omniscience principle and no choice principle.

Not claimed

One integer-valued map on all 1-chains, open paths included; it costs exactly the WLPO function (Theorem 23.40 and Theorem 22.65). The first homology group as a quotient.

Proof. Swept angles of paths on the circle exist and are unique, add along chains, are kept by the moves, and vanish on boundaries, exactly as in Lemmas 23.28, 23.30 and 23.32; the loop z of Lemma 23.33, traced on the circle itself, is a cycle of swept angle 4, hence of label 1. The remaining work is the integrality and the direction "label 0 implies boundary". Subdivide every path of a cycle into pieces that each lie in one chart (Lemma 23.10, item 4). Replace each piece by a *chord*: the path between the same end points that moves linearly in the chart coordinate $y'/(x' + |y'|)$. The straight-line homotopy in that coordinate stays in the chart and fixes the end points, so by Lemma 23.11 the piece and its chord differ by a boundary (constant paths are boundaries). The *fundamental loop* runs by eight chords through the axis points $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$ and the points $(\pm\frac{3}{5}, \pm\frac{4}{5})$ between them; it is a cycle with swept angle 4. For a point x of chart j, the *ray* to x runs from $(1, 0)$ along the fundamental loop to the axis point of chart j, then by one chord to x. A chord from p to q in chart j is homologous to the ray to q minus the ray to p, by a triangle inside the chart. If x lies in charts j and $j + 1$, the two rays to x differ by a boundary, or, across charts 3 and 0, by one fundamental loop: with m_j the point of the fundamental loop between the axis points of charts j and $j + 1$, a triangle in chart j through its axis point, m_j and x, and one in chart $j + 1$ through m_j, its axis point and x, leave the two chords from m_j to x read in charts j and $j + 1$; these coincide, since the overlap is an interval of the chart coordinate on which $\theta_{j+1} - \theta_j$ is constant (Lemma 23.26). Now give every term of the boundary of the cycle the chart of the chord it comes from, so that the sum of the rays to its terms, with their coefficients, is homologous to the cycle, and follow the moves that reduce this boundary to the empty list. An exchange changes nothing; a merge or split of two terms at the same point changes the sum of rays by whole fundamental loops, as their charts are equal or neighbours (charts j and $j + 2$ do not meet); a term with coefficient 0 contributes nothing. By induction, the cycle is homologous to k fundamental loops and its swept angle is $4k$. Its label is k. If the label is 0, then $k = 0$ and the cycle is a boundary. Every step is a finite construction from moduli and grids. □

A label is attached to cycles only. An open path has a swept angle, but an integer for it would need a branch of the angle at its end points, a choice of which whole turn to count from, and nothing in the path supplies one. Whether the label can be extended to one integer-valued function on all chains is the question of Section 23.10, and it has a price.

So the three endings of Table 23.1 explain the theorem. A linked circle in S^D needs a 3-cycle of S^D that does not bound, and the descent derives the vanishing of the third homology of S^D from that of the first homology of S^{D-2}; its steps are implications in one direction. Among the spheres only the circle has cycles that wind: $H_1(S^0) = 0$ because paths in two points are constant, and $H_1(S^m) = 0$ for $m \geq 2$ because S^{m-1} is connected (Proposition 23.19). The sphere S^{D-2} is the circle exactly when $D = 3$, so three is the one dimension the descent cannot settle; there the linked equatorial circle (Theorem 23.34) shows directly that $H_3(S^3)$ does not vanish. Recognition Science reads this as the reason space has three dimensions; with $D = 3$, the count $2^D = 8$ of binary states in Chapter 22 is arithmetic.

23.10 The winding number as data, and its price

Theorem 23.38 gives every cycle of the circle an integer label, as a relation. The classical isomorphism $H_1(S^1) \cong \mathbb{Z}$ is read as more: one homomorphism from chains to integers. Under the reading of maps of Parts VI–VIII (Section 21.1) a map is a rule given as data, and from "each cycle has exactly one label" we do not form a map. So the *data form* of the circle's first homology is a separate statement, and it asks for a value on every chain, open paths included. It is row R2-35d of the table of Chapter 22, and it costs exactly the WLPO function (Theorem 22.65). Table 23.3 sets it among the results of this chapter.

Definition 23.39 (The data form and the principles around it)**.** A *path* is a uniformly continuous map $[0,1] \to S^1$, and 1-chains of S^1 are combinations of paths. Write $ix = (-x_1, x_0)$ for the anticlockwise quarter turn of a point $x = (x_0, x_1)$ of the circle.

1. The *data form* of $H_1(S^1) = \mathbb{Z}$: a map W from 1-chains of S^1 to $\mathbb{Z}$ such that equivalent chains have equal values, W of c followed by c' is $W(c) + W(c')$, $W(\partial L) = 0$ for every 2-chain L, every cycle c with $W(c) = 0$ is a boundary, and some cycle g has $W(g) = 1$.
2. A *quarter-turn colouring*: a map $q : S^1 \to \mathbb{Z}$ with $q(ix) \equiv q(x) + 1 \pmod 4$ for every x.
3. An *antipodal two-colouring*: a map $b : S^1 \to \{0,1\}$ with $b(-x) \neq b(x)$ for every x.
4. The *non-constant unit principle*: a map $F : [0,1] \to \mathbb{Z}$ with $F(0) = 0$ and $F(1) \neq 0$.
5. A *branch of the angle*: a map $\theta : S^1 \to \mathbb{R}_\delta$ such that for every x and every chart $j < 4$ containing x, the diamond angle $\theta_j(x)$ is $\theta(x) + 4m$ for some integer m.
6. *Unique choice of path labels*: every relation between paths and integers that relates each path to exactly one integer contains a map from paths to integers.
7. *Borsuk–Ulam for two-colourings* (BU_2): every map $b : S^1 \to \{0,1\}$ gives some pair of antipodal points the same colour, $b(-x) = b(x)$; its *weak form* says that no map b has $b(-x) \neq b(x)$ for every x.
8. *Continuity on Cantor space*: for every map d from binary sequences to $\{0,1\}$ and every binary sequence α there is n such that $d(\beta) = d(\alpha)$ for every β that agrees with α below n.

The WLPO function is that of Definition 21.1: one map from binary sequences to $\{0,1\}$ that answers "is the sequence identically 0?".

Table 23.3. The native statements of this chapter and their prices. "Free" means proved with no omniscience principle and no choice principle.

Statement	Price	Result
A circle can be linked in S^D exactly when $D = 3$	free	Theorem 23.8
The equatorial circle of S^3 is linked	free	Theorem 23.34
$H_3(S^D) = 0$ exactly when $D \neq 3$, for $D \geq 2$	free	Corollary 23.37
$H_1(S^m) = 0$ for $m \geq 2$; $H_2(S^m) = 0$ for $m = 1$, $m \geq 3$	free	Proposition 23.19
Cycles of S^D apart from a point bound	free	Lemma 23.12
Loops apart from an arc bound in its complement	free	Lemma 23.20
No circle in S^0; no room beside a circle in S^1	free	Proposition 23.24
The equatorial circle of S^2 and of S^4 is not linked, by explicit homotopies	free	Proposition 23.36
$H_1(S^1) = \mathbb{Z}$ as a relation	free	Theorem 23.38
$H_1(S^1) = \mathbb{Z}$ as a function on all chains	the WLPO function; refuted by BU_2	Theorems 23.40, 22.65
The WLPO function	refuted by continuity on Cantor space	Theorem 23.40

Theorem 23.40 (The price of the winding number as data)**.** *On* $\mathbb{R}_\delta$*:*

1. *the data form implies a quarter-turn colouring, a quarter-turn colouring implies an antipodal two-colouring, and an antipodal two-colouring implies the non-constant unit principle;*
2. *the WLPO function gives a branch of the angle; a branch of the angle with unique choice of path labels gives the data form; so the WLPO function with unique choice of path labels gives the data form, and the WLPO function alone gives the non-constant unit principle;*
3. BU_2 *refutes the antipodal two-colouring, hence the quarter-turn colouring and the data form; the nonexistence of an antipodal two-colouring is exactly the weak form of* BU_2*; and continuity on Cantor space refutes the WLPO function.*

Every implication is proved with no omniscience principle and no choice principle.

Conditions and scope.

Base and premises

The line and the native circle; the reading of maps of Section 21.1. Each part is an implication between the named statements.

Conclusion

The data form lies between the quarter-turn colouring, which it implies, and the WLPO function with unique choice of path labels, which imply it. Both ends are refuted by principles that hold in Brouwer's intuitionistic mathematics. Theorem 22.65 closes the bracket: each of the four data forms is exactly the WLPO function, and unique choice of path labels is not needed.

Not claimed

Within this theorem, the converses of parts 1 and 2; they are Theorem 22.65. That BU_2 or continuity on Cantor space is consistent with the base: that is the standard fact about Brouwerian models cited below, not a model built here.

Proof. Part 1. Let W be the data form. Then $W(\varnothing) = 0$, since $W(\varnothing) = W(\varnothing) + W(\varnothing)$, and $W(-c) = -W(c)$, since $-c$ followed by c is equivalent to the empty list. The *quarter chord* at x is the path $s \mapsto ((1-s)x + six)/\sqrt{(1-s)^2 + s^2}$, defined because $(1-s)^2 + s^2 \geq 1/2$; it sweeps exactly one quarter turn. To see this, read it on the halves $s \leq \frac{1}{2}$ and $s \geq \frac{1}{2}$. With x in chart j and $\rho^j(x) = (\alpha, \gamma)$, $\alpha > 0$, the chart coordinates of the chord are proportional to $((1-s)\alpha - s\gamma, (1-s)\gamma + s\alpha)$, and cotransitivity gives $\gamma < 0$, when the whole chord lies in chart j, or $\gamma > -\alpha$ and then $\gamma > 0$, when it lies in chart $j+1$, or $\gamma < \alpha$, when the halves lie in charts j and $j+1$. In each case the readings add up to $\theta_{j+1}(ix) - \theta_j(x) = 1$, as $\rho^{j+1}(ix) = (\alpha, \gamma)$ and $\theta_{j+1} = \theta_j$ on the overlap. The quarter chords at x, ix, $-x$, $-ix$ in turn form a cycle of label 1. Let g be the cycle with $W(g) = 1$, and let k_0 be its label (Theorem 23.38). For a cycle c of label n, the cycle $k_0 c - ng$ has label 0, so it is a boundary and W vanishes on it: $k_0 W(c) = n$. Put $G(x) = W(\text{quarter chord at } x)$. The cycle of four quarter chords gives

$$k_0\big(G(x) + G(ix) + G(-x) + G(-ix)\big) = 1,$$

so $k_0 = \pm 1$. Put $q(x) = k_0\big(3G(x) + 2G(ix) + G(-x)\big)$. With a, b, c, d the values of G at $x, ix, -x, -ix$,

$$q(ix) - q(x) - 1 = k_0(3b + 2c + d - 3a - 2b - c) - 1 = k_0(a + b + c + d) - 1 - 4k_0 a = -4k_0 a,$$

a multiple of 4. So q is a quarter-turn colouring. Applying it twice, $q(-x) \equiv q(x) + 2 \pmod 4$, so $b(x) = 1$ when $q(x)$ is 0 or 1 modulo 4, and $b(x) = 0$ otherwise, is an antipodal two-colouring; residues of integers are decided. Finally, let w parametrize the right half of the circle from $(0, -1)$ to $(0, 1)$ as in Lemma 23.20, and put $F(s) = 0$ if $b(w(2s-1)) = b(0, -1)$ and $F(s) = 1$ otherwise; equality of bits is decided. Then $F(0) = 0$, and $F(1) = 1$ because $(0, 1) = -(0, -1)$.

Part 2. With the WLPO function every comparison of a point with a ratio by $\leq$, in either order, is decided as data: $x \leq a$ says that a binary sequence computed from the stages of x is identically 0, and $a \leq x$ is $-x \leq -a$. At a point x of the circle, decide first whether $|x_0| \geq 1/4$. If so, decide whether $x_0 \leq 0$: then x lies in chart 2 or in chart 0, because $|x_0| \geq 1/4$ makes x_0 apart from 0. If not, then $x_1^2 = 1 - x_0^2 \geq 15/16$, so $|x_1| > 1/4$, and deciding whether $x_1 \leq 0$ places x in chart 3 or in chart 1. Let $\theta(x)$ be the diamond angle in the chart so picked. By Lemma 23.26, in any other chart containing x the angle differs by 0 or ± 4: θ is a branch of the angle. Only comparisons with $\leq$ were decided, so the LPO function is not needed.

Now let θ be a branch. For a path γ with swept angle w, read on an admissible grid, replace each chart reading by θ: each changes by a multiple of 4, and the θ readings telescope. So $w + \theta(\gamma(0)) - \theta(\gamma(1)) = 4n$ for exactly one integer n. This relation between paths and integers relates each path to one integer, and unique choice of path labels turns it into a map $n(\gamma)$. Put $W(c) = \sum_i n_i\, n(\gamma_i)$ for $c = (n_1, \gamma_1), \ldots, (n_m, \gamma_m)$. It respects the moves and adds. On a cycle the end-point terms cancel, because θ is a map and the 0-chain of end points is equivalent to the empty list; so $4W(c)$ is the swept angle of c and $W(c)$ is its label. Theorem 23.38 then gives

the remaining properties: boundaries have label 0, a cycle of label 0 is a boundary, and some cycle has label 1. The WLPO function also gives the non-constant unit principle directly: the map $F(s) = 0$ if $s \leq 1/2$ and $F(s) = 1$ otherwise is defined by one comparison of s with $1/2$.

Part 3. An antipodal two-colouring is a map b that BU_2 says cannot exist, and the negation of "some b has $b(-x) \neq b(x)$ for every x" is, by logic alone, "no b has it". Let d be a WLPO function, so $d(\bar{0}) = 1$ for the sequence $\bar{0}$ of zeros. Continuity at $\bar{0}$ gives n with $d(\beta) = 1$ for every β that agrees with $\bar{0}$ below n. The sequence β with a single 1, at place n, is such a sequence, yet it is not identically 0, so $d(\beta) = 0$. This contradiction refutes the WLPO function. □

Both separating principles hold in Brouwer's intuitionistic mathematics, where every map from Cantor space to $\{0, 1\}$ is continuous, and every map from the connected circle to $\{0, 1\}$ is constant [4], [40]; there, by part 3, neither the data form nor the WLPO function exists. Both principles fail classically: the colouring $b(x) = 1$ exactly when $x_1 > 0$, or $x_1 = 0$ and $x_0 > 0$, is antipodal, and excluded middle with unique choice gives a WLPO function. The classical Borsuk–Ulam theorem says that every continuous map from the circle to the line takes equal values at two antipodal points [192]; BU_2 is its form for two-valued maps, made a principle by dropping the continuity hypothesis.

So the free relation of Theorem 23.38 is all that the proof of Theorem 23.8 needs, and the function costs more. Read physically: one integer label on all records of motion around a circle, read as a single function, fixes at every point a phase in $\mathbb{Z}/4$ that advances by one under a quarter turn, a discontinuous field of quarter phases; a world whose two-valued readings of the circle obey BU_2 cannot carry it.

Questions to work through

1. (Lemma 23.17.) Show that a native embedding of the circle into S^0 cannot exist, using only Lemma 23.17 and the inverse modulus, and explain why no decision of the sign of a point is needed to show that two points of S^0 closer than 1 are equal.

2. (Lemma 23.12.) Verify $|2tx + kp|^2 = Q^2$ from the definitions of A, Q and k. For $D = 1$, $p = (1, 0)$ and $x = (0, 1)$, compute $H(\frac{1}{2}, x)$, check that it lies on the circle, and check that it is farther from p than x is.

3. (Lemma 23.26, Theorem 23.38.) The fundamental loop runs through $(1, 0)$, $(\frac{3}{5}, \frac{4}{5})$, $(0, 1)$, $(-\frac{3}{5}, \frac{4}{5})$, $(-1, 0)$, $(-\frac{3}{5}, -\frac{4}{5})$, $(0, -1)$, $(\frac{3}{5}, -\frac{4}{5})$ and back. Compute the diamond angle swept by each of its eight chords and show that the total is 4.

4. (Lemma 23.16.) Show that the slide of Lemma 23.16 does not reach the equator from points near a pole, and explain why the proof never needs it to: which property of the cycle z makes $H(1, \cdot)$ land on the equator at every value of z?

5. (Theorem 23.40, part 1.) Given a quarter-turn colouring q, show that $b(x) = 1$ exactly when $q(x)$ is 0 or 1 modulo 4 is an antipodal two-colouring, and that the identity $q(ix) - q(x) - 1 = -4k_0 G(x)$ holds for the colouring built from W.

6. (Theorem 23.40, part 3.) Prove that continuity on Cantor space refutes the WLPO function. Then show classically that both BU_2 and continuity on Cantor space are false.

Chapter 24

Ledgers, rates and the light cone

An account kept by double entry changes in one way only: a post adds one unit to its debit side or one unit to its credit side. After some posts the account's record is a pair of counts (D, C). What does one account's history of debits and credits already contain?

More than bookkeeping suggests. Take the eight posts

$$d\,c\,d\,d\,c\,d\,c\,d,$$

where d is a debit and c a credit. After each post write down the record (D, C), the number of posts $t = D + C$, the balance $x = D - C$ and the product DC:

post	d	c	d	d	c	d	c	d
(D, C)	$(1,0)$	$(1,1)$	$(2,1)$	$(3,1)$	$(3,2)$	$(4,2)$	$(4,3)$	$(5,3)$
t	1	2	3	4	5	6	7	8
x	1	0	1	2	1	2	1	2
DC	0	1	2	3	6	8	12	15

Every post raises t by one and moves x by one, up for a debit and down for a credit. In the plane of x and t each post is a step at speed one, and the record never leaves the cone $|x| \le t$. The product DC is a quarter of $t^2 - x^2$: at the end, $(64 - 4)/4 = 15 = 5 \cdot 3$.

The cost $J(r) = \frac{1}{2}(r + r^{-1}) - 1 = (r-1)^2/(2r)$ of Chapter 21 reads this motion. The final ratio $D/C = 5/3$ costs $J(5/3) = 2^2/(2 \cdot 5 \cdot 3) = 2/15$. With the velocity $\beta = x/t = 1/4$ and $\gamma = (1 - \beta^2)^{-1/2} = 4/\sqrt{15}$, that is exactly $2(\gamma\beta)^2$. The square root of the ratio costs $J(\sqrt{5/3}) = 8/(2\sqrt{15}) - 1 = \gamma - 1$, the excess of the Lorentz factor over one.

Now split the history after four posts into two legs, *dcdd* with counts $(3,1)$ and *cdcd* with counts $(2,2)$. The number $\tau = 2\sqrt{DC}$ is $2\sqrt{3}$ for the first leg, 4 for the second and $2\sqrt{15}$ for the whole history. The whole exceeds the sum of its parts, $2\sqrt{15} \approx 7.746 > 2\sqrt{3} + 4 \approx 7.464$, as the proper time of the stay-at-home twin exceeds that of the traveller, and the excess is again a value of J:

$$(2\sqrt{15})^2 - (2\sqrt{3} + 4)^2 = 32 - 16\sqrt{3} = 2 \cdot 2\sqrt{3} \cdot 4 \cdot J\big(\sqrt{3/1}\big).$$

The integer core of this identity needs no square roots. With $s_1 = 3 \cdot 1$, $s_2 = 2 \cdot 2$ and $s = 5 \cdot 3$, $(s - s_1 - s_2)^2 - 4s_1 s_2 = 64 - 48 = 16 = (3 \cdot 2 - 2 \cdot 1)^2$. Figure 24.1 draws the history.

This chapter proves that these coincidences are laws, and finds where they stop. One account's records fill exactly the future light cone of the interval $DC = (t^2 - x^2)/4$, and DC is the only quadratic form, up to a factor, that vanishes on the two directions a single post can take (Section 24.1). J of the ratio is twice the squared proper velocity, and J of its square root is $\gamma - 1$ (Section 24.2). The number $\tau = 2\sqrt{DC}$ is superadditive over the legs of a history, and

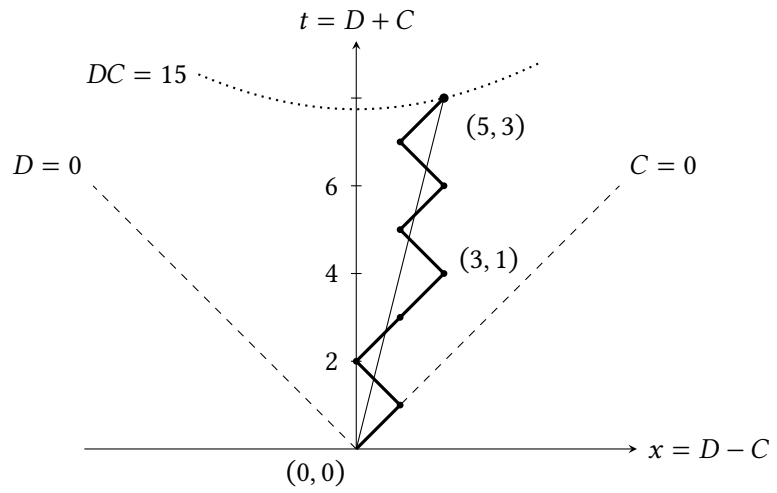


Figure 24.1. The history *dcddcdcd* in the plane of balance x and post count t (thick). A debit steps up and to the right, a credit up and to the left, so every post runs along the light cone (dashed). The thin chord is the inertial history with the same end record $(5, 3)$; the dotted hyperbola is $t^2 - x^2 = 60$, on which $DC = 15$.

the deficit is J of the relative ratio, so a history has the greatest τ for its end record exactly when it does not change ratio (Section 24.3). The posting word of the recognition hierarchy, the history above continued, approaches the inertial line of velocity $\varphi^{-3} = \sqrt{5} - 2$, bending at every step by the least amount an integer history can bend, and its time dilation is exactly $2\varphi^{-3/2} \approx 0.9717$ (Section 24.4).

None of this survives to more accounts. Several accounts posted one unit at a time reach exactly a taxicab cone, which is the cone of a quadratic form only in one space dimension (Section 24.5). The cone of one time and three space dimensions appears instead as the record cone of a single binary distinction carried coherently, by 2×2 Hermitian matrices whose diagonal is the ledger's debit and credit and whose coherent changes of determinant one are exactly the proper orthochronous Lorentz group (Section 24.6). Such a distinction can sit in any finite complex register, single posts are light-like under weak premises, and among coherent carriers only the complex one lets record counts multiply (Section 24.7). A history's comparison cost against its own frame is its proper-time deficit, so the least comparison cost is the greatest proper time and is attained exactly by inertial histories; counting shows that, over a split into two legs of fixed proportions, most posting orders have a first leg posting at nearly the rate of the whole (Section 24.8). The clock of Recognition Science has an exact resolution and a band in which it recognizes itself, and a change of a register's configuration has an action floor of sharp sine form (Section 24.9). We price the questions one can ask about an unending balance history: solvency forever, closing, waiting, alarms and deadlines (Section 24.10). Finally we climb from posting toward a clock, complex amplitudes and the square law, conserved charges and a space of cells, proving what posting forces and naming the premise at each step where it stops deciding (Section 24.11).

Three kinds of statement occur, and Table 24.1 sorts them. Statements about counts are statements about $\mathbb{N}$, $\mathbb{Z}$ and the golden integers $\mathbb{Z}[\varphi]$ of Definition 21.24, and we prove them by integer arithmetic. Those that the table marks free we claim with no omniscience principle and no choice principle. Those it marks *integer arithmetic* are proved in the same way, but

Table 24.1. Where the results of this chapter live.

Results	Objects	Logic
Reachability, interval, forms vanishing on the rays, common futures; twin identity in integers; Fibonacci counts and golden identities; Cassini and the kink identity; reach of separate accounts	$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}[\varphi]$	free
Cone inequality, light rays, diamond count, the frame; forms reversed by a join; the kink bound; taxicab reach; record counts; binomial word counts	$\mathbb{N}$, $\mathbb{Z}$	integer arithmetic, claimed classically
J and velocity; proper time; limits of the recognition word; dilation; quadratic cones; Hermitian records and $\mathrm{SO}^{+}(1,3)$; carriers; comparison cost; counting bounds; clocks and the action floor	classical $\mathbb{R}$ and $\mathbb{C}$	classical
Prices of solvency, closing, waiting, alarms, revision, deadlines	binary sequences, balances in $\mathbb{Z}_{\delta}$	choice-free; each price named
Refutable and confirmable: solvency and closing; computable deadlines (Kleene's tree)	recorders, programs	classical
One-bit flip of a post; symmetries and symmetric charges of three accounts	$\{0,1\}^3$, $\mathbb{Z}^d$	free
Uniform item times, finest changes and joint ticks; conservation under writing rules and transfer laws; gluings, transport and linked traces	patterns, ledgers, $\mathbb{Z}^D$	classical
Tick registers, the square law, contextual reads	classical $\mathbb{R}^8$ and $\mathbb{C}$	classical

we claim them only in classical mathematics: the label says how they are proved, not what they cost, and we make no claim about their constructive strength. The scope note of each result says which of its parts are free. Statements about square roots, limits, real quadratic forms and complex matrices are proved in classical mathematics over the classical real and complex numbers, and each says so; we do not claim them on $\mathbb{R}_{\delta}$. The prices of Section 24.10, the equivalences of questions about balance histories, alarms and recorders with principles on binary sequences, are proved with no choice principle; the refutable and confirmable classification of solvency and closing, and the statements there about computable runs, are proved classically. Readings of counts as time, position or spacetime are identifications. We name each one where it enters, and no theorem assumes one unless its statement says so.

24.1 One account fills a light cone

We first fix the objects of the opening.

Definition 24.1 (Accounts, posts and cone coordinates)**.** A *side* is a debit d or a credit c. A *posting word* is a finite string w of sides; $D(w)$ and $C(w)$ are its numbers of debits and of credits, so that its length is $|w| = D(w) + C(w)$. The *record* of an account is a pair $s = (D, C)$ of natural numbers. Posting w to s gives $s \cdot w = (D + D(w),\ C + C(w))$, and a record s' is *reachable* from s when $s' = s \cdot w$ for some posting word w. The *cone coordinates* of s are the integers

$$t(s) = D + C, \qquad x(s) = D - C,$$

the number of posts since the empty record and the balance. The *proper time* of a posting word is the number $\tau(w) = 2\sqrt{D(w)\,C(w)}$. A *ledger of d accounts* gives each account k a debit count D_k and a credit count C_k; a post at account k on a side adds one to that count of account k and changes no other count.

So each account of a ledger carries its own record, posted by its own words. The name "proper time" is a name for a function of two counts. Reading it as the time shown by a clock carried along the history is an identification, and so is reading t and x as time and position; we keep both apart from what is proved.

Theorem 24.2 (The light cone of one account)**.** *Let $s = (D, C)$ and $s' = (D', C')$ be records of one account, and put $\Delta t = t(s') - t(s)$ and $\Delta x = x(s') - x(s)$.*

(a) s' is reachable from s if and only if $D \le D'$ and $C \le C'$.

(b) Equivalently, s' is reachable from s if and only if $|\Delta x| \le \Delta t$.

(c) For integers a and b: $a \ge 0$ and $b \ge 0$ if and only if $4ab \ge 0$ and $a + b \ge 0$; and $(a+b)^2 - (a-b)^2 = 4ab$. Hence the reachable displacements are exactly those with $\Delta t \ge 0$ and $\Delta t^2 - \Delta x^2 \ge 0$, and $\Delta t^2 - \Delta x^2 = 4(D' - D)(C' - C)$.

(d) A posting word w has $|D(w) - C(w)| = D(w) + C(w)$ if and only if all its letters are debits or all are credits.

(e) For integers α, β, γ, the form $q(a, b) = \alpha a^2 + \beta ab + \gamma b^2$ vanishes on both axes, $q(n, 0) = q(0, n) = 0$ for every integer n, if and only if $q(a, b) = \beta ab$ for all integers a, b.

(f) A record s is reachable from both $s_1 = (D_1, C_1)$ and $s_2 = (D_2, C_2)$ if and only if it is reachable from $(\max(D_1, D_2), \max(C_1, C_2))$.

(g) If s' is reachable from s, the records r with $s \le r \le s'$ in the componentwise order number exactly

$$\frac{\Delta t^2 - \Delta x^2}{4} + \Delta t + 1.$$

Conditions and scope.

Base and premises

Natural numbers and integers; posting as in Definition 24.1. Every clause is proved below by integer arithmetic; clauses (a), (c), (e) and (f) with no omniscience principle and no choice principle, while (b), (d) and (g) are claimed in classical mathematics.

Conclusion

Posting fills exactly the future cone $\{\Delta t \geq |\Delta x|\}$; its light rays are the one-sided words; up to a factor, $DC = (t^2 - x^2)/4$ is the only integer quadratic form vanishing on both rays; common futures are futures of the componentwise maximum; a causal diamond holds $(\tau/2)^2 + \Delta t + 1$ records, τ its proper time.

Identification

That t is elapsed time and x position on the account's worldline, counted from the empty record, is an identification; the theorem does not supply it. So is the reading that a comparison of two records is a record in the future of both, which by (f) places it in the future of their componentwise maximum.

Not claimed

Any relation between different accounts: each account has its own plane.

Proof. (a) If $s' = s \cdot w$, then $D' = D + D(w) \geq D$ and $C' = C + C(w) \geq C$. Conversely, if $D \leq D'$ and $C \leq C'$, the word of $D' - D$ debits followed by $C' - C$ credits takes s to s'.

(b) Put $a = D' - D$ and $b = C' - C$, so that $\Delta t = a + b$ and $\Delta x = a - b$. The inequality $|a - b| \leq a + b$ says $-(a + b) \leq a - b \leq a + b$, that is, $b \geq 0$ and $a \geq 0$. By (a) this is reachability.

(c) If $a, b \geq 0$, then $4ab \geq 0$ and $a + b \geq 0$. Conversely, let $4ab \geq 0$ and $a + b \geq 0$. If $a < 0$, then $b > 0$ would give $ab < 0$, and $b \leq 0$ would give $a + b < 0$; so $a \geq 0$, and likewise $b \geq 0$. The identity is expansion. With $a = D' - D$ and $b = C' - C$ it gives $\Delta t^2 - \Delta x^2 = 4ab$, and the first part with (a) gives the description of reachable displacements.

(d) Write $D = D(w)$ and $C = C(w)$, both at least 0. The equation $|D - C| = D + C$ says $D - C = D + C$ or $C - D = D + C$, that is, $C = 0$ or $D = 0$. A word has no credit exactly when all its letters are debits, and no debit exactly when all are credits.

(e) Evaluating at $(1, 0)$ gives $\alpha = 0$, and at $(0, 1)$ gives $\gamma = 0$; then $q(a, b) = \beta ab$. Conversely βab vanishes when $a = 0$ or $b = 0$.

(f) By (a), reachability from s_1 and from s_2 says $D_1 \leq D$, $D_2 \leq D$, $C_1 \leq C$ and $C_2 \leq C$, which is $\max(D_1, D_2) \leq D$ and $\max(C_1, C_2) \leq C$; by (a) again this is reachability from the maximum.

(g) The records between s and s' are the pairs (i, j) with $D \leq i \leq D'$ and $C \leq j \leq C'$. There are $(a + 1)(b + 1) = ab + (a + b) + 1$ of them, and $4ab = \Delta t^2 - \Delta x^2$ by (c). □

The interval is thus not chosen. Among integer quadratic forms in the displacement $(D' - D, C' - C)$, those that vanish on the two kinds of single post are the multiples of $(D' - D)(C' - C)$, a quarter of $\Delta t^2 - \Delta x^2$. The integer records also fix a frame: no rescaling of debit against credit maps the record order to itself.

Proposition 24.3 (Counts single out a frame)**.** *Let f be a bijection of $\mathbb{N}\times\mathbb{N}$ such that $p\le q$ if and only if $f(p)\le f(q)$, in the componentwise order. Then f is the identity or the exchange $(D,C)\mapsto(C,D)$.*

Proof. An order bijection keeps every notion defined from the order alone. The least element $(0,0)$ is kept. Call p a *ray point* when any two records below p are comparable. The ray points are exactly the records with $D=0$ or $C=0$: below $(a,0)$ lie only the records $(i,0)$, which are comparable, while below (a,b) with $a,b\ge 1$ lie $(1,0)$ and $(0,1)$, which are not. Say that q *covers* p when $p<q$ and nothing lies strictly between; q covers (a,b) exactly when $q=(a+1,b)$ or $q=(a,b+1)$. The two records covering $(0,0)$ are $(1,0)$ and $(0,1)$, and f permutes them. Composing with the exchange if necessary, we may assume $f(1,0)=(1,0)$, and then $f(0,1)=(0,1)$.

Suppose $f(a,0)=(a,0)$ with $a\ge 1$. Then $f(a+1,0)$ covers $(a,0)$ and is a ray point, so it is $(a+1,0)$ or $(a,1)$; the second is not a ray point because $a\ge 1$. By induction f fixes every $(a,0)$, and in the same way every $(0,b)$. Finally, for $a,b\ge 1$ the record (a,b) is the least upper bound of $(a-1,b)$ and $(a,b-1)$, and an order bijection keeps least upper bounds; by induction on $a+b$, $f(a,b)=(a,b)$. Every comparison used is a comparison of natural numbers. □

A Lorentz boost of the plane would stretch D by a factor k and shrink C by $1/k$. On integer records no such map with $k\neq 1$ preserves the order, so the counts carry a preferred frame: the one in which they are counts. Boosts act on ratios of counts, which is where J enters.

24.2 The cost of a ratio reads the motion

A record (D,C) with $D,C\ge 1$ has the ratio $r=D/C$ and the velocity $\beta=x/t=(D-C)/(D+C)$. The two are tied by the map

$$v(r)=\frac{r-1}{r+1},\qquad r>0,$$

for $v(D/C)=(D-C)/(D+C)$. Write $\gamma=(1-\beta^2)^{-1/2}$ for the Lorentz factor and $\tau=2\sqrt{DC}$ for the proper time of any word with these counts.

Theorem 24.4 (J reads velocity and the Lorentz factor)**.**

(a) For natural numbers $D,C\ge 1$,

$$J\Big(\frac{D}{C}\Big)=\frac{2(D-C)^2}{(D+C)^2-(D-C)^2}=\frac{2x^2}{t^2-x^2}=2\gamma^2\beta^2.$$

(b) For natural numbers $D,C\ge 1$,

$$J\Big(\sqrt{D/C}\Big)=\frac{D+C}{2\sqrt{DC}}-1=\frac{t}{\tau}-1=\gamma-1.$$

(c) For every real $r>0$: $|v(r)|<1$, $J(r)=2v(r)^2/(1-v(r)^2)$, and $J(r)\ge\frac14$ if and only if $|v(r)|\ge\frac13$.

(d) For all real $r_1, r_2 > 0$,

$$v(r_1 r_2) = \frac{v(r_1) + v(r_2)}{1 + v(r_1)v(r_2)}.$$

(e) For every posting word w with $t = D(w) + C(w) > 0$ and $x = D(w) - C(w)$, $\tau(w) = t\sqrt{1 - (x/t)^2}$.

Conditions and scope.

Base and premises

The classical real numbers, for the square roots and the real ratio r. Proved in classical mathematics.

Conclusion

The cost of a record's ratio is twice its squared proper velocity $\gamma\beta$; the cost of the square root of the ratio, the Doppler factor, is $\gamma - 1$; the threshold $J \geq \frac{1}{4}$ is the speed $|\beta| \geq \frac{1}{3}$, a ratio outside $(\frac{1}{2}, 2)$; velocities of multiplied ratios add relativistically; proper time is time dilated by $\sqrt{1 - \beta^2}$.

Identification

Velocity, Lorentz factor and Doppler factor are names for functions of the counts. Their physical reading rests on the reading of t and x as time and position.

Not claimed

That any ledger operation multiplies record ratios: (d) is a property of the map v. Nor any statement on $\mathbb{R}_\delta$.

Proof. For $a, b > 0$,

$$J\Big(\frac{a}{b}\Big) = \frac{a^2 + b^2}{2ab} - 1 = \frac{(a-b)^2}{2ab}.$$

(a) Take $a = D$ and $b = C$. Since $(D + C)^2 - (D - C)^2 = 4DC$, the middle expression equals $2(D - C)^2/(4DC) = (D - C)^2/(2DC) = J(D/C)$. Dividing numerator and denominator of $2x^2/(t^2 - x^2)$ by t^2 gives $2\beta^2/(1 - \beta^2) = 2\gamma^2\beta^2$.

(b) Take $a = \sqrt{D}$ and $b = \sqrt{C}$: $J(\sqrt{D}/\sqrt{C}) = (D + C)/(2\sqrt{DC}) - 1$. This is $t/\tau - 1$, and $t/\tau = t/\sqrt{t^2 - x^2} = (1 - \beta^2)^{-1/2} = \gamma$.

(c) For $r > 0$, $|r - 1| < r + 1$, so $|v(r)| < 1$. Next $1 - v(r)^2 = 4r/(r + 1)^2$, so $2v(r)^2/(1 - v(r)^2) = 2(r - 1)^2/(4r) = (r - 1)^2/(2r) = J(r)$. Since $1 - v^2 > 0$, $2v^2/(1 - v^2) \geq \frac{1}{4}$ says $8v^2 \geq 1 - v^2$, that is, $v^2 \geq \frac{1}{9}$.

(d) With $v_i = v(r_i)$,

$$v_1 + v_2 = \frac{2(r_1 r_2 - 1)}{(r_1 + 1)(r_2 + 1)}, \qquad 1 + v_1 v_2 = \frac{2(r_1 r_2 + 1)}{(r_1 + 1)(r_2 + 1)},$$

and the quotient is $(r_1 r_2 - 1)/(r_1 r_2 + 1) = v(r_1 r_2)$.

(e) $2\sqrt{DC} = \sqrt{(D + C)^2 - (D - C)^2} = \sqrt{t^2 - x^2} = t\sqrt{1 - (x/t)^2}$. □

In rapidities the theorem is a statement about the hyperbolic cosine. With $\eta = \frac{1}{2}\ln r$ one has $v(r) = \tanh\eta$, $J(r) = \cosh 2\eta - 1$ and $J(\sqrt{r}) = \cosh\eta - 1 = \gamma - 1$; multiplying ratios adds rapidities, which is (d). The Doppler factor $\sqrt{D/C}$ is Bondi's k [138], and the velocity addition law is that of special relativity [147]. Part (c) translates the threshold $J \geq \frac{1}{4}$ that Chapter 21 attaches to a certain change into a speed: a third of the speed of a single post.

24.3 Proper time is superadditive, and the deficit is J

A *history* of an account is a finite list of legs $w_1, \dots, w_n$, posting words posted one after another; the whole history is the word $w_1 w_2 \cdots w_n$, whose counts are the sums of the legs' counts. It is *inertial* when all legs post in the same proportion, $D(w_i)C(w_j) = D(w_j)C(w_i)$ for all i and j. The opening split the history *dcddcdcd* into the legs *dcdd* and *cdcd* and found the whole longer in proper time than its parts. That is general, and the deficit has an exact form.

Theorem 24.5 (The twin paradox is J). *Let w_1, w_2 be posting words with counts (D_1, C_1) and (D_2, C_2), and write $s_i = D_i C_i$ and $s = (D_1 + D_2)(C_1 + C_2)$.*

(a) In integers, $(s - s_1 - s_2)^2 - 4 s_1 s_2 = (D_1 C_2 - D_2 C_1)^2$, and $s_1 + s_2 \leq s$.

(b) $\tau(w_1) + \tau(w_2) \leq \tau(w_1 w_2)$, with equality if and only if $D_1 C_2 = D_2 C_1$.

(c) For every finite list of words $w_1, \dots, w_n$, $\sum_i \tau(w_i) \leq \tau(w_1 \cdots w_n)$, with equality if and only if $D(w_i)C(w_j) = D(w_j)C(w_i)$ for all i and j.

(d) If $D_1, C_1, D_2, C_2 \geq 1$ and $r_i = D_i/C_i$, then

$$\tau(w_1 w_2)^2 - \big(\tau(w_1) + \tau(w_2)\big)^2 = 2\,\tau(w_1)\,\tau(w_2)\,J\big(\sqrt{r_1/r_2}\big).$$

(e) A word of $2n$ posts has $\tau \leq 2n$, with equality if and only if it has n debits and n credits. In particular a window of eight posts has $\tau \leq 8$, with equality exactly for four debits and four credits.

Proposition 24.6 (The deficit of many legs). *Let $w_1, \dots, w_n$ be posting words with $D(w_i)$, $C(w_i) \geq 1$, and write $\tau_i = \tau(w_i)$ and $r_i = D(w_i)/C(w_i)$. Then*

$$\tau(w_1 \cdots w_n)^2 - \Big(\sum_i \tau_i\Big)^2 = \sum_{i<j} 2\,\tau_i\,\tau_j\,J\big(\sqrt{r_i/r_j}\big).$$

Conditions and scope.

Base and premises

Part (a) is an identity of integers, proved with no omniscience principle and no choice principle. Parts (b) to (e) and the proposition involve square roots and are proved in classical mathematics over the real numbers.

Conclusion

Proper time $2\sqrt{DC}$ is superadditive over legs; the inertial history has the greatest proper time for its end record and is the only history that attains it; the squared deficit of two legs is $2\tau_1\tau_2$ times the cost J of their relative Doppler factor, and the deficit of many legs is the sum of these pair costs.

Identification

That the physical proper time of a posting history is $2\sqrt{DC}$ is an identification, and so is the reading of one eight-tick cycle of Recognition Science as eight posts at one account; under the latter, (e) says a cycle has proper time at most 8, reached exactly by the balanced cycle.

Proof. For legs $i \neq j$ put $p_{ij} = D(w_i)C(w_j)$. The whole history has counts $\sum_i D(w_i)$ and $\sum_j C(w_j)$, so

$$\frac{1}{4}\tau(w_1 \cdots w_n)^2 = \sum_{i,j} D(w_i)C(w_j) = \sum_i s_i + \sum_{i<j}(p_{ij} + p_{ji}),$$

with $s_i = D(w_i)C(w_i)$. Also $\frac{1}{4}\tau_i^2 = s_i$ and $\frac{1}{4}\tau_i\tau_j = \sqrt{s_i s_j} = \sqrt{p_{ij}}\sqrt{p_{ji}}$. Hence

$$\tau(w_1 \cdots w_n)^2 - \Big(\sum_i \tau_i\Big)^2 = 4\sum_{i<j}\big(\sqrt{p_{ij}} - \sqrt{p_{ji}}\big)^2. \tag{24.1}$$

(a) For two legs, $s - s_1 - s_2 = p_{12} + p_{21}$ and $4s_1s_2 = 4p_{12}p_{21}$, so $(s-s_1-s_2)^2 - 4s_1s_2 = (p_{12}-p_{21})^2$, and $s = s_1 + s_2 + p_{12} + p_{21} \geq s_1 + s_2$. Only integer arithmetic is used.

(b) and (c). The right side of (24.1) is at least 0, so $(\sum_i \tau_i)^2 \leq \tau(w_1 \cdots w_n)^2$, and both sides are square roots of nonnegative numbers, so $\sum_i \tau_i \leq \tau(w_1 \cdots w_n)$. Equality holds exactly when every term $(\sqrt{p_{ij}} - \sqrt{p_{ji}})^2$ vanishes, that is, when $p_{ij} = p_{ji}$ for all $i < j$. Part (b) is the case $n = 2$.

(d) and the proposition. When all counts are at least 1, p_{ij} and p_{ji} are positive and $p_{ij}/p_{ji} = r_i/r_j$. By the formula for $J(a/b)$ in the proof of Theorem 24.4,

$$J\big(\sqrt{r_i/r_j}\big) = J\Big(\frac{\sqrt{p_{ij}}}{\sqrt{p_{ji}}}\Big) = \frac{(\sqrt{p_{ij}} - \sqrt{p_{ji}})^2}{2\sqrt{p_{ij}p_{ji}}},$$

and $2\tau_i\tau_j = 8\sqrt{p_{ij}p_{ji}}$. So each term of (24.1) is $2\tau_i\tau_j J(\sqrt{r_i/r_j})$. Part (d) is the case $n = 2$.

(e) If $D + C = 2n$, then $4DC = (D+C)^2 - (D-C)^2 \leq 4n^2$, with equality exactly when $D = C$; so $\tau = 2\sqrt{DC} \leq 2n$, with equality exactly when $D = C = n$. □

The theorem says that "no cost between the legs" and "inertial" are the same condition. A change of posting ratio between two legs costs $J(\sqrt{r_1/r_2})$, and that cost, times $2\tau_1\tau_2$, is exactly the squared proper time the history gives up. In special relativity the same statement is the reverse triangle inequality for timelike vectors [147]; here it is an identity among counts, and its integer core (a) holds with nothing beyond arithmetic.

24.4 The recognition worldline

The history of the opening was not chosen at random. It is the fourth recognition word.

Definition 24.7 (The recognition words)**.** Let σ be the substitution of posting words that replaces each debit by dc and each credit by d, so that $\sigma(uv) = \sigma(u)\sigma(v)$. The *recognition words* are $w_0 = d$ and $w_{k+1} = \sigma(w_k)$:

$$w_1 = dc, \quad w_2 = dcd, \quad w_3 = dcddc, \quad w_4 = dcddcdcd, \ \ldots$$

The Fibonacci numbers are $F(0) = 0$, $F(1) = 1$ and $F(k+2) = F(k+1) + F(k)$. In the golden integers $\mathbb{Z}[\varphi]$ of Definition 21.24, whose elements are $a + b\varphi$ with integers a, b and $\varphi^2 = \varphi + 1$, we write $\psi = 1 - \varphi$ and $N(a + b\varphi) = a^2 + ab - b^2$.

In Recognition Science the substitution is one recognition join acting on an account's posting history. On census coordinates a join sends (a, b) to $(b, a + b)$, which is multiplication by φ in $\mathbb{Z}[\varphi]$: $\varphi(a+b\varphi) = b+(a+b)\varphi$. The recognition words are the Fibonacci words of combinatorics on words.

Theorem 24.8 (The recognition worldline)**.**

(a) w_k *has* $F(k+1)$ *debits and* $F(k)$ *credits.*

(b) In $\mathbb{Z}[\varphi]$, $F(k+1) - \varphi F(k) = \psi^k$ *and* $F(k+2) = \varphi^2 F(k) + \psi^k$. *Hence* $s_k = D(w_k)C(w_k)$ *satisfies* $s_{k+1} = \varphi^2 s_k + F(k+1)\psi^k$.

(c) $\tau(w_{k+2})/\tau(w_{k+1}) \to \varphi$ *as* $k \to \infty$.

(d) The velocity $(D(w_k) - C(w_k))/(D(w_k) + C(w_k))$ *tends to* $\sqrt{5} - 2 = \varphi^{-3}$.

(e) $J(\varphi) = (\sqrt{5} - 2)/2 = \varphi^{-3}/2$, *and with* $\beta = \sqrt{5} - 2$, $J(\varphi) = 2\beta^2/(1 - \beta^2)$.

(f) For integers λ, μ, ν *let* $q(a, b) = \lambda a^2 + \mu ab + \nu b^2$. *Then* $q(b, a + b) = -q(a, b)$ *for all integers* a, b *if and only if* $\mu = \lambda$ *and* $\nu = -\lambda$, *that is,* $q = \lambda N$; *and* $q(b, a + b) = q(a, b)$ *for all* a, b *if and only if* $\lambda = \mu = \nu = 0$.

(g) For real $C > 0$ *and* D: $C^2 + CD - D^2 = 0$ *and* $D > 0$ *if and only if* $D = \varphi C$.

Conditions and scope.

Base and premises

Parts (a) and (b) are statements about natural numbers and $\mathbb{Z}[\varphi]$, proved with no omniscience principle and no choice principle; part (f) is a statement about integer forms, proved by integer arithmetic and claimed in classical mathematics. Parts (c), (d), (e) and (g) are proved in classical mathematics over the real numbers.

Conclusion

The recognition words post Fibonacci counts; their squared half proper time grows by exactly φ^2 up to the alternating correction $F(k+1)\psi^k$; each recognition dilates proper time by φ in the limit; the velocity tends to φ^{-3}, at which $J(\varphi)$ is twice the squared proper velocity; the norm of $\mathbb{Z}[\varphi]$ is, up to a factor, the only integer form a join reverses, no nonzero form is preserved by a join, and the norm's null ray is the ratio φ.

Identification

Reading the words as the worldline of a recognizing account carries the identifications of Section 24.1.

Proof. (a) If a word has counts (D, C), its image under σ has $D + C$ debits, one for each letter, and D credits, one for each debit. Starting from $(1, 0) = (F(1), F(0))$, the counts $(F(k+1), F(k))$ go to $(F(k+2), F(k+1))$.

(b) In $\mathbb{Z}[\varphi]$, $\varphi\psi = \varphi - \varphi^2 = -1$. For $k = 0$, $F(1) - \varphi F(0) = 1 = \psi^0$. If $F(k+1) - \varphi F(k) = \psi^k$, then

$$F(k+2) - \varphi F(k+1) = (1-\varphi)F(k+1) + F(k) = \psi F(k+1) - \varphi\psi F(k)$$
$$= \psi\big(F(k+1) - \varphi F(k)\big) = \psi^{k+1}.$$

Then $F(k+2) = F(k+1) + F(k) = (\varphi+1)F(k) + \psi^k = \varphi^2 F(k) + \psi^k$, and multiplying by $F(k+1)$ gives $s_{k+1} = F(k+2)F(k+1) = \varphi^2 s_k + F(k+1)\psi^k$. Each step is a computation with pairs of integers.

(c) and (d). Over the reals $F(k+1)/F(k) \to \varphi$ (Binet's formula, or (b) with $|\psi| < 1$). Then

$$\frac{\tau(w_{k+2})}{\tau(w_{k+1})} = \sqrt{\frac{F(k+3)}{F(k+2)} \cdot \frac{F(k+2)}{F(k+1)}} \to \varphi,$$

and with $\rho_k = F(k)/F(k+1) \to \varphi^{-1}$ the velocity is $(1-\rho_k)/(1+\rho_k) \to (1-\varphi^{-1})/(1+\varphi^{-1})$. Since $\varphi^{-1} = (\sqrt{5}-1)/2$, this is $(3-\sqrt{5})/(1+\sqrt{5}) = \sqrt{5}-2$, as $(\sqrt{5}-2)(1+\sqrt{5}) = 3-\sqrt{5}$. Finally $\varphi^3 = 2\varphi + 1 = 2 + \sqrt{5}$ and $(2+\sqrt{5})(\sqrt{5}-2) = 1$.

(e) $J(\varphi) = \frac{1}{2}(\varphi + \varphi^{-1}) - 1 = \frac{1}{2}\sqrt{5} - 1$. With $\beta = \sqrt{5}-2$, $\beta^2 = 9 - 4\sqrt{5}$ and $1 - \beta^2 = 4\sqrt{5} - 8 = 4\beta$, so $2\beta^2/(1-\beta^2) = \beta/2 = J(\varphi)$.

(f) Suppose $q(b, a+b) = -q(a, b)$ for all a, b. At $(a, b) = (1, 0)$ this reads $\nu = -\lambda$; at $(0, 1)$ it reads $\lambda + \mu + \nu = -\nu$, so $\mu = -\lambda - 2\nu = \lambda$. Conversely, $N(b, a+b) = b^2 + b(a+b) - (a+b)^2 = -(a^2 + ab - b^2)$. Suppose instead $q(b, a+b) = q(a, b)$ for all a, b. Then $(1, 0)$ gives $\nu = \lambda$, $(0, 1)$ gives $\lambda + \mu + \nu = \nu$, so $\mu = -\lambda$, and $(1, 1)$ gives $\lambda + 2\mu + 4\nu = \lambda + \mu + \nu$, so $\mu + 3\nu = 0$, whence $2\lambda = 0$ and all three vanish.

(g) $C^2 + CD - D^2 = -(D - \varphi C)(D - \psi C)$, and $\psi C < 0 < D$ excludes the second factor. □

In the worked history w_4 the velocity was already $1/4$; the limit is $\varphi^{-3} \approx 0.2361$. The words never become inertial, and they bend toward the line of velocity φ^{-3} by the least amount an integer history can bend.

Proposition 24.9 (Each recognition bends by the least integer kink)**.**

(a) (Cassini) $F(k+2)F(k) - F(k+1)^2 = (-1)^{k+1}$ *for every* k.

(b) $w_{k+2} = w_{k+1}w_k$, *and with* $s_k = D(w_k)C(w_k)$,

$$(s_{k+2} - s_{k+1} - s_k)^2 - 4s_{k+1}s_k = 1.$$

(c) *If posting words u, v have $D(u)C(v) \neq D(v)C(u)$, then $(D(u)C(v) - D(v)C(u))^2 \geq 1$.*

(d) *$\tau(w_{k+1}) + \tau(w_k) < \tau(w_{k+2})$ for every k.*

Parts (a) and (b) are proved with no omniscience principle and no choice principle; part (c) is proved by integer arithmetic and claimed in classical mathematics; part (d) uses real square roots and is proved classically.

Proof. (a) For $k = 0$, $F(2)F(0) - F(1)^2 = -1$. The step: $F(k+3)F(k+1) - F(k+2)^2 = (F(k+2)+F(k+1))F(k+1) - F(k+2)(F(k+1)+F(k)) = F(k+1)^2 - F(k+2)F(k)$, the negative of the previous value.

(b) $w_2 = dcd = w_1 w_0$, and if $w_{k+2} = w_{k+1}w_k$ then applying σ gives $w_{k+3} = w_{k+2}w_{k+1}$. By Theorem 24.5(a) with the legs w_{k+1} and w_k, the left side is the square of $D(w_{k+1})C(w_k) - D(w_k)C(w_{k+1}) = F(k+2)F(k) - F(k+1)^2$, which is ± 1 by (a).

(c) A nonzero integer has square at least 1.

(d) By Theorem 24.5(b), equality would force the cross term of (b) to vanish, and it is ± 1. □

By Theorem 24.5(a), the quantity in (b) is the squared cross term of the two legs, and by (c) no two legs of different ratio have a smaller one. So each recognition is the closest an integer history can come to an inertial step without being one.

Along the recognition worldline the velocity tends to $\beta = \varphi^{-3}$, and the dilation of proper time at that velocity has an exact golden form. The value is $2\varphi^{-3/2} = 2/\varphi^{3/2}$, about 0.9717.

Theorem 24.10 (The recognition dilation). *Let $\beta = \sqrt{5} - 2$ and $\gamma = (1 - \beta^2)^{-1/2}$.*

(a) *In $\mathbb{Z}[\varphi]$ the element $b = 2\varphi - 3$ satisfies $b\varphi^3 = 1$, $b^2 + 4b = 1$ and $1 + b = \varphi(1 - b)$; over the reals $b = \beta$.*

(b) *$\beta = \varphi^{-3}$; $1 - \beta^2 = 4\beta$; the time dilation is $\sqrt{1 - \beta^2} = 2\sqrt{\beta} = 2\varphi^{-3/2} \approx 0.97174$; and $\gamma = \varphi^{3/2}/2 \approx 1.02909$.*

(c) *The Doppler factor is $\sqrt{(1+\beta)/(1-\beta)} = \sqrt{\varphi}$, with rapidity $\frac{1}{2}\ln\varphi$ and $\cosh(\frac{1}{2}\ln\varphi) = \gamma$; $J(\sqrt{\varphi}) = \gamma - 1 \approx 0.02909$; and $1 - \beta^2 = 8J(\varphi)$.*

(d) *The proper time per post of the recognition words tends to the dilation:*

$$\tau(w_k)/(D(w_k) + C(w_k)) \to 2\sqrt{\beta}.$$

Conditions and scope.

Base and premises

Part (a) is a computation in $\mathbb{Z}[\varphi]$ with no omniscience principle and no choice principle. Parts (b) to (d) are proved in classical mathematics over the real numbers.

Conclusion

At the recognition velocity φ^{-3} the squared dilation is $4\varphi^{-3} = 8J(\varphi)$, the dilation is $2\varphi^{-3/2}$, and the Lorentz excess $\gamma - 1$ is the cost of the Doppler factor $\sqrt{\varphi}$.

Not claimed

Any physical link between this dilation and other quantities of Recognition Science: the theorem is a set of equalities between numbers.

Proof. (a) In $\mathbb{Z}[\varphi]$, $\varphi^2 = \varphi + 1$ and $\varphi^3 = 2\varphi + 1$. Then $(2\varphi - 3)(2\varphi + 1) = 4\varphi^2 - 4\varphi - 3 = 1$; $(2\varphi - 3)^2 + 4(2\varphi - 3) = (4\varphi^2 - 12\varphi + 9) + (8\varphi - 12) = 4(\varphi + 1) - 4\varphi - 3 = 1$; and $1 + b = 2\varphi - 2$, while $\varphi(1 - b) = 4\varphi - 2\varphi^2 = 2\varphi - 2$. Over the reals $2\varphi - 3 = \sqrt{5} - 2$.

(b) $\beta = \varphi^{-3}$ by (a). As in the proof of Theorem 24.8(e), $1 - \beta^2 = 4\beta$, so $\sqrt{1 - \beta^2} = 2\sqrt{\beta} = 2\varphi^{-3/2}$ and $\gamma = 1/(2\varphi^{-3/2}) = \varphi^{3/2}/2$. Numerically $\sqrt{\beta} = 0.48587$ and $\varphi^{3/2} = 2.05817$.

(c) By (a), $(1+\beta)/(1-\beta) = \varphi$. Next $\cosh(\frac{1}{2}\ln\varphi) = \frac{1}{2}(\sqrt{\varphi} + 1/\sqrt{\varphi})$, whose square is $\frac{1}{4}(\varphi + 2 + \varphi^{-1}) = \frac{1}{4}(2 + \sqrt{5}) = \varphi^3/4$; so it is $\varphi^{3/2}/2 = \gamma$, and $J(\sqrt{\varphi}) = \frac{1}{2}(\sqrt{\varphi} + 1/\sqrt{\varphi}) - 1 = \gamma - 1$. Finally $8J(\varphi) = 4(\sqrt{5} - 2) = 4\beta = 1 - \beta^2$ by Theorem 24.8(e).

(d) $\tau(w)/(D + C) = 2\sqrt{DC}/(D + C) = \sqrt{1 - v^2}$ with $v = (D - C)/(D + C)$, which tends to β by Theorem 24.8(d); the map $v \mapsto \sqrt{1 - v^2}$ is continuous. □

24.5 Many accounts reach only the taxicab cone

One account gives a light cone in one space dimension. The obvious way to get more space dimensions is to take more accounts: let the total number of posts be time and the balances of the accounts be the coordinates of position. This fails, and it fails for a reason that can be stated exactly.

Definition 24.11 (Ledgers of several accounts)**.** A ledger state of d accounts is a list $r = (r_1, \dots, r_d)$ of records $r_i = (D_i, C_i)$. A *single post* changes one account's record by posting one side to it; r' is *reachable* from r when finitely many single posts lead from r to r'. The *total post count* is $T(r) = \sum_i (D_i + C_i)$ and the *potentials* are $x_i = D_i - C_i$. The *taxicab cone*, or ℓ^1 cone, in $\mathbb{R}^{1+d}$ is $\{(t, x) : \sum_i |x_i| \le t\}$. A set $K \subseteq \mathbb{R}^{1+n}$ is a *quadratic cone* when $K = \{p : p_0 \ge 0 \text{ and } p^{\mathsf{T}} M p \ge 0\}$ for some real $(1 + n) \times (1 + n)$ matrix M.

Theorem 24.12 (Separate accounts give the taxicab cone)**.** *Let $d \ge 1$.*

(a) A ledger state r' is reachable from r if and only if every count of every account in r' is at least the corresponding count in r.

(b) From the empty ledger there is a reachable state with total post count T and potentials $x \in \mathbb{Z}^d$ if and only if $\sum_i |x_i| \le T$ and $T - \sum_i x_i$ is even.

(c) The state with total post count $2T$ and potentials $2x$ is reachable from the empty ledger if and only if $\sum_i |x_i| \le T$, that is, if and only if (T, x) lies in the taxicab cone.

(d) The taxicab cone in $\mathbb{R}^{1+d}$ is a quadratic cone if and only if $d \le 1$. (This part holds for $d = 0$ as well.)

Conditions and scope.

Base and premises

Part (a) is a statement about natural numbers, proved with no omniscience principle and no choice principle; parts (b) and (c) are statements about integers, proved by integer arithmetic and claimed in classical mathematics. Part (d) is about real quadratic forms and is proved in classical mathematics.

Conclusion

Posting d separate accounts one unit at a time reaches, after coarse graining by two, exactly the taxicab cone, and that cone is the future cone of a quadratic form only when $d \leq 1$.

Identification

Reading the total post count as time and the potentials as position. Under that reading, posting of separate accounts is Lorentzian only in one space dimension.

Not claimed

Other posting laws: steps of other sizes, links between accounts, or other ways of comparing accounts. For any finite set of steps the coarse-grained cone would be a polytope; we do not prove this.

Proof. (a) A single post raises one count, so counts never decrease along a sequence of posts. Conversely, if every count of r' is at least that of r, raise the counts one account at a time, one post at a time.

(b) For each account, $|x_i| \leq D_i + C_i$ and $D_i + C_i \equiv x_i \pmod 2$. Summing over the accounts gives $\sum_i |x_i| \leq T$ and $T \equiv \sum_i x_i$. Conversely, put $D_i = \max(x_i, 0)$ and $C_i = \max(-x_i, 0)$, so that $D_i + C_i = |x_i|$ and $D_i - C_i = x_i$. The remaining $T - \sum_i |x_i|$ posts number an even amount, since $|x_i| \equiv x_i \pmod 2$; post them to account 1 as debit–credit pairs, which leaves its potential unchanged. By (a) the result is reachable from the empty ledger.

(c) For the doubled point the parity condition holds automatically, and $\sum_i |2x_i| \leq 2T$ exactly when $\sum_i |x_i| \leq T$.

(d) For $d = 0$ the cone is $\{t \geq 0\}$, the cone of $M = 0$. For $d = 1$, $|x| \leq t$ holds exactly when $t \geq 0$ and $t^2 - x^2 \geq 0$.

Let $d \geq 2$, and suppose the taxicab cone equals $\{p_0 \geq 0,\ p^{\mathsf{T}} M p \geq 0\}$. Restricting to points whose coordinates after the first two space coordinates vanish, we obtain a real quadratic form $Q(t, a, b) = q_1 t^2 + q_2 a^2 + q_3 b^2 + q_4 ta + q_5 tb + q_6 ab$ with

$$|a| + |b| \leq t \iff \big(t \geq 0 \text{ and } Q(t, a, b) \geq 0\big) \qquad \text{for all real } t, a, b.$$

First a boundary lemma: if p lies in the cone and, for every $\varepsilon > 0$, $p + \varepsilon v$ lies outside it with nonnegative time coordinate, then $Q(p) = 0$. Indeed $Q(p + \varepsilon v) < 0$ for every $\varepsilon > 0$, and $Q(p + \varepsilon v) = Q(p) + \varepsilon L + \varepsilon^2 M'$ for numbers L, M', so letting $\varepsilon \to 0$ gives $Q(p) \leq 0$; and $Q(p) \geq 0$ because p is in the cone.

Apply the lemma at $p = (1, 1, 0)$ with $v = (0, 1, 0)$: $q_1 + q_2 + q_4 = 0$. At $p = (1, 0, 1)$ with $v = (0, 0, 1)$: $q_1 + q_3 + q_5 = 0$. At $p = (2, 1, 1)$ with $v = (0, 1, 0)$: $4q_1 + q_2 + q_3 + 2q_4 + 2q_5 + q_6 = 0$. Twice the first plus twice the second minus the third gives $q_2 + q_3 - q_6 = 0$, which says

$Q(0,1,-1)=0$. So the point $(0,1,-1)$ has $t\geq 0$ and $Q\geq 0$, yet $|1|+|-1|=2>0=t$: it is not in the cone. This contradiction proves the claim. □

The parity condition in (b) is the reason for coarse graining by two: a single account can reach $x=1$ only at odd times. The quadratic no-go in (d) does not depend on the parity at all. The taxicab cone is a pyramid with flat faces, and a quadratic form whose cone it were would have to vanish at three of its boundary points; that already forces the form to vanish at $(0,1,-1)$, outside the cone.

24.6 The record cone of one coherent binary distinction

The cone of one time and three space dimensions comes from a different place: from a single binary distinction whose two values are carried coherently. The ledger reappears as the diagonal of its record.

Definition 24.13 (Coherent records and coherent changes)**.** The Pauli matrices are

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

For $t\in\mathbb{R}$ and $x=(x_1,x_2,x_3)\in\mathbb{R}^3$ put

$$R(t,x)=\tfrac{1}{2}\big(t\,1+x_1\sigma_1+x_2\sigma_2+x_3\sigma_3\big)=\frac{1}{2}\begin{pmatrix}t+x_3&x_1-ix_2\\x_1+ix_2&t-x_3\end{pmatrix}.$$

The *coordinates* of a complex 2×2 matrix ρ are $t=\mathrm{Re}(\rho_{00}+\rho_{11})$, $x_1=2\,\mathrm{Re}\,\rho_{01}$, $x_2=-2\,\mathrm{Im}\,\rho_{01}$ and $x_3=\mathrm{Re}(\rho_{00}-\rho_{11})$. The Minkowski form is $Q(t,x)=t^2-|x|^2$ and the future cone is $K=\{(t,x)\,:\,t\geq 0,\ Q(t,x)\geq 0\}$. A *coherent record* is a Hermitian 2×2 complex matrix ρ with $v^\dagger\rho v\geq 0$ for every $v\in\mathbb{C}^2$. A *coherent change* is a map $\rho\mapsto A\rho A^\dagger$ with A a complex 2×2 matrix; its action on coordinates, $p\mapsto$ the coordinates of $AR(p)A^\dagger$, is a real linear map of $\mathbb{R}^4$, whose matrix we write Λ_A.

A coherent record is an unnormalized state of a two-level system, the smallest system that carries a binary distinction coherently. Its two diagonal entries are the weights of the two values; its off-diagonal entry is the coherence between them.

Theorem 24.14 (The record cone of a coherent binary distinction)**.** *For all real t and $x\in\mathbb{R}^3$:*

(a) $R(t,x)$ is Hermitian and has coordinates (t,x); every Hermitian 2×2 complex matrix is R of its coordinates. So coherent records have four real parameters, one plus three.

(b) $\det R(t,x)=(t^2-|x|^2)/4$.

(c) $R(t,x)$ is a coherent record if and only if $t\geq 0$ and $|x|^2\leq t^2$, that is, if and only if $(t,x)\in K$.

(d) $R(D+C,0,0,D-C)=\mathrm{diag}(D,C)$, with determinant DC: the records of one account are the diagonal slice, and the determinant restricts to the interval of Theorem 24.2.

(e) If $\det A=1$ and ρ is a coherent record, then $A\rho A^\dagger$ is a coherent record with the same determinant.

(f) For $k > 0$ the matrix $B_k = \operatorname{diag}(\sqrt{k}, 1/\sqrt{k})$ has determinant 1, and $B_k R(t,x) B_k^\dagger = R(t', x_1, x_2, x_3')$ with

$$t' = \tfrac{1}{2}\big(k(t+x_3) + (t-x_3)/k\big), \qquad x_3' = \tfrac{1}{2}\big(k(t+x_3) - (t-x_3)/k\big);$$

on the diagonal it sends (D, C) to $(kD, C/k)$.

(g) If A is unitary with $\det A = 1$, then $AR(t,x)A^\dagger$ has time coordinate t and spatial length $|x|$.

(h) All entries of $R(t,x)$ are real if and only if $x_2 = 0$.

Conditions and scope.

Base and premises

Complex 2×2 matrices over the classical real and complex numbers. Proved in classical mathematics.

Conclusion

The coherent records of one binary distinction form exactly the future light cone of one time and three space dimensions, with the determinant as a quarter of the interval; the ledger is the diagonal slice; the Doppler rescaling of one account, which is no symmetry of integer counts (Proposition 24.3), is a determinant-one change of records; unitary changes are rotations; real amplitudes leave one space dimension out.

Identification

That the record of an account is the diagonal of a coherent record, and that physical spacetime is this cone, are identifications. The theorem does not supply them.

Proof. (a) The formula is Hermitian because the Pauli matrices are, and reading the coordinates off it returns (t, x). If ρ is Hermitian, ρ_{00} and ρ_{11} are real and $\rho_{10} = \overline{\rho_{01}}$, so ρ is R of its coordinates.

(b) $\det R = \frac{1}{4}\big((t+x_3)(t-x_3) - (x_1 - ix_2)(x_1 + ix_2)\big) = \frac{1}{4}(t^2 - x_3^2 - x_1^2 - x_2^2)$.

(c) A Hermitian matrix $\begin{pmatrix} a & \bar{z} \\ z & b \end{pmatrix}$ has $v^\dagger \rho v = a|v_0|^2 + 2\,\mathrm{Re}(\bar{z}\bar{v}_0 v_1) + b|v_1|^2$. It is positive semidefinite exactly when $a \geq 0$, $b \geq 0$ and $|z|^2 \leq ab$. If these hold, $v^\dagger \rho v \geq a|v_0|^2 - 2\sqrt{ab}|v_0||v_1| + b|v_1|^2 = (\sqrt{a}|v_0| - \sqrt{b}|v_1|)^2 \geq 0$. Conversely, the basis vectors give $a, b \geq 0$; the vector $v = (-\bar{z}, a)$ gives $a(ab - |z|^2) \geq 0$, which settles $a > 0$; and if $a = 0$, the vectors $(-s\bar{z}, 1)$ give $b - 2s|z|^2 \geq 0$ for every $s > 0$, so $z = 0$. For $R(t,x)$, $a = (t + x_3)/2$, $b = (t - x_3)/2$ and $|z|^2 = (x_1^2 + x_2^2)/4$, so $ab - |z|^2 = (t^2 - |x|^2)/4$, and $a, b \geq 0$ says $t \geq |x_3|$. The three conditions together say $t \geq 0$ and $|x|^2 \leq t^2$.

(d) Substitute: $t + x_3 = 2D$ and $t - x_3 = 2C$.

(e) $A\rho A^\dagger$ is Hermitian, $v^\dagger A\rho A^\dagger v = (A^\dagger v)^\dagger \rho (A^\dagger v) \geq 0$, and $\det(A\rho A^\dagger) = |\det A|^2 \det \rho = \det \rho$.

(f) $\det B_k = 1$. Conjugation by B_k multiplies the upper diagonal entry by k, the lower by $1/k$, and the off-diagonal entries by $\sqrt{k} \cdot (1/\sqrt{k}) = 1$. Reading coordinates gives the formulas, and on the diagonal $D = (t + x_3)/2$ and $C = (t - x_3)/2$ go to kD and C/k.

(g) The time coordinate is the trace, and $\operatorname{tr}(ARA^\dagger) = \operatorname{tr}(RA^\dagger A) = \operatorname{tr} R$. By (b) and $\det(ARA^\dagger) = |\det A|^2 \det R$ the determinant is also kept, so $t^2 - |x'|^2 = t^2 - |x|^2$.

(h) The imaginary parts of the off-diagonal entries are $\mp x_2/2$. □

Which linear maps of $\mathbb{R}^4$ come from coherent changes of determinant one? Exactly the proper orthochronous Lorentz transformations: the maps of determinant 1 that keep the Minkowski form and the future cone. And an invertible linear map that merely keeps the future cone, in both directions, keeps the Minkowski form up to a positive factor.

Theorem 24.15 (The Lorentz group of a coherent binary distinction)**.**

(i) Let Λ be an invertible real 4×4 matrix with $p \in K$ if and only if $\Lambda p \in K$, for every p. Then there is a real $c > 0$ with $Q(\Lambda p) = cQ(p)$ for every p.

(ii) A real 4×4 matrix Λ satisfies $Q(\Lambda p) = Q(p)$ for all p, maps K into K, and has $\det\Lambda = 1$, if and only if $\Lambda = \Lambda_A$ for some complex 2×2 matrix A with $\det A = 1$.

Proof. (i) Since $p \in -K$ exactly when $-p \in K$, the map Λ also keeps $-K$, hence the double cone $K \cup (-K) = \{Q \geq 0\}$: $Q(p) \geq 0$ if and only if $Q(\Lambda p) \geq 0$. Next, Λ maps null vectors to null vectors. Let $Q(p) = 0$; then $Q(\Lambda p) \geq 0$. If $p = 0$ there is nothing to show. If $p = (t, x) \neq 0$, then $t \neq 0$, and $p_\varepsilon = ((1-\varepsilon)t, x)$ has $Q(p_\varepsilon) < 0$ for $0 < \varepsilon < 1$, hence $Q(\Lambda p_\varepsilon) < 0$; letting $\varepsilon \to 0$ gives $Q(\Lambda p) \leq 0$. So $Q(\Lambda p) = 0$.

Write $Q(\Lambda p) = \sum_{j,k} N_{jk} p_j p_k$ with N symmetric. The null vectors $(1, \pm e_i)$ give $N_{00} \pm 2N_{0i} + N_{ii} = 0$, so $N_{0i} = 0$ and $N_{ii} = -N_{00}$. For $i \neq j$ the null vector $(5, 3e_i + 4e_j)$, null because $25 = 9+16$, gives $25N_{00} + 9N_{ii} + 16N_{jj} + 24N_{ij} = 24N_{ij} = 0$. Hence $Q(\Lambda p) = N_{00} Q(p)$. Finally $N_{00} = Q(\Lambda e_0) \geq 0$ because $e_0 \in K$, and $N_{00} \neq 0$ because Λ is onto: some p has $\Lambda p = e_0$, and then $N_{00} Q(p) = 1$.

(ii) First let $\det A = 1$. Since $\det(A\rho A^\dagger) = |\det A|^2 \det\rho$ for every ρ (proof of Theorem 24.14(e)), part (b) shows that Λ_A keeps $Q = 4\det$, and (c) with (e) shows that it maps K into K. Keeping Q means $\Lambda_A^{\mathsf{T}} \eta \Lambda_A = \eta$ for $\eta = \operatorname{diag}(1,-1,-1,-1)$, so $\det\Lambda_A = \pm 1$. To see the sign, note that $A \mapsto \Lambda_A$ is multiplicative, $\Lambda_{AB} = \Lambda_A \Lambda_B$, and that $\Lambda_A = \Lambda_{-A}$. Replacing A by $-A$ if necessary we may assume $\operatorname{tr} A \neq -2$. Choose a complex s with $s^2 = \operatorname{tr} A + 2$ and put $B = (A+1)/s$. By the Cayley–Hamilton theorem $A^2 = (\operatorname{tr} A)A - 1$, so

$$B^2 = \frac{A^2 + 2A + 1}{\operatorname{tr} A + 2} = \frac{(\operatorname{tr} A)A - 1 + 2A + 1}{\operatorname{tr} A + 2} = A.$$

Hence $\det\Lambda_A = (\det\Lambda_B)^2 \geq 0$, and so $\det\Lambda_A = 1$.

Conversely, let Λ keep Q and K and have determinant 1, and put $u = \Lambda e_0$. Then $Q(u) = 1$ and $u \in K$, so $u_0 \geq 1$. The matrix $P = 2R(u)$ is Hermitian with $\det P = Q(u) = 1$ and $\operatorname{tr} P = 2u_0$, so $P^2 = 2u_0 P - 1$. Put $B = (P+1)/\sqrt{2u_0 + 2}$. Then $B^2 = (P^2 + 2P + 1)/(2u_0 + 2) = P$, and $\det B = (\det P + \operatorname{tr} P + 1)/(2u_0 + 2) = 1$, using $\det(P+1) = \det P + \operatorname{tr} P + 1$ for 2×2 matrices. Since B is Hermitian, $BR(e_0)B^\dagger = \frac{1}{2}B^2 = \frac{1}{2}P = R(u)$: the change B carries e_0 to u. The matrix $\Lambda' = \Lambda_{B^{-1}}\Lambda$ keeps Q, fixes e_0 and has determinant 1 by the first part. Keeping the bilinear form of Q and fixing e_0, it keeps the time coordinate and maps the space $\{t = 0\}$ to itself, where $Q = -|x|^2$; so $\Lambda' = \operatorname{diag}(1, S)$ with S a real orthogonal 3×3 matrix of determinant 1.

Such an S is a rotation about an axis. Indeed $\det(S-1)=\det(S-SS^{\mathsf T})=\det S\det(1-S^{\mathsf T})=\det(1-S)=-\det(S-1)$, so S fixes a unit vector n; it maps the plane orthogonal to n to itself with determinant 1, so it turns that plane by an angle θ, and

$$Sx=\cos\theta\,x+(1-\cos\theta)(n\cdot x)\,n+\sin\theta\,(n\times x)$$

for a suitable orientation of n (Rodrigues' formula). Put $U=\cos\frac{\theta}{2}\,1-i\sin\frac{\theta}{2}\,(n\cdot\sigma)$, which is unitary with $\det U=1$. The Pauli identity $(a\cdot\sigma)(b\cdot\sigma)=(a\cdot b)\,1+i(a\times b)\cdot\sigma$ gives $(n\cdot\sigma)(x\cdot\sigma)-(x\cdot\sigma)(n\cdot\sigma)=2i(n\times x)\cdot\sigma$ and $(n\cdot\sigma)(x\cdot\sigma)(n\cdot\sigma)=2(n\cdot x)(n\cdot\sigma)-x\cdot\sigma$, and expanding,

$$U(x\cdot\sigma)U^{\dagger}=\big(\cos\theta\,x+(1-\cos\theta)(n\cdot x)n+\sin\theta\,n\times x\big)\cdot\sigma=(Sx)\cdot\sigma.$$

With $UU^{\dagger}=1$ this says $\Lambda_U=\operatorname{diag}(1,S)=\Lambda'$. Hence $\Lambda=\Lambda_B\Lambda_U=\Lambda_{BU}$ with $\det(BU)=1$. □

Part (i) is a linear form of Zeeman's theorem that causality implies the Lorentz group [150], and part (ii) is the classical spinor description of $\mathrm{SO}^{+}(1,3)$ by $\mathrm{SL}(2,\mathbb{C})$ [145]; A and $-A$ give the same change. The proof needs no angles beyond the rotation angle of S, and it shows how the boost part arises: as the positive square root of the record $2R(u)$.

Why two values carried by complex amplitudes, and not by real ones or by quaternions? Count the real parameters of a record of a distinction with n values. The n diagonal weights are real; each of the $n(n-1)/2$ coherences between two values has k real components, where $k=0$ for a classical carrier, 1 for real amplitudes, 2 for complex, 4 for quaternionic and 8 for octonionic ones.

Definition 24.16 (Record counts)**.** For natural numbers k and n put $P_k(n)=n+k\,n(n-1)/2$. Record counts *multiply* for k when $P_k(nm)=P_k(n)P_k(m)$ for all n and m.

Multiplication of record counts says that a record of a composite of two systems is fixed by the parameters of the parts and their products, the condition known as local tomography [142, 149].

Theorem 24.17 (Record counts multiply only for two carriers)**.** *Record counts multiply for k if and only if $k=0$ or $k=2$. In particular, among $k\in\{1,2,4,8\}$ only $k=2$ multiplies. If $k>0$ and record counts multiply for k, a binary distinction has $P_k(2)=4=1+3$ real parameters; the classical carrier has $P_0(2)=2=1+1$.*

Proof. $P_k(2)=2+k$ and $P_k(4)=4+6k$, so multiplication at $n=m=2$ requires $(2+k)^2=4+6k$, that is, $k^2=2k$, that is, $k\in\{0,2\}$. Conversely $P_0(n)=n$ and $P_2(n)=n+n(n-1)=n^2$ are multiplicative. The rest is arithmetic. Everything is a computation with natural numbers. □

Conditions and scope.

Base and premises

Theorem 24.15 is proved in classical mathematics over the real and complex numbers; Theorem 24.17 is proved by natural-number arithmetic.

Conclusion

The coherent changes of determinant one of one binary distinction act on its record cone as exactly $\mathrm{SO}^+(1,3)$ (A and $-A$ giving the same map), and an invertible linear map that keeps the record cone in both directions keeps the Minkowski form up to a positive factor. Among carriers of a binary distinction, the classical one gives $1+1$ parameters and the only coherent one whose record counts multiply gives $1+3$.

Identification

That the physical carrier's records multiply, that the carrier is coherent ($k>0$), and that physical spacetime is the record cone, are identifications. The whole causal order of postings is not a record-cone order in any chart (Theorem 24.84), so the last identification is not one of the posting order itself. "Three is selected" means exactly what the theorem says: $k \in \{0,2\}$ are the only multiplicative carriers, and $k=2$ gives $1+3$.

Not claimed

That the dimension of the record cone equals the parameter count for carriers other than the complex one; the cone is computed above for $k=2$ only.

This route to three space dimensions is independent of the linking route of Chapter 23. The two agree in the number: the space of the complex record cone has dimension $P_2(2)-1=3$, the dimension in which that chapter finds linking. The agreement is a match of numbers, not an identification of the two spaces.

24.7 Where the distinction is carried

Recognition Science's least costly change of a ledger is a single post, and the binary datum of a post is its side. This section asks where that datum can be held so that its records form the cone of Section 24.6, and how much of the cone a post of definite side actually uses.

Definition 24.18 (Registers, compressions and run records)**.** A *register* is a finite set $\mathscr{R}$ of basis states; its states are complex matrices indexed by $\mathscr{R}\times\mathscr{R}$. For two distinct basis states $e(0), e(1)$, the *compression* of a matrix ρ on $\mathscr{R}$ is the 2×2 matrix $(\rho_{e(i)e(j)})_{i,j\in\{0,1\}}$. For $v\in\mathbb{C}^2$ the *pure record* of v is $vv^\dagger$. The *side vector* of a debit is $(1,0)$ and of a credit $(0,1)$, and the *run record* of a posting word is the sum, over its posts, of the pure records of their side vectors. A change $L\to L'$ of a ledger of d accounts is *monotone* when no count decreases, and Recognition Science charges it the *step cost*

$$\sum_k G(D'_k-D_k)+\sum_k G(C'_k-C_k), \qquad G(u)=\cosh u-1,$$

the logarithmic form of J (Chapter 21) applied to each change of count.

Theorem 24.19 (The carrier of the record cone)**.**

(a) Let $d \geq 1$. If a monotone change $L \to L' \neq L$ has step cost at most that of every monotone change $L \to L'' \neq L$, then L' arises from L by a single post at one account. The debit post and the credit post at one account give different ledgers.

(b) For every register and distinct basis states $e(0), e(1)$, the compression of every positive semidefinite matrix is a coherent record, and every coherent record is the compression of a positive semidefinite matrix.

(c) For $v \in \mathbb{C}^2$, $vv^\dagger$ is a coherent record whose coordinates are null, $Q = 0$, with time coordinate $|v_0|^2 + |v_1|^2$. Conversely, every coherent record of determinant 0 is $vv^\dagger$ for some v.

(d) Every coherent record is $vv^\dagger + ww^\dagger$ for some $v, w \in \mathbb{C}^2$.

(e) The pure record of a debit has coordinates $(1, 0, 0, 1)$ and that of a credit $(1, 0, 0, -1)$. The run record of a word w is $\mathrm{diag}(D(w), C(w))$; its coordinates have $x_1 = x_2 = 0$.

(f) The Hermitian $n \times n$ complex matrices form a real vector space of dimension $n^2 = P_2(n)$, and $(nm)^2 = n^2 m^2$. The real symmetric $n \times n$ matrices form one of dimension $n(n+1)/2 = P_1(n)$, and $P_1(4) = 10 \neq 9 = P_1(2)^2$.

Conditions and scope.

Base and premises

The classical real and complex numbers. Proved in classical mathematics.

Conclusion

A post's side, held on two basis states of any finite complex register, has exactly the record cone of Theorem 24.14; its pure records are exactly the null records, every record is the sum of two of them, and classical posts are the two null rays of the account's plane. Complex record counts multiply and real ones do not.

Identification

That a physical run's record is its run record, with the two sides stored on two basis states, and that a post can carry a coherent amplitude, which is needed for the directions x_1 and x_2, are identifications. Recognition Science's native register is complex by definition; the theorem counts its records and does not derive that it is complex.

Proof. (a) For an integer change $u \geq 0$, $G(0) = 0$, $G(1) > 0$, and $G(u) > G(1)$ for $u \geq 2$, since cosh increases on $[0, \infty)$. A monotone change $L \to L'' \neq L$ raises some count, so its step cost is at least $G(1)$, and a single post at any account costs exactly $G(1)$. So the least change L' costs at most $G(1)$. If it raised two counts, or one count by two or more, it would cost more. Hence it raises exactly one count by exactly one. A debit post raises a debit count and a credit post does not.

(b) Let ρ be positive semidefinite on $\mathscr{R}$. For $v \in \mathbb{C}^2$ let $\tilde{v}$ be the vector on $\mathscr{R}$ with $\tilde{v}_{e(i)} = v_i$ and zeros elsewhere; then $v^\dagger(\text{compression})v = \tilde{v}^\dagger \rho \tilde{v} \geq 0$, and the compression is Hermitian. Conversely a coherent record M is the compression of the matrix equal to M on the block $\{e(0), e(1)\}$ and zero elsewhere, which is positive semidefinite.

(c) $vv^\dagger$ is Hermitian with $u^\dagger vv^\dagger u = |v^\dagger u|^2 \geq 0$, its determinant is $|v_0|^2|v_1|^2 - |v_0 v_1|^2 = 0$, so $Q = 4\det = 0$, and its trace is $|v_0|^2 + |v_1|^2$. Conversely let $M = \begin{pmatrix} a & \bar{z} \\ z & b \end{pmatrix}$ be a coherent record with $ab = |z|^2$. If $a > 0$, take $v = (\sqrt{a}, z/\sqrt{a})$; then $vv^\dagger$ has entries a, $\bar{z}$, z and $|z|^2/a = b$. If $a = 0$, then $|z|^2 \leq ab = 0$, and $v = (0, \sqrt{b})$ serves.

(d) In the notation of (c), $|z|^2 \leq ab$. If $a = 0$, then $z = 0$ and $M = ww^\dagger$ with $w = (0, \sqrt{b})$ and $v = 0$. If $a > 0$, take $v = (\sqrt{a}, z/\sqrt{a})$ and $w = (0, \sqrt{b - |z|^2/a})$.

(e) $\mathrm{diag}(1,0) = R(1,0,0,1)$ and $\mathrm{diag}(0,1) = R(1,0,0,-1)$; summing over the posts gives $\mathrm{diag}(D, C)$, whose off-diagonal entries vanish.

(f) A Hermitian matrix is fixed by its n real diagonal entries and its $n(n-1)/2$ complex entries above the diagonal, $n + n(n-1) = n^2$ real numbers; a real symmetric one by $n + n(n-1)/2$ real numbers. The rest is arithmetic. □

Part (c) is the flagpole picture of a null vector as a spinor, and (d) is Penrose's zig-zag: a timelike record is two null ones [145]. Classical runs stay in the plane $x_1 = x_2 = 0$, the plane of Section 24.1. Two questions remain: why the two sides should be stored on two basis states, and whether a post can point anywhere else in the cone. The first can be answered from weaker premises, in two ways.

Theorem 24.20 (An exact debit readout makes every post light-like)**.** *Let ρ assign to every posting word w a complex 2×2 matrix $\rho(w)$, with ρ of the empty word 0, $\rho(ww') = \rho(w) + \rho(w')$, and each single post's matrix $\rho(d)$, $\rho(c)$ a coherent record of trace 1.*

(a) Suppose some E with E and $1 - E$ coherent records reads the debit count exactly: $\mathrm{tr}(E\rho(w)) = D(w)$ *for every w. Then there is a unit vector $a \in \mathbb{R}^3$ with*

$$\rho(w) = R\big(D(w) + C(w),\ (D(w) - C(w))\, a\big) \qquad \text{for every } w.$$

In particular every single post's record is null with time coordinate 1.

(b) Suppose instead that $\rho(w)_{00} = D(w)$ for every w. Then $\rho(w)$ is the run record $\mathrm{diag}(D(w), C(w))$ *for every w.*

Proof. Additivity gives $\rho(w) = D(w)P_d + C(w)P_c$ with $P_d = \rho(d)$ and $P_c = \rho(c)$. Both are Hermitian of trace 1, so $P_d = R(1, a)$ and $P_c = R(1, b)$ with $|a|, |b| \leq 1$ by Theorem 24.14(c).

(a) Write $E = R(e_0, \varepsilon)$. Since $\mathrm{tr}\,\sigma_i = 0$ and $\mathrm{tr}(\sigma_i\sigma_j) = 2\delta_{ij}$, one has $\mathrm{tr}(R(t,x)R(s,y)) = \frac{1}{2}(ts + x\cdot y)$. Exact readout of one debit and of one credit gives $e_0 + \varepsilon\cdot a = 2$ and $e_0 + \varepsilon\cdot b = 0$. Positivity of E and $1 - E = R(2 - e_0, -\varepsilon)$ gives $|\varepsilon| \leq e_0$ and $|\varepsilon| \leq 2 - e_0$. By the Cauchy–Schwarz inequality,

$$2 - e_0 = \varepsilon\cdot a \leq |\varepsilon| \leq e_0, \qquad e_0 = -\varepsilon\cdot b \leq |\varepsilon| \leq 2 - e_0.$$

So $e_0 = 1$ and $|\varepsilon| = 1$, and both chains are equalities: $\varepsilon\cdot a = 1 = |\varepsilon||a|$ with $|a| \leq 1$ forces $a = \varepsilon$, and likewise $b = -\varepsilon$. Hence $\rho(w) = D\,R(1,a) + C\,R(1,-a) = R(D + C, (D - C)a)$, and a single post has $Q = 1 - |a|^2 = 0$.

(b) The upper diagonal entries give $(1 + a_3)/2 = 1$ and $(1 + b_3)/2 = 0$. With $|a|, |b| \leq 1$ this forces $a = (0,0,1)$ and $b = (0,0,-1)$, so $P_d = \mathrm{diag}(1,0)$, $P_c = \mathrm{diag}(0,1)$ and $\rho(w) = \mathrm{diag}(D, C)$. □

The second way starts from how a ledger is written into a register. Fix a bound $N \geq 1$, and let a *ledger code* be a ledger of d accounts with every count in $\{0, \dots, N\}$. A post at account k on a side raises that count by one modulo $N+1$. In the *census encoding*, a register with a reference part $\mathscr{R}_0$ and a code part whose basis states are the codes holds, after the ledger with code y, the state $\rho_y \otimes |y\rangle\langle y|$ with $\operatorname{tr} \rho_y = 1$. The *side record* of a post at account k from code y is the state after the post, with the reference traced out, compressed to the two possible outcome codes: first the code after a debit post, then the code after a credit post.

Proposition 24.21 (Census encoding stores the side on two basis states)**.** *In the census encoding:*

(a) if the posted count is less than N, the side record of a post is $\operatorname{diag}(1,0)$ *for a debit and* $\operatorname{diag}(0,1)$ *for a credit, so it lies on the axis* $x_1 = x_2 = 0$*;*

(b) the two outcome codes of a post are distinct basis states;

(c) for a posting word w *with* $D(w), C(w) \leq N$*, posted at one account from the empty code, the side records of its posts add up to the run record* $\operatorname{diag}(D(w), C(w))$*.*

Proof. (b) A debit post changes the debit count of account k and a credit post does not; since $N \geq 1$, raising a count by one modulo $N+1$ changes it. (a) After the post the state is $\rho_{y'} \otimes |y'\rangle\langle y'|$, where y' is the outcome code of the posted side. Tracing out the reference leaves $(\operatorname{tr} \rho_{y'})|y'\rangle\langle y'| = |y'\rangle\langle y'|$, and its compression to the two outcome codes, which are distinct by (b), has a single entry 1, in the place of the posted side. (c) Along the word no count exceeds N, so each post has a posted count below N, and (a) applies post by post; the sum of the side records is the run record. □

So the storage of sides on two basis states follows either from an exact readout of the debit count or from the census encoding of the ledger. Both routes also show the limit of posting. A post of definite side lands on the axis of the account's plane. A coherent post, one whose side record is a pure record off that axis, would be a superposition of two posted ledgers, and posting does not produce one. Such posts can be prepared, however, in any register.

Theorem 24.22 (Coherent posts can be prepared)**.**

(a) In every register with distinct basis states $e(0), e(1)$*, for every unit vector* $v \in \mathbb{C}^2$ *there is a unitary* U *of the register such that the compression of* $U|e(0)\rangle\langle e(0)|U^\dagger$ *is* $vv^\dagger$*.*

(b) For every unit $v \in \mathbb{C}^2$ *some unitary* A *with* $\det A = 1$ *has* $A \operatorname{diag}(1,0) A^\dagger = vv^\dagger$*; and for every nonzero coherent record* M *of determinant* 0 *some* A *with* $\det A = 1$ *has* $A \operatorname{diag}(1,0) A^\dagger = M$*.*

(c) A change A *keeps the debit post on the axis, that is, the coordinates* x_1, x_2 *of* $A \operatorname{diag}(1,0) A^\dagger$ *vanish, if and only if* $A_{00} = 0$ *or* $A_{10} = 0$*.*

(d) Every set of possible post records that contains the debit post $\operatorname{diag}(1,0)$ *and is closed under* $\rho \mapsto A \rho A^\dagger$ *for unitary* A *with* $\det A = 1$ *contains every pure record of trace* 1*.*

Conditions and scope.

Base and premises

Theorems 24.20 and 24.22 and Proposition 24.21 are proved in classical mathematics over the complex numbers. The readout route assumes additive records, coherent posts of trace 1 and an exact debit readout; the encoding route assumes the census encoding and stays within the count bound.

Conclusion

The storage of a post's side on two basis states follows from either set of premises. Single posts are light-like: classical posts are the two null rays, and under an exact readout every single post is null, along whatever axis. Every pure record of a side superposition can be prepared by a unitary of any register, and invariance of the possible post records under rotations would force all of them.

Identification

The directions x_1, x_2 of the record cone still rest on a premise that is not derived: either that nature prepares posts in side superpositions, or that the set of possible post records is invariant under rotations of the frame. Neither follows from posting; by (d) the second, with the debit post, makes every side superposition a possible post record.

Proof. (a) Let $u = \begin{pmatrix} v_0 & -\bar{v}_1 \\ v_1 & \bar{v}_0 \end{pmatrix}$, a unitary of determinant $|v_0|^2 + |v_1|^2 = 1$ with first column v. Let V embed $\mathbb{C}^2$ into the register along e, so $V^\dagger V = 1$, and put $U = 1 + V(u-1)V^\dagger$. Then $U^\dagger U = 1 + V\big((u-1) + (u^\dagger - 1) + (u^\dagger - 1)(u-1)\big)V^\dagger = 1 + V(u^\dagger u - 1)V^\dagger = 1$, and $V^\dagger U V = u$. The compression of $U|e(0)\rangle\langle e(0)|U^\dagger$ is $(V^\dagger U V|0\rangle)(V^\dagger U V|0\rangle)^\dagger = (u|0\rangle)(u|0\rangle)^\dagger = vv^\dagger$.

(b) The matrix u of (a) has $u\,\mathrm{diag}(1,0)u^\dagger = vv^\dagger$. For a nonzero null record, $M = vv^\dagger$ with $v \neq 0$ by Theorem 24.19(c); take $A = u\,\mathrm{diag}(|v|, 1/|v|)$ with u built from $v/|v|$.

(c) $A\,\mathrm{diag}(1,0)A^\dagger = cc^\dagger$ with $c = (A_{00}, A_{10})$, whose off-diagonal entry is $A_{00}\overline{A_{10}}$; it vanishes exactly when one factor does.

(d) Apply the unitaries u of (a) to the debit post. □

24.8 Least comparison cost is greatest proper time

Section 24.3 found that the inertial history has the greatest proper time for its end record. Does a least-cost principle select it? It depends on which cost. Charged leg by leg, Recognition Science's step cost prefers the opposite: its cheapest histories have no proper time at all. Charged against the frame of the whole history, the comparison cost is exactly the proper-time deficit, and its least value selects exactly the inertial histories.

Definition 24.23 (Step cost and comparison cost of a history)**.** Let $(w_1, \dots, w_n)$ be a history with whole word $W = w_1 \cdots w_n$. Its *step cost* charges each leg as one step of a single account, $\sum_i \big(G(D(w_i)) + G(C(w_i))\big)$, with $G(u) = \cosh u - 1$ as in Definition 24.18. When $D(W), C(W) \geq 1$, the *comparison cost* of a leg w against W is

$$c_W(w) = \frac{\left(\sqrt{D(w)C(W)} - \sqrt{C(w)D(W)}\right)^2}{\sqrt{D(W)C(W)}},$$

and the comparison cost of the history is $\sum_i c_W(w_i)$.

Theorem 24.24 (The step cost is not proper time)**.**

(a) On a ledger of one account, the step cost of a change is $G(D' - D) + G(C' - C)$. On a ledger of any number of accounts, a run of n single posts at one account, each charged as one step, costs $nG(1)$ whatever the order of its sides.

(b) For natural numbers with $a + b = a' + b'$: $G(a) + G(b) \le G(a') + G(b')$ if and only if $a'b' \le ab$.

(c) For every history, splitting it into single posts never raises its step cost, and the split has total proper time 0.

(d) The statement "between histories with the same end record, the lower step cost belongs to the greater total proper time" is false. For every word W with $D(W), C(W) \ge 1$, the split of W into single posts has the least step cost among all histories with the end record of W, and total proper time 0, while $\tau(W) > 0$.

Proof. (a) The first statement is the definition of the step cost for one account. A single post changes one count by 1 and the other by 0, and $G(0) = 0$, so each post costs $G(1)$.

(b) $\cosh a + \cosh b = 2\cosh\frac{a+b}{2}\cosh\frac{a-b}{2}$. With $a + b$ fixed this increases with $|a - b|$, and $(a - b)^2 = (a + b)^2 - 4ab$ decreases as ab increases.

(c) G is superadditive on $[0, \infty)$: $G(a) + G(b) \le G(a + b)$ is $\cosh a + \cosh b - 1 \le \cosh a \cosh b + \sinh a \sinh b$, that is, $0 \le (\cosh a - 1)(\cosh b - 1) + \sinh a \sinh b$, true for $a, b \ge 0$. By induction $mG(1) \le G(m)$, so a leg with counts (D, C) costs at least $(D + C)G(1)$, the cost of its $D + C$ single posts. A single post is one-sided, so its proper time is 0.

(d) By (c), every history with the end record of W costs at least $(D(W) + C(W))G(1)$, which is the cost of the split of W into single posts; that split has total proper time $0 < 2\sqrt{D(W)C(W)}$. A concrete instance: the legs dd and dc cost $G(2) + 2G(1) \approx 3.848$ and have total proper time $0 + 2 = 2$, while the single leg $dddc$ costs $G(3) + G(1) \approx 9.611$ and has proper time $2\sqrt{3} \approx 3.464$. The cheaper history has the smaller proper time. □

The step cost prefers the null zigzag: a history that never posts both sides in one step. Proper time measures the opposite, how much the two sides are mixed. The comparison cost repairs this by charging each leg against the frame of the whole.

Theorem 24.25 (The comparison cost is the proper-time deficit)**.** *Let $(w_1, \dots, w_n)$ be a history whose whole word W has $D(W), C(W) \ge 1$.*

(a) $\tau(W) - \sum_i \tau(w_i) = \sum_i c_W(w_i)$.

(b) For every leg w, $c_W(w) \ge 0$, with equality if and only if $D(w)C(W) = D(W)C(w)$. If $D(w), C(w) \ge 1$, then

$$c_W(w) = \tau(w)\, J\Big(\sqrt{\frac{D(w)/C(w)}{D(W)/C(W)}}\Big).$$

(c) For two histories with the same end record, the comparison cost of the first is at most that of the second if and only if the total proper time of the second is at most that of the first. The one-leg history (W) has comparison cost 0.

(d) A history has the least comparison cost among all histories with its end record if and only if every leg is parallel to the whole: $D(w_i)C(W) = D(W)C(w_i)$ for every i.

Conditions and scope.

Base and premises

Real square roots and cosh; both theorems are proved in classical mathematics.

Conclusion

"Least cost equals greatest proper time" is false for the step cost, which is minimized by a null zigzag, and true for the comparison cost against the history's own frame, which equals the proper-time deficit exactly. Least comparison cost means inertial.

Identification

That physical histories are those of least comparison cost is an identification; by (d) it says exactly that physical histories are inertial. The comparison cost refers to the frame of the whole history, so it is not a cost charged step by step.

Proof. Write $D = D(W)$, $C = C(W)$, $D_i = D(w_i)$ and $C_i = C(w_i)$, so that $\sum_i D_i = D$ and $\sum_i C_i = C$.

(a) Expanding the square, $c_W(w_i) = (D_i C + C_i D - 2\sqrt{D_i C_i DC})/\sqrt{DC}$. Summing over i gives $2DC/\sqrt{DC} - 2\sum_i \sqrt{D_i C_i} = \tau(W) - \sum_i \tau(w_i)$.

(b) $c_W(w)$ is a square divided by a positive number, and it vanishes exactly when $D(w)C = C(w)D$. If $D(w), C(w) \geq 1$, put $p = D(w)C$ and $q = C(w)D$. Then p/q is the ratio inside the square root, and, as in the proof of Theorem 24.5, $J(\sqrt{p/q}) = (\sqrt{p} - \sqrt{q})^2/(2\sqrt{pq})$. Since $\sqrt{pq} = \sqrt{D(w)C(w)}\sqrt{DC}$, we get $\tau(w)J(\sqrt{p/q}) = (\sqrt{p} - \sqrt{q})^2/\sqrt{DC} = c_W(w)$.

(c) By (a), each comparison cost is $\tau(W)$ minus the total proper time, and $\tau(W)$ is the same for both histories. The one-leg history has $c_W(W) = 0$.

(d) Every history with this end record has comparison cost at least 0, by (b), and the one-leg history attains 0. So a history is least exactly when its cost is 0, that is, by (b), when every leg is parallel to W. □

By Theorem 24.5(b) a leg parallel to the whole has the same ratio as the whole, so (d) says that least comparison cost means inertial. The identity (a) is the posting form of the special-relativistic formula for proper time as a sum of dilated pieces. It is linear in the deficit, where Theorem 24.5(d) was quadratic.

Why would a history be inertial? Counting gives a partial answer. Most posting orders with a given end record are nearly inertial over any split into two legs of fixed proportions.

Theorem 24.26 (Most posting orders are nearly inertial). *Let $n = n_1 + n_2$ and $0 \le D \le n$, and write $\binom{m}{a}$ for binomial coefficients, zero when $a < 0$ or $a > m$.*

(a) There are $\binom{n}{a}$ posting words of length n with a debits.

(b) Among the words of length n with D debits, those whose first n_1 posts contain a debits number $\binom{n_1}{a}\binom{n_2}{D-a}$, and $\sum_a \binom{n_1}{a}\binom{n_2}{D-a} = \binom{n}{D}$.

(c) $n\sum_a a\binom{n_1}{a}\binom{n_2}{D-a} = n_1 D\binom{n}{D}$: on average the first leg posts debits at the rate of the whole.

(d) $(n-1)\sum_a (an - n_1 D)^2 \binom{n_1}{a}\binom{n_2}{D-a} = n_1 n_2 D(n-D)\binom{n}{D}$.

(e) Let $\varepsilon > 0$, $n_1 \ge 1$ and $n \ge 2$. Among the words of length n with D debits, the share of those whose first leg's debit fraction a/n_1 differs from D/n by at least ε is at most

$$\frac{n_2}{4\varepsilon^2 n_1 (n-1)}.$$

If $n_1 = p_1 m$ and $n_2 = p_2 m$ with fixed $p_1, p_2 \ge 1$, this share tends to 0 as $m \to \infty$, for every choice of $D \le n$.

(f) If a weighting of posting words gives equal weight to words of the same length and debit count, then the weighted share of any set of words with D debits equals its counted share.

Conditions and scope.

Base and premises

Parts (a) to (d) are identities of natural numbers and integers, proved below with binomial identities. Parts (e) and (f) are proved over the real numbers in classical mathematics.

Conclusion

Over a split into two legs of fixed proportions, the share of posting orders whose first leg deviates from the whole's rate by ε tends to 0.

Identification

That all posting orders with the same length and debit count are equally weighted is an identification.

Not claimed

Approximate inertia of histories with more than two legs, or a derivation of the inertial reading: counting supports it and does not derive it.

Proof. (a) A word is fixed by the set of positions of its debits. (b) Choose the a debit positions among the first n_1 places and the other $D-a$ among the last n_2; summing over a counts every word once.

(c) and (d). Write $S = \binom{n}{D}$ and $\Sigma_f = \sum_a f(a)\binom{n_1}{a}\binom{n_2}{D-a}$. From $a\binom{n_1}{a} = n_1\binom{n_1-1}{a-1}$ and $a(a-1)\binom{n_1}{a} = n_1(n_1-1)\binom{n_1-2}{a-2}$, with (b) applied to $n-1$ and $n-2$ places, and from $n\binom{n-1}{D-1} = D\binom{n}{D}$ and $n(n-1)\binom{n-2}{D-2} = D(D-1)\binom{n}{D}$:

$$n\,\Sigma_a = n_1 D S, \qquad n(n-1)\,\Sigma_{a(a-1)} = n_1(n_1-1)D(D-1)S.$$

The first is (c). For (d), expand $(an - n_1 D)^2 = n^2 a(a-1) + n^2 a - 2nn_1 Da + n_1^2 D^2$ and multiply by $n - 1$:

$$(n-1)\Sigma_{(an-n_1D)^2} = S\big(nn_1(n_1-1)D(D-1) + (n-1)nn_1 D - 2(n-1)n_1^2 D^2 + (n-1)n_1^2 D^2\big).$$

The bracket equals $n_1 D\big(n^2 - nn_1 - nD + n_1 D\big) = n_1 D(n-n_1)(n-D)$, and $n - n_1 = n_2$.

(e) The first leg's fraction deviates by at least ε exactly when $|an - n_1 D| \geq \varepsilon n_1 n$. On those words $(an - n_1 D)^2 \geq (\varepsilon n_1 n)^2$, so by (d) (Chebyshev's argument) the number of deviating words is at most $n_1 n_2 D(n-D)S/((n-1)\varepsilon^2 n_1^2 n^2)$. Divide by S and use $D(n-D) \leq n^2/4$. With $n_1 = p_1 m$ and $n_2 = p_2 m$ the bound is $p_2/(4\varepsilon^2 p_1(n-1))$, which tends to 0 and does not depend on D.

(f) All words with D debits have the same length and debit count, hence the same weight. □

The mean and variance in (c) and (d) are those of the hypergeometric distribution [140]. The theorem concerns the rate of the first leg; it says nothing about how often the first leg is exactly parallel to the whole.

24.9 Clocks: resolution, self-recognition and the action floor

A ledger counts posts; a clock counts turns. Recognition Science's clock has eight modes, and mode k turns $k/8$ of a turn per tick. Records of the clock are phases, and a phase does not show how many whole turns it has made. So two questions arise. How finely do phase records identify the energies that drive the clock? And what does it cost, in drive, to change a register's configuration with certainty? The answers are exact, and the first one has a twist: what limits resolution is not what a record reads but when it is taken.

Definition 24.27 (Clock generators and records)**.** Fix $\hbar > 0$ and a tick $\tau_0 > 0$. A *generator* is a list $E = (E_0, \dots, E_7)$ of real mode energies. The *record* of mode k at time t is the phase $e^{-iE_k t/\hbar}$. Two generators have the *same tick map* when their records agree on every mode at every whole tick $n\tau_0$, $n \in \mathbb{Z}$. The *source generator* is $E_k^{\mathrm{src}} = \pi\hbar k/(4\tau_0)$, under which mode k turns by $-k/8$ of a turn per tick. The *extra turns* of E against E' on mode k are $x_k = (E_k - E'_k)\tau_0/(2\pi\hbar)$; the energy $2\pi\hbar/\tau_0$ of one extra turn per tick is the *branch quantum.* A set T of real numbers *identifies* E' when every generator whose records agree with those of E' on every mode at every time $s\tau_0$, $s \in T$, equals E'. The *native coefficient field* is $\mathbb{Q}(i, \sqrt{2}) = \mathbb{Q}(\zeta_8)$, $\zeta_8 = e^{2\pi i/8} = (1+i)/\sqrt{2}$.

Lemma 24.28 (The record law)**.** *The records of E and E' on mode k agree at time $s\tau_0$ if and only if $x_k s \in \mathbb{Z}$.*

Proof. $e^{-iE_k s\tau_0/\hbar} = e^{-iE'_k s\tau_0/\hbar}$ exactly when $(E_k - E'_k)s\tau_0/\hbar \in 2\pi\mathbb{Z}$, that is, when $x_k s \in \mathbb{Z}$. □

Theorem 24.29 (The resolution of the clock)**.**

(a) Let $M \geq 1$. If $0 \leq E_k\tau_0/\hbar < 2\pi M$ for every k and the records of E agree with those of E^{src} on every mode at time τ_0/M, then $E = E^{\mathrm{src}}$. This identification is not robust: for every $\varepsilon > 0$ some $E \neq E^{\mathrm{src}}$ in the same window has records at τ_0/M within ε of the source's on every mode.

(b) The window is sharp. For $M \geq 1$, the generator $E^{(M)}$ with M extra turns per tick on mode 1 and the source energies on the other modes differs from E^{src}, has the same tick map (hence the same cost of the one-tick map), nonnegative energies and $E_0^{(M)} = 0$, has $E_1^{(M)}\tau_0/\hbar = \pi/4 + 2\pi M$, and agrees with E^{src} on every mode at every time $j\tau_0/M$, $j \in \mathbb{Z}$.

(c) No finite set of rational times identifies the source: for every finite set S of rationals some $E \neq E^{\mathrm{src}}$ with the same tick map, nonnegative energies and $E_0 = 0$ agrees with E^{src} on every mode at every time $q\tau_0$, $q \in S$. Yet every $E \neq E^{\mathrm{src}}$ disagrees with E^{src} on some mode at some rational time $q\tau_0$. And if s is irrational, the set $\{1, s\}$ identifies E^{src}.

(d) The records of E^{src} at whole ticks lie in $\mathbb{Q}(i, \sqrt{2})$ and repeat with period 8 ticks. For rational q, the record of mode 1 at time $q\tau_0$ lies in $\mathbb{Q}(i, \sqrt{2})$ if and only if q is an integer.

Proof. (a) By the record law with $s = 1/M$, each $(E_k - E_k^{\mathrm{src}})\tau_0/\hbar$ lies in $2\pi M\mathbb{Z}$. Both $E_k\tau_0/\hbar$ and $E_k^{\mathrm{src}}\tau_0/\hbar = \pi k/4$ lie in $[0, 2\pi M)$, so their difference lies in $(-2\pi M, 2\pi M) \cap 2\pi M\mathbb{Z} = \{0\}$. For the instability, let $0 < \delta \leq \min(\pi, \varepsilon M/2)$ and change only mode 0, to $E_0\tau_0/\hbar = 2\pi M - \delta$. Its record at τ_0/M is $e^{i\delta/M}$, within $\delta/M < \varepsilon$ of the source's record 1.

(b) The extra turns are $x_1 = M$ and $x_k = 0$ otherwise. At $s = j/M$ the record law holds, since $Mj/M = j$; whole ticks are the case $M \mid j$. The cost of the one-tick map is a function of the tick map. The remaining statements are read off the definition.

(c) Let L be the product of the denominators of the elements of S, and give E exactly L extra turns on mode 1: then $Lq \in \mathbb{Z}$ for every $q \in S$, and E has the same tick map, nonnegative energies and $E_0 = 0$. If $E \neq E^{\mathrm{src}}$, some $x_k \neq 0$. If $x_k \notin \mathbb{Z}$, the time τ_0 already disagrees; otherwise $q = 1/(|x_k| + 1)$ gives $x_k q \in (-1, 1) \setminus \{0\}$, not an integer. Finally, agreement at τ_0 gives $x_k \in \mathbb{Z}$, and agreement at $s\tau_0$ gives $x_k s \in \mathbb{Z}$; if $x_k \neq 0$, then $s = x_k s/x_k$ would be rational.

(d) The record of mode k at $n\tau_0$ is $e^{-i\pi kn/4} = \zeta_8^{-kn}$, which lies in $\mathbb{Q}(\zeta_8)$ and depends on n only modulo 8. For rational r, $e^{2\pi ir} \in \mathbb{Q}(\zeta_8)$ exactly when $8r \in \mathbb{Z}$: if $r = a/b$ in lowest terms, $e^{2\pi ir}$ is a primitive bth root of unity, and if it lies in $\mathbb{Q}(\zeta_8)$ then $\mathbb{Q}(\zeta_{\mathrm{lcm}(b,8)}) = \mathbb{Q}(\zeta_b, \zeta_8) = \mathbb{Q}(\zeta_8)$ has degree 4 over $\mathbb{Q}$; the cyclotomic field $\mathbb{Q}(\zeta_m)$ has degree $\phi(m)$ [148], and among multiples of 8 only $m = 8$ has $\phi(m) = 4$, so $b \mid 8$. The converse is $e^{2\pi ir} = \zeta_8^{8r}$. The record of mode 1 at $q\tau_0$ is $e^{2\pi i(-q/8)}$, and $8 \cdot (-q/8) = -q$. □

Conditions and scope.

Base and premises

Real generators and complex phases; the theorem and Proposition 24.30 are proved in classical mathematics.

Conclusion

One record per mode at τ_0/M identifies the source within a band of M turns per tick, and no better; finitely many rational sample times never identify it, while each wrong generator is caught at some rational time; the tick and one irrational time identify it with no band. Records can be exact native numbers below the tick: the continuum enters through when a record is taken, not through what it reads.

Not claimed

Robustness of (a) to noise in the records, which fails.

Which sets of times identify the source? Exactly those that generate a dense group.

Proposition 24.30 (Identifying times generate a dense group)**.**

(a) A set T of reals identifies E^{src} if and only if the additive subgroup of $\mathbb{R}$ generated by T is dense.

(b) For a finite set S of rationals, with L the least common multiple of their denominators: the records of E and E' on mode k agree at τ_0 and at every $q\tau_0$, $q \in S$, if and only if x_k is an integer multiple of L.

(c) If T identifies E^{src} and the mode-1 source record at $s\tau_0$ lies in $\mathbb{Q}(i, \sqrt{2})$ for every $s \in T$, then T contains an irrational number. The Pythagorean time *$s_P = -4\arccos(3/5)/\pi$ is irrational, $\{1, s_P\}$ identifies E^{src}, and every source record at $s_P\tau_0$ lies in $\mathbb{Q}(i)$: on mode k it is $\big((3+4i)/5\big)^k$.*

Proof. (a) For a real $x \neq 0$, the times s with $xs \in \mathbb{Z}$ form the group $x^{-1}\mathbb{Z}$, which has no element in $(0, 1/|x|)$. By the record law, the records of E and E^{src} on mode k agree on all of T exactly when $x_k s \in \mathbb{Z}$ for all s in the group generated by T. If that group is dense, it meets $(0, 1/|x_k|)$ when $x_k \neq 0$, so every x_k vanishes. If it is not dense, it is $a\mathbb{Z}$ for some $a \geq 0$, since every additive subgroup of $\mathbb{R}$ is dense or cyclic; a generator with $1/a$ extra turns per tick on mode 1 (any nonzero number of extra turns if $a = 0$) agrees on T and is not the source.

(b) Agreement at τ_0 is $x_k \in \mathbb{Z}$. For $q = u/v$ in lowest terms, $x_k u/v \in \mathbb{Z}$ means $v \mid x_k u$, that is, $v \mid x_k$. All denominators divide x_k exactly when L does.

(c) By the proof of Theorem 24.29(d), a rational s with native mode-1 record is an integer. If every element of T were rational, T would generate a subgroup of $\mathbb{Z}$, which is not dense, and T would not identify the source by (a). Next, mode k at $s_P\tau_0$ has record $e^{ik\arccos(3/5)} = (3/5 + 4i/5)^k$. The number $z = (3+4i)/5$ is not a root of unity: it satisfies $z^2 - \frac{6}{5}z + 1 = 0$, and as it is not real this is its minimal polynomial, which does not have integer coefficients, so z is not an algebraic integer. Hence $\arccos(3/5)/\pi$ is irrational, since $\arccos(3/5)/\pi = u/v$ would give $z^{2v} = 1$; so s_P is irrational, and $\{1, s_P\}$ identifies the source by Theorem 24.29(c). □

A clock can also recognize itself. Mode k of the source returns to phase 1 for the first time at $8\tau_0/k$. Records taken at the tick and at these seven first returns identify the generator up to a fixed band.

Theorem 24.31 (The self-recognition band is 105)**.**

(i) For $k = 1, \dots, 7$, the record of source mode k is 1 at time $8\tau_0/k$ and differs from 1 at every time t with $0 < t < 8\tau_0/k$.

(ii) A generator E agrees with E^{src} on every mode at the eight times $q\tau_0$ with

$$q \in \{1,\ 8,\ 4,\ \tfrac{8}{3},\ 2,\ \tfrac{8}{5},\ \tfrac{4}{3},\ \tfrac{8}{7}\}$$

if and only if, mode by mode, E differs from E^{src} by an integer multiple of $105 = 3 \cdot 5 \cdot 7$ branch quanta.

(iii) These eight times generate the group $\frac{1}{105}\mathbb{Z}$, since $1 + \frac{8}{3} + \frac{8}{7} - 3 \cdot \frac{8}{5} = \frac{1}{105}$; agreement at all of them is equivalent to agreement at the single time $\tau_0/105$ on every mode.

Proof. (i) $e^{-i\pi kt/(4\tau_0)} = 1$ exactly when $kt/(8\tau_0) \in \mathbb{Z}$, and the least positive such t is $8\tau_0/k$.

(ii) By the record law the condition is $x_j q \in \mathbb{Z}$ for each listed q. The tick gives $x_j \in \mathbb{Z}$; then $\frac{8}{3}x_j \in \mathbb{Z}$ gives $3 \mid x_j$, $\frac{8}{5}x_j$ gives $5 \mid x_j$ and $\frac{8}{7}x_j$ gives $7 \mid x_j$, so $105 \mid x_j$. Conversely $105q$ is an integer for each listed q: the values are $105, 840, 420, 280, 210, 168, 140, 120$.

(iii) $(105+280+120-504)/105 = 1/105$, and every listed time is a multiple of $1/105$. Agreement at $\tau_0/105$ says $x_j/105 \in \mathbb{Z}$, which is (ii). □

By Proposition 24.30(a) the clock's own returns never single out its source: the generator with 105 extra turns per tick on mode 1 passes all of them. Reading a comparison against mode k as stopping at that mode's first return is an identification; the band itself needs none.

The second question concerns registers rather than clocks: what a single change of configuration costs. Recognition Science executes register operations by words of tokens, each driven for a duration ε while the register's own holding Hamiltonian acts. Section 21.7 met the same question for comparisons, on registers of binary channels (Definition 21.35); here the channels have eight states, the paid token is the same, and the object is the register itself.

Definition 24.32 (Tokens, words and paid words)**.** Let a register be given, with its configurations as basis and the operator norm $\|\cdot\|$. A *token* t carries a skew-Hermitian *generator* G_t; a *word* is a finite list of tokens, and its *drive* is $\sum_t \|G_t\|$, in units of $\hbar$. Given a self-adjoint *holding* H and $\hbar, \varepsilon > 0$, the *paid token* is $e^{-i\|G_t\|} \exp(G_t - i\varepsilon H/\hbar)$, the token driven for time ε with the holding on, and the *paid word* W is the product of its paid tokens; with $H = 0$ it is the *ideal word*, up to a phase. In Recognition Science's native alphabet, given in full in Definition 30.19, the register consists of eight-state channels; a *braid token* on a channel has generator θB for a real angle θ and one of the 21 triad generators B of that definition, fixed nonzero real antisymmetric matrices on that channel, and an *entangling token* on an ordered pair (k, l) of distinct channels has generator $\frac{\pi}{4} i S_k F_l$, where S_k multiplies each configuration by the sign $(-1)^b$, b the number of bits 1 in the state of channel k, and F_l flips the low bit of the state of channel l. Recognition Science's *native register* is a register with this alphabet and the holding of Section A.8, equation (A.13): it has $20 + 2N + 4a + 4b$ channels for parameters $N, a, b \geq 0$ and $m, \beta \in \mathbb{Z}$, and its holding is built there from a positive interior holding $H_{\mathrm{old}}(N, m)$ that the construction takes as supplied.

The operator $S_k F_l$ is self-adjoint with square 1, so the entangling token's ideal action is $\exp(\frac{\pi}{4} i S_k F_l) = (1 + i S_k F_l)/\sqrt{2}$, the ideal entangler. The cost of certainty rests on one geometric fact: overlap angles add.

Lemma 24.33 (Overlap angles add)**.** *Let p, q, r be unit vectors with $|\langle p, q\rangle| \geq \cos\alpha$ and $|\langle q, r\rangle| \geq \cos\beta$, where $\alpha, \beta \in [0, \pi/2]$. Then $|\langle p, r\rangle| \geq \cos(\alpha + \beta)$.*

Proof. Multiplying p and r by unit phases, we may assume $a = \langle q, p\rangle$ and $b = \langle q, r\rangle$ are real and nonnegative. Write $p = aq + p'$ and $r = bq + r'$ with p', r' orthogonal to q, so $\|p'\| = \sqrt{1 - a^2}$ and $\|r'\| = \sqrt{1 - b^2}$. Then $|\langle p, r\rangle| = |ab + \langle p', r'\rangle| \geq ab - \sqrt{1 - a^2}\sqrt{1 - b^2} = \cos(\alpha' + \beta')$ with $a = \cos\alpha'$, $b = \cos\beta'$ and $\alpha', \beta' \in [0, \pi/2]$. Since $a \geq \cos\alpha$ and $b \geq \cos\beta$, $\alpha' \leq \alpha$ and $\beta' \leq \beta$, and cos decreases on $[0, \pi]$. □

Theorem 24.34 (The action floor of a native step)**.** *Let a register, a self-adjoint holding H, $\hbar, \varepsilon > 0$ and a word w of n tokens be given, with paid word W, and put $\theta = \mathrm{drive}(w) + n\varepsilon\|H\|/\hbar$.*

(a) For distinct configurations $x \neq y$, $|\langle y, Wx\rangle| \leq \sin\big(\min(\pi/2, \theta)\big)$.

(b) Hence a word that moves x to $y \neq x$ with probability $|\langle y, Wx\rangle|^2 \geq 1 - 2\delta$, $0 \leq \delta \leq \frac{1}{2}$, pays drive

$$\hbar\,\mathrm{drive}(w) \geq \hbar \arcsin\sqrt{1 - 2\delta} - n\varepsilon\|H\|,$$

and a certain move, $|\langle y, Wx\rangle| = 1$, pays at least $\hbar\pi/2 - n\varepsilon\|H\|$.

(c) The bound is attained. For distinct channels k, l, the word of two entangling tokens on (k, l) has drive exactly $\pi/2$ and ideal word iS_kF_l, and with zero holding its paid word moves every configuration x to the configuration x' with the low bit of channel l flipped, with $|\langle x', Wx\rangle| = 1 = \sin(\pi/2)$.

(d) On Recognition Science's native register, under its holding H, the same word moves x to x' with $|\langle x', Wx\rangle| \geq 1 - 2\varepsilon\|H\|/\hbar$.

Conditions and scope.

Base and premises

Complex matrices over the classical numbers; proved in classical mathematics. Parts (a) and (b) hold for every register, holding and word; their proof uses only that the token generators are skew-Hermitian. Part (c) holds on every register of at least two channels with the native alphabet; its proof uses the entangling generator itself.

Conclusion

The amplitude of any change of configuration is bounded by the sine of the drive plus the holding drift, so a change with certainty pays at least a quarter turn, $\hbar\pi/2$, less the drift; the doubled entangler attains the bound.

Identification

For a comparison clocked against a reference cycle whose action $A_0 = 2\pi\hbar$ uses the register's $\hbar$, the floor reads $A_0/4$ less the drift; that reading is an identification.

Not claimed

That the native register attains the quarter turn exactly under its holding: (d) gives the amplitude $1 - 2\varepsilon\|H\|/\hbar$, not 1.

Proof. (a) Each paid token is, up to a phase, e^{Y_t} with $Y_t = G_t - i\varepsilon H/\hbar$ skew-Hermitian and $\|Y_t\| \le \|G_t\| + \varepsilon\|H\|/\hbar =: \theta_t$. First, for skew-Hermitian Y with $\|Y\| \le \pi/2$ and unit v, $|\langle v, e^Y v\rangle| \ge \cos\|Y\|$: in an orthonormal eigenbasis of Y, with eigenvalues $i\lambda_j$ and $|\lambda_j| \le \|Y\|$, the real part of $\langle v, e^Y v\rangle$ is $\sum_j |v_j|^2 \cos\lambda_j \ge \cos\|Y\|$. So $|\langle v, P_t v\rangle| \ge \cos(\min(\pi/2, \theta_t))$ for every unit v, the bound being trivial when $\theta_t \ge \pi/2$. Next, if unitaries U_1, U_2 satisfy $|\langle v, U_i v\rangle| \ge \cos\Theta_i$ for all unit v, with $\Theta_i \in [0, \pi/2]$, then Lemma 24.33 with $p = v$, $q = U_1 v$ and $r = U_2 U_1 v$ gives $|\langle v, U_2 U_1 v\rangle| \ge \cos(\Theta_1 + \Theta_2)$, and the bound $\cos(\pi/2) = 0$ is trivial. By induction over the tokens, $|\langle v, Wv\rangle| \ge \cos\Theta$ with $\Theta = \min(\pi/2, \theta)$. Finally W is unitary and $x \perp y$, so $|\langle x, Wx\rangle|^2 + |\langle y, Wx\rangle|^2 \le \|Wx\|^2 = 1$, and $|\langle y, Wx\rangle|^2 \le 1 - \cos^2\Theta = \sin^2\Theta$.

(b) If $\theta \ge \pi/2$ the claims hold because arcsin $\le \pi/2$. Otherwise (a) gives $\sqrt{1 - 2\delta} \le \sin\theta$, so $\arcsin\sqrt{1 - 2\delta} \le \theta$; multiply by $\hbar$. Certainty is $\delta = 0$.

(c) Write $S = S_k F_l$. As S is self-adjoint with $S^2 = 1$, $\exp(\frac{\pi}{4} iS) = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4} S$, and the square is $\exp(\frac{\pi}{2} iS) = iS$. Each generator has norm $\pi/4$, so the drive is $\pi/2$. With $H = 0$ the paid word is $e^{-i\pi/2} \cdot iS = S$, and $Sx = \pm x'$.

(d) For skew-Hermitian A and B, Duhamel's formula $e^{A+B} - e^A = \int_0^1 e^{s(A+B)} B e^{(1-s)A}\, ds$ gives $\|e^{A+B} - e^A\| \le \|B\|$. So each paid token is within $\varepsilon\|H\|/\hbar$ of its zero-holding version, the two-token word is within $2\varepsilon\|H\|/\hbar$ of the word of (c), and the amplitude drops by at most that much. □

The theorem has the form of a quantum speed limit, in the line of Mandelstam and Tamm [103] and Margolus and Levitin [104], stated for drive rather than energy: a quarter turn of drive buys one certain change. Part (a) is the step bound of Lemma 21.36(2), proved there for binary channels with the Fubini–Study distance and here for every register by adding overlap angles, and the word of (c) is the one that attains the threshold of Corollary 21.38(2). Whether the native register can make a certain change at all under its holding is a different question, and it is open. We can say exactly what it reduces to.

Definition 24.35 (Phased words, density, exactness and entangler reach)**.** Fix a register, a holding H and $\hbar, \varepsilon > 0$. A *phased paid word* is cW with $|c| = 1$ and W a paid word. The setting is *dense* when every unitary of the register is an operator-norm limit of phased paid words, and *exact* when every unitary is a phased paid word. It has *entangler reach* when every ideal entangler $(1 + iS_k F_l)/\sqrt{2}$ is a limit of phased paid words.

Proposition 24.36 (Certain flips reduce to reaching the entangler)**.** *Consider a register of at least two channels with Recognition Science's native alphabet, a self-adjoint holding H and $\hbar, \varepsilon > 0$.*

(a) If the setting is dense, it is exact.

(b) Every ideal braid rotation $\exp(\alpha B)$ *is a limit of phased paid words.*

(c) *Suppose the ideal native alphabet is universal on the register: the closed group generated by the ideal braid rotations and the ideal entanglers contains every unitary of determinant* 1. *Then entangler reach, density and exactness are equivalent.*

(d) *Under the universality of (c) on every native register, if entangler reach holds at every setting of the native register under its holding, then at every setting some paid native word moves some configuration to a different one with amplitude of modulus exactly* 1.

(e) *Adding a real multiple of the identity to H does not change entangler reach, and $(\hbar, \varepsilon)$ enter paid words only through $\varepsilon/\hbar$. In Recognition Science's native register with parameters (N, a, b, m), the exterior flux β enters the holding only as such a multiple, the energy $c(\beta)$ of one exterior edge. So entangler reach at $(N, a, b, m, \beta, \hbar, \varepsilon)$ is equivalent to entangler reach at $(N, a, b, m, 0, 1, \varepsilon/\hbar)$.*

Proof. Let $\mathscr{P}$ be the set of phased paid words. It is closed under products, and so is its closure $\overline{\mathscr{P}}$, a closed set of unitaries that is closed under products contains the inverse of each element g, because the powers of g have a convergent subsequence g^{m_j} in the compact unitary group, and then $g^{m_{j+1}-m_j-1} \to g^{-1}$.

(a) Fix a braid token and write $T(\theta) = \exp(\theta B + D)$ with $D = -i\varepsilon H/\hbar$; each $T(\theta)$ is in $\mathscr{P}$. Some θ_0 has a derivative $T'(\theta_0)T(\theta_0)^{-1}$ that is not a multiple of the identity: otherwise $T(\theta)T(0)^{-1}$ would be scalar for all θ, the differences of eigenvalues of the normal matrix $\theta B + D$ would stay in a countable set, and their spread $\sum_{j,k} |\lambda_j - \lambda_k|^2 = 2d\|\theta B + D\|_F^2 - 2|\operatorname{tr}(\theta B + D)|^2$ would be constant, whereas it is a quadratic polynomial in θ with leading coefficient $2d\|B\|_F^2 > 0$. Consider words $F(\theta_1, \dots, \theta_m) = U_m T(\theta_m) \cdots U_1 T(\theta_1)$ with fixed $U_i \in \mathscr{P}$. At $\theta_i = \theta_0$ their partial derivatives are conjugates of that non-scalar direction by the partial products. The unitary conjugates of a non-scalar skew-Hermitian matrix span all traceless skew-Hermitian matrices, and $\mathscr{P}$ is dense, so the U_i can be chosen one at a time until these directions, together with the phase direction, span the tangent space of the unitary group. By the inverse function theorem the map $(c, \theta_1, \dots, \theta_m) \mapsto cF(\theta)$ covers a neighbourhood O of $g_0 = c_0 F(\theta_0, \dots, \theta_0)$, so $O \subseteq \mathscr{P}$. For any unitary V, density gives $A \in \mathscr{P}$ so close to $V g_0^{-1}$ that $A^{-1}V \in O$; then $V = A \cdot A^{-1}V \in \mathscr{P}$.

(b) Put $B_\Theta = B + D/\Theta$ for $\Theta > 0$, so $T(\Theta) = \exp(\Theta B_\Theta)$ and $\exp((\Theta + \alpha)B_\Theta) = \exp(\alpha B_\Theta)T(\Theta)$. By Duhamel's bound, $T(\Theta + \alpha)$ is within $|\alpha|\|D\|/\Theta$ of $\exp((\Theta + \alpha)B_\Theta)$, and $\exp(\alpha B_\Theta)$ is within the same of $\exp(\alpha B)$. Hence $\|T(\Theta + \alpha)T(\Theta)^{-1} - \exp(\alpha B)\| \le 2|\alpha|\|D\|/\Theta \to 0$, and $T(\Theta)^{-1} \in \overline{\mathscr{P}}$ by the remark on inverses.

(c) Exactness gives density, and density gives entangler reach. Conversely, with entangler reach and (b), $\overline{\mathscr{P}}$ contains the ideal braid rotations and entanglers; it is a closed group, so by universality it contains every unitary of determinant 1, and with phases every unitary. So the setting is dense, and exact by (a).

(d) At each setting the setting is exact by (c). The permutation matrix that exchanges two distinct configurations x and y is unitary, so it is cW for a paid word W, and $|\langle y, Wx\rangle| = 1$.

(e) Adding $\lambda 1$ to H multiplies each paid token by the phase $e^{-i\varepsilon\lambda/\hbar}$, which phased words absorb; and the exponent $G_t - i(\varepsilon/\hbar)H$ involves $\hbar$ and ε only through their ratio. In the native register, β occurs in the placed holding (A.11) only in the term $c(\beta)I$, and (A.13) carries that term to $c(\beta)$ times the identity of the whole register; so passing from β to 0 adds the real multiple $c(0) - c(\beta)$ of the identity. □

Conditions and scope.

Base and premises

Classical analysis of the unitary group: the inverse function theorem, compactness, and the irreducibility of the conjugation action on traceless skew-Hermitian matrices. Parts (c) and (d) also assume the universality of the ideal native alphabet, as stated. That universality holds on every register of at least two channels with Recognition Science's native alphabet, by Theorem A.6: the braid rotations about the 21 triads are the rotations $\exp(tC_{ab7}/\sqrt{3})$, t real, which supply all the local cyclic rotations of that theorem, and the ideal entanglers include its couplings $(I + iS \otimes F)/\sqrt{2}$ and $(I + iF \otimes S)/\sqrt{2}$ on adjacent channels. So the only open premise of (d) is entangler reach. Part (e) reads off equations (A.11) and (A.13) that β enters the native holding only as the multiple $c(\beta)$ of the identity.

Conclusion

Exact certainty on a register is exactly as hard as approximate universality, and approximate universality reduces to one gate type, the entangler; the class of settings where it holds depends only on (N, a, b, m) and the dimensionless duration $\varepsilon/\hbar$.

Not claimed

Entangler reach itself. It is the open statement below.

Open problem (entangler reach on the native register). Let $N, a, b \geq 0$ and $m \in \mathbb{Z}$, let $H_{\mathrm{old}}(N, m)$ be a positive interior holding supplied as in Section A.8, and let H be the native holding (A.13) built from it at $\beta = 0$. For every $\varepsilon/\hbar > 0$: is every ideal entangler $(1 + iS_kF_l)/\sqrt{2}$ of the register an operator-norm limit of phased paid native words, each token driven for the duration $\varepsilon/\hbar$ under H with $\hbar = 1$? The question is asked for each such $H_{\mathrm{old}}(N, m)$. Neither a proof nor a refutation is known. The question is whether the drift of the native holding can be refocused in the limit.

24.10 What one balance history can be asked

Section 17.7 asked what finite records of a repeatedly examined system can confirm, refute or settle, and what deciding a question about it costs. An account under posting is such a system, and its own histories supply every binary world. Let the posts go on without end, and ask about the balance: will the account stay solvent forever? Will it ever close at zero? Each question has an exact price, and so do the detectors and deadlines one might build to answer it. Here *deciding* a question means proving, for every history, that it holds or fails; a map that returns the answer would be the corresponding data form (Section 21.1), which we do not use. Every price in this section, that is every equivalence with a principle, is proved with no choice principle; Theorem 24.39(i), Theorem 24.43(b) and the last statement of Theorem 24.43(d) reason classically, as they say. The principles are those of Definition 7.8, with Markov's principle in its binary form $\mathrm{MP}_{\mathrm{bin}}$.

Definition 24.37 (Balance histories and their questions)**.** A *balance history* is a map h from the natural numbers to sides, $h(n)$ being the side of post n. Its *balance* $b_h(n)$ is the number of debits minus the number of credits among the first n posts, an integer of $\mathbb{Z}_\delta$; so $b_h(0) = 0$ and $b_h(n+1) = b_h(n) \pm 1$, with $+$ for a debit. The history is *solvent forever*, $S(h)$, when $b_h(n) \geq 0$ for every n; it *closes*, $Z(h)$, when $b_h(n) = 0$ for some $n > 0$; it is *overdrawn*, $O(h)$, when $b_h(n) < 0$ for some n. Its *transfer partner* $\bar{h}$ posts the opposite side at every step. Examining a history n times yields its first n posts. A question Q about histories is *decided* when $Q(h)$ or $\neg Q(h)$ holds for every h.

The balance $b_h(n)$ is the cone coordinate x of the account's record after n posts (Definition 24.1). The prices come from one lemma about questions of the two simplest logical shapes.

Lemma 24.38 (Prices of one-quantifier questions)**.** *Let worlds be examined as in Section 17.7, and let $P(w, n)$ be decidable for every world w and number n.*

(a) If $Q(w) \Leftrightarrow \forall n\, P(w, n)$ and the zero test reduces to Q (Definition 17.14), then deciding Q is equivalent to WLPO, and the statement

$$\forall w, w'\ \big(\neg(\neg Q(w) \wedge \neg Q(w')) \to Q(w) \vee Q(w')\big)$$

is equivalent to LLPO.

(b) If $Q(w) \Leftrightarrow \exists n\, P(w, n)$ and the event test reduces to Q, then deciding Q is equivalent to LPO, and $\forall w(\neg\neg Q(w) \to Q(w))$ is equivalent to $\mathrm{MP}_{\mathrm{bin}}$.

None of the four equivalences uses a choice principle.

Proof. For a world w let $\alpha_w(n) = 1$ when $P(w, n)$ fails, in case (a), or holds, in case (b), and 0 otherwise; α_w is a binary sequence.

(a) WLPO at α_w gives $\alpha_w \equiv 0$, hence $Q(w)$, or $\neg(\alpha_w \equiv 0)$, hence $\neg Q(w)$. Conversely, if Φ reduces the zero test to Q, deciding $Q(\Phi(\alpha))$ decides $\alpha \equiv 0$.

For LLPO, let w, w' satisfy $\neg(\neg Q(w) \wedge \neg Q(w'))$. Interleave the failures of w at even places and of w' at odd places, and keep only the first: $\beta(2n) = 1$ when $P(w, n)$ fails and no earlier place of the interleaving records a failure, and likewise $\beta(2n+1)$ for w'. Then β has at most one entry 1. Suppose LLPO gives $\beta(2k) = 0$ for all k. If $P(w, n)$ failed, the first recorded failure would come at or before place $2n$, at an odd place, so w' would fail somewhere too, and $\neg Q(w) \wedge \neg Q(w')$ would hold, which is excluded; as $P(w, n)$ is decidable, it holds. So $Q(w)$; the odd case gives $Q(w')$. Conversely, given α with at most one entry 1, let $\alpha_e(k) = \alpha(2k)$ and $\alpha_o(k) = \alpha(2k+1)$, and apply the statement to $w = \Phi(\alpha_e)$ and $w' = \Phi(\alpha_o)$. Its hypothesis holds: suppose $\neg(\alpha_e \equiv 0)$ and $\neg(\alpha_o \equiv 0)$. If some $\alpha(2k) = 1$, every odd entry is 0, against $\neg(\alpha_o \equiv 0)$; so every $\alpha(2k)$ is 0, against $\neg(\alpha_e \equiv 0)$. Its conclusion is $\alpha_e \equiv 0$ or $\alpha_o \equiv 0$, which is LLPO at α.

(b) LPO at α_w gives $\alpha_w \equiv 0$, hence $\neg Q(w)$, or an n with $P(w, n)$, hence $Q(w)$. Conversely, deciding $Q(\Phi(\alpha))$ decides $\exists n\, \alpha(n) = 1$, and the negative case gives $\alpha \equiv 0$ because each entry is 0 or 1. If $\neg\neg Q(w)$, then $\neg(\alpha_w \equiv 0)$, and $\mathrm{MP}_{\mathrm{bin}}$ gives an n with $P(w, n)$. Conversely, if $\neg(\alpha \equiv 0)$, then $\neg\neg Q(\Phi(\alpha))$, so $Q(\Phi(\alpha))$, which gives an n with $\alpha(n) = 1$. □

Theorem 24.39 (The prices of one balance history)**.** *For balance histories of one account:*

(a) deciding solvency forever is equivalent to WLPO;

(b) deciding closing is equivalent to LPO;

(c) "every history that cannot fail to close closes", $\forall h\,(\neg\neg Z(h) \to Z(h))$, is equivalent to $\mathrm{MP}_{\mathrm{bin}}$;

(d) naming a solvent history of two that are not both insolvent, $\forall h, h'\,(\neg(\neg S(h) \wedge \neg S(h')) \to S(h) \vee S(h'))$, is equivalent to LLPO;

(e) deciding overdraft is equivalent to LPO, and $\forall h\,(\neg\neg O(h) \to O(h))$ to $\mathrm{MP}_{\mathrm{bin}}$;

(f) deciding closing is equivalent to deciding solvency together with (c);

(g) $b_h(n) + b_{\bar{h}}(n) = 0$ for every h and n, so this conservation is decided with no principle, while deciding "h is solvent forever and its transfer pair conserves" is equivalent to WLPO;

(h) for every N, "$b_h(n) \geq 0$ for all $n \leq N$" and "$b_h(n) = 0$ for some $0 < n \leq N$" are decided with no principle;

(i) solvency forever is refutable and not confirmable, and closing is confirmable and not refutable.

Parts (a) to (h) use no choice principle. Part (i) reasons classically about recorders, as Section 17.7 does.

Conditions and scope.

Base and premises
Unending posting histories of one account and integer balances; the omniscience principles on binary sequences.

Conclusion
Each classical price is a question about one balance history: solvency forever is WLPO, closing is LPO, waiting for a closing that cannot fail is $\mathrm{MP}_{\mathrm{bin}}$, naming a solvent account of two is LLPO. Closing costs solvency plus waiting, the ledger form of LPO = WLPO+$\mathrm{MP}_{\mathrm{bin}}$; a conserved total costs nothing, and so does every question with a horizon.

Identification
That the nth examination of an account yields its nth post is the examination model, a reading of what observing an account means.

Not claimed
Which history nature posts. The coding histories in the proof are devices showing that the ledger's own histories contain every binary world.

Proof. Solvency is $\forall n\,(b_h(n) \geq 0)$, closing is $\exists n\,(n > 0 \wedge b_h(n) = 0)$ and overdraft is $\exists n\,(b_h(n) < 0)$, each with a decidable test, since $b_h(n)$ is computed from the first n posts. Two coding histories give the reductions. The *solvency coder* of a binary sequence α posts $d\,c$ for each entry $\alpha(n) = 0$ and $c\,d$ for each entry $\alpha(n) = 1$. Its balance is 0 after every pair, +1 inside a pair

for 0, and -1 inside a pair for 1; so it is solvent forever exactly when $\alpha \equiv 0$, and overdrawn exactly when some $\alpha(n) = 1$. The *closing coder* posts one debit and then the solvency coder of α; its balance is 1 after every pair, 2 inside a pair for 0 and 0 inside a pair for 1, so it closes exactly when some $\alpha(n) = 1$. Each post of either coder depends on one entry of α, so both maps have finite information.

Parts (a) and (d) are Lemma 24.38(a) with the solvency coder; (b), (c) and (e) are Lemma 24.38(b) with the closing and solvency coders. For (f), (a)–(c) reduce it to LPO $\Leftrightarrow$ WLPO $\wedge$ $\mathrm{MP}_{\mathrm{bin}}$: LPO gives both principles, and conversely WLPO decides $\alpha \equiv 0$ or $\neg(\alpha \equiv 0)$, and in the second case $\mathrm{MP}_{\mathrm{bin}}$ finds a 1 (Proposition 7.35). For (g), $b_{\bar{h}} = -b_h$ by induction on n; so conservation holds outright, and the conjunction says the same as solvency, priced by (a). For (h), a bounded quantifier over a decidable test is decided by inspecting finitely many cases. For (i), solvency is Π_1 and closing is Σ_1 in the sense of Definition 17.12, and the coders reduce the zero test to solvency and the event test to closing; apply Theorems 17.13 and 17.15. □

A record of an overdraft refutes solvency, and no record confirms it; a record of a closing confirms closing, and no record refutes it. The prices say more: they are what it costs to know the answer for every history at once.

Detectors that revise their verdict have exact prices too. We use the guessing rules of Definition 17.19, writing 1 for *yes* and 0 for *no*.

Definition 24.40 (Sound rules and alarms)**.** The *verdict stream* of a guessing rule g on a world w is $\sigma(n) = g(w{\restriction}n)$. The stream *changes* at $n+1$ when $\sigma(n+1) \neq \sigma(n)$, and it *settles* on b when $\sigma(n) = b$ for all large n. The rule is *sound* for a question Q when, on every world, verdicts that settle on 1 imply $Q(w)$ and verdicts that settle on 0 imply $\neg Q(w)$. A stream is an *alarm from* b when $\sigma(0) = b$ and it changes at most once. A *refuting alarm* is a rule whose stream is an alarm from 1 on every world, a *confirming alarm* one whose stream is an alarm from 0 on every world.

An alarm latches: once it has left its starting value it never returns, for a return would be a second change. A refuting alarm trips to refute; a confirming alarm trips to confirm.

Theorem 24.41 (The alarm table)**.** *The following equivalences and implications hold for all sets of worlds, examinations, guessing rules and questions, with no choice principle.*

(a) Every question with a sound refuting alarm is decided if and only if WLPO holds.

(b) Every question with a sound confirming alarm is decided if and only if LPO holds.

(c) Every question with a sound rule whose verdict stream, on each world, changes at most some number of times (a bound depending on the world and not given as data) is decided if and only if LPO holds.

(d) In particular, if g is sound for Q and each world's verdict stream has a bound on its changes, then LPO decides Q on every world.

(e) The statement "LPO decides every question that some guessing rule decides in the limit" is equivalent to the statement that LPO implies $\neg\neg P \to P$ for every proposition P.

(f) Further: $\neg\neg Q \to Q$ for every question with a sound confirming alarm, and the trip of every alarm from 0 that is not silent forever, are each equivalent to $\mathrm{MP}_{\mathrm{bin}}$; whether an alarm from b stays at b forever, for every alarm from b, is equivalent to WLPO, for either b; $Q(w) \vee R(w)$ on every world, for sound refuting alarms of Q and of R that never both trip on one world, is equivalent to LLPO; that every alarm from b settles, for either b, and that every stream with a bound on its changes settles, are each equivalent to LPO; and with a bound B on the changes fixed in advance for all worlds, sound rules decide their questions with no principle when $B = 0$, while for every $B \geq 1$ this is equivalent to LPO.

Conditions and scope.

Base and premises

Arbitrary worlds, examinations, guessing rules and questions; implications between decisions and principles, proved without excluded middle and without choice.

Conclusion

WLPO, LPO and LLPO are each exactly the cost of reading an end-of-time property of latching detectors: one alarm's silence, its trip or final verdict, which of two never-both-tripping alarms is silent. $\mathrm{MP}_{\mathrm{bin}}$ is not a reading but a wait. A refuting alarm costs WLPO, a confirming one LPO, and no question with a sound detector that revises finitely often costs more than LPO.

Not claimed

The ceiling (e) as a theorem: it holds exactly when LPO gives double-negation elimination for every proposition, a statement far beyond the binary sequences that LPO speaks about. The proved ceiling is (d).

Proof. Take W to be the binary runs, each read by itself, for the directions that yield a principle.

(a) Given WLPO and a sound refuting alarm on w, apply WLPO to $n \mapsto 1 - \sigma(n)$. If $\sigma \equiv 1$, the verdicts settle on 1 and $Q(w)$. Otherwise suppose $Q(w)$: any n with $\sigma(n) = 0$ would latch the stream at 0, settling it on 0 and refuting $Q(w)$; so $\sigma \equiv 1$, a contradiction, and $\neg Q(w)$. Conversely, for $Q(\alpha) \Leftrightarrow \alpha \equiv 0$ the rule "1 while no entry 1 has been seen" is a sound refuting alarm, and deciding Q is WLPO.

(b) Given LPO, apply it to σ: an n with $\sigma(n) = 1$ latches the stream at 1 and gives $Q(w)$; $\sigma \equiv 0$ settles on 0 and gives $\neg Q(w)$. Conversely "1 once an entry 1 has been seen" is a sound confirming alarm for $\exists n\, \alpha(n) = 1$.

(c) and (d). An alarm has at most one change, so (c) contains (b) and gives LPO. Given LPO and a bound on the changes of σ, prove, by induction on r, that if σ changes at most r times after place k then $Q(w)$ or $\neg Q(w)$. Apply LPO to $j \mapsto [\sigma(k + j) \neq \sigma(k)]$. If it is identically 0, the verdicts settle on $\sigma(k)$ and soundness decides $Q(w)$. Otherwise σ changes between k and some $k + j$, so it changes at most $r - 1$ times after $k + j$, and the induction hypothesis applies; for $r = 0$ the first case is the only one. The bound need not be given as data: the goal is a disjunction, so its existence suffices.

(e) Suppose the ceiling, and let LPO hold and $\neg\neg P$. Take one world and the question $Q = P$, with the rule that always says 1. It decides Q in the limit: if P, it settles on 1, and $\neg P$ is

impossible. So $P \vee \neg P$, and with $\neg\neg P$ this gives P. Conversely, if LPO yields double-negation elimination, apply it to $\neg\neg(Q(w) \vee \neg Q(w))$, which holds outright.

(f) For a sound confirming alarm with $\neg\neg Q(w)$, its stream cannot be identically 0, which would settle on 0; $\mathrm{MP}_{\mathrm{bin}}$ finds a trip, which latches and gives $Q(w)$. Waiting for a trip is $\mathrm{MP}_{\mathrm{bin}}$ applied to the stream. Conversely the stream "some entry 1 seen before place n" of a binary sequence α is an alarm from 0, silent forever exactly when $\alpha \equiv 0$; it yields $\mathrm{MP}_{\mathrm{bin}}$, and the rule of (b) does the same for stability. Silence of an alarm from b is WLPO applied to $n \mapsto [\sigma(n) \neq b]$, and conversely the stream that starts at b and switches once an entry 1 of α has been seen is silent exactly when $\alpha \equiv 0$. The two-alarm statement is the argument of Lemma 24.38(a) for the interleaved trips, with the silence-so-far alarms of the even and odd parts of α for the converse. An alarm from b settles on b when it never trips and on the other value when it does, so LPO applied to its trips settles it; conversely the switching stream of α settles on b exactly when $\alpha \equiv 0$, and settles on the other value only after an entry 1, which a bounded search then finds. Bounded streams settle by the induction of (c). With $B = 0$ fixed in advance, the verdict is the constant g of the empty record, which settles, and soundness decides Q; for $B \geq 1$ the induction of (c) uses LPO, and the confirming alarm of (b) has $B = 1$. □

Finally, deadlines. Section 17.7 showed, classically, that a question about binary runs settled by a recorder is fixed by a prefix of bounded length (Theorem 17.16). What does that bound cost, and when does a bound fixed in advance exist? A recorder in the sense of Definition 17.12 *announces* b at stage n on a run α when its answer at $\alpha{\restriction}n$ is b and it reads on at every shorter record. It *always answers* when it announces at some stage on every binary run, and it has a *deadline* N when it announces at a stage $n \leq N$ on every binary run.

Definition 24.42 (Finite-memory recorders)**.** A *recorder with $s+1$ memory states* has states $\{0, \dots, s\}$, an initial state, a transition map from a state and a bit to a state, and an output map from states to *yes*, *no* or "read on"; its answer at a record is the output of the state reached by reading the record from the initial state.

Theorem 24.43 (A deadline costs the fan principle)**.**

(a) "Every recorder that always answers on binary runs has a deadline" is equivalent to the decidable fan theorem FAN_{Δ} *(Definition 13.2).*

(b) There is a computable recorder that announces on every computable binary run while, for every N, some computable run leaves it reading on through stage N. So the computable form of the deadline principle is false.

(c) "Every question about binary runs settled by a recorder is fixed by the first N examinations, for some N" is equivalent to $\mathrm{FAN}_{\Delta} \wedge \mathrm{MP}_{\mathrm{bin}}$*. For questions settled by a recorder that always answers, the same statement is equivalent to* FAN_{Δ}*; for recorders that only ever answer* yes *and cannot fail to answer on any run, the statement that runs agreeing up to some N receive the same answer is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$*.*

(d) A recorder with $s+1$ memory states that always answers announces at a stage at most s on every run, and for every s some recorder with $s+1$ states announces yes *exactly at stage s on every run. If such a recorder announces on every computable run, it announces at a stage at most s on every run.*

Parts (a), (c) and the first two statements of (d) are proved with no choice principle and no omniscience principle beyond those named. Part (b) and the last statement of (d) concern computable runs and are proved classically.

Conditions and scope.

Base and premises

Binary runs, recorders as rules on finite records, and the decidable fan theorem of Chapter 13.

Conclusion

A deadline fixed in advance for every always-answering experiment costs exactly what Heine–Borel for $[0, 1]$ costs (Theorem 13.6); the classical passage from "settled" to "fixed by a prefix" costs $\mathrm{FAN}_\Delta \wedge \mathrm{MP}_{\mathrm{bin}}$, the fan principle paying for experiments that always answer and Markov's principle for those that only ever say yes. Experiments with finite memory carry their deadline with them, and the bound is sharp.

Identification

A run is a system under repeated binary examination, a recorder an experiment with a stopping rule, and a deadline a number of examinations fixed before the run.

Not claimed

Whether FAN_Δ alone gives the passage of (c); by (c) it does exactly when FAN_Δ implies $\mathrm{MP}_{\mathrm{bin}}$, a question about the two principles alone that we leave open.

Proof. (a) Let R always answer. The strings at which R answers form a decidable bar, so FAN_Δ gives N such that every run has an answering record of length at most N; a search through the shorter records finds the first answer, at a stage at most N. Conversely, for a decidable bar D, the recorder that answers *yes* at the strings of D always answers, and a deadline for it is a uniform bound for D.

(b) Let T_K be Kleene's tree (Lemma 16.29), and let R read on at the members of T_K and answer *yes* elsewhere. Membership is decided by a program, so R is computable. Every computable run leaves T_K, so R announces on it. For each N, T_K has a member g_N of length N; the run that begins with g_N and continues with zeros is computable, and its records of length at most N are initial segments of g_N, members of T_K, where R reads on.

(c) Suppose FAN_Δ and $\mathrm{MP}_{\mathrm{bin}}$, and let R settle Q. On each run R cannot fail to answer: with no answer, neither Q nor $\neg Q$ could hold. So the binary sequence "R has answered by stage n" is not identically 0, and $\mathrm{MP}_{\mathrm{bin}}$ finds a stage: R always answers. By (a) it has a deadline N, and runs that agree up to N receive the same first answer, hence agree on Q. If R always answers, FAN_Δ alone suffices.

For FAN_Δ from the forward statement, let D be a decidable bar, and let the question ask for the value $\alpha(k)$ of a run at its first hit k, the least k with $\bar{\alpha}k \in D$. The recorder that announces $\alpha(k)$ at stage $k + 1$ always answers and settles this question. Let N fix it. If some run had its first hit at $k \geq N$, flipping its entry k would keep its first N entries and its first hit and change the answer. So every first hit comes before N, and N is a uniform bound.

For $\mathrm{MP}_{\mathrm{bin}}$, let ξ be a binary sequence that is not identically 0, and let $Q(\alpha)$ say that ξ or α has an entry 1. The recorder that answers *yes* at a record of length n once ξ or the record shows a 1 below n, and reads on otherwise, confirms Q; it never answers *no*, and $\neg Q(\alpha)$ never holds, because it would give $\xi \equiv 0$. So it settles Q, it only answers *yes*, and it cannot fail to answer. Let N fix Q, or the answer. The run of zeros agrees up to N with the run whose only 1 is at N, for which Q holds. So Q holds for the run of zeros, and ξ has an entry 1. Conversely, given $\mathrm{MP}_{\mathrm{bin}}$, a recorder that only answers *yes* and cannot fail to answer does answer on every run, so all runs receive the same answer and $N = 0$ serves.

(d) Let a recorder with $s + 1$ states always answer, and let α be a run. Whether it reads on at every stage up to s is decided by inspection; suppose it does. Among the $s + 2$ states reached after $0, 1, \dots, s + 1$ entries, two coincide, at stages $i < j \leq s + 1$. The run that copies α up to place j and then repeats the block of places i to $j - 1$ forever passes only through the states of stages $0, \dots, j - 1$, all of them reading on, so it never answers, a contradiction. Hence the recorder answers at some stage at most s. The counter that moves from state q to $\min(q + 1, s)$ and answers *yes* exactly in state s announces at stage s on every run. The repeating runs used are eventually periodic, hence computable, which gives the last statement. □

24.11 From records to time, quantum, matter and space

How far does a post reach toward a clock that returns, complex amplitudes read by a square law, conserved charges, and a space of cells with a speed limit? We go one step at a time. Either posting decides the step, and we prove the answer, or the step is a premise: we name it, every result above it carries it, and where we can we exhibit two models that agree below the step and differ on it. Results we call free are proved with no omniscience principle and no choice principle; the others in classical mathematics, and we do not claim them on $\mathbb{R}_\delta$. Counts are integers here, so that a count can also be lowered.

A post flips one bit

For a ledger L of d accounts (Definition 24.1) let $x(L) = (D_k - C_k)_k \in \mathbb{Z}^d$ be its balance vector and $p(L) \in \{0, 1\}^d$ its parity pattern, $p_k(L) = x_k(L) \bmod 2$. Two patterns are *adjacent* when they differ in exactly one bit, as along an edge of the d-cube. For $d = 3$ the *weight* $|x|$ of a pattern is its number of ones, e_0, e_1, e_2 are the patterns of weight one, and a *signed permutation* is a map $x \mapsto (x_{\pi(0)} \oplus t_0,\ x_{\pi(1)} \oplus t_1,\ x_{\pi(2)} \oplus t_2)$ with π a permutation of $\{0, 1, 2\}$, $t \in \{0, 1\}^3$ and $\oplus$ addition modulo 2.

Proposition 24.44 (One post flips one bit)**.** *Let L be a ledger of d accounts.*

(a) A debit at account k adds the unit vector e_k to $x(L)$ and a credit subtracts it; no other balance changes. From every ledger, each of the $2d$ unit steps $\pm e_k$ of $\mathbb{Z}^d$ is the change of x made by a post.

(b) A post at account k changes bit k of the parity pattern and no other bit.

(c) For every pattern q, some post takes L to a ledger with pattern q if and only if q is adjacent to $p(L)$.

(d) Let $d = 3$. The bijections f of $\{0,1\}^3$ for which $f(x), f(y)$ are adjacent exactly when x, y are adjacent are the 48 *signed permutations, and they carry every pattern to every pattern. Exactly six of them fix* 000*, the six permutations of the accounts. A signed permutation fixes* 000 *if and only if it keeps the weight of every pattern, and two patterns are carried to each other by one of the six if and only if they have equal weight; the four weight classes have* 1*,* 3*,* 3 *and* 1 *members.*

Proof. (a), (b) A debit at k raises D_k by one and changes no other count, so x_k rises by one and its parity changes; a credit lowers x_k by one. The step $+e_k$ is a debit at k, and $-e_k$ a credit. (c) follows from (b): if q differs from $p(L)$ in bit k only, a debit at k reaches it.

(d) A signed permutation is a bijection and keeps the number of places in which two patterns differ, so it preserves adjacency both ways. It sends 000 to t, and the images of e_0, e_1, e_2, shifted by t, give π; so the $6 \cdot 8 = 48$ signed permutations are distinct, and the maps $x \mapsto x \oplus t$ carry 000 anywhere. Let f preserve adjacency both ways, put $t = f(000)$ and $g(x) = f(x) \oplus t$. Then g fixes 000 and keeps the length of a shortest path from 000, which is the weight, so $g(e_i) = e_{\sigma(i)}$ for a permutation σ. As $e_i \oplus e_j$ is the only pattern of weight two adjacent to both e_i and e_j, $g(e_i \oplus e_j) = e_{\sigma(i)} \oplus e_{\sigma(j)}$, and $g(111) = 111$. So g moves the ones of every pattern by σ, and f is a signed permutation; it fixes 000 exactly when $t = 000$, which leaves the six permutations. With $t \neq 000$ the image of 000 has positive weight, while permutations keep weight and carry any pattern with k ones to any other; there are $\binom{3}{k}$ of them. □

The proposition is free. Eight, here, is a count of patterns; that a recognizer passes through all eight in turn and returns is the premise of Theorem 24.47. Reading the weight as the cost of a state charges the three accounts equally, and that is a premise; taking one account per axis of space is an identification.

Proposition 24.45 (Symmetric charges count odd accounts)**.** *Let G be a commutative monoid, written additively. Call $x, y \in \{0,1\}^3$* independent *when no bit is* 1 *in both, and let $x \vee y$ be their bitwise union. A map $Q : \{0,1\}^3 \to G$ is an* additive charge *when $Q(000) = 0$ and $Q(x \vee y) = Q(x) + Q(y)$ for all independent x, y.*

(a) Q is an additive charge if and only if $Q(x) = x_0 q_0 + x_1 q_1 + x_2 q_2$ for some $q_0, q_1, q_2 \in G$; then $q_i = Q(e_i)$.

(b) An additive charge is unchanged by the six permutations of the accounts if and only if $q_0 = q_1 = q_2$, and then $Q(x) = |x|\, q_0$. In that case $Q(111) = 3Q(x)$ when $|x| = 1$ and $2Q(111) = 3Q(x)$ when $|x| = 2$; for an integer charge, 3 *divides* $Q(111)$.

Proof. (a) A pattern is the union of the pairwise independent e_i with $x_i = 1$; conversely the bits of independent patterns add. (b) The transposition of accounts i and j sends e_i to e_j; conversely $|x| q_0$ is invariant, and $|111| = 3$. □

The proposition is free. The pattern of a ledger marks its accounts with odd balance, so a symmetric additive charge counts the odd accounts. The thirds are arithmetic; we identify these charges with no physical charge.

A process returns exactly when its item time is uniform

Definition 24.46 (Processes and items)**.** A *process* on d accounts is a sequence $w = (w_n)_{n\geq 0}$ in $\{0,1\}^d$. The pass of T ticks from n, $(w_n, \dots, w_{n+T-1})$, is *read as one item* when it contains every pattern. The process *advances by finest changes* when w_n and w_{n+1} are adjacent for every n; it *completes* when from every n some pass is read as one item; it is *state-local* when $w_i = w_j$ implies $w_{i+1} = w_{j+1}$. It has a *uniform item time* L when for every n the pass of L ticks from n is read as one item and no shorter pass from n is.

Theorem 24.47 (A process returns after 2^d ticks exactly when its item time is uniform)**.** *For a process w on d accounts the following are equivalent: (i) w has a uniform item time; (ii) every pass of 2^d ticks is read as one item; (iii) $w_{n+2^d} = w_n$ for every n, and $(w_0, \dots, w_{2^d-1})$ contains every pattern; (iv) w completes and is state-local. Then the item time is 2^d and the least period of w is 2^d. The Gray process, which repeats the patterns of three accounts in the order $0, 1, 3, 2, 6, 7, 5, 4$ (a pattern read as $x_0 + 2x_1 + 4x_2$), advances by finest changes and meets these conditions. The process on two accounts that repeats $00, 01, 11, 10, 11, 01, 00, 10$ advances by finest changes and reads every pass of four ticks from a multiple of four as one item, and meets none of them.*

Proof. A pass read as one item has at least 2^d ticks, and a pass of 2^d ticks is read as one item exactly when it repeats no pattern. (i)⇒(ii): let the item time be L. If $w_n = w_{n+j}$ with $1 \leq j < L$, the pass of $L-1$ ticks from $n+1$ would contain every pattern, which the item time from $n+1$ forbids. So windows of L ticks repeat no pattern, $L \leq 2^d$, and $L = 2^d$. (ii)⇒(iii): w_n does not recur among $w_{n+1}, \dots, w_{n+2^d-1}$, and the pass from $n+1$ contains it, so $w_{n+2^d} = w_n$. (iii)⇒(ii): the pass from n is a cyclic shift of the first. (ii)⇒(i) with $L = 2^d$. (iii)⇒(iv): the first pass repeats no pattern, so $w_i = w_j$ gives $i \equiv j \pmod{2^d}$ and $w_{i+1} = w_{j+1}$. (iv)⇒(iii): state-locality gives a map s with $w_{n+1} = s(w_n)$. Completion from 1 makes w_0 recur; let $m > 0$ be least with $w_m = w_0$. Then w has period m, and $w_0, \dots, w_{m-1}$ are distinct (a repeat $w_i = w_j$, $i < j < m$, would give $w_{i+m-j} = w_0$). By completion they are all 2^d patterns, so $m = 2^d$. A period $p < 2^d$ would repeat a pattern inside the first pass. The Gray order changes one bit per step, also from 4 back to 0, and lists all eight patterns. In the two-account process each block of four contains every pattern, and $w_4 = 11 \neq 00 = w_0$. □

Conditions and scope.

Base and premises

Sequences of patterns; finite counting, in classical mathematics.

Conclusion

A uniform item time, return after 2^d ticks with a complete first pass, and completion with a next state fixed by the present one are one premise in four forms; under it the period is 2^d, eight for three accounts.

Not claimed

That a recognizer's ticks have a uniform item time; the two-account process shows posting does not force it.

The finest changes, and the ones that count

A change $L \to L'$ of a ledger of d accounts is *finest* when $\sum_k |D'_k - D_k| + \sum_k |C'_k - C_k| = 1$. The *unrecord* at account k on a side lowers that count by one, and the *record total* is $|L| = \sum_k (D_k + C_k)$. For a finite set V of recognizers, each v with its own ledger M_v, a *joint tick* $M \to M'$ is a change in which every recognizer either holds or makes a finest change raising its own record total by one; it is *serial* when exactly one recognizer changes.

Theorem 24.48 (The finest changes are posts and unrecords, and only posts count)**.** *Let L, L' be ledgers of d accounts.*

(a) $L \to L'$ is finest if and only if L' is a post or an unrecord of L. Every unrecord has step cost $G(1)$, by the formula of Definition 24.18, as a debit does; posts and unrecords each change exactly one bit of the parity pattern, and each is a bijection of the set of ledgers.

(b) A finest change raises the record total by one if and only if it is a post, and then it lowers no count. A change is monotone with $\sum_k |D'_k - D_k| + \sum_k |C'_k - C_k| = 1$ if and only if it is finest and raises the record total by one.

(c) After a list of h finest changes the record total has risen by exactly h if and only if every change in the list is a post; then no count was lowered.

(d) The joint ticks are exactly the changes in which each recognizer holds or posts one unit. If m recognizers post, the joint total $\sum_v |M_v|$ rises by m, and the individual unit counts sum to m. So a joint tick raises the joint total by one, and is a finest change of the joint ledger, exactly when it is serial; over a history of joint ticks in each of which someone posts, the joint total rises by the number of ticks exactly when every tick is serial.

Proof. (a) A sum of absolute values of integer differences is 1 exactly when one difference is ± 1 and the others vanish. A change of one count by ± 1 costs $G(\pm 1) = G(1)$, moves one balance by ± 1 and so one parity bit, and is inverted by the opposite change. (b) A post raises the total by one, an unrecord lowers it by one, and a monotone change of unit size is a post. (c) Each change moves the total by ± 1. (d) By (b) a recognizer that does not hold posts; each poster adds one to its total and to its unit count. Over a history the joint total rises by the sum of the numbers of posters, each at least one, which equals the number of ticks exactly when each is one. □

The theorem is proved by integer arithmetic and claimed in classical mathematics. Counting occurrences selects the posts among the finest changes, and a joint record counts events, not ticks, unless the ticks are serial. That nature's finest changes are posts is the posting premise itself, read on counts (the count reading); the ledgers of Proposition 24.44 change by posts, and cost, the one-bit flip and invertibility do not tell a post from an unrecord. Seriality is a premise: one poster and two posters both make joint ticks.

Two ticks are the imaginary unit

A *tick register* is a real vector space with a linear map T, the tick. A recognizer whose state is a function on its eight patterns has the tick register $\mathbb{R}^8$ of functions on $\mathbb{Z}/8$ with the shift $(Sf)(t) = f(t+1)$, and its *core* is $\mathscr{C} = \{f : f(t+4) = -f(t) \text{ for all } t\}$.

Theorem 24.49 (Four ticks reverse, and two ticks are i). *Let T be a linear map of a vector space V over a field in which $2 \neq 0$, with $T^8 = 1$.*

(a) $T^4 v = -v$ for every v if and only if no nonzero v has $T^n v = v$ with $0 < n < 8$.

(b) If $T^4 = -1$, then $(T^2)^2 = -1$. Over $\mathbb{R}$, V is then a complex vector space with i acting as T^2, T is complex linear, and every v is uniquely $a + b$ with $Ta = \zeta a$ and $Tb = -\zeta b$, where $\zeta = (1 + i)/\sqrt{2}$, $\zeta^2 = i$ and $\zeta^8 = 1$.

(c) On $\mathbb{R}^8$, $S^4 f = -f$ exactly when $f \in \mathscr{C}$; a subspace invariant under S has no nonzero state returning before eight ticks exactly when it lies in $\mathscr{C}$; no linear J on $\mathbb{R}^8$ with $J^2 = -1$ commutes with S; and for $k < 8$, S^k squares to -1 on $\mathscr{C}$ exactly when $k = 2$ or $k = 6$.

Proof. (a) Let $T^4 = -1$ and $T^n v = v$ with $0 < n < 8$. Some multiple jn is congruent to $g = \gcd(n, 8)$ modulo 8, so $T^g v = v$; as g divides 4, $T^4 v = v = -v$, and $v = 0$. Conversely $w = v + T^4 v$ has $T^4 w = w$, so $w = 0$. (b) $(T^2)^2 = T^4 = -1$, so $(a + bi)v = av + bT^2 v$ makes V complex, and T commutes with T^2. As operators $\zeta^2 = i = T^2$, so $(T - \zeta)(T + \zeta) = 0$; then $a = (T + \zeta)v/(2\zeta)$ and $b = -(T - \zeta)v/(2\zeta)$ split v. If $a + b = 0$, then $\zeta a = Ta = -Tb = \zeta b = -\zeta a$, so $a = b = 0$. (c) $(S^4 f)(t) = f(t + 4)$; the second claim is (a) on the subspace. If J commutes with S, then $J\mathbf{1}$ is fixed by S, so $J\mathbf{1} = c\mathbf{1}$ with c real and $J^2\mathbf{1} = c^2\mathbf{1} \neq -\mathbf{1}$. On $\mathscr{C}$, $S^4 = -1$, and S^2 is not ± 1 (it sends $e_0 - e_4$ to $e_6 - e_2$), so $S^{2k} = -1$ exactly when $2k \equiv 4 \pmod 8$. □

The theorem is proved in classical mathematics, the identity $(T^2)^2 = -1$ with no choice principle. That a recognizer's state lives on this tick register is a premise, and so is the half turn: the shift fixes the constants of $\mathbb{R}^8$ and reverses every state of $\mathscr{C}$. Which complex structure a recognizer uses is not decided either: S^2 and S^6 both square to -1 on the core.

Proposition 24.50 (Pairs of recognizers and the sign of the pass).

(a) Let T_1, T_2 be commuting linear maps of a space U over a field in which $2 \neq 0$, with $T_1^4 = T_2^4 = -1$. Then $(T_1 T_2)^4 = 1$. Every v is uniquely the sum of $\frac{1}{2}(v - T_1^2 T_2^2 v)$ in $\ker(T_1^2 - T_2^2)$, where the imaginary units agree, and $\frac{1}{2}(v + T_1^2 T_2^2 v)$ in $\ker(T_1^2 + T_2^2)$, where they are opposite; $(T_1 T_2)^2$ is -1 on the first and $+1$ on the second.

(b) For signs $s(t) \in \{\pm 1\}$ let $(T_s f)(t) = s(t) f(t + 1)$ on $\mathbb{R}^8$. Then $T_s^8 = (\prod_t s(t)) \cdot 1$. Conjugating by multiplication with signs $g(t)$ replaces $s(t)$ by $g(t)s(t)g(t + 1)$, which keeps $\prod s$; for $g(t) = s(0) \cdots s(t - 1)$ the new signs are 1 except at $t = 7$, where the sign is $\prod s$. Eight ticks reverse every state exactly when $\prod s = -1$.

Proof. (a) With $A = T_1^2$, $B = T_2^2$: $A^2 = B^2 = -1$, $AB = BA$, $(T_1 T_2)^4 = A^2 B^2 = 1$, $(A - B)(1 - AB) = A - A^2 B - B + AB^2 = 0$ and likewise $(A + B)(1 + AB) = 0$. If $Av = Bv$ and $Av = -Bv$, then $Bv = 0$ and $v = -B^2 v = 0$. On the first kernel $ABv = A^2 v = -v$, on the second $ABv = -A^2 v = v$. (b) $(T_s^8 f)(t) = s(t) \cdots s(t + 7) f(t + 8)$; the product of $g(t)s(t)g(t + 1)$ is $\prod s \prod g^2$; and the cumulative g has $g(t + 1) = g(t)s(t)$ for $t < 7$. □

So two half-turn recognizers ticking together split their joint tick register into a part where their imaginary units agree, on which two joint ticks act as -1, and a part where the units

are opposite and two joint ticks act as +1. Up to sign convention there are two eight-tick passes, the plain one and the twisted one, which reverses every state after a full pass as a spin one-half state does under a full turn (an identification); both are passes on the same eight states, so the sign is a premise. Whether the exchange sign of two identical recognizers is tied to the twist, as in the spin-statistics connection, is open. The identity $(T_1T_2)^4 = 1$ and the values ∓ 1 of $(T_1T_2)^2$ are free; the rest is proved in classical mathematics.

The square law from phase blindness

In the complex structure $i = S^2$ of the core, multiplication by $\zeta = (1+i)/\sqrt{2}$ is $(Pf)(t) = \big(f(t) + f(t+2)\big)/\sqrt{2}$. What can a read of a tick register state be?

Definition 24.51 (Reads and the five read premises)**.** A *read* assigns a real number $\mu(f, A)$ to every state $f \in \mathbb{R}^8$ and every set $A \subseteq \mathbb{Z}/8$ of positions. The *read premises* are: (R1) there is one $\varphi : \mathbb{R} \to \mathbb{R}$ with $\mu(f, \{t\}) = \varphi(f(t))$ for all f and t; (R2) $\mu(f, A) = \sum_{t \in A} \mu(f, \{t\})$; (R3) $\varphi(0) = 0$; (R4) $|x| \le |y|$ implies $\varphi(x) \le \varphi(y)$; (R5) $\mu(Pf, \mathbb{Z}/8) = \mu(f, \mathbb{Z}/8)$ for every $f \in \mathscr{C}$: the total read of a core state does not see the phase made of ticks.

Lemma 24.52 (Monotone quadratic functions are squares)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ satisfy $g(0) = 0$, $g(x+y) + g(x-y) = 2g(x) + 2g(y)$ for all real x, y, and $g(x) \le g(y)$ whenever $|x| \le |y|$. Then $g(x) = g(1)x^2$ for every real x.*

Proof. Monotonicity gives $g(-x) = g(x)$ and $g(1) \ge g(0) = 0$. Putting nx and x for x and y gives $g((n+1)x) = 2g(nx) + 2g(x) - g((n-1)x)$, so $g(nx) = n^2 g(x)$ by induction, $g(x/m) = g(x)/m^2$, and $g(q) = q^2 g(1)$ for rational $q \ge 0$. For real $x \ge 0$ and rationals $0 \le p \le x \le q$, $p^2 g(1) \le g(x) \le q^2 g(1)$; let $p, q \to x$. □

The equation is the parallelogram law of Jordan and von Neumann [143]; some regularity is needed, since $g = A^2$ solves it for every additive, possibly nonlinear, A [97].

Theorem 24.53 (The square law from phase blindness)**.** *If a read meets (R1) to (R5), then for every state $f \in \mathbb{R}^8$ and every set A of positions*

$$\mu(f, A) = \varphi(1) \sum_{t \in A} f(t)^2.$$

The square reads, $\varphi(x) = x^2$, meet (R1) to (R5). The magnitude reads, $\mu(f, A) = \sum_{t \in A} |f(t)|$, meet (R1) to (R4) and fail (R5).

Proof. By (R4), φ is even. For real a, b the core state $f_{a,b} = a(e_0 - e_4) + b(e_2 - e_6)$ has entries $a, b, -a, -b$ at $0, 2, 4, 6$ and 0 elsewhere, so its total read is $2\varphi(a) + 2\varphi(b)$, and $Pf_{a,b} = f_{(a+b)/\sqrt{2},\,(b-a)/\sqrt{2}}$. So (R5) gives $\varphi(a) + \varphi(b) = \varphi\big(\frac{a+b}{\sqrt{2}}\big) + \varphi\big(\frac{a-b}{\sqrt{2}}\big)$. With $b = 0$ this is $\varphi(\sqrt{2}u) = 2\varphi(u)$; with $a = x+y$, $b = x-y$ it becomes $\varphi(x+y)+\varphi(x-y) = \varphi(\sqrt{2}x)+\varphi(\sqrt{2}y) = 2\varphi(x)+2\varphi(y)$. Lemma 24.52 and (R1), (R2) give the formula, with $\varphi(1) \ge \varphi(0) = 0$. The square reads meet (R5) because $\sum_t (Pf)(t)^2 = \sum_t f(t)^2 + \sum_t f(t)f(t+2)$, and on $\mathscr{C}$ the last sum is its own negative under $t \mapsto t+2$. For the magnitude reads, $f_{1,0}$ has total 2 and $Pf_{1,0}$ has total $2\sqrt{2}$. □

Conditions and scope.

Base and premises

The classical real numbers; the read premises of Definition 24.51, with (R5) for the unit $i = S^2$; in classical mathematics. Applied to a recognizer, the theorem also needs its state to live on $\mathbb{R}^8$ with the shift.

Conclusion

A read that depends on the amplitude at each position by one rule, adds over positions, vanishes on empty positions, grows with the magnitude and does not see the phase made of ticks is proportional to the squared amplitude, the form of Born's rule [139]. No normalization, homogeneity or continuity is assumed.

Not claimed

A probability: no ratio form is proved, and $\varphi(1)$ is not fixed ($\varphi = 0$ meets every premise). The read premises remain premises.

Proposition 24.54 (Two recognizers: the square law and product records)**.** *On functions F on $(\mathbb{Z}/8)^2$ let $\mathscr{C}_2 = \{F : F(s{+}4,t) = -F(s,t) = F(s,t{+}4)\}$ and $(P_1F)(s,t) = \big(F(s,t){+}F(s{+}2,t)\big)/\sqrt{2}$, the first recognizer's phase. Let a read μ of $\mathscr{C}_2$ on sets of pairs of positions meet (R1) to (R4) for single pairs and (R5) in the form $\mu(P_1F, all) = \mu(F, all)$. Then $\mu(F, A) = \varphi(1)\sum_{p\in A} F(p)^2$ for every $F \in \mathscr{C}_2$, and for $f, g \in \mathscr{C}$, the product state $F(s,t) = f(s)g(t)$ and all sets A_1, A_2,*

$$\mu(F, A_1 \times A_2)\,\mu(F, all) = \mu(F, A_1 \times all)\,\mu(F, all \times A_2).$$

Proof. The product of $f_{a,b}$ with $e_0 - e_4$ has entries $\pm a$ at four pairs and $\pm b$ at four, total read $4\varphi(a) + 4\varphi(b)$, and P_1 acts on it as P on $f_{a,b}$; the proof of Theorem 24.53 applies. For a product state $\sum_{A_1\times A_2} f(s)^2 g(t)^2 = \sum_{A_1} f^2 \sum_{A_2} g^2$. □

Reading $\mathscr{C}_2$ as the joint tick register of two recognizers is a premise. Independence holds in this cross-multiplied form, for real product states, in classical mathematics; we do not prove that recognizers compose by the complex tensor product, nor independence at single positions for complex states without entanglement.

The premise (R1), that the read of a position depends only on the amplitude there, is the one the square law still rests on, and (R2) and (R5) do not imply it (Proposition 24.55). Such noncontextual reads also start Gleason's theorem, which needs complex dimension at least three [141]; the core has two.

Proposition 24.55 (A contextual read)**.** *For $f \in \mathscr{C}$ put $\|f\|^2 = \sum_t f(t)^2$ and, for $f \neq 0$,*

$$\kappa(f, \{t\}) = f(t)^2 + \frac{f(t)^2\big(f(t+1)^2 - f(t-1)^2\big)}{2\|f\|^2},$$

with $\kappa(0, \{t\}) = 0$, and $\kappa(f, A) = \sum_{t\in A} \kappa(f, \{t\})$. Then κ vanishes at positions where $f(t) = 0$, is nonnegative, and has total $\|f\|^2$, so it meets (R2) and (R5). No φ satisfies $\kappa(f, \{t\}) = \varphi(f(t))$ for all $f \in \mathscr{C}$ and t.

Proof. Since $f(t-1)^2 \le \|f\|^2$, $\kappa(f,\{t\}) \ge f(t)^2/2$. In the total the terms $f(t)^2 f(t+1)^2$ and $f(t)^2 f(t-1)^2$ cancel under $t \mapsto t+1$, and P keeps $\|f\|^2$ on $\mathscr{C}$ (proof of Theorem 24.53). Finally $e_0 - e_4$ and $e_0 + e_1 - e_4 - e_5$ have amplitude 1 at position 0, where κ reads 1 and $\frac{9}{8}$. □

Read across states, (R4) says that $|f(t)| \le |g(s)|$ implies $\mu(f,\{t\}) \le \mu(g,\{s\})$, and this is exactly (R1) with (R4): applied both ways to equal amplitudes it gives equal reads, so $\varphi(x) = \mu\big(x(e_0 - e_4),\{0\}\big)$ meets (R1) and (R4); the converse is immediate. The contextual read fails it at position 0. So the square law rests on (R1) with (R4), and the proposition shows that (R2), (R5), vanishing at empty positions and nonnegativity do not imply (R1). The proposition is proved in classical mathematics.

Conserved totals and charges

A *writing rule* on d accounts gives, for accounts $s \ne t$ and an integer n (a comparison of s with t taken n times), an increment $R(s,t,n) \in \mathbb{Z}^d$ of the balance vector x. It is *additive* when $R(s,t,n+m) = R(s,t,n) + R(s,t,m)$, *local* when $R(s,t,n)_k = 0$ for $k \notin \{s,t\}$, and *name-independent* when $R(t,s,n)_k = R(s,t,n)_{\tau(k)}$, τ the transposition of s and t. It *keeps the balance total* $\sum_k x_k$ when $\sum_k R(s,t,n)_k = 0$, and has the *reversal symmetry* when $R(s,t,n) = R(t,s,-n)$, for all $s \ne t$ and n. A *transfer* from t to s is $e_s - e_t$; made by a debit at s and a credit at t, it keeps the balance total and raises the record total $|L|$ by two.

Theorem 24.56 (Noether on the ledger)**.**

(a) An additive local rule has $R(s,t,n) = n(a\,e_s + b\,e_t)$ with $a = R(s,t,1)_s$ and $b = R(s,t,1)_t$, and changes the balance total by $n(a+b)$.

(b) For an additive, local, name-independent rule the following are equivalent: it keeps the balance total; it has the reversal symmetry; every event is an integer multiple of a transfer; $R(s,t,n) = n\,a(s,t)(e_s - e_t)$ for all $s \ne t$ and n.

(c) If $d \ge 2$, the one-sided rule $R(s,t,n) = n\,e_s$ and the two-sided rule $R(s,t,n) = n(e_s - e_t)$ are additive, local and name-independent; the second has the reversal symmetry and keeps the balance total, the first has neither.

Proof. (a) Additivity makes $n \mapsto R(s,t,n)$ a homomorphism from $\mathbb{Z}$, and locality leaves two entries. (b) Name-independence gives $R(t,s,1)_t = a$ and $R(t,s,1)_s = b$, so $R(t,s,-n) = -n(a\,e_t + b\,e_s)$. The reversal symmetry, the kept balance total and being a multiple of $e_s - e_t$ each say $a = -b$, which is the last form. (c) Direct substitution; the one-sided rule changes the balance total by n. □

Conditions and scope.

Base and premises

Integer vectors; additive, local, name-independent writing; in classical mathematics.

Conclusion

One discrete symmetry, reading a comparison the same from either end, is equivalent to one conserved quantity, the balance total, and to double entry: every event is a multiple

of one transfer. It is an equivalence of Noether's type [144] for one symmetry and one quantity.

Not claimed

The reversal symmetry: both rules of (c) meet everything else, so whether an event has a distinguished end is a premise, and with it double entry and conservation.

Above the reversal symmetry every event is a transfer. Let $N \geq 1$ recognizers each have one account per axis, so that accounts are pairs (s, a) of a site s and an axis $a \in \{0, 1, 2\}$, and let their ledgers sit side by side. A *transfer law* E is a set of ordered pairs of accounts between which a unit may move; a linear charge with integer weights q is $\sum_u q(u)x_u$. The law is *axis-covariant* when the six permutations of the axes, applied at every site at once, carry E into itself, and *site-connected* when a property of sites that holds at one end of each transfer exactly when it holds at the other, and holds at one site, holds at all.

Proposition 24.57 (One conserved charge or three, never two)**.** *A linear charge is unchanged by every transfer of E, from every ledger, exactly when $q(u) = q(v)$ for every $(u, v) \in E$. Let E be axis-covariant and site-connected. If every transfer keeps its axis, the conserved charges are exactly those with $q(s, a) = c(a)$: the three axis totals and their combinations. If some transfer changes axis, they are exactly the constant weights: the balance total. Both cases occur for every $N \geq 1$.*

Proof. A transfer along (u, v) changes the charge by $q(u) - q(v)$. Let q be conserved and $((s, a), (s', b)) \in E$; by covariance $q(s, \pi a) = q(s', \pi b)$ for the six permutations π. If $a = b$ this gives $q(s, c) = q(s', c)$ for all c; if $a \neq b$ it gives $q(s, c) = q(s', c')$ for all $c \neq c'$, so all six weights at s and s' are equal. Either way "$q(s, c) = q(s_0, c)$ for all c" passes along transfers and holds at every site, and an axis-changing transfer makes the three weights equal. The converses are immediate. The law of all axis-keeping transfers conserves the total of axis 0; the law of all transfers changes axes. □

The proposition is proved in classical mathematics and carries the reversal symmetry and the side-by-side composition of ledgers as premises. Which case holds depends on whether an exchange moves a unit between axes, a property of the joint law that posting does not decide.

Cells, a speed limit and linking

A *face gluing* of a set V of voxels is a triple of bijections $\sigma_0, \sigma_1, \sigma_2$ of V: the $+k$ face of v is glued to the $-k$ face of $\sigma_k(v)$. It is *connected* when every voxel is reached from every voxel across faces, *flat* when the σ_k commute, and *free* when no nonzero $x \in \mathbb{Z}^3$ fixes a voxel under $\sigma^x = \sigma_0^{x_0}\sigma_1^{x_1}\sigma_2^{x_2}$. Under *local transport* a debit on account k carries a located unit from v to $\sigma_k(v)$ and a credit to $\sigma_k^{-1}(v)$.

Proposition 24.58 (Flat gluings are lattices, and local transport has speed one)**.**

(a) Let a gluing be flat and connected, and $v_0 \in V$. The stabilizer $\Lambda = \{x : \sigma^x v_0 = v_0\}$ is the same for every v_0, and $x + \Lambda \mapsto \sigma^x v_0$ is a bijection $\mathbb{Z}^3/\Lambda \to V$ carrying $x + e_k$ to σ_k of the image of x; Λ is the only subgroup with such a bijection, and V has $[\mathbb{Z}^3 : \Lambda]$ voxels, both sides possibly infinite. A flat, connected, free gluing is $\mathbb{Z}^3$ with unit translations, unique up to a translation.

(b) Four voxels with $\sigma_0 = \sigma_2$ the rotation $(0\,1\,2\,3)$ and σ_1 the reflection $(0\,1)(2\,3)$ form a connected gluing that is not flat.

(c) Under local transport n posts move a unit across at most n faces; on $\mathbb{Z}^3$ at most ℓ^1 distance n, and n debits on one axis move it exactly n. The rule that carries the unit across two faces per post gives every unit the same balance history as local transport, and is not local.

Proof. (a) Flatness makes $x \mapsto \sigma^x$ an action of $\mathbb{Z}^3$, transitive by connectedness. For $w = \sigma^y v_0$, $\sigma^x w = w$ exactly when $\sigma^x v_0 = v_0$, since the σ^x commute; so all stabilizers are equal, and the map is well defined, injective, onto and equivariant. An equivariant bijection from $\mathbb{Z}^3/\Lambda'$ sends 0 to a voxel with stabilizer Λ', so $\Lambda' = \Lambda$. A free gluing has $\Lambda = 0$, and two equivariant bijections from $\mathbb{Z}^3$ differ by a translation. (b) $\sigma_0\sigma_1(0) = 2 \neq 0 = \sigma_1\sigma_0(0)$, and σ_0 alone connects the voxels. (c) Each post crosses one face, and on $\mathbb{Z}^3$ changes one coordinate by one; the balance moves by the same unit step under either rule. □

The proposition is proved in classical mathematics, the bound in faces crossed, on every gluing, with no choice principle. Consistently glued voxels form the cubic lattice or a periodic quotient, and a locally carried unit moves at most one face per post. Flatness is a premise, and (b) shows that posting does not decide it. Local transport and the reading of a voxel's six posts as its six faces are identifications: by (c) the balances do not see how a unit is carried.

Last, three dimensions. In the cubical cell structure of $\mathbb{R}^D$ a cell is a unit cube spanned by a set of coordinate directions at an integer base point, and its *dual cell* is the unit cube spanned by the other directions, centred at the cell's centre. A configuration of a posted pair is a primal trace b, a finite integer combination of edges with zero boundary, and a dual trace $a = \delta c$, the coboundary of a finite integer combination c of $(D-2)$-cells. A *move* is $a \mapsto a \pm \delta \mathbf{1}_f$ for a $(D-2)$-cell f whose dual cell meets no edge of b, or $b \mapsto b \pm \partial p$ for a plaquette p that meets the dual cell of no cell q with $a(q) \neq 0$; configurations are *deformable* into each other when moves and reversed moves join them. The *elementary pair* is $b_0 = \partial p_0$, p_0 the unit square at the origin in directions $0, 1$, with $a_0 = \delta \mathbf{1}_{f_0}$, f_0 the $(D-2)$-cell at the origin spanning every direction except 1 and 2; the *split* configuration is $(0, 0)$.

Proposition 24.59 (A posted pair stays linked exactly in three dimensions)**.** *Let $D \geq 3$.*

(a) For $D = 3$ the pairing $\langle c, b \rangle$ depends only on the traces $a = \delta c$ and b, and no move changes it. It is 1 on the elementary pair and 0 on the split configuration, so the elementary pair is not deformable into the split configuration.

(b) For every $D \geq 4$ the elementary pair reaches the split configuration in two moves.

Proof. A cell meets the dual cell of τ only if it spans every direction of τ, and every cell meets its own dual at its centre. (a) If $\delta c = \delta c'$, then $c - c'$ is a cocycle on the edges of $\mathbb{Z}^3$, hence δg for a function g on vertices, since every closed edge path in the cubical lattice bounds a sum of plaquettes; so $\langle c - c', b\rangle = \langle g, \partial b\rangle = 0$. A move of a by the edge f changes the pairing by $\pm b(f)$, and $b(f) = 0$, or f would be an edge of b meeting its own dual. A move of b by p changes it by $\pm\langle c, \partial p\rangle = \pm a(p)$, and $a(p) = 0$ since p meets its own dual. On the elementary pair f_0 is the edge from the origin in direction 0, which ∂p_0 traverses once. (b) For $D \geq 4$, f_0 spans $D - 2 \geq 2$ directions, so no edge meets its dual cell and $a_0 \mapsto 0$ is a move; then $b_0 \mapsto 0$ is a move. □

Conditions and scope.

Base and premises

The integer cubical lattice of $\mathbb{R}^D$; in classical mathematics. Reading a posted pair as two traces, space as $\mathbb{Z}^D$, and free motion as exactly the unblocked one-cell moves are identifications.

Conclusion

The linking number of the traces [146] is kept by every move in three dimensions, and the smallest linked pair cannot be separated; from four dimensions up it comes apart in two moves, because an edge meets the dual cell of a $(D-2)$-cell only when $D - 2 \leq 1$. So the premise that free motion cannot erase a posted pair holds for this realization exactly when $D = 3$.

Not claimed

Why traces are curves, why crossing is forbidden, why space is a lattice, and dimensions below three.

Posting decides the first step outright, for ledgers that change by posts: the count reading. Above it the premises are a uniform item time, serial ticks, the tick register with the sign of its pass and its half turn, the read premises, the joint tick register of two recognizers (commuting ticks, and $\mathscr{C}_2$ for the square law), equal account costs, the reversal symmetry, the side-by-side composition of ledgers and flatness; one account per axis, local transport, the six faces and the trace kinematics are identifications. For a uniform item time, serial ticks, the half turn, the sign of the pass, (R1), (R5), the reversal symmetry and flatness, the results above exhibit two models that differ on the premise.

24.12 Which values the world keeps real

Chapter 20 named the identification of access: the physically real collections are those whose members each leave a record from which they are read back. That identification is a statement about a world, about which histories of acts can occur in it. This section makes the world explicit and asks the concrete question: in which worlds are the natural numbers physically real? The answer is a statement about storage. A world in which comparisons reach a patch of records keeps the natural numbers real exactly when the patch's capacity grows without bound; for the patch of the Hubble sphere, exactly when the expansion rate at the acts comes arbitrarily close to 0. Deciding that, for every capacity, costs exactly Σ_2 excluded middle.

Definition 24.60 (Acts, worlds and real collections)**.**

1. A *record* is a finite binary word. An *act* is one of: post a record p; retain (advance the stage, keeping every record); compare the records at positions i and j. A *history* is a finite list of acts. Its *ledger* lists the posted records in order, the outcomes of the comparisons (a comparison of two present records registers whether they agree, and one with an absent term registers nothing), and the stage.

2. A *world* W is a set of histories, its *permitted* ones, that contains the empty history and every prefix of a permitted history.

3. A *recognition observable* in W on a set X assigns to each $x \in X$ a permitted history $o(x)$ such that for $x \neq y$ some continuation c and some act a, permitted after both $o(x)c$ and $o(y)c$, give different results. X is *real in* W when such an observable on X exists. Let $\mathscr{A}$ be the world in which every history is permitted. *Access holds in* W when every set real in $\mathscr{A}$ is real in W and conversely.

4. The *window cost* of a list of records is the number of their bits plus one per record. Position p of a list is *within* N when it holds a record and the records from p on have window cost at most N.

5. For a *capacity* $\mathrm{cap} : \mathbb{N} \to \mathbb{N}$ the *patch world* W_{cap} permits the histories whose posts are single bits and whose every comparison, made after t acts, addresses two positions within $\mathrm{cap}(t)$.

Three statements about a world are identifications, physical premises that the theorems take as hypotheses. *Patch reach*: every history of W_{cap} is permitted in W; the patch can carry out every comparison its records allow, after any number of acts that post nothing. *Locality* (the time window): in W, every comparison made after t acts addresses two positions within $\mathrm{cap}(t)$. *Posting reach*: every history of W_{cap} that consists of one-bit posts followed by one comparison is permitted in W. Access holds in W exactly when the natural numbers are real in W; this is the form in which the theorems below state realness.

Theorem 24.61 (The capacity decides the natural numbers)**.** *Let W be a world and* cap *a capacity.*

1. *With patch reach, if for every N some t has $\mathrm{cap}(t) \geq N$, the natural numbers are real in W and access holds.*

2. *With posting reach, if for every T some $t \geq T$ has $2t \leq \mathrm{cap}(t)$, access holds.*

3. *With locality, if* cap *is bounded, every set real in W is finite, and access fails.*

4. *With patch reach and locality, access holds exactly when* cap *is not bounded.*

Parts 1 and 2, with the growth stated as in their hypotheses, are free. Parts 3 and 4 are proved in classical mathematics.

Proof. 1. Let the value n occur as $n+1$ one-bit marks and a blank, followed by retain acts up to the first stage $t(n) \geq n+2$ with $\mathrm{cap}(t(n)) \geq 2(n+2)$. The search for $t(n)$ terminates: by hypothesis some t has $\mathrm{cap}(t)$ above $2(n+2)$ and above $\mathrm{cap}(s)$ for every $s < n+2$, and such a t is at least $n+2$. So nothing is chosen, and the whole record, of window cost $2(n+2)$, is within the capacity at that stage. Patch reach permits these histories and, at their end, every comparison of two of their positions. For $n < m$ compare positions $n+1$ and 0, with the empty continuation: in the run of n they hold the blank and a mark and disagree; in the run of m they hold two marks and agree. So the natural numbers are real in W. Access follows: a set real with every act available is the image of an injection into the natural numbers, through its records, and a recognition observable restricts along an injection.

2. The same, with the occurrence of n continued by marks, not retain acts, up to the first stage $t(n) \geq n+2$ with $2t(n) \leq \mathrm{cap}(t(n))$; the record then has window cost $2t(n)$.

3. Let $\mathrm{cap} \leq N$. If two runs hold record counts that differ by more than N, no position lies within N in both after any common continuation, so no comparison is permitted after both, and posts and retains register nothing: the two values are never told apart. So the values of a real set have record counts within N of one another, and two of them with equal counts and equal records within the window are never told apart, since every comparison after a common continuation reads positions that are either in that window or posted by the continuation. The counts of a real set lie in an interval of length N and the windows hold boundedly many bits, so finitely many classes remain, and the natural numbers are not real.

4. Parts 1 and 3, with "not bounded" turned into "exceeds every bound", a step whose price Theorem 24.62 computes. □

Whether a capacity is unbounded is a Π_2 question about the infinite future: no finite list of capacities decides it, since every finite list extends both to a bounded and to an unbounded capacity. Its price is exact.

Theorem 24.62 (The price of availability)**.** *With no choice principle in either direction:*

1. Deciding availability, *"every capacity* $\mathrm{cap} : \mathbb{N} \to \mathbb{N}$ *exceeds every bound at some act, or stays below some bound", is equivalent to* Σ_2 *excluded middle (Definition 29.26), and implies LPO.*

2. The unbounded witness, *"every capacity that is not bounded exceeds every bound at some act", is equivalent to Markov's principle.*

No omniscience principle and no choice principle is used in any of the implications.

Proof. The Π_2 *split* says: for every binary double sequence β, every row has a 1, or some row is all 0.

Deciding availability gives LPO. For a binary sequence α let the flag capacity be 0 at stage t until α has shown a 1, and $t+1$ from then on. It is unbounded exactly when α has a 1; if it stays below N, α has no 1, since a 1 at place n would give $t+1 \geq N$ at $t = \max(n, N)$.

Deciding availability gives the split. For β let $\mathrm{cap}(t)$ count the leading rows $0, 1, \ldots$ that have each shown a 1 by stage t. If every row has a 1, the capacity exceeds every bound. If it stays

below N, the rows $0, \dots, N-1$ do not all have a 1; LPO, available by the previous step, decides each of these finitely many rows, and the first row without a 1 is all 0.

The split gives availability. Apply it to $\beta(N,t) = 1$ exactly when $N \le \mathrm{cap}(t)$.

The split is Σ_2 excluded middle. The split gives the excluded middle: a row with a 1 in every row refutes "some row is all 0". Conversely, Σ_2 excluded middle applied to $\beta(N,m) = 1 - \alpha(N)$ is LPO for α; then if no row is all 0, LPO finds a 1 in each row.

The unbounded witness. Given Markov's principle and a capacity that is not bounded: for each N, "some t has $\mathrm{cap}(t) \ge N$" is not false, since otherwise N would bound it, and Markov's principle finds t. Conversely, given the witness and a binary sequence α that is not all 0, the flag capacity of α is not bounded, by the argument above, so it reaches 1 at some stage, and α has shown a 1 by then. □

So the question whether the natural numbers are physically real in a world with patch reach and locality is decided, for every capacity, at exactly Σ_2 excluded middle, one level above LPO; and the step from "no bound holds" to "every bound is exceeded" is Markov's principle. Stated positively, as in Theorem 24.61, parts 1 and 2, the step from growth to realness costs nothing.

The Hubble patch

For the patch of the Hubble sphere the capacity is fixed by the expansion. Let the expansion rate at scale factor $a \ge 1$ (today is $a = 1$) be

$$H(a) = H_0 \sqrt{\Omega_m a^{-3} + \Omega_\Lambda \rho(a)}, \qquad H_0 > 0,\ \Omega_m \ge 0,\ \Omega_\Lambda > 0,$$

where the dark-energy density ρ has $\rho(1) = 1$ and $\rho'(a) = -3\,\Delta w(a)\rho(a)/a$ for the deviation $\Delta w = 1 + w$ of its equation of state. The record count of the Hubble sphere at rate H is $N(H) = 4\pi c^2/(H^2 \lambda_{\mathrm{rec}}^2)$, one bit per λ_{rec}^2 of horizon area, and the *Hubble capacity* at the tth act, which occurs at scale factor $a_t \ge 1$, is $\lceil N(H(a_t)) \rceil$. *Count-only aging* says that the deviation $\Delta w(a)$ is a nondecreasing function of the number of records posted by the epoch a, along histories that extend as a grows; Recognition Science's present deviation is $\Delta w(1) = A_{\mathrm{RS}} > 0$.

Theorem 24.63 (Storage decides the horizon)**.** *In classical mathematics, for the Hubble capacity:*

1. *With patch reach and locality, access holds exactly when for every $\varepsilon > 0$ some act has $H(a_t) < \varepsilon$.*

2. *Let $a_t \to \infty$. With patch reach, $H \to 0$ gives access; with locality, a positive asymptotic rate, $H(a) \to H_\infty > 0$, excludes it; and with patch reach, count-only aging with $\Delta w(1) = A_{\mathrm{RS}}$ gives access.*

3. Storage. *Suppose a world posts bits forever (after every permitted history either bit can be posted) and keeps its whole ledger within its capacity (after every permitted history of length t the ledger's window cost is at most $\mathrm{cap}(t)$). Then $\mathrm{cap}(n) \ge 2n$ for every n. For the Hubble capacity the rate at the acts tends to 0, and if $a_t \to \infty$ there is no positive asymptotic rate. With posting reach, access holds.*

4. Posting branch. *Let $a_t \to \infty$, let at most one act fall in each tick of a clock of any positive tick length, run along the cosmic time of the expansion, and let the future age count-only with $\Delta w(1) = A_{\mathrm{RS}}$. Then the Hubble capacity eventually exceeds twice the number of acts, and with posting reach access holds.*

5. Volume acts. *If instead the acts keep pace with the Hubble volume, $\kappa H(a_t)^{-3} \le t + 1$ for every t and some $\kappa > 0$, then in every future the Hubble capacity at the tth act is eventually below $2t$, and no world that posts bits forever keeps its whole ledger within it.*

The step from a rate tending to 0 to access is free when it is read on $\mathbb{R}_\delta$ (Proposition 24.65).

Proof. 1. $N(H)$ is a decreasing function of $H > 0$ and unbounded as $H \to 0$, so the capacity is unbounded exactly when the rate comes arbitrarily close to 0; Theorem 24.61, part 4.

2. The first two claims are parts 1 and 3 of Theorem 24.61. For the third: posted records accumulate, so $\Delta w(a) \ge \Delta w(1) = A_{\mathrm{RS}}$ for $a \ge 1$, and $\rho(a) = \exp\bigl(-3\int_1^a \Delta w(s)\, ds/s\bigr) \le a^{-3A_{\mathrm{RS}}}$, which tends to 0; so $H(a) \to 0$.

3. The history of n one-bit posts is permitted and has window cost $2n$. Then $\lceil N(H(a_t))\rceil \ge 2t$ gives $H(a_t)^2 \le 4\pi c^2/(\lambda_{\mathrm{rec}}^2(2t-1))$, which tends to 0; along $a_t \to \infty$ this excludes a positive limit. The last claim is Theorem 24.61, part 2.

4. Under count-only aging H decreases and $H(a) \le Ca^{-3\min(1,A_{\mathrm{RS}})/2}$, by the bound on ρ in part 2 and the a^{-3} fall of matter. The elapsed cosmic time is $\int_1^a ds/(sH(s)) \le \log a/H(a)$, and one act per tick bounds the number of acts by that time over the tick, while $N(H(a)) = 4\pi c^2/(\lambda_{\mathrm{rec}}^2 H(a)^2)$. The ratio of the capacity to the elapsed time is at least a constant times $1/(H(a)\log a)$, which tends to infinity; so $\mathrm{cap}(t) \ge 2t$ eventually, and part 2 of Theorem 24.61 applies.

5. $\kappa H(a_t)^{-3} \le t + 1$ gives $H(a_t)^{-2} \le ((t+1)/\kappa)^{2/3}$, so the capacity is at most a constant times $(t+1)^{2/3}$, below $2t$ for large t; and part 3 needs $\mathrm{cap}(t) \ge 2t$. □

Conditions and scope.

Base and premises

Classical real numbers; the identifications patch reach, locality and posting reach; the record law $H^2 N_H = 4\pi c^2/\lambda_{\mathrm{rec}}^2$; count-only aging; one act per tick in part 4.

Conclusion

The natural numbers are physically real in the Hubble patch exactly when the expansion slows without bound at the acts. A patch that keeps its whole ledger while posting forever needs capacity that grows at least linearly in its acts, so it forbids a Hubble sphere of bounded area. One act per tick with count-only aging gives that growth; acts in proportion to the volume never fit.

Not claimed

Which act-rate law nature follows (Remark 24.75). Locality itself is not derived from the light cone of Sections 24.1 and 24.5: the light cone orders postings and does not fix which retained records a later comparison addresses.

The "permanent horizon" of part 2 means a Hubble sphere of bounded area; the count-only future still has a cosmological event horizon. A future in which dark energy is static from today is excluded in a precise sense.

Proposition 24.64 (A future static from today jumps at today)**.** *In classical mathematics, with $H_0 > 0$, $\Omega_m \geq 0$, $\Omega_\Lambda > 0$:*

1. *If the rate from today on equals the static rate, $H(a) = H_0\sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$ for $a \geq 1$, then $\Delta w(a) \to 0$ as $a \to 1^+$; so Δw jumps at today from the past value A_{RS}, and the static future is excluded by the premise that w has no jump at the present epoch.*

2. *Some future has $\Delta w(a) \to A_{\mathrm{RS}}$ as $a \to 1^+$, one deviation function equal to $A_{\mathrm{RS}}/(1+z)$ for redshifts $z \geq 0$ and to Δw in the future, and a positive asymptotic rate.*

Proof. 1. The static rate forces $\rho \equiv 1$ for $a \geq 1$, so $\rho' = 0$ and $\Delta w = 0$ for $a > 1$. 2. Take $\Delta w(a) = A_{\mathrm{RS}}/a$; then $\rho(a) = \exp(3A_{\mathrm{RS}}(a^{-1} - 1)) \to e^{-3A_{\mathrm{RS}}} > 0$, and $A_{\mathrm{RS}}/a = A_{\mathrm{RS}}/(1+z)$ at $a = (1+z)^{-1}$ joins the past law continuously. □

So continuity at today excludes a static future, but the continuity of the deviation alone does not exclude a Hubble sphere of bounded area; count-only aging does (Theorem 24.63, part 2).

On $\mathbb{R}_\delta$ the same criterion can be stated without square roots and without choice.

Proposition 24.65 (The Hubble criterion on the line)**.** *Let $H_t > 0$ be points of $\mathbb{R}_\delta$ (the rate at the tth act), $K > 0$ a point and* cap *a capacity. Say the capacity is* calibrated from below *when $NH_t^2 < K$ gives $N \leq \mathrm{cap}(t)$, and* from above *when $N < \mathrm{cap}(t)$ gives $NH_t^2 < K$; the rate* comes near 0 *when for every n some t has $(n+1)H_t < 1$.*

1. *With patch reach and the calibration from below, a rate that comes near 0 gives access. Free.*

2. *Given the calibration from above, a capacity that exceeds every bound makes the rate come near 0. Free.*

3. *With patch reach, locality and both calibrations, access holds exactly when the rate comes near 0.*

4. *For $a \geq 0$, $a < 1$ exactly when $a^2 < 1$, free; so "the rate comes near 0" is the same statement read on H_t^2, and in particular on $H_0^2(\Omega_m a_t^{-3} + \Omega_\Lambda \rho_t)$: no square root of a measured quantity is needed. Part 3 holds with calibrations stated on any positive G_t in place of H_t^2, and the step from G to access is free.*

5. *The rate $H_t = 1/(t+1)$ with capacity $(t+1)^2$ meets both calibrations at $K = 1$, comes near 0, and gives access with patch reach.*

Parts 1 and 2, the first two claims of part 4 and its step from G to access are free; part 3, the rest of part 4 and part 5 are proved in classical mathematics.

Proof. 1. For each N pick n with $(n+1)^2 \geq N/K$ computed from a bound for $1/K$; a t with $(n+1)H_t < 1$ has $NH_t^2 < K$, so $\mathrm{cap}(t) \geq N$, and Theorem 24.61, part 1, applies. 2. Take a natural number $K' \geq K$, $N = (n+1)^2K'$ and t with $\mathrm{cap}(t) > N$; then $NH_t^2 < K$ gives $(n+1)^2H_t^2 < 1$, and part 4 gives $(n+1)H_t < 1$. 3. Parts 1, 2 and Theorem 24.61. 4. For $a \geq 0$, $a^2 - 1 = (a-1)(a+1)$ with $a+1 \geq 1$. 5. $N/(t+1)^2 < 1$ exactly when $N < (t+1)^2$, which gives $N \leq (t+1)^2$; and $N < (t+1)^2$ gives $N/(t+1)^2 < 1$. □

24.13 The eight-tick comparison at its exact scope

Section 21.7 took the reference action of a comparison to be one full cycle, $A_0 = 2\pi\hbar$, as an identification. What physical statement is that identification? A comparison is timed by its reference record: it stops when the reference has accumulated the action A_0.

Definition 24.66 (Self-timed comparisons)**.** In classical mathematics, a *stop rule* assigns to every residual $d \geq 0$ and reference energy $r > 0$ a duration $S(d,r) \geq 0$. It is *reference-clocked* with action A_0 when $rS(d,r) = A_0$ for all $d, r > 0$. It *stops at the first return* at the scale $\hbar'$ when, for all $d, r > 0$, $S(d,r) > 0$, the phase $e^{-\mathrm{i}rS(d,r)/\hbar'}$ equals 1, and $e^{-\mathrm{i}rt/\hbar'} \neq 1$ for $0 < t < S(d,r)$. The *fundamental clock mode* has energy $E_1 = \hbar\pi/(4\tau_0)$, a quarter turn per tick τ_0.

Theorem 24.67 (A comparison lasts eight ticks exactly when $A_0 = 2\pi\hbar$)**.** *In classical mathematics, let S be reference-clocked with action A_0.*

1. *For every $d > 0$: $S(d, E_1) = 2^3\tau_0$ exactly when $A_0 = 2\pi\hbar$.*
2. *Some $d > 0$ has $S(d, E_1) = 2^3\tau_0$ exactly when S stops at the first return at the scale $\hbar$.*
3. *The duration does not fix the scale at which the comparison's own phase is counted: at $\hbar' = \hbar/9$ the clock mode has the same phase at every whole tick as at $\hbar$, and a rule with $A_0 = 2\pi\hbar$ lasts eight ticks, but it does not stop at the first return at $\hbar'$. If S stops at the first return at its own scale $\hbar'$ and some $d > 0$ has $S(d, E_1) = 2^3\tau_0$, then $\hbar' = \hbar$.*
4. *For every $A_0 > 0$ there is a comparison controller with reference action A_0 (Definition 21.35: pooled records drive the sum, a record read against itself drives the reference), and some reference-clocked rule with $A_0 > 0$ lasts other than eight ticks for every $d > 0$. So those laws leave A_0 free, and the eight-tick duration is not a consequence of them.*

Proof. 1. $E_1 S(d, E_1) = A_0$, and $E_1 \cdot 2^3\tau_0 = 2\pi\hbar$. 2. At the scale $\hbar$ the phase $rS/\hbar = A_0/\hbar$ is the same for every d, r; it first returns to 1 when it reaches 2π, so the first-return stop is $A_0 = 2\pi\hbar$, which is part 1. 3. At $\hbar' = \hbar/9$ the phase of the clock mode per tick is $9\pi/4 \equiv \pi/4$ modulo 2π. The phase of the rule at its stop, read at $\hbar'$, is $e^{-2\pi\mathrm{i}\hbar/\hbar'} = 1$, but it has already returned to 1 at the time $S/9$, so the stop is not the first return. With the first-return stop at $\hbar'$, the phase $rS/\hbar'$ first reaches 2π at the stop, so $A_0 = 2\pi\hbar'$, and the eight-tick duration with part 1 gives $A_0 = 2\pi\hbar$. 4. The controller that selects, for records d and r, a phase loop of generator cost $A_0 d/(\hbar r)$ obeys both laws at every $A_0 > 0$ (Corollary 21.38, proof of part 2), and the rule $S(d,r) = \hbar/r$, reference-clocked with $A_0 = \hbar$, lasts $\hbar/E_1 = 4\tau_0/\pi$. □

Conditions and scope.

Base and premises
Classical real numbers; a reference-clocked comparison; the fundamental clock mode of energy $\hbar\pi/(4\tau_0)$.

Conclusion
$A_0 = 2\pi\hbar$, Planck's constant h in native units, is exactly the statement that one comparison against the fundamental clock mode lasts eight ticks. That statement is an independent physical premise: the laws of the controller leave A_0 free, and the posed derivation of the eight-tick duration from those laws is false.

Not claimed
The scope is the duration of one comparison. The theorem does not derive, or use, the cycle of 2^3 states of three binary channels; and the duration does not fix the scale at which the comparison's own phase is counted, which records taken at whole ticks cannot tell from $\hbar/(1+8j)$. Read at a scale $\hbar'$, the phase at the stop of a rule with $A_0 = 2\pi\hbar$ is 1 exactly when $\hbar/\hbar'$ is a whole number.

Read with Theorem 21.37: a certain change costs a quarter of the reference cycle, and the reference cycle is h exactly when one comparison of the clock mode lasts eight ticks. The binary gain sits at the threshold.

24.14 What a store with few addresses keeps real

Locality bounds what a comparison can reach by recency. A different law bounds it by address: a comparison reaches only certain positions. Which address laws keep the natural numbers real? The answer turns on whether the addresses depend only on how many records are held, and on whether the store forgets.

Definition 24.68 (Count-only stores)**.** A *count-only store* is a map addr from natural numbers to finite sets of positions. Its world permits the histories whose posts are single bits and whose every comparison, made when L records are held, addresses two present positions in $\mathrm{addr}(L)$. The store *settles* when from some L_0 on $\mathrm{addr}(L)$ is one fixed set of positions below L_0; it *keeps* when $\mathrm{addr}(L) \subseteq \mathrm{addr}(L+1)$, every position in $\mathrm{addr}(L)$ is below L and $|\mathrm{addr}(L)| \le k$ for a fixed k. It *serves the lag d beyond P* when some L and positions $P \le j, i < j < L$ have $i, j \in \mathrm{addr}(L)$ and $i, j \in \mathrm{addr}(L+d)$. It has a *persistent anchor* when for all B, D some $p < s$ with $s \ge B$ have $p, s \in \mathrm{addr}(s+1)$ and $p, s \in \mathrm{addr}(s+1+D)$. Its *depth* is bounded by w when every addressed present position is among the newest w records.

Theorem 24.69 (Count-only stores)**.** *For count-only stores, and the world of each, with one store whose addresses depend on content for comparison in part 5:*

1. Sufficient. *If for some increasing sequence of counts $s_0 < s_1 < \cdots$ every difference $s_b - s_a$ is served beyond s_a, the natural numbers are real and access holds. In particular this holds when every positive lag is served beyond every bound, and when the store has a persistent anchor. Free.*

2. Necessary. *If the natural numbers are real, then for all P and D some lag of at least D is served beyond P; for every P an infinite set of counts has all its differences served beyond P; the store does not settle; and its depth is unbounded. A store that serves no lag of D or more, or has depth bounded by w, or settles, keeps only finite collections real.*

3. Keeping truncates. *A store that keeps keeps only finite collections real, and values that share a record count number at most 2^k.*

4. Examples. *The store of the first record and the largest square position below the count, $\{0, \lfloor\sqrt{L-1}\rfloor^2\}$, has a persistent anchor and keeps the natural numbers real. The store of the first and the newest record, $\{0, L-1\}$, does not settle and has unbounded depth, yet exactly four values are real: two bits are real, five values are not. The store of the first two records, the positions $p < \min(L, 2)$, keeps, and exactly five values are real. The newest record alone keeps at most one value real, and the two newest at most four. The store $\{L-1, L \bmod 3\}$, which has two addresses and does not keep, tells apart all eight three-record words. The store $\{0, 1, L-1, L-2\}$ serves lag one beyond every bound and every lag beyond 0, yet no lag of two or more beyond 2, and the natural numbers are not real there. The store $\{0, L-1, L-2, L-3\}$ serves the lags one and two beyond every bound and none of three or more, so serving some lag beyond every bound is not enough, and it keeps only finite collections real.*

5. Content beats count. *The store that addresses the first record, the newest record and the last blank, which is not count-only, and which after each act keeps only positions it held before and the newest one, keeps the natural numbers real; the address of the blank carries the value.*

Parts 1 and 5 are free; the rest is proved in classical mathematics.

Proof. 1. Let the value n occur as s_n blanks. For $n < m$ the lag $d = s_m - s_n$ is served beyond s_n: some count L and positions $i < j < L$ with $j \geq s_n$ have i, j addressed at L and at $L + d$. Continue both runs with blanks, one mark at the post that lands at position j of the shorter run, and blanks up to count L of the shorter run; then compare i with j. In the shorter run, at count L, position j holds the mark and i a blank: they disagree. The longer run is d records ahead, at count $L + d$, where i and j are again addressed; its mark sits at $j + d$, so i and j are both blank and agree. If every positive lag is served beyond every bound, take $s_n = n$. A persistent anchor serves every lag D beyond every bound, at the counts $s + 1$ and $s + 1 + D$.

2. If no lag $d = b - a$ is served beyond P, two values with record counts $a < b$, both at least P, that agree on their first P bits are never told apart: every comparison the two runs can both make reads positions that agree. Grouping values by their first P bits, an infinite group has its counts in an infinite set whose differences are all served beyond P, and so arbitrarily large lags are served. A store of depth w is a window of $2w$ bits, and a settled store reads a fixed finite set of positions.

3. A keeping store's address sets grow and are bounded in size, so they settle; values sharing a count and agreeing at the settled set agree on every test.

4. Direct computations from parts 1–3 and the definition; for $\{0, L-1\}$, values told apart must share their record count, and then only the first and newest bits can differ.

5. Value n is $n+1$ marks and a blank; for $n < m$ post $m-n$ further marks and compare position $m+1$ with the first record: the newest record, a mark, in the shorter run, and the blank in the longer one. □

Whether every count-only store with an infinite difference set beyond every bound keeps the natural numbers real is open. With the stores of the theorem, what limits realness is a bound on stored information or a window of addresses, not the number of addresses.

A different count concerns one cell of the cosmogenesis lattice.

Proposition 24.70 (A cell holds exactly three bits)**.** *In classical mathematics: the basis measurement of an eight-level site is a valid instrument, and decoding its outcome as the basis index separates the 2^3 basis preparations with certainty; and no reader that decodes the outcome of a valid instrument on the site separates more than 2^3 normalized preparations with certainty. So a cell holds exactly three bits.*

Proof. The outcome effects of the basis measurement are the diagonal projections, which sum to the identity. If a reader separates preparations $\rho_1, \dots, \rho_M$ with certainty, its decoded effects $E_1, \dots, E_M$ have $\operatorname{tr}(E_i\rho_i) = 1$ and sum to at most the identity; since $\operatorname{tr}(E\rho) \le \operatorname{tr} E$ for a normalized ρ, $M \le \sum_i \operatorname{tr} E_i \le 8$. □

That a cosmogenesis cell's preparations and readers are these is an identification.

24.15 Certifying a law at finite precision

Theorem 17.22 shows that a closed law, built from equations and $\le$ by conjunctions, holds exactly when it passes the test at every tolerance, and that the law $x \le y \vee y \le x$ passes every test at every state but holds at every state only under LLPO. What does passing every test certify, for laws with disjunctions?

Theorem 24.71 (Certification by tolerance tests)**.** *Call a law* non-strict *when it has no atom $x < y$; it may use $=$, $\le$, $\wedge$, $\vee$ and finite conjunctions and disjunctions. A class V of points is* tolerance complete *when every non-strict law whose quantities take values in V holds at every state at which it passes the test at every tolerance.*

1. *The whole line is tolerance complete exactly when LLPO holds.*
2. *A class on which $x < y$, $x = y$ or $y < x$ is decided for any two of its points, in particular $\mathbb{Z}[\varphi]$, $\mathbb{Q}(\varphi)$ and the ratios, whose order is decided by integer arithmetic, is tolerance complete. Free.*
3. *Every tolerance complete class has the dichotomy $x \le y \vee y \le x$ for its points. Conversely, a class that contains 1 and is closed under sums and negatives is tolerance complete exactly when it has the dichotomy, and that direction is free.*

No choice principle is used.

Proof. A free step. A law that passes every test is not refuted: a strict refutation has a margin $a < b$, and at a tolerance below $b - a$ the refuted atom fails the test, and smaller tolerances keep it failing.

1. The dichotomy law passes every test, so completeness gives LLPO by Theorem 17.22, part 4. Conversely, by the free step it suffices that "not refuted" gives "holds" for non-strict laws. At an atom $x \leq y$ it is the definition of $\leq$, and at $x = y$ tightness of the line. At $\vee$ it is de Morgan's law for two statements each certified by a record, which LLPO gives: the binary sequence that marks the first record of each statement, the first at even places and the second at odd places, has at most one mark.

2. On such a class a non-strict law holds or is refuted, decided by integer arithmetic on the values, and by the free step it is not refuted.

3. The dichotomy law of two constant quantities passes every test. Conversely, under the dichotomy the class is closed under the maximum, which is one of the two points. Every law has a margin in the class: the difference of the sides of an inequality, the larger of an equation's two differences, the maximum over conjunctions and the minimum over disjunctions, with -1 and 1 for empty families. The law is refuted exactly when its margin is positive. Two margins are ordered by the dichotomy, so "not both refuted" splits into "one is not refuted", which is the de Morgan step of part 1 with no principle. □

On the values the ledger carries, the golden integers and rationals, passing every finite-precision test certifies an exact law, disjunctions included, at no price. Over the whole line the same certification costs exactly LLPO.

A window on records changes what a fixed apparatus can settle.

Proposition 24.72 (A window loses the first examination)**.** *In classical mathematics, consider the question whether the first and the Kth examination of a binary world agree, settled by the first $K + 1$ entries. A* passive *apparatus answers from the ledger of one fixed continuation after the examinations.*

1. *In a world whose comparisons reach only within a window of size $N \leq K$, every passive apparatus answers the same for two ledgers with the same record count and the same recent tail, and no passive apparatus settles the question.*

2. *In every world that extends every run and permits the examinations, in particular in $\mathscr{A}$, the apparatus that compares positions 0 and K once $K + 1$ records are in is passive and settles it.*

Proof. 1. After a common continuation every comparison outcome agrees for two ledgers with the same count and recent tail, by induction along the continuation. On the all-blank world the apparatus must answer after the Kth entry; on the world with a mark at the first entry it cannot answer earlier; and from the $(K + 1)$st entry on the recent tail no longer holds the first entry, so it confirms agreement where the entries differ. 2. The outcome of the comparison is whether the two records agree, which for injective records is whether the examinations agree. □

24.16 Crossings and postings

A patch holds located units on cubes of $\mathbb{Z}^3$, and its accounts are books: a map from cubes to finitely many accounts. When does the patch act once per tick?

Definition 24.73 (Crossing ledgers)**.** A *unit path* gives a cube at every tick, moving to an adjacent cube or staying. For a family of unit paths and a book map, the *crossing ledger* posts at each tick, for each unit whose book changes, a debit at the book it leaves and a credit at the book it enters. A ledger on several accounts is *pair-serial* when at each tick it holds or takes exactly one double entry, a debit at one account and a credit at another. A family is a *unit system* when at most one unit moves at each tick. A *legal atomic tick* posts exactly one entry.

Proposition 24.74 (Crossings, double entries and ticks)**.** *In classical mathematics:*

1. *The crossing ledger's record total grows by twice the number of crossings, and the net credit at an account is its change of occupancy.*
2. *A crossing ledger is pair-serial exactly when at most one unit crosses at each tick; the crossing ledger of a unit system is pair-serial. Two units that are each a unit system on their own and cross at the same tick have a crossing ledger that is not pair-serial, under every reading on accounts. A ledger that posts one fixed double entry at every tick is pair-serial and is the crossing ledger of no finite family of units.*
3. *A double entry is two postings and never a legal atomic tick, and a legal atomic tick never joins two ledgers whose columns balance. So the statement that the patch's ledger advances by one legal atomic tick per tick is false for every stream of events read with balanced columns; the corrected statement, proved, is that under the source and target reading (each paired event posts a debit at its source's account and a credit at its target's) it advances by exactly one double entry per tick.*
4. *A pair-serial ledger gains at most two entries per tick; if each act posts one double entry and acts are clocked, there is at most one act per tick.*
5. *If each move of a unit is a post at a site of one recognition structure and a post moves exactly one unit, then Recognition Science's atomic tick (exactly one site posts per tick) makes the units a unit system; conversely every unit system carries such an atomic tick. The atomic tick alone constrains nothing: it is a schedule, and two units crossing at one tick are not a unit system under any schedule.*
6. *When the units' own ledgers are run jointly, each tick posting on one unit's ledger, which moves that unit one hop, the crossing ledger is the fold of the schedule through the book step: at each tick it holds, or takes one double entry from the old book to the new, and the two cubes are adjacent.*

Proof. 1. Each crossing posts one debit and one credit, and the occupancy of a book changes by the units entering minus those leaving. 2. One crossing gives one double entry, two give four entries at one tick, which no hold or double entry is; a fixed double entry would lower the debited book's occupancy below 0 after more ticks than there are units. 3. A legal atomic

tick changes one column by one, so it cannot preserve column balance. 4. Count entries. 5. Two units that move at one tick post at sites the atomic tick identifies, so they are one unit; a unit system's moving unit, or an idle site, gives the schedule. 6. A post moves its unit one hop, so its old and new cubes are adjacent, and every other unit stays; part 1 and induction along the schedule. □

So the one-tick question of the patch is reduced to one physical identification: the patch's units are carried by one recognition structure, with their moves as its posts. That identification is the joint net-flow position in which each unit sits at the balance of its own ledger; with it the book ledger is a faithful image of the units' own posting run, and the patch moves one unit per tick, recorded at two books.

Remark 24.75 (What stays open in the world's records)**.** The following are stated exactly and are open.

(a) *Locality from the light cone.* No accepted result fixes which retained records a comparison addresses; the missing law is chronological coverage, equivalent to locality (the time window). The light cone of Sections 24.1 and 24.5 orders postings and does not supply it.

(b) *The act-rate law.* Whether the Hubble patch acts at most once per tick of the cosmic clock or in proportion to its volume is fixed by no accepted result; Theorem 24.63 proves what each gives.

(c) *One crossing per tick.* Reduced to the identification that the patch's units form one unit system, read on the patch's accounts (Proposition 24.74), and that in turn to the identification that moves are posts.

(d) *The price of access from locality.* In classical mathematics, access with locality makes the capacity exceed every bound; the price of that step on the line is open.

(e) *Certain flips.* Whether some native word flips a register configuration with certainty, at zero boundary flux and unit action scale; and whether the integer parameter m of the register drops out of that question (if a certain flip at $m = 0$ gives one at every m, the question reduces to $m = 0$ and zero flux).

(f) *Count-only stores.* Whether a count-only store with, beyond every bound, an infinite set of counts whose differences are all served keeps the natural numbers real.

24.17 Controlled posts and the coherence of a side

Section 24.7 left one premise open: the directions x_1, x_2 of the record cone need posts whose side is held coherently, and posting does not produce them. The obvious way to try is to let a coherent register choose the side. Put a *side register* in a superposition of debit and credit, and let it control which post is written to the ledger. Does the side then stay coherent? The answer is no, for an exact reason: the ledger records which side was posted. This section proves that, says what erasure can and cannot repair, and identifies the one kind of controlled law that keeps the side coherent: the kind whose ledger records nothing. The results are proved in classical mathematics over the complex numbers.

Definition 24.76 (Controlled posts)**.** A *code register* has the basis states $1, \dots, n$; in the census encoding of Section 24.7 they are the ledger codes, and a post at account k on side s is the map $y \mapsto \mathrm{bump}_{k,s}(y)$ that raises that count by one modulo $N+1$. A *controlled law* is a pair $f = (f_0, f_1)$ of maps of the codes, one for each side. For amplitudes $v \in \mathbb{C}^2$ of the side register and a code state $c \in \mathbb{C}^n$, the *controlled state* is $\sum_{s,k} v_s\, c_k\, |s\rangle|f_s(k)\rangle$, and the *side record* is its reduction to the side register, the 2×2 matrix $\rho_{ss'} = v_s \bar{v}_{s'} \langle f_{s'}c, f_s c\rangle$ with $f_s c = \sum_k c_k |f_s(k)\rangle$. The *overlap* of the law on c is $\langle f_1 c, f_0 c\rangle$. When c is a single code k we say the code is *definite*.

The side record is a coherent record (Definition 24.13); it is null, a coherent post, exactly when its determinant vanishes, and timelike, a classical mixture of the two posts, when its determinant is positive.

Proposition 24.77 (A controlled post records its side)**.** *In classical mathematics, for every controlled law f on n codes, every definite code k and all amplitudes v:*

(a) with a definite side, $v = e_s$, the controlled state is the classical post $|s\rangle|f_s(k)\rangle$;

(b) $\rho_{ss'} = v_s \bar{v}_{s'}$ if $f_s(k) = f_{s'}(k)$, and $\rho_{ss'} = 0$ otherwise;

(c) the side record is null exactly when $v_0 = 0$, $v_1 = 0$ or $f_0(k) = f_1(k)$;

(d) if $f_0(k) \neq f_1(k)$ and $v_0, v_1 \neq 0$, the side record is timelike, while the joint state of side and ledger stays pure; if $f_0(k) = f_1(k)$ it is the pure record $v v^\dagger$;

(e) δ's two posts at one account, $f_s = \mathrm{bump}_{k,s}$ on the codes with count bound $N \geq 1$, never leave the same code: so a controlled post whose branches are δ's two posts leaves every superposed side a timelike record.

Proof. (a) and (b) are the partial trace of $\sum_s v_s |s\rangle|f_s(k)\rangle$: the entry (s, s') is $v_s \bar{v}_{s'} \langle f_{s'}(k)|f_s(k)\rangle$. (c) The determinant is $|v_0|^2|v_1|^2(1 - [f_0(k) = f_1(k)])$. (d) follows from (b) and (c). (e) A debit post changes the debit count of the account and a credit post does not, and raising a count by one modulo $N+1$ changes it (Proposition 24.21(b)). □

It is locality and census certainty together that make the two codes differ: the census count is the same after a debit and after a credit. Can the coherence be restored afterwards?

Proposition 24.78 (Erasure, and what it writes)**.** *In classical mathematics, for controlled laws whose maps f_0, f_1 are permutations of the codes:*

(a) the side record has the off-diagonal entry $v_0 \bar{v}_1$ times the overlap and the diagonal entries $|v_s|^2 \|c\|^2$;

(b) no unitary acting on the code register alone changes the side record;

(c) on a definite code that the two branches move apart the overlap is 0; on the uniform code state, $c_k = 1$ for all k, it is n, and the side record is $n\, v v^\dagger$;

(d) if each branch moves every code apart, $\|c\| = 1$ and some code has probability above one half, then the overlap has modulus below 1: coherence cannot be full when the code is that certain;

(e) a second controlled post of the other side restores full coherence on a definite code: since δ's debit and credit posts commute, the branches debit-then-credit and credit-then-debit end in the same code. The side register returns to its prepared record $vv^\dagger$, but each branch has written one debit and one credit at the account: below the count bound both counts rise by one and nothing else changes, the same in both branches. The result is a balanced pair, not a coherent unit post.

Proof. (a) As in Proposition 24.77, with $\langle f_s c, f_s c\rangle = \|c\|^2$ for a permutation. (b) The partial trace over the code register is invariant under unitaries of that register. (c) The two posted code states of a definite code are distinct basis states; on the uniform state each $f_s c$ is the uniform state. (d) By Cauchy–Schwarz the overlap has modulus at most $\|c\|^2 = 1$, with equality only if $f_0 c = \mu f_1 c$, that is $c(f_0^{-1}(z)) = \mu\, c(f_1^{-1}(z))$ for every code z and some $|\mu| = 1$. At $z = f_0(y)$ this gives the code $y' = f_1^{-1}(f_0(y)) \neq y$ the same probability as y, and two codes of probability above one half are impossible. (e) $\mathrm{bump}_{k,d}$ and $\mathrm{bump}_{k,c}$ change different counts, so they commute; then apply Proposition 24.77(b), and count the changes. □

Proposition 24.79 (Which controlled laws keep the side coherent)**.** *In classical mathematics:*

(a) on a definite ledger code, no controlled law writes a coherent single post whose side follows the side register; a debit-versus-credit post at one account keeps the coherence strictly below full on every legal ledger state; changes that do not depend on the side keep it coherent;

(b) for permutations f_0, f_1, a code state $c \neq 0$ and amplitudes with $v_0, v_1 \neq 0$, the side record is null exactly when $c \circ f_0^{-1} = \mu\, c \circ f_1^{-1}$ for some $\mu \in \mathbb{C}$; then $|\mu| = 1$, the joint state is a product, the ledger ends in one state whatever the side, and the side picks up the phase μ.

Proof. (b) The determinant of the side record is $|v_0|^2|v_1|^2(\|c\|^4 - |\langle f_1 c, f_0 c\rangle|^2)$, which vanishes exactly in the equality case of Cauchy–Schwarz, $f_0 c = \mu f_1 c$; equal norms give $|\mu| = 1$, and then the controlled state is $(v_0\mu|0\rangle + v_1|1\rangle) \otimes f_1 c$. (a) A change that does not depend on the side gives the side record $\|c\|^2 vv^\dagger$. On a definite code (b) asks for $f_0(k) = f_1(k)$, and two single posts leave the same code only if they are the same post. For a debit-versus-credit post at one account that is legal, the debit being allowed on every code of the state, take the code y^* of the support with the largest debit count: the debit branch puts weight on a code with a larger debit count, which the credit branch never reaches, so the two posted states are not proportional and the inequality of (b) is strict. There is no uniform gap: a state spread evenly over a chain of L codes has overlap $(L-1)/L$. □

So a controlled law keeps the side coherent exactly when its ledger does not record the side. Branches may post differently on every code and stay coherent, but only if the code state is spread so that the two posted states agree up to a phase; a ledger in a definite state always records the side. The open premise of Section 24.7 is therefore not supplied by controlling δ's posts: the coherent directions of the record cone need a carrier that does not record the side.

24.18 Charts of the causal order of postings

Section 24.5 proved that separate accounts reach only the taxicab cone, and Section 24.6 that the record cone of one coherent distinction is the round cone of one time and three space dimensions. The two cones are different shapes. This section asks how far the order of postings can still be drawn in the round cone, if the coordinates are allowed to bend. The answer has three parts: no affine chart works anywhere; the finite pieces listed in Theorem 24.83 have exact cone charts, while large grids have none [201]; and the whole order has none, already in its integer part, while long runs crowd every chart toward the light cone. First the record cone's own scale.

Proposition 24.80 (The record cone is a calibration of the light speed)**.** *In classical mathematics, let a world meet the foundation's facts with a light speed $c > 0$ in lengths of J per tick and the causal bound $c^2 \geq \frac{3}{4}$, so that its lattice parameter is $\kappa = \frac{1}{4} + c^2$. Read record time t as t ticks and record space x as σx lengths. The classical posts' records are null directions of time one, and every null record direction is a light-like displacement of the world exactly when $c = \sigma$. At the posting scale $\sigma = \frac{\sqrt{3}}{2}$ this is the statement $\kappa = 1$; at $\sigma = 1$ it holds in the world with $\kappa = \frac{5}{4}$; and every such world meets it at its own scale $\sigma = c$.*

Proof. A null record direction $(1, x)$, $|x| = 1$, is read as a displacement of σ lengths in one tick, which is light-like exactly when $\sigma = c$. $\sigma = \frac{\sqrt{3}}{2}$ is the space leg of one posting, and $\frac{1}{4} + \frac{3}{4} = 1$. □

So the record cone makes the posts null; it does not fix κ, which stays the named premise of Section 27.8. The objects of the rest of this section are the stage events of the lattice of postings.

Definition 24.81 (Stage events, the physical causal order and record-cone charts)**.**

1. A *stage event* is an event of the four-account ledger whose position spends no tick; its *content* $c(e) \in \mathbb{R}^4$ is what the accounts spend from a reference event to it, and the content of a pair is $c(f) - c(e)$. The *physical causal order* puts f after e exactly when the pair spends a nonnegative content on every account: in content space it is the *account orthant*, a cone with four flat faces whose edges are the single postings. The *posting lattice* is the set of stage events with integer content from the origin.

2. For a set P with an order $\preceq$, a *record-cone chart* is a map $\varphi : P \to \mathbb{R}^4$ with $p \preceq q$ exactly when $\varphi(q) - \varphi(p)$ lies in the future cone K of Definition 24.13. A *cone chart in* a real inner product space E is a map $p \mapsto (t(p), x(p)) \in \mathbb{R} \times E$ with $p \preceq q$ exactly when $\|x(q) - x(p)\| \leq t(q) - t(p)$.

3. For sides $a, b, c, d \geq 1$ the *grid* $a \times b \times c \times d$ is the product of the chains $\{0, \dots, a-1\}$ and so on, ordered componentwise; three-account grids have $d = 1$; $[n] = \{0, \dots, n-1\}$ and $\mathbb{N}$ are chains. A grid is *read on stage events* when stage events realize it, with contents the grid points and the physical causal order equal to the grid order.

Proposition 24.82 (No affine chart; the orthant is two light cones)**.** *In classical mathematics:*

(a) no linear map of $\mathbb{R}^4$ *carries the account orthant onto the record cone K, so no coordinates on stage events whose differences are a linear function of the content make the physical causal order the record-cone order;*

(b) an explicit invertible linear map sends every nonnegative content into K and the four single postings to the null directions $(1,1,0,0)$, $(1,-1,0,0)$, $(1,0,1,0)$, $(1,0,0,1)$*: physical precedence implies record-cone precedence of the mapped content, and not conversely;*

(c) the account orthant is exactly the intersection of the two pair cones: with $\pi_{ij}(v) = (v_i + v_j,\ v_i - v_j)$*, a content v is nonnegative exactly when* $\pi_{01}(v)$ *and* $\pi_{23}(v)$ *both lie in the* $1+1$ *light cone; so on all stage events physical precedence is the conjunction of two* $1+1$ *light-cone readings, one for each pair of accounts.*

Proof. (a) Such coordinates are a linear map L with $v \geq 0$ exactly when $Lv \in K$; L is injective, since a nonzero v with $Lv = 0$ would make both v and $-v$ nonnegative, so L carries the orthant onto K. A linear image of the orthant is the cone spanned by four vectors, hence polyhedral, and a polyhedral cone of dimension four with no line in it has a two-dimensional face, a flat sector of its boundary spanned by two independent vectors. K is not polyhedral, since ∂K has no two-dimensional face: if u, w and $u + w$ are all null then $\langle u, w\rangle = 0$, and two orthogonal null vectors are parallel. (b) The map is explicit and checked on the four edges, which span. (c) $|v_i - v_j| \leq v_i + v_j$ exactly when $v_i, v_j \geq 0$. □

So the influence order of four accounts is two $1 + 1$ light cones, not one $1 + 3$ cone. Curved charts can still draw some finite pieces of it in one round cone.

Theorem 24.83 (Some finite pieces have exact curved charts)**.** *In classical mathematics, each of the following orders has a record-cone chart with integer coordinates, and each is read on stage events, so on those events the physical causal order is a record-cone order:*

(a) the sixteen corners of a unit posting cube, $2 \times 2 \times 2 \times 2$*, where no affine chart exists;*

(b) the three-account grids $3 \times 3 \times 3$ *and* $4 \times 4 \times 4$*, the latter with a chart in one plane;*

(c) the truncations $2 \times 3 \times [n]$ *of* $2 \times 3 \times \mathbb{N}$*, for every* $n \leq 20$*;*

(d) every four-account grid with all sides at least two and at most forty-eight events, in every order of the accounts, and $2 \times 2 \times 2 \times m$ *for every* $m \leq 8$*.*

Conditions and scope.

Base and premises

Finite orders and integer charts; classical real numbers.

Conclusion

Each of these finite pieces of the posting lattice can be drawn exactly in one round light cone, with curved coordinates; affine coordinates never suffice (Proposition 24.82).

Not claimed

Any grid not listed: $2 \times 3 \times 3 \times 3$ (fifty-four events) and $3 \times 3 \times 3 \times 3$ are open, and so are the truncations $2 \times 3 \times [n]$ for $n > 20$.

Proof. Each chart is an explicit table of integer points, and the claim that the grid order equals the record-cone order of the table is a finite check over all ordered pairs (256 pairs for the cube, 729 and 4096 for the three-account grids, 1296 for $2 \times 2 \times 3 \times 3$), each pair an integer comparison of t with $|x|$ through $t \geq 0$ and $t^2 \geq |x|^2$, done by exact computation. A chart of a grid restricts to a chart of every subgrid, which gives every $n \leq 20$ from $n = 12, 16, 20$ and every $m \leq 8$ from $m = 8$. Exchanging two neighbouring accounts carries a chart to a chart, by relabelling the grid points, which gives every order of the accounts. Realization on stage events uses that the content is a linear bijection of stage differences. For (a), the chart puts the empty corner at time -8, the single postings at time -3 in four tetrahedral directions, the pairs at time 0, the triples at time 3 and the full corner at time 8. No affine chart exists: with u_d the image of account d's posting and $s = \sum_d u_d$, the Minkowski form Q satisfies $\sum_d Q(s - 2u_d) = 4 \sum_d Q(u_d) \geq 0$, so some pair of a triple and the fourth single posting would be causal although neither contains the other. □

The charts were found numerically and their best margins fall quickly as the pieces grow: for the truncations $2 \times 3 \times [n]$ the least distance of an incomparable pair from the light cone was 0.0118 at $n = 12$, 0.0029 at 16 and 0.00025 at 20, and searches at $n = 22$ and 24 found none. Those numbers are not part of any proof. The next theorem shows that the whole order has no chart at all.

Theorem 24.84 (No chart of the whole posting order)**.** *In classical mathematics:*

(a) no map of $2 \times 2 \times \mathbb{R}$ *into* $\mathbb{R} \times E$*, for any real inner product space* E*, is a cone chart; the stage order with real contents contains* $2 \times 2 \times \mathbb{R}$ *(contents* $(i, j, t, 0)$*), so no chart of all stage events reads the physical causal order as the record cone;*

(b) no cone chart of $2 \times 3 \times \mathbb{N}$ *with the order of* $\mathbb{N}$ *reversed has times bounded below, in any complete real inner product space;*

(c) $2 \times 3 \times \mathbb{N}$ *has no cone chart in any complete real inner product space and no record-cone chart; it is read on stage events of the posting lattice (contents* $(i, j, k, 0)$*), so the posting lattice has no record-cone chart.*

Conditions and scope.

Base and premises

Classical real analysis on inner product spaces.

Conclusion

The physical causal order is not the round light cone in any chart, curved or not, and the obstruction does not need the continuum: the integer posting lattice already fails. This agrees with the theorem of Felsner, Fishburn and Trotter that the grids $n \times n \times n$ are not sphere orders for large n [201], the record-cone order on finite sets being containment of balls with time as radius; the proofs here do not use it.

Not claimed

The smallest finite piece without a chart; whether one exists inside $2 \times 3 \times \mathbb{N}$ is the sphere-order form of a question of Felsner, Fishburn and Trotter, and it is open.

Proof. (a) Along each line $t \mapsto (i, j, t)$ the chart's time is monotone, hence continuous off a countable set; take a common point t of continuity of the four lines. There every comparable pair of the four points (i, j, t) is null-separated: a positive margin would let (i, j, s), s slightly above t, still precede (i', j', t). Then $(1, 0, t)$ and $(0, 1, t)$ lie on the null segment from $(0, 0, t)$ to $(1, 1, t)$, by equality in the triangle inequality, and so are comparable, but they are not. The argument needs the real line. (b) Each chain $k \mapsto (i, j, k)$ decreases in time and is bounded below, so its space part is Cauchy, since $\|x(k) - x(k')\| \le t(k) - t(k')$, and it converges. Write L_c for the limit of the chain $c = (i, j)$; it precedes every point of its chain. Four facts are used. (i) If $c \le c'$ then $L_c \preceq L_{c'}$ and the pair is null: $(c, k) \preceq (c', k)$ for every k passes to the limit, and a positive margin $t(L_{c'}) - t(L_c) - \|x(L_{c'}) - x(L_c)\|$ would give $(c, k) \preceq (c', k')$ for some large $k < k'$, which the reversed order forbids. (ii) If the chains c, c' are incomparable and $L_c \preceq L_{c'}$, every point P of chain c' is null from L_c: P lies above $L_{c'}$, hence above L_c, and a positive margin would put P above the points of chain c near L_c, none of which it follows. (iii) If $a \preceq b \preceq d$ and the pair (a, d) is null, then b lies on the null segment from a to d: the triangle inequality $\|x(d) - x(a)\| \le \|x(d) - x(b)\| + \|x(b) - x(a)\| \le t(d) - t(a)$ is an equality, and in an inner product space that forces parallel steps. Any two points of one null line are comparable. (iv) By (i) and (iii), the limits of the chains $(0, 0) \le (0, 1) \le (0, 2) \le (1, 2)$ and $(0, 0) \le (1, 0) \le (1, 1) \le (1, 2)$ all lie on the null segment from L_{00} to L_{12}, so any two of the six limits are comparable. Now there are three cases. *Case 1:* $L_{10} \preceq L_{01}$. Then also $L_{10} \preceq L_{02}$, and by (ii) every point of the chains $(0, 1)$ and $(0, 2)$ is null from L_{10}. The points $(0, 1, 0)$ and $(0, 2, 1)$ both precede $(0, 2, 0)$, so by (iii) both lie on the null segment from L_{10} to $(0, 2, 0)$ and are comparable, but they are incomparable in the order. *Case 2:* $L_{02} \preceq L_{10}$. Then also $L_{02} \preceq L_{11}$, every point of the chains $(1, 0)$ and $(1, 1)$ is null from L_{02}, and the incomparable points $(1, 0, 0)$ and $(1, 1, 1)$, both below $(1, 1, 0)$, are comparable by (iii). *Case 3:* neither. By (iv), $L_{01} \preceq L_{10} \preceq L_{02}$ with strictly increasing times, on one null line. By (ii) the point $(1, 0, 0)$ is null from L_{01} and lies above L_{10}, so by (iii) L_{10} lies on the null segment from L_{01} to $(1, 0, 0)$, which therefore lies on that line; likewise $(0, 2, 0)$ is null from L_{10} and lies above L_{02}, so it lies on the same line. Then $(1, 0, 0)$ and $(0, 2, 0)$ are comparable, but they are not. (c) Minkowski inversion, $u \mapsto u/(t^2 - \|x\|^2)$ on the open future cone, reverses the cone order there. A chart of $2 \times 3 \times \mathbb{N}$, moved so that its least point lies one time unit above the origin, inverted, and relabelled by reversing the two finite coordinates, is a chart of the reversed order with positive times, which (b) excludes. The record cone is the case $E = \mathbb{R}^3$. □

How fast do charts of long runs degrade? A pair of events is at *angular margin* $\theta \in (0, 1]$ when $\theta(\|\Delta x\| + |\Delta t|) \le \|\Delta x\| - |\Delta t|$: it is spacelike by a definite angle, independent of scale. Two chains $A_0 \preceq \cdots \preceq A_{n-1}$ and $B_0 \preceq \cdots \preceq B_{n-1}$ with $A_k \preceq B_k$, where every A_k is incomparable with every earlier B_j, $j < k$, are a *ladder of length n*.

Theorem 24.85 (Long runs crowd every chart)**.** *In classical mathematics:*

(a) in every map of $2\times[n]$, $n\ge 2$, into $\mathbb{R}\times E$, E a real inner product space, that sends comparable pairs into the cone, some incomparable pair has $\|\Delta x\|-|\Delta t|\le 2\,\mathrm{span}/(n-1)$, where span *is the time from the least to the greatest point; so the same holds for every map of $2\times 3\times[n]$, through its subgrid $2\times 1\times[n]$, and for the record cone;*

(b) in a ladder of length n in $\mathbb{R}\times E$, E any real normed space, whose incomparable pairs (A_k,B_j), $j<k$, all have angular margin at least $\theta>0$, $\theta^2(n-1)\le 2$;

(c) in one space dimension the bound is $\theta^2(n-1)\le 1$, and it is sharp: a ladder of length 5 with every margin at least $\frac{1}{2}$ attains it.

Proof. (a) Write the points of $2\times[n]$ as $(0,0,k)$ and $(1,0,k)$, as in the subgrid $2\times 1\times[n]$; only these two chains are used. The pair $(0,0,k+1)$, $(1,0,k)$ is incomparable for every k. At a level k where the time step of the chain $(0,0,\cdot)$ is at most its average, at most $\mathrm{span}/(n-1)$, both points lie within that step of $(0,0,k)$ along cone relations, and the triangle inequality bounds the pair's spacelike margin $\|\Delta x\|-|\Delta t|$ by twice the step. (b) Only the triangle inequality is used; let $n\ge 2$. For a pair (A_k,B_j) put $G=\|x(A_k)-x(B_j)\|$, $\tau=t(A_k)-t(B_j)$ and the two null gaps $p(k,j)=G+\tau$, $q(k,j)=G-\tau$. For $j<k$ the margin $\theta(G+|\tau|)\le G-|\tau|$ gives $\theta q\le p$ and $\theta p\le q$, since $|\tau|\ge\pm\tau$. Put $a_k=q(k,0)$, $b_k=p(k,0)$ and $s_k=t(A_{k+1})-t(A_k)$, so that $\|x(A_{k+1})-x(A_k)\|\le s_k$. Five facts follow. (1) $a_{k+1}\le a_k$ and $b_{k+1}\ge b_k$: the distance to B_0 changes by at most s_k while τ grows by exactly s_k. (2) $(b_{k+1}-b_k)+(a_k-a_{k+1})=2s_k$, since $p-q=2\tau$. (3) $p(k+1,k)\le 2s_k$: from $A_k\preceq B_k$, $\|x(A_{k+1})-x(B_k)\|\le s_k+t(B_k)-t(A_k)$; add $t(A_{k+1})-t(B_k)$. (4) $q(k+1,k)\ge a_{k+1}$: from $B_0\preceq B_k$, $\|x(A_{k+1})-x(B_0)\|\le\|x(A_{k+1})-x(B_k)\|+t(B_k)-t(B_0)$; subtract $t(A_{k+1})-t(B_0)$. (5) $g=a_{n-1}>0$: if $q(n-1,0)=0$ the margin gives $p(n-1,0)=0$, so A_{n-1} and B_0 coincide and are comparable. By (3), the margin of the incomparable pair (A_{k+1},B_k) and (4), $\theta a_{k+1}\le\theta q(k+1,k)\le p(k+1,k)\le 2s_k$, so by (2)

$$\theta a_{k+1}\le(b_{k+1}-b_k)+(a_k-a_{k+1})\qquad(0\le k\le n-2).$$

Twice the step, not the step, is at least θa_{k+1}. Now take square roots $r=\sqrt{g}\le\sigma=\sqrt{a_{k+1}}\le\sigma'=\sqrt{a_k}$, by (1) and (5), and put $\Delta=b_{k+1}-b_k\ge 0$; the claim is $\theta g\le\Delta+2(1+\theta)r(\sigma'-\sigma)$. If $a_k\ge(1+\theta)a_{k+1}$, then $4(1+\theta)^2\sigma^2\le 4(1+\theta)\sigma'^2\le(2+\theta)^2\sigma'^2$, so $(2+2\theta)\sigma\le(2+\theta)\sigma'$, that is $\theta\sigma'\le 2(1+\theta)(\sigma'-\sigma)$, and $\theta g\le\theta r\sigma'\le 2(1+\theta)r(\sigma'-\sigma)$. Otherwise $\sigma'^2<(1+\theta)\sigma^2$, so $\sigma'+\sigma\le 2(1+\theta)\sigma$; multiplying the displayed inequality $\theta\sigma^2\le\Delta+(\sigma'-\sigma)(\sigma'+\sigma)$ by $r^2/\sigma^2\le 1$ gives $\theta g\le\Delta+(\sigma'-\sigma)\,r(r/\sigma)\,2(1+\theta)\le\Delta+2(1+\theta)r(\sigma'-\sigma)$. Adding the $n-2$ steps $k=1,\dots,n-2$, the sums telescope:

$$\theta g(n-2)\le b_{n-1}-b_1+2(1+\theta)\sqrt{g}\,(\sqrt{a_1}-\sqrt{g}).$$

The margins of the pairs (A_{n-1},B_0) and (A_1,B_0) give $b_{n-1}\le g/\theta$ and $b_1\ge\theta a_1$, and the inequality of arithmetic and geometric means gives $2(1+\theta)\sqrt{g}\sqrt{a_1}\le\theta a_1+(1+\theta)^2 g/\theta$. So $\theta g(n-2)\le g/\theta+(1+\theta)^2 g/\theta-2(1+\theta)g=g(2/\theta-\theta)$; dividing by $g/\theta>0$ gives $\theta^2(n-2)\le 2-\theta^2$, that is $\theta^2(n-1)\le 2$. (c) In light-cone coordinates $u=t+x$, $v=t-x$ the cone order of one space dimension is componentwise, and the margin of a spacelike pair is $\min(|\Delta u|,|\Delta v|)/\max(|\Delta u|,|\Delta v|)$. The pair (A_1,B_0) fixes an orientation for all pairs; each pair (A_k,B_{k-1}) has v-gap at least $g=v(B_0)-v(A_{n-1})$, so its u-gap is at least θg; the u-gaps add up

to at least $(n-1)\theta g$ on the pair (A_{n-1}, B_0), whose v-gap is g, so $\theta \cdot (n-1)\theta g \le g$. The ladder $A_k = (k, 0)$, $B_j = (j, 2)$ in light-cone coordinates, $n = 5$, meets every hypothesis with $\theta = \frac{1}{2}$ and equality. □

So no chart of a long run keeps a fixed angular distance from the light cone: the exponent two of one space dimension holds in every dimension, and extra dimensions buy at most a factor $\sqrt{2}$ in θ. The crowding bound does not explain the much faster fall of the margins observed in the numerical charts above.

Remark 24.86 (What is open, exactly)**.** Four statements are open, neither proved nor refuted. (i) The grid $2 \times 3 \times 3 \times 3$ has a record-cone chart read on its stage events. (ii) So does $3 \times 3 \times 3 \times 3$. (iii) Every truncation $2 \times 3 \times [n]$ has one; every $n > 20$ is open. (iv) In every real normed space, a ladder of length n with all angular margins at least $\theta > 0$ has $\theta^2(n-1) \le 1$. Searches found no chart in (i)–(iii) and no ladder above the bound $1/\sqrt{n-1}$ in small cases of (iv); a search that finds nothing proves nothing.

Questions to work through

1. (Theorem 24.5 and Proposition 24.6.) Split the history *dcddcdcd* of the opening into the three legs *dc*, *ddc* and *dcd*. Compute the three proper times, check that their sum is less than $2\sqrt{15}$, and verify the deficit formula of the proposition term by term. Which pair of legs contributes nothing, and why?

2. (Proposition 24.3 and Theorem 24.14(f).) Show that $(D, C) \mapsto (2D, C)$ preserves the componentwise order of $\mathbb{N}\times\mathbb{N}$ but is not an order bijection, and that $(D, C) \mapsto (kD, C/k)$ maps $\mathbb{N} \times \mathbb{N}$ into itself only for $k = 1$. Why is this consistent with the Doppler rescaling being a coherent change of determinant 1?

3. (Theorem 24.12.) With two accounts, which of the pairs of total post count and potentials $(5, (2, -1))$, $(4, (2, -1))$ and $(6, (3, 3))$ are reachable from the empty ledger? Give a posting for each reachable one.

4. (Theorems 24.17 and 24.19(f).) For the quaternionic carrier, $k = 4$, compute $P_4(n)$ for $n = 2, 3, 4, 6$, and show that record counts fail to multiply both at $2 \cdot 2$ and at $2 \cdot 3$.

5. (Theorem 24.31.) Show that agreement at the four times τ_0, $\frac{8}{3}\tau_0$, $\frac{8}{5}\tau_0$ and $\frac{8}{7}\tau_0$ already forces the band 105, and that the other four times of the theorem add nothing.

6. (Theorem 24.39 and Theorem 24.43(d).) Write the first eight posts and balances of the closing coder of $\alpha = (0, 1, 0, 0, \ldots)$, and find where it closes. Then consider the recorder with states $\{0, 1, 2\}$ that starts in 0, moves to state 1 after reading a 0 and to state 2 after reading a 1 (from any state), answers *yes* in state 2 and reads on otherwise. Does it always answer? Follow the proof of Theorem 24.43(d) on the run of zeros and say what it finds.

7. (Theorem 24.53 and Proposition 24.55.) For the core states $f = e_0 + 2e_2 - e_4 - 2e_6$ and $g = e_0 + e_1 - e_4 - e_5$, check that P keeps $\sum_t f(t)^2$ but not $\sum_t |f(t)|$, compute $\kappa(g, \{t\})$ and $\kappa(Pg, \{t\})$ with their totals, and find two positions of g showing that κ is not a function of the amplitude.

Part VII

Classical physics on the line

Part VII asks whether classical physics survives on a line where every number is a construction. It does, and each result says what it uses: some hold on the line with no omniscience principle and no choice, decisions that classical mechanics makes without comment carry a price, exact or between two named principles, and the many results proved only in classical mathematics say so. Chapter 25 moves the mechanics of Kepler and Newton onto $\mathbb{R}_\delta$: bound non-radial orbits return, N bodies keep ten integrals, the perihelion of a relativistic test orbit in a weak-field window advances by $6\pi m/p + \frac{3\pi}{2}(18 + e^2)(m/p)^2$ up to an explicitly bounded remainder, the pulsar-timing observables of such an orbit follow in a distant observer's time in the test-mass limit, every long-wave solution of the lattice's own field equations, in the weak sense of Section 25.12, with the jets of a point mass has $\beta = \gamma = 1$, and deciding, for every body, whether a given principal axis is its axis of largest moment costs exactly LPO. Chapter 26 treats lattice fields, entropy and the second law. It proves that the tick of Recognition Science exists and is unique, and gives the rate at which it approaches equilibrium from five numerical inputs, (T1)–(T5), which the chapter states and does not prove. Chapter 27 summarizes what the comparison cost says about gravity: on transverse-traceless strains its long-wave form is the Fierz–Pauli form of linearized gravity, the locality of a lattice cost selects the Fierz–Pauli completion, the cost built from J agrees with Regge's action, read as the gravitational action, to second order only, and at the Euclidean background its cubic term breaks gauge invariance, so it has no post-Newtonian parameter β of its own, and the lattice parameter $\kappa = 1$ is equivalent to a named premise about records and is not derived. Regge's own long-wave vertices are one half of Einstein's through quartic order on every Euclidean background of the Kuhn family; among hinge charges built from one function of the area, gauge consistency selects Regge's action through cubic order, and for shape costs it fixes Regge's action up to its coupling; four gravitons scatter at tree level as in general relativity on every flat background g_κ of the lattice family off its poles, at long wavelength; and at order K^4 every background with $\kappa > \frac{1}{2}$, $\kappa = 1$ included, is birefringent. The full treatments are the preprints (Zenodo, 2026) [95, 196, 197]. The chapter proves its results on its own pages; where a proof gives a route, or ends in a long finite computation that it describes, the result's scope box says so.

Chapter 25

Orbits and bodies

A planet moves on an ellipse about the Sun and comes back to where it was after one period. On $\mathbb{R}_\delta$ every number is a construction: a position is a regular sequence of ratios, a time is another one, and no comparison of two points is decided unless a rule decides it. The classical proof that the planet returns uses more than that. It solves Kepler's equation by an intermediate value argument, it integrates the time along the orbit, it names the direction of the perihelion, and it extends a local solution of a differential equation to all times. Each of these steps can carry a price on $\mathbb{R}_\delta$. Does a planet on $\mathbb{R}_\delta$, where nothing is decided for free, still return to where it was? And when the force law carries the first relativistic correction, by how much does its orbit turn per revolution?

Here is a small case. Put the gravitational parameter at $m = 1$ and start a body at $x = (1, 0, 0)$ with velocity $v = (0, \frac{6}{5}, 0)$. The energy, the angular momentum and the Laplace–Runge–Lenz vector are

$$E = \tfrac{1}{2}|v|^2 - \frac{m}{|x|} = -\frac{7}{25}, \qquad L = x \times v = \big(0, 0, \tfrac{6}{5}\big), \qquad A = v \times L - m\frac{x}{|x|} = \big(\tfrac{11}{25}, 0, 0\big),$$

and they satisfy $|A|^2 = 2E|L|^2 + m^2$: indeed $\frac{121}{625} = -\frac{504}{625} + 1$. The orbit is the conic $r(m + |A|\cos\theta) = |L|^2$, here $r(1 + \frac{11}{25}\cos\theta) = \frac{36}{25}$, with θ the angle from A: the body is at its perihelion $r = 1$ now and reaches its aphelion $r = \frac{18}{7}$ half a turn later. The semi-major axis is $a = \frac{1}{2}(1 + \frac{18}{7}) = \frac{25}{14} = m/(-2E)$, and Kepler's third law gives the period T by $T^2(-2E)^3 = (2\pi_\delta)^2 m^2$, so $T = 2\pi_\delta(\frac{25}{14})^{3/2} \approx 14.99$. Every number here is a ratio or is computed from ratios, so nothing needs deciding. The questions of this chapter begin when the initial state is a general point of $\mathbb{R}_\delta^6$. Then A may be so small that no finite stage of its regular sequences tells which way the perihelion points, and the classical inverse of Kepler's time map is an intermediate value search.

We prove that none of this stops the planet. Every bound, non-radial motion of Newton's inverse-square field on $\mathbb{R}_\delta$ is in the same state one period later, at every time, and the proof uses no omniscience principle and no choice principle (Theorems 25.20 and 25.21). Kepler's equation is solved with no choice (Theorem 25.10). The perihelion direction is where a price appears: assigning one to every planar state implies LLPO, and LLPO with countable choice provides it (Theorem 25.13).

The relativistic question has a sharp answer too. For the motion of a test body in proper time under the field $\ddot{x} = -mx/|x|^3 - 3mL^2x/|x|^5$, in a weak-field window, the polar angle advances per radial period by

$$\Phi = 2\pi_\delta + 6\pi_\delta\frac{m}{p} + \frac{3\pi_\delta}{2}\,(18 + e^2)\Big(\frac{m}{p}\Big)^2 + R, \qquad 0 \le R \le 160\,\pi_\delta\Big(\frac{m}{p}\Big)^3,$$

where p and e are read off the orbit's own two apsidal radii (Theorem 25.42). This, the proper-time form of Kepler's third law and the period seen by a far observer are proved for the

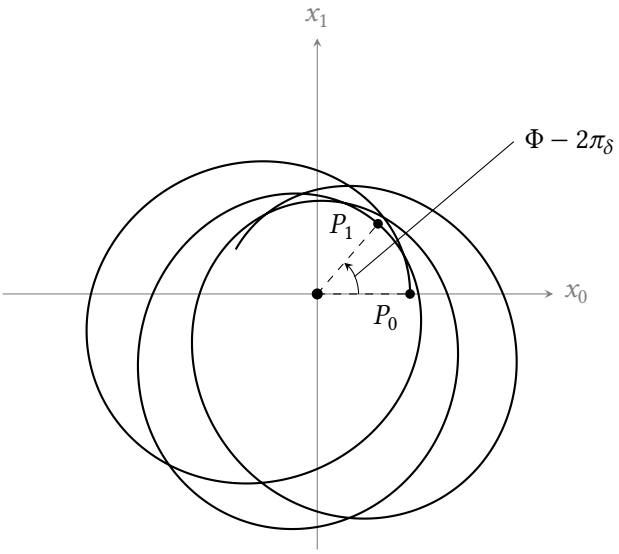


Figure 25.1. An orbit whose inverse radius oscillates in the polar angle with frequency $\kappa = 0.88$: after one radial period the polar angle has advanced by $\Phi = 2\pi_\delta/\kappa$, so the pericentre moves from P_0 to P_1. Newton's field has $\kappa = 1$ and closes; an inverse-cube term gives $\kappa = \sqrt{1 - \epsilon/L^2}$ (Theorem 25.23), and the relativistic field gives the advance of Theorem 25.42. The advance drawn is more than ten times the largest one the weak-field window of Section 25.6 allows.

actual motion, with explicit constants, and again with no omniscience and no choice. As a numerical illustration of the window: the orbit equation $u'' + u = 1 + u^2/100$ of Section 25.6, started at an apside $u = 1$ with $u' = 0$, has an apsidal period Φ with $2.02\,\pi_\delta \le \Phi \le 2.0208\,\pi_\delta$ (Theorem 25.35): after the polar angle turns by Φ, the inverse radius and its rate are back at their starting values. So the apsides advance by between 3.6 and 3.75 degrees per revolution. Figure 25.1 shows such an advancing orbit, with the correction made large enough to see.

Table 25.1 collects the prices. Most entries are free, which is the point: the classical mechanics of orbits and bodies holds on $\mathbb{R}_\delta$ as it stands, and the prices sit exactly at the decisions classical mechanics makes without comment, such as which way the perihelion points, whether an orbit is bound, or which axis of a body has the largest moment of inertia.

The chapter opens Part VII. It uses the line and its prices from Chapter 12, the certified solver of Chapter 15, and the circle functions and π_δ of Section 21.2. Maps are read as in Section 21.1: a map is a rule given as data, computed from regular sequences naming its arguments. Section 25.1 defines motions of polynomial fields and proves that they are determined by their initial states. Sections 25.2 to 25.5 treat Newton's mechanics of orbits: Kepler's laws, other central forces, the return of bound orbits, and N bodies. Sections 25.6 and 25.7 treat the relativistic orbit, first in the polar angle and then in proper time, and Section 25.8 says which of its inputs Recognition Science derives. Section 25.9 shows that the certified solver of Chapter 15 produces motions, and Section 25.10 treats the rigid body.

Two conventions hold throughout. The cosine, the sine and π_δ are those of Definition 21.6. A *quarter period* is a point q with $0 \le q \le 2$ and $\cos q = 0$; by Theorem 21.8 there is exactly one, $q = \pi_\delta/2$, so $4q = 2\pi_\delta$, $\cos q = 0$ and $\sin q = 1$, and the addition formulas of Section 21.2 give $\cos 2q = -1$, $\sin 2q = 0$ and period $4q$ for both functions. Physical readings are identifications: reading x as a position, m as GM and t as a time is classical physics laid over the mathematics,

and no theorem below supplies it.

25.1 Motions on the line

Every force law of this chapter is a polynomial in the state, once the inverse distance is carried as a coordinate of its own. That makes the laws exact objects with ratio coefficients and their derivatives formal.

Definition 25.1 (Polynomial fields and first integrals)**.** A *polynomial field* p on d coordinates gives each coordinate $i < d$ a polynomial p_i in $x_0, \dots, x_{d-1}$ with coefficients in $\mathbb{Q}_\delta$, as in Definition 15.27; we evaluate it at points of $\mathbb{R}_\delta^d$. For a polynomial F with coefficients in $\mathbb{Q}_\delta$ the *Lie derivative* is the polynomial

$$\mathscr{L}_p F = \sum_{i<d} \frac{\partial F}{\partial x_i}\, p_i,$$

formed formally. F is a *first integral* of p when $(\mathscr{L}_p F)(x) = 0$ for every $x \in \mathbb{R}_\delta^d$. A polynomial D is *invariant with rate* G when $\mathscr{L}_p D = G\,D$ at every point.

Definition 25.2 (Motions)**.** A *motion* of a polynomial field p is a map x from $\mathbb{R}_\delta$ to $\mathbb{R}_\delta^d$ such that for every natural number T and every k there is a j with

$$\bigl|x_i(t) - x_i(s) - p_i\bigl(x(s)\bigr)(t-s)\bigr| \le \frac{|t-s|}{k+1}$$

for every $i < d$ and all $s, t \in [-T, T]$ within $1/(j+1)$ of each other. A *motion on* $[0, T]$, for a natural number T, satisfies the same with $s, t \in [0, T]$.

So $p(x(s))$ is a uniform derivative of x in the sense of Definition 12.80 on every bounded interval, with one modulus for all its points, and x has derivative $p(x(s))$ at each s. This is Bishop's notion of a solution [38]. It asks for a modulus, not for a limit of approximations, so a motion carries its own evidence that it solves the equation.

Theorem 25.3 (A motion is determined by its initial state)**.** *Let p be a polynomial field on d coordinates.*

(a) If x and y are motions of p on $[0, T]$ with $x(0) = y(0)$, then $x(t) = y(t)$ for every $t \in [0, T]$.

(b) If x and y are motions of p with $x(0) = y(0)$, then $x(t) = y(t)$ for every $t \in \mathbb{R}_\delta$.

No omniscience principle and no choice principle is used.

Proof. (a) *A motion is bounded.* Take the modulus j_0 of x at $k = 0$ and $h = 1/(j_0 + 1)$. For a natural bound B let $P(B)$ be the sum of the absolute values of the coefficients of all p_i times $(B+1)^{\deg p}$, so that $|p_i(z)| \le P(B)$ when every $|z_i| \le B$. If $|x_i(nh)| \le B_n$ for all i, then $|x_i(t)| \le B_n + h(P(B_n) + 1) =: B_{n+1}$ for $nh \le t \le (n+1)h$. After $(j_0 + 1)T$ steps this bounds x on $[0, T]$ by a natural number, and likewise y; let B bound both.

Table 25.1. What the mechanics of this chapter costs on $\mathbb{R}_\delta$. "Free" means no omniscience principle and no choice principle; "classical" means proved with excluded middle and choice.

Statement	Price
Two motions with one initial state agree; first integrals are constant along motions	free
Kepler's first integrals, the vector A, the conic; Kepler's equation; the third law	free
Every bound, non-radial Kepler motion returns after one period, at every time	free
Every planar physical state has a perihelion direction	implies LLPO; LLPO with AC_ω gives it
Bound, parabolic or unbound, decided for every state	LPO
A vector $v \times L - g\,x$ is conserved only for the inverse square; Fradkin's tensor	free
"No precession" decided for every real frequency $\kappa > 0$	WLPO
Binet's clock for an inverse-cube correction; precession exactly when $\epsilon \neq 0$	free
Ten integrals of N bodies; two bodies; Lagrange's rotating triangle is the motion	free
The physical state $u_{ab} = 1/\lvert x_a - x_b\rvert$ from non-collision	$\mathrm{MP}_{\mathrm{bin}}$
Post-Newtonian advance, second order, proper-time Kepler law, far observer's period	free
Certified runs are motions; schedule independence; Lipschitz data; short times	free
Exact conservation along certified runs; invariant families preserved	free
Completeness of the region check; one fixed machine for short times	classical
One conversion certificate for a whole polynomial theory	classical (one clause)
Rigid body: integrals, stability, Chetaev's escape from the middle axis	free
Deciding whether an axis is a body's axis of largest moment	LPO

The field is Lipschitz on the ball. Expanding monomials, $|p_i(z) - p_i(z')| \le K \max_l |z_l - z'_l|$ for $|z_l|, |z'_l| \le B$, with an explicit natural number K.

The two motions stay together. Fix k, let j serve both motions at the tolerance k, fix $t \in [0,T]$, take a natural number $N \ge (j+1)T$ and put $t_n = nt/N$, so that consecutive points are within $1/(j+1)$. With $e_n = \max_i |x_i(t_n) - y_i(t_n)|$ and $h = t/N$,

$$x(t_{n+1}) = x(t_n) + p\big(x(t_n)\big)h + \rho_n, \qquad y(t_{n+1}) = y(t_n) + p\big(y(t_n)\big)h + \rho'_n,$$

with $|\rho_n|, |\rho'_n| \le h/(k+1)$. Hence $e_{n+1} \le (1 + Kh)e_n + 2h/(k+1)$ and, since $e_0 = 0$,

$$e_N \le (1+Kh)^N \frac{2Nh}{k+1} \le \frac{2Te^{KT}}{k+1},$$

using $1 + u \le e^u$ (Section 21.2). This holds for every k, so $x(t) = y(t)$.

(b) An equation between points of $\mathbb{R}_\delta$ is the negation of apartness, so it may be proved by cases that hold only under double negation. Under $t \ge 0$, restrict both motions to $[0,T]$ with $T \ge t$ natural and apply (a). Under $t \le 0$, the curves $s \mapsto x(-s)$ and $s \mapsto y(-s)$ are motions of the field $-p$ on $[0,T]$ with $T \ge -t$, and (a) applies again. Since $\neg\neg(t \le 0 \vee t \ge 0)$ holds, $x(t) = y(t)$. □

The discrete Gronwall step is the one of Chapter 15 (Theorem 15.8) and goes back to Gronwall [159]. Uniqueness costs nothing; what the classical theory adds is existence, and on $\mathbb{R}_\delta$ each existence statement below comes with a construction.

Lemma 25.4 (Calculus along motions)**.** *Let x be a motion of p and F a polynomial.*

(a) For every point c the curve $s \mapsto x(s + c)$ is a motion of p.

(b) The function $t \mapsto F(x(t))$ has the uniform derivative $(\mathscr{L}_p F)(x(t))$ on every bounded interval.

(c) A function with uniform derivative c, a constant, on every bounded interval satisfies $f(t) = f(0) + ct$ for every t.

The same holds for motions on $[0,T]$, with $[0,T]$ in place of the line. No omniscience principle and no choice principle is used.

Proof. (a) If $|s| \le T$ then $|s + c| \le T + C$ for a natural $C \ge |c|$; use the modulus of x on $[-T-C, T+C]$. (b) On a ball bounding x over the interval, Taylor's formula for the polynomial F gives $|F(z') - F(z) - \nabla F(z)\cdot(z'-z)| \le C|z'-z|^2$, and $|x(t) - x(s)| \le (P+1)|t-s|$ once $|t-s|$ is within the modulus at $k=0$; the derivative of x is $p(x)$, and $\nabla F \cdot p = \mathscr{L}_p F$. (c) Divide $[0,t]$ into N steps within the modulus at tolerance k; the errors add up to at most $|t|/(k+1)$, for every k. □

Proposition 25.5 (First integrals are constant along motions)**.** *If F is a first integral of p, then $F(x(t)) = F(x(0))$ for every motion x of p and every t, and for every motion on $[0,T]$ and every $t \in [0,T]$. No omniscience principle and no choice principle is used.*

Proof. By Lemma 25.4(b) the derivative of $F(x(t))$ is $(\mathscr{L}_p F)(x(t)) = 0$, and (c) with $c = 0$ gives the claim. □

Two constructions recur in the relativistic sections, where a time must be computed from an angle and an angle from a time. Proposition 12.85 forms the integral of a uniformly continuous function with countable choice, for the limit. When the integrand is a map with a Lipschitz constant, the limit has an explicit rate and no choice is needed.

Lemma 25.6 (Antiderivatives of Lipschitz maps)**.** *Let g be a map on $\mathbb{R}_\delta$ and K a natural number with $|g(t) - g(s)| \le K|t - s|$ for all s, t. There is a map G with $G(0) = 0$ and*

$$\big|G(t) - G(s) - g(s)(t - s)\big| \le K|t - s|^2 \qquad \text{for all } s, t,$$

so G has the uniform derivative g on every bounded interval. No omniscience principle and no choice principle is used.

Proof. For a natural number N take the grid $a_i = i/(N+1) - N$, $0 \le i \le 2N(N+1)$, of step $1/(N+1)$ from $-N$ to N. The clamp $c(t;\alpha,\beta) = \frac{1}{2}(\alpha + \beta + |t - \alpha| - |t - \beta|)$ equals α left of α, t between and β right of β, so $\ell_i(s,t) = c(t;a_i,a_{i+1}) - c(s;a_i,a_{i+1})$ is the signed length of the part of $[s,t]$ in the i-th cell. Put $G_N(t) = \sum_{i<2N(N+1)} g(a_i)\,\ell_i(0,t)$: a map with $G_N(0) = 0$ which, for $|t| \le N$ and t of either sign, is the integral from 0 to t of the step function equal to $g(a_i)$ on the i-th cell. The identity $c(t;\alpha,\beta) + c(t;\beta,\gamma) = c(t;\alpha,\gamma) + \beta$ is a ring identity, so for $|s|, |t| \le N$ the $\ell_i(s,t)$ add up to $t - s$; the clamp is nondecreasing, so they are all ≥ 0 if $s \le t$ and all ≤ 0 if $t \le s$, and either way their absolute values add up to $|t - s|$. Hence $G_N(t) - G_N(s) - g(s)(t - s) = \sum_i (g(a_i) - g(s))\,\ell_i(s,t)$, and $\ell_i(s,t)$ is 0 unless the cell meets $[s,t]$, where $|a_i - s| \le \frac{1}{N+1} + |t - s|$. Each case split here concludes an equation or an inequality, which holds once it holds up to double negation, so nothing is decided, and

$$\big|G_N(t) - G_N(s) - g(s)(t - s)\big| \le K\Big(\frac{1}{N+1} + |t - s|\Big)|t - s| \qquad \text{for } |s|, |t| \le N.$$

For $|t| \le R \le N, M$, the difference $D = G_N - G_M$ therefore changes by at most $K(\frac{1}{N+1} + \frac{1}{M+1} + 2|t' - s'|)|t' - s'|$ between points $s', t' \in [-R, R]$; adding this over n equal steps from 0 to t, with $D(0) = 0$, gives $|D(t)| \le K(\frac{1}{N+1} + \frac{1}{M+1})|t| + 2K|t|^2/n$ for every n, so $|G_N(t) - G_M(t)| \le K(\frac{1}{N+1} + \frac{1}{M+1})|t|$. A regular sequence naming t gives a natural number $B \ge |t|$, and then the points $G_{\mu(k)}(t)$, $\mu(k) = 2KB(k+1) + B$, form a Cauchy family. Theorem 12.25 gives its limit $G(t)$ with no choice; $G_N(t)$ tends to it, and limits are unique (Theorem 12.26), so $G(t)$ does not depend on the sequence naming t: G is a map, and $G(0) = 0$. For N at least natural bounds of $|s|$ and $|t|$ the remainder of G_N is at most $K|t - s|^2 + K|t - s|/(N+1)$, and the last term tends to 0, which gives the inequality for G. □

Lemma 25.7 (Inverting a clock)**.** *Let θ be a map on $\mathbb{R}_\delta$ with $\theta(0) = 0$ and a natural number K with $t - s \le \theta(t) - \theta(s) \le K(t - s)$ whenever $s \le t$. There is a map ψ with $\theta(\psi(y)) = y$ and $\psi(\theta(t)) = t$ for all y and t. If θ has the uniform derivative H on every bounded interval, and $HG = 1$ with $|G| \le 1$, then ψ has the uniform derivative $G \circ \psi$ on every bounded interval. No omniscience principle and no choice principle is used.*

Proof. Given y, a regular sequence naming it gives a natural $B \ge |y|$, and the root of $t \mapsto \theta(t) - y$ lies in $[-B, B]$ because $|\theta(t)| \ge |t|$. On $[-B, B]$ the function is uniformly continuous with modulus from K, it has the modulus of uniqueness $\mu(k) = k$ (two near-roots t, t' satisfy $|t - t'| \le |\theta(t) - \theta(t')|$), and its values at grid points are computed from those of θ. Theorem 12.92, after rescaling $[-B, B]$ to $[0, 1]$, computes the root. Two names of y give roots with the same image, and θ is injective, so ψ is a map. For the derivative, with $\xi = \psi(s)$ and $\xi' = \psi(t)$, $\psi(t) - \psi(s) - G(\xi)(t - s) = -G(\xi)\big(\theta(\xi') - \theta(\xi) - H(\xi)(\xi' - \xi)\big)$, and $|\xi' - \xi| \le |t - s|$. □

25.2 Kepler from Newton

Newton's inverse-square force $\ddot{x} = -mx/|x|^3$ is not a polynomial in the position. It becomes one when the inverse distance is carried as a coordinate u with its own equation of motion, and the gravitational parameter m as a coordinate that does not move.

Definition 25.8 (The Kepler field)**.** The *Kepler field* acts on states $(x, v, u, m) \in \mathbb{R}^8_\delta$, with $x, v \in \mathbb{R}^3_\delta$:

$$\dot{x} = v, \qquad \dot{v} = -m\,u^3 x, \qquad \dot{u} = -u^3\,(x \cdot v), \qquad \dot{m} = 0.$$

Its quantities are the energy $E = \frac{1}{2}|v|^2 - mu$, the angular momentum $L = x \times v$, the Laplace–Runge–Lenz vector $A = v \times L - mu\,x$ and the invariant $I = u^2|x|^2$. A state is *physical* when $I = 1$, $u > 0$ and $m \ge 0$; then $u = 1/|x|$, and the field is Newton's.

The coordinate u obeys the equation that $1/|x|$ obeys: $\frac{d}{dt}|x|^{-1} = -|x|^{-3}(x \cdot v)$. On physical states the two agree, and the next theorem shows that the field keeps $I - 1$ at zero at the rate $I - 1$ itself.

Theorem 25.9 (Kepler's integrals, the vector A and the conic)**.** *For the Kepler field p and every state of $\mathbb{R}^8_\delta$:*

(a) E and the three components of L are first integrals;

(b) $\mathscr{L}_p A_k = m\,u\,v_k\,(I - 1)$ for $k = 0, 1, 2$, and $\mathscr{L}_p(I - 1) = -2u^2(x \cdot v)(I - 1)$;

(c) $u\,(A \cdot x) - u|L|^2 + mI = 0$ and $|A|^2 = 2E|L|^2 + m^2 I$.

In particular, on states with $I = 1$ the rate of A vanishes, and with $r = 1/u$,

$$A \cdot x + m\,r = |L|^2, \qquad |A|^2 = 2E|L|^2 + m^2.$$

No omniscience principle and no choice principle is used.

Proof. All three parts are identities of polynomials with ratio coefficients, and $\mathbb{R}_\delta$ is a commutative ring (Proposition 12.19), so ordinary vector algebra proves them.

(a) $\mathscr{L}_p E = v\cdot\dot v - m\dot u = -mu^3(x\cdot v) + mu^3(x\cdot v) = 0$, and $\mathscr{L}_p L = v\times v + x\times(-mu^3 x) = 0$.

(b) With $x\times(x\times v) = x(x\cdot v) - v|x|^2$ and $\dot L = 0$,

$$\begin{aligned}\dot A &= \dot v\times L - m\dot u\, x - mu\, v = -mu^3 x\times(x\times v) + mu^3(x\cdot v)x - mu\, v\\ &= mu^3|x|^2 v - mu\, v = mu\, v(I-1).\end{aligned}$$

And $\dot I = 2u\dot u|x|^2 + 2u^2(x\cdot v) = -2u^4|x|^2(x\cdot v) + 2u^2(x\cdot v) = -2u^2(x\cdot v)(I-1)$.

(c) $A\cdot x = (v\times L)\cdot x - mu|x|^2 = L\cdot(x\times v) - mu|x|^2 = |L|^2 - mu|x|^2$; multiplying by u gives the first identity. Since v is orthogonal to L and $x\cdot(v\times L) = |L|^2$,

$$|A|^2 = |v|^2|L|^2 - 2mu\,|L|^2 + m^2u^2|x|^2 = 2E|L|^2 + m^2 I.$$

On $I = 1$ divide the first identity by u. □

When A has a direction n, so that $A = |A|n$, and θ is the angle from n to x, the identity $A\cdot x + mr = |L|^2$ is the conic $r(m + |A|\cos\theta) = |L|^2$ with eccentricity $e = |A|/m$, which is an ellipse when $E < 0$. Kepler found the ellipse and the equal areas [160] and the third law [161]; Newton derived them from the inverse-square force [166]. The vector A goes back to Laplace and was used by Runge and by Lenz [164, 167, 165]; Goldstein traces its history [158].

The classical step from position to time is Kepler's equation $M = \eta - e\sin\eta$ between the mean anomaly M, which grows uniformly in time, and the eccentric anomaly η. Its inverse is classically an intermediate value argument, and on $\mathbb{R}_\delta$ the intermediate value theorem costs LLPO with countable choice (Theorem 12.54). The map $\eta \mapsto \eta - e\sin\eta$ increases at rate at least $1-e$, and that rate removes the price.

Theorem 25.10 (Kepler's equation is solved with no choice)**.** *Let h be a map on $\mathbb{R}_\delta$ with $|h(\theta)| \le 1$ and $|h(\alpha) - h(\beta)| \le |\alpha - \beta|$ for all points. For all points e and M with $0 \le e < 1$ there is exactly one η with $\eta - e\,h(\eta) = M$. In particular Kepler's equation $\eta - e\sin\eta = M$ has exactly one solution. No omniscience principle and no choice principle is used.*

Proof. Put $k(\eta) = \eta - e\,h(\eta)$. For $\alpha \le \beta$, $k(\beta) - k(\alpha) \ge (\beta-\alpha) - e|h(\beta) - h(\alpha)| \ge (1-e)(\beta-\alpha)$.

Uniqueness. If $k(\eta) = k(\eta') = M$ and $\eta < \eta'$, then $k(\eta') - k(\eta) \ge (1-e)(\eta'-\eta) > 0$, which is false; so $\eta' \le \eta$, and likewise $\eta \le \eta'$. Hence $\eta = \eta'$.

Existence. On $[0,1]$ put $f(t) = k(M + 2t - 1) - M$. Then $f(0) = -1 - e\,h(M-1) \le -1 + e \le 0$ and $f(1) = 1 - e\,h(M+1) \ge 1 - e \ge 0$, and $|f(t) - f(t')| \le 4|t - t'|$, a modulus of uniform continuity given as a function. Since $1 - e > 0$, a positive stage gives a natural N with $1 - e \ge 1/(N+1)$ (Lemma 12.16). The inequality $|t - t'| \le \frac{N+1}{2}|f(t) - f(t')|$ may be proved by cases on the order of t and t', and each case follows from the rate of k; so $\mu(k) = (N+1)(k+1)$ is a modulus of uniqueness. Finally h is a map in the sense of Section 21.1: its value at each grid point $M + 2j/3^n - 1$ is named by a regular sequence computed from n, j and the sequences naming M and e. Theorem 12.92 computes a root t^* of f, and $\eta = M + 2t^* - 1$ solves the equation.

The sine qualifies. $|\sin| \le 1$ by Section 21.2. For the Lipschitz bound, $\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$ by the addition formulas, so it suffices that $|\sin\theta| \le |\theta|$. By cotransitivity $|\theta| > 1$ or $|\theta| < 2$. In the first case $|\sin\theta| \le 1 < |\theta|$. In the second $\sin\theta = \theta\,\mathrm{S}(\theta^2)$ with $0 \le \theta^2 \le 4$, and $\mathrm{S}(s) - 1 = \sum_{i\ge 0}\big(-s^{2i+1}/(4i+3)! + s^{2i+2}/(4i+5)!\big)$ has nonpositive brackets for $s \le 20$, while $\mathrm{S}(s) \ge 1 - s/6 \ge 0$ for $s \le 4$; so $0 \le \mathrm{S}(\theta^2) \le 1$. □

Without the rate the same root would need the presented intermediate value theorem of Theorem 12.92 with a modulus of uniqueness from elsewhere, or LLPO with countable choice. Kepler's map supplies the modulus itself: two near-solutions are near each other because the map increases at rate $1 - e$.

The ellipse of Kepler's first law, run in the eccentric anomaly, solves the field exactly, and its period is $2\pi_\delta$ in η.

Theorem 25.11 (Kepler's third law, with no choice)**.** *Let m, a, b, n, e be points with $a > 0$, $0 \le e < 1$, $b \ge 0$, $b^2 = a^2(1-e^2)$, $n > 0$ and $n^2a^3 = m$. Put $\rho(\eta) = 1 - e\cos\eta$,*

$$X(\eta) = \Big(a(\cos\eta - e),\ b\sin\eta,\ 0,\ -\frac{na\sin\eta}{\rho},\ \frac{nb\cos\eta}{\rho},\ 0,\ \frac{1}{a\rho},\ m\Big), \qquad t(\eta) = \frac{\eta - e\sin\eta}{n}.$$

Then for every η:

(a) $X(\eta)$ is a physical state with $I = 1$ and $2a\,E = -m$;

(b) t has derivative $\rho(\eta)/n$ at η, and each coordinate X_i has derivative $\frac{\rho(\eta)}{n}\,p_i(X(\eta))$ at η, where p is the Kepler field;

(c) $X(\eta + 2\pi_\delta) = X(\eta)$ and $t(\eta + 2\pi_\delta) = t(\eta) + 2\pi_\delta/n$;

(d) the period $T = 2\pi_\delta/n$ satisfies $T^2 m = (2\pi_\delta)^2 a^3$.

No omniscience principle and no choice principle is used.

Proof. $\rho \ge 1 - e > 0$, so $1/\rho$ exists. Write $c = \cos\eta$ and $s = \sin\eta$.

(a) $|x|^2 = a^2(c-e)^2 + b^2s^2 = a^2(c-e)^2 + a^2(1-e^2)(1-c^2) = a^2(1-ec)^2 = a^2\rho^2$, so $u = 1/(a\rho)$ gives $I = 1$. Next $a^2s^2 + b^2c^2 = a^2(1-e^2c^2) = a^2\rho(1+ec)$, so $|v|^2 = n^2a^2(1+ec)/\rho$ and

$$E = \frac{n^2a^2(1+ec)}{2\rho} - \frac{m}{a\rho} = \frac{n^2a^2\big((1+ec)-2\big)}{2\rho} = -\frac{n^2a^2}{2} = -\frac{m}{2a},$$

using $m = n^2a^3$ and $(1+ec) - 2 = -\rho$.

(b) The sine and cosine have derivatives cos and $-\sin$ at every point, with no choice: by the addition formulas, $\sin(\eta+h) - \sin\eta - h\cos\eta = \sin\eta(\cos h - 1) + \cos\eta(\sin h - h)$, which is at most $h^2 + |h|^3$ in absolute value for $|h| \le 1$ (Section 21.2), and likewise for the cosine. Sums, products and quotients by the positive ρ follow the usual rules, with the usual proofs. Then $t' = (1-ec)/n = \rho/n$, and the coordinates check one by one. For instance $\frac{d}{d\eta}a(c-e) = -as = \frac{\rho}{n}\cdot v_0$ with $v_0 = -nas/\rho$. With $\rho' = es$,

$$\frac{dv_0}{d\eta} = -na\,\frac{c\rho - s\cdot es}{\rho^2} = -na\,\frac{c-e}{\rho^2},$$

because $c(1-ec)-es^2 = c-e$; and $\frac{\rho}{n}(-mu^3x_0) = \frac{\rho}{n}\cdot(-n^2a^3)\frac{a(c-e)}{a^3\rho^3} = -na\frac{c-e}{\rho^2}$. For u: $\frac{d}{d\eta}\frac{1}{a\rho} = -\frac{es}{a\rho^2}$, and $x\cdot v = \big(-na^2s(c-e)+nb^2sc\big)/\rho = na^2es(1-ec)/\rho = na^2es$, so $\frac{\rho}{n}(-u^3x\cdot v) = -\frac{\rho}{n}\cdot\frac{na^2es}{a^3\rho^3} = -\frac{es}{a\rho^2}$. The coordinates x_1, v_1 are similar, and x_2, v_2, m are constant.

(c) The sine and cosine have period $4q = 2\pi_\delta$.

(d) $T^2m = (2\pi_\delta)^2m/n^2 = (2\pi_\delta)^2a^3$. □

Conditions and scope.

Base and premises

The line $\mathbb{R}_\delta$, its circle functions and π_δ, and the Kepler field. Nothing is decided and nothing is chosen.

Conclusion

Kepler's ellipse, run in the eccentric anomaly, is a solution of Newton's field whose period obeys $T^2m = (2\pi_\delta)^2a^3$, and every mean anomaly has exactly one eccentric anomaly.

Identification

Reading (x, v) as a body's position and velocity, m as GM, and t as time is classical physics; the theorem does not supply it. Under the classical reading π_δ is π.

Not claimed

That every motion of the field lies on such an ellipse; that is Theorem 25.20.

The first price in classical mechanics appears at the direction of the perihelion. On a circular orbit $A = 0$, and near one A is a vector whose regular sequences may stay near 0 for as long as one cares to read them.

Definition 25.12 (Directions and perihelion directions)**.** *Every plane vector has a direction* when for all points a, b there are points n_0, n_1, l with $n_0^2 + n_1^2 = 1$, $l \geq 0$, $a = ln_0$ and $b = ln_1$. *Every planar physical state has a perihelion direction* when every physical Kepler state with $x_2 = v_2 = 0$ has a unit vector (n_0, n_1) and $l \geq 0$ with $(A_0, A_1) = l(n_0, n_1)$. *Every physical state has a perihelion direction* when every physical Kepler state has a unit vector n with $n \cdot L = 0$ and $l \geq 0$ with $A = ln$.

Theorem 25.13 (The perihelion direction lies between LLPO and LLPO with countable choice)**.**

(a) Every planar physical state has a perihelion direction if and only if every plane vector has a direction, and if and only if every physical state has a perihelion direction.

(b) If every plane vector has a direction, then LLPO holds.

(c) Assuming AC_ω*, LLPO implies that every plane vector has a direction.*

Apart from the countable choice named in (c), no choice principle is used.

Proof. (a) Every plane vector (a, b) is the vector A of a planar physical state: take $x = (1, 0, 0)$, $u = 1$, $t = a^2 + 1$, $v = (-b/t,\ t,\ 0)$ and $m = t^2 - a$. Then $L = (0, 0, t)$, $v \times L = (t^2,\ b,\ 0)$ and $A = (t^2, b, 0) - m(1, 0, 0) = (a, b, 0)$; and $m \geq 0$ because $2m = 2a^4 + 3a^2 + (a - 1)^2 + 1$. So perihelion directions for planar states give directions for plane vectors. Conversely a direction of (A_0, A_1) is a perihelion direction.

For the spatial form, the state just built has $L = (0, 0, t)$ with $t \geq 1$, so a unit n with $n \cdot L = 0$ has $n_2 = 0$, and (n_0, n_1) is a direction of (a, b). Conversely, assume that plane vectors have directions. Then space vectors have directions: write $(a, b) = l_1(m_0, m_1)$ and $(l_1, c) = l(k_0, k_1)$; then $(k_0 m_0, k_0 m_1, k_1)$ is a unit vector and $(a, b, c) = l(k_0 m_0, k_0 m_1, k_1)$. By (b) LLPO holds, and with it $yz = 0$ implies $y = 0$ or $z = 0$: LLPO decides $|y| \leq |z|$ or $|z| \leq |y|$, and in the first case $y^2 \leq |yz| = 0$. Now let a physical state be given, and let $A = ln$ with n a unit vector. The identities of Theorem 25.9 give $A \cdot L = (v \times L) \cdot L - mu\, x \cdot (x \times v) = 0$, so $l(n \cdot L) = 0$. If $n \cdot L = 0$ we are done. If $l = 0$, then $A = 0$, and any unit vector orthogonal to L serves: write $L = \lambda k$ with k a unit vector; by cotransitivity $k_0^2 < \frac{1}{2}$ or $k_0^2 > \frac{1}{4}$, so $k \times (1, 0, 0)$ or $k \times (0, 1, 0)$ has squared length above $\frac{1}{4}$, and its direction p satisfies $p \cdot L = \lambda\,(p \cdot k) = 0$.

(b) Let α be a binary sequence with at most one entry 1. Put $a = \sum_{n\ \mathrm{even}} \alpha(n) 2^{-n}$ and $b = \sum_{n\ \mathrm{odd}} \alpha(n) 2^{-n}$, and let $(a, b) = l(n_0, n_1)$ with $n_0^2 + n_1^2 = 1$. By cotransitivity $n_0^2 > \frac{1}{3}$ or $n_0^2 < \frac{1}{2}$, and in the second case $n_1^2 > \frac{1}{2}$. Suppose $n_0^2 > \frac{1}{3}$. If $\alpha(n) = 1$ for an odd n, then $a = 0$, so $l n_0 = 0$ with n_0 apart from 0; hence $l = 0$ and $b = 0$, although $b = 2^{-n}$. So α vanishes at every odd place. If $n_1^2 > \frac{1}{3}$ the same argument shows that α vanishes at every even place. That is LLPO.

(c) Let (a, b) be given. On $[0, 1]$ the path $P(t) = 1 - |2t - 1|$, $Q(t) = 2t - 1$ runs from $(0, -1)$ through $(1, 0)$ to $(0, 1)$, and $P(t)^2 + Q(t)^2 \geq \frac{1}{2}$. The cross product $f(t) = P(t)\, b - Q(t)\, a$ is uniformly continuous with $f(0) = a$ and $f(1) = -a$. LLPO decides $a \geq 0$ or $a \leq 0$ (Corollary 12.18); in the first case use $f(1 - t)$, in the second $f(t)$, so that the values at the ends have the order the intermediate value theorem asks. With countable choice, LLPO gives that theorem (Theorem 12.54), hence a t^* with $f(t^*) = 0$. Put $(P, Q) = (P(t^*), Q(t^*))$, $N = \sqrt{P^2 + Q^2} > 0$ and $D = aP + bQ$. Since the cross product vanishes, $(a, b) = (D/N^2)(P, Q)$. LLPO decides $D \geq 0$ or $D \leq 0$: in the first case $n = (P, Q)/N$ and $l = D/N$, in the second $n = -(P, Q)/N$ and $l = -D/N$. □

Near a circular orbit the vector A is nearly 0, and no finite reading of the state tells which way the perihelion points. The theorem makes that precise: the direction is not merely hard to compute, it is equivalent to a decision about a binary sequence. Everything the orbit does in time, as the next sections show, needs no such decision.

The type of an orbit is the sign of its energy, and the four prices of signs on the line (Corollary 12.18) reappear exactly.

Theorem 25.14 (What deciding the orbit type costs)**.** *Every point of $\mathbb{R}_\delta$ is the energy of a physical Kepler state. Consequently, made for every physical state, the assertion $E < 0 \vee E = 0 \vee E > 0$ is equivalent to LPO; the assertion $E = 0 \vee E \neq 0$ to WLPO; the assertion $E \leq 0 \vee E \geq 0$ to LLPO; and the assertion $E \neq 0 \to E < 0 \vee E > 0$ to $\mathrm{MP}_{\mathrm{bin}}$. No choice principle is used.*

Proof. Given a point z, take $x = (1, 0, 0)$, $u = 1$, $m = |z| + 1$ and $v = (0, \sqrt{2(z + m)}, 0)$, using the free square root of the nonnegative point $2(z + m) \geq 2$ (Proposition 12.19). Then the state is

physical and $E = (z + m) - m = z$. Each assertion for all physical states is therefore the same assertion for all points, and Corollary 12.18 names its price. □

25.3 Other central forces

Newton's force is one member of a family. Replace u^3 in the force by u^k: on physical states the acceleration becomes $-m\,x/|x|^k$, a central force of strength m/r^{k-1}. Which of the Kepler identities survive?

Definition 25.15 (Central fields)**.** For a natural number k the *central field* c_k acts on states $(x, v, u, m) \in \mathbb{R}^8_\delta$ by

$$\dot{x} = v, \qquad \dot{v} = -m\,u^k x, \qquad \dot{u} = -u^3(x \cdot v), \qquad \dot{m} = 0.$$

So c_3 is the Kepler field, c_{n+3} has the force m/r^{n+2} on physical states, and c_0 is the harmonic force $\ddot{x} = -mx$. For a polynomial g in u and m alone, the *vector* $A_g = v \times L - g\,x$; the Kepler vector is A_{mu}. The *Fradkin tensor* is $F_{ij} = v_i v_j + m\,x_i x_j$ for $i, j < 3$.

Theorem 25.16 (Integrals of central forces)**.**

(a) For every k the three components of $L = x \times v$ are first integrals of c_k.

(b) For every n, the energy $\tfrac{1}{2}|v|^2 - m\,u^{n+1}/(n+1)$ is a first integral of c_{n+3}.

(c) $\tfrac{1}{2}|v|^2 + \tfrac{1}{2}m|x|^2$ is a first integral of c_0, and so is every component F_{ij} of the Fradkin tensor.

No omniscience principle and no choice principle is used.

Proof. (a) $\dot{L} = x \times (-mu^k x) = 0$. (b) The derivative is $-mu^{n+3}(x \cdot v) - mu^n \dot{u} = -mu^{n+3}(x \cdot v) + mu^{n+3}(x \cdot v) = 0$. (c) $v \cdot (-mx) + m\,x \cdot v = 0$, and $\dot{F}_{ij} = -mx_i v_j - m v_i x_j + m(v_i x_j + x_i v_j) = 0$. □

The harmonic oscillator's tensor is due to Fradkin [157]; its eigenvectors are the axes of the ellipse that every harmonic orbit traces. For the vector A_g the answer is the opposite of generous.

Theorem 25.17 (A conserved vector only for the inverse square)**.** *Let k be a natural number. There is a polynomial g in u and m alone such that every component of $A_g = v \times L - g\,x$ has Lie derivative 0 along c_k at every state with $I = 1$ if and only if $k = 3$. For $k = 3$, every such g satisfies $g = mu$ at every state with $u > 0$. No omniscience principle and no choice principle is used.*

Proof. The rate. Since L is conserved, $\frac{d}{dt}(v \times L) = -mu^k x \times (x \times v) = -mu^k\big((x \cdot v)x - |x|^2 v\big)$ and $\frac{d}{dt}(gx) = \partial_u g\,\dot{u}\,x + g\,v = -\partial_u g\,u^3(x \cdot v)x + g\,v$, so for each component l

$$\mathscr{L}_{c_k}(A_g)_l = (x \cdot v)\,x_l\big(u^3 \partial_u g - mu^k\big) + v_l\big(mu^k|x|^2 - g\big). \tag{25.1}$$

If $k = 3$. With $g = mu$ the first bracket is $mu^3 - mu^3 = 0$ and the second is $mu(u^2|x|^2 - 1) = mu(I - 1)$, which vanishes when $I = 1$.

Only if. For natural numbers $N \geq 1$ and M take the states $x = (1/N, 0, 0)$, $u = N$, $m = M$, which have $I = 1$. With $v = (0, 1, 0)$ and $l = 1$, (25.1) gives $MN^{k-2} - g(N, M) = 0$; with $v = (1, 0, 0)$ and $l = 0$ it gives $N^{-2}\big(N^3 \partial_u g(N, M) - MN^k\big) + MN^{k-2} - g(N, M) = 0$, that is $N\partial_u g(N, M) = g(N, M)$. So the polynomials $u^2 g - mu^k$ and $u\,\partial_u g - g$, which have ratio coefficients, vanish at every point of $\{1, 2, \dots\} \times \{0, 1, \dots\}$. Such a polynomial is zero: written as $\sum_j P_j(u) m^j$, for each fixed N it has infinitely many roots in m, so every $P_j(N) = 0$, and then every P_j has infinitely many roots. Hence $u^2 g = mu^k$ as polynomials. For $k \leq 1$ the right side has no factor u^2, which is impossible; so $k \geq 2$ and $g = mu^{k-2}$. Then $u\,\partial_u g = (k - 2)mu^{k-2}$ equals $g = mu^{k-2}$, so $k - 2 = 1$.

Uniqueness. For $k = 3$ the same rows give $u^2 g = mu^3$ as polynomials, so $g = mu$ as polynomials, and $g(u, m) = mu$ at every state. □

Bertrand proved that among central forces only the inverse square and the harmonic force have all bounded orbits closed [152]. The theorem is the algebraic core of that statement on $\mathbb{R}_\delta$: a conserved vector pointing at the pericentre, built from the state with a coefficient depending only on the distance, exists only for the inverse square. The harmonic force has a tensor instead of a vector, because its orbits are centred ellipses with two pericentres.

What happens when Newton's force is perturbed? The simplest perturbation that keeps the orbit equation linear is an inverse-cube term.

Definition 25.18 (The inverse-cube correction)**.** For a ratio ϵ, the field c^ϵ is $\dot{x} = v$, $\dot{v} = -mu^3 x - \epsilon u^4 x$, $\dot{u} = -u^3(x \cdot v)$, $\dot{m} = 0$; on physical states the force is $m/r^2 + \epsilon/r^3$. Write $P = -(x \cdot v)\,u$ for the radial rate. A point Θ is a *first apsis* of frequency κ when $\Theta > 0$, $\sin \kappa\Theta = 0$ and $\sin \kappa\theta > 0$ for every θ with $0 < \theta < \Theta$. The oscillator has *no precession* when $2q$ is a first apsis, that is, when the inverse radius returns to its extreme after a polar half-turn.

Theorem 25.19 (Binet's oscillator for an inverse-cube correction)**.**

(a) The angular momentum and $\frac{1}{2}|v|^2 - mu - \frac{1}{2}\epsilon u^2$ *are first integrals of* c^ϵ*. On states with* $I = 1$, $\mathscr{L} u = u^2 P$ *and* $\mathscr{L} P = u^2\big(m - (|L|^2 - \epsilon)u\big)$.

(b) Let $m, \epsilon, L, \kappa, w_0, C$ *be points with* $\kappa^2 L^2 = L^2 - \epsilon$ *and* $w_0 \kappa^2 L^2 = m$*, and put* $W(\theta) = w_0 + C\cos\kappa\theta$*. Then* W *has derivative* $W'(\theta) = -C\kappa \sin\kappa\theta$*, the function* LW' *has derivative* $-L\kappa^2(W - w_0)$*, and* $L^2\big(-\kappa^2(W - w_0)\big) = m - (L^2 - \epsilon)W$.

(c) For $\kappa > 0$, Θ *is a first apsis of frequency* κ *if and only if* $\Theta > 0$ *and* $\kappa\Theta = 2q$.

(d) For $\kappa > 0$ *there is no precession if and only if* $\kappa = 1$*; when* L *is apart from* 0 *and* $\kappa^2 L^2 = L^2 - \epsilon$*, this holds if and only if* $\epsilon = 0$.

(e) The assertion that for every point $\kappa > 0$ *there is no precession or there is, is equivalent to WLPO.*

Parts (a) to (d) use no omniscience principle and no choice principle; (e) uses no choice principle.

Proof. (a) As in Theorem 25.16, the energy has derivative $-(mu^3+\epsilon u^4)(x\cdot v)+(m+\epsilon u)u^3(x\cdot v) = 0$. Next $\mathscr{L}u = -u^3(x\cdot v) = u^2 P$. For P,

$$\mathscr{L}P = -|v|^2 u + (mu^3 + \epsilon u^4)|x|^2 u + (x\cdot v)^2 u^3.$$

On $I = 1$ the middle term is $mu^2 + \epsilon u^3$, and Lagrange's identity $|L|^2 = |x|^2|v|^2 - (x\cdot v)^2$ gives $|v|^2 = |L|^2u^2 + (x\cdot v)^2u^2$; substituting, $\mathscr{L}P = u^2(m - (|L|^2 - \epsilon)u)$.

(b) The derivatives of the cosine and sine (Theorem 25.11) and the chain rule for $\theta \mapsto \kappa\theta$ give W' and $(LW')' = -L\kappa^2 C\cos\kappa\theta = -L\kappa^2(W - w_0)$. Finally $-L^2\kappa^2(W - w_0) = -\kappa^2L^2W + w_0\kappa^2L^2 = -(L^2 - \epsilon)W + m$.

(c) The sine is positive on $(0, 2q)$: for $0 < \theta \le 2$, $\sin\theta \ge \theta/3 > 0$ (Section 21.2), and $\sin(2q - \theta) = \sin\theta$ by the addition formulas; since $2q - 2 < 2$, cotransitivity places every $\theta \in (0, 2q)$ in $(0, 2)$ or in $(2q - 2, 2q)$. And $\sin 2q = 0$. So $\Theta = 2q/\kappa$ is a first apsis. Conversely, let Θ be a first apsis. If $\kappa\Theta < 2q$, then $\sin\kappa\Theta > 0$; if $\kappa\Theta > 2q$, then $\theta = 2q/\kappa$ lies in $(0, \Theta)$ with $\sin\kappa\theta = 0$. Both are refuted, so $\kappa\Theta = 2q$.

(d) By (c), $2q$ is a first apsis exactly when $2q\kappa = 2q$, that is $\kappa = 1$, since $q > 0$. If $\kappa = 1$ then $\epsilon = L^2 - \kappa^2L^2 = 0$. If $\epsilon = 0$ then $(\kappa^2 - 1)L^2 = 0$ with L^2 invertible, so $\kappa^2 = 1$ and $\kappa = 1$.

(e) By (d) the assertion is: $\kappa = 1$ or $\kappa \ne 1$, for every $\kappa > 0$. Given any point z, apply it to $\kappa = 1 + |z|$: this decides $z = 0$ or $z \ne 0$, which for every z is WLPO (Corollary 12.18). Conversely WLPO decides $\kappa - 1 = 0$ or not. □

For a ratio ϵ the test $\epsilon = 0$ is integer arithmetic, so the price in (e) belongs to frequencies given only as points. The theorem is still about the oscillator, a function of the polar angle. That the actual motion in time is this oscillator, read in a polar angle constructed along the motion, is Theorem 25.23 below.

25.4 Bound orbits return

Theorem 25.11 shows that Kepler's ellipse is a solution. The question of the opening is the converse: does every motion with a bound state return? Classically one fits the ellipse to the initial state, inverts Kepler's time map and invokes uniqueness. On $\mathbb{R}_\delta$ the inversion is avoided by running both curves in the eccentric anomaly.

Theorem 25.20 (Every bound Kepler motion returns)**.** *Let x be a motion of the Kepler field whose state at time 0 has $I = 1$, $u > 0$, $E < 0$ and $|L|^2 > 0$, and let $T \ge 0$ satisfy $T^2(-2E)^3 = (2\pi_\delta)^2m^2$, with E and m taken at time 0. Then $x(T) = x(0)$. No omniscience principle and no choice principle is used.*

Proof. A larger field. On $\mathbb{R}^{12}_\delta$, with a Kepler state z and four more coordinates C, S, e, ι, let

$$z' = \iota(1 - eC)\,p(z), \qquad C' = -S, \qquad S' = C, \qquad e' = 0, \qquad \iota' = 0,$$

with p the Kepler field. This is a polynomial field, so Theorem 25.3 applies to it.

The fit. Let $s = x(0) = (x, v, u, m)$. Since $2mu = |v|^2 - 2E > 0$ and $u > 0$, $m > 0$; by Theorem 25.9, $m^2 - |A|^2 = -2E|L|^2 > 0$. Put $\kappa = \sqrt{-2E}$, $\alpha = |A|$ and $\lambda = |L|$, free square roots, and

$$a = \frac{m}{\kappa^2}, \quad e = \frac{\alpha}{m} < 1, \quad b = \frac{\lambda}{\kappa}, \quad n = \frac{\kappa^3}{m},$$

so that $b^2 = a^2(1-e^2)$ and $n^2a^3 = m$. The equation $x(T) = x(0)$ may be proved under double negation, and $\neg\neg(|A|^2 > 0 \vee |A|^2 = 0)$ holds. If $|A|^2 > 0$, put $C_0 = (m-\kappa^2/u)/\alpha$, $S_0 = (x\cdot v)\kappa/\alpha$ and take the frame $f_1 = A/\alpha$, $f_2 = (L\times A)/(\lambda\alpha)$, $f_3 = L/\lambda$. If $|A|^2 = 0$, put $e = 0$, $b = a$, $(C_0, S_0) = (1, 0)$ and take the frame $f_1 = ux$, $f_2 = u(L\times x)/\lambda$, $f_3 = L/\lambda$; here $x\cdot v = 0$, $|v|^2 = mu$ and $\kappa/u = \lambda$. In both cases $C_0^2 + S_0^2 = 1$, for instance $(m-\kappa^2 r)^2 + \kappa^2(x\cdot v)^2 = m^2 - \kappa^2\big(r^2|v|^2 - (x\cdot v)^2\big) = m^2 - \kappa^2|L|^2 = \alpha^2$ with $r = 1/u$, and expanding the definitions shows

$$x = a(C_0 - e)f_1 + bS_0 f_2, \qquad v = \frac{-naS_0}{1-eC_0}f_1 + \frac{nbC_0}{1-eC_0}f_2, \qquad u = \frac{1}{a(1-eC_0)}.$$

The orbit as a motion. Put $C(\eta) = C_0\cos\eta - S_0\sin\eta$ and $S(\eta) = S_0\cos\eta + C_0\sin\eta$, and let $o(\eta)$ be the state of Theorem 25.11 with $\cos\eta, \sin\eta$ replaced by $C(\eta), S(\eta)$, its position and velocity written in the frame f_1, f_2, f_3. The computation of that theorem, with uniform remainders on bounded intervals, and $p(Rz) = Rp(z)$ for the orthonormal frame R, show that $V(\eta) = (o(\eta), C(\eta), S(\eta), e, 1/n)$ is a motion of the larger field. By the fit, $V(0) = (s, C_0, S_0, e, 1/n)$.

The given motion in anomaly time. Put $\tau(\eta) = \big(\eta - e(S(\eta) - S_0)\big)/n$, so that $\tau' = (1-eC)/n$ and $\tau(0) = 0$. By the chain rule $W(\eta) = \big(x(\tau(\eta)), C(\eta), S(\eta), e, 1/n\big)$ is a motion of the larger field, and $W(0) = V(0)$.

Return. By Theorem 25.3, $W = V$. At $\eta = 4q$ the functions C, S are back at C_0, S_0, so $o(4q) = s$ and $x(\tau(4q)) = x(0)$, with $\tau(4q) = 4q/n = 2\pi_\delta/n$. Finally $(1/n)^2(-2E)^3 = m^2/\kappa^6\cdot\kappa^6 = m^2$, so $(2\pi_\delta/n)^2(-2E)^3 = (2\pi_\delta)^2m^2 = T^2(-2E)^3$; both T and $2\pi_\delta/n$ are nonnegative and $(-2E)^3 > 0$, hence $T = 2\pi_\delta/n$. □

No inverse of Kepler's equation and no polar angle enter the proof: the eccentric anomaly is the clock of both curves, and uniqueness compares them. The perihelion direction appears only through the frame, inside a case split that a goal of the form $x(T) = x(0)$ permits.

Theorem 25.21 (Kepler motion is periodic at every time)**.** *Under the hypotheses of Theorem 25.20, $x(t+T) = x(t)$ for every $t \in \mathbb{R}_\delta$. No omniscience principle and no choice principle is used.*

Proof. By Lemma 25.4(a) the curve $t \mapsto x(t+T)$ is a motion, and its state at time 0 is $x(T) = x(0)$ by Theorem 25.20. Theorem 25.3(b) makes it equal to x. □

Conditions and scope.

Base and premises

A motion of the Kepler field on all of $\mathbb{R}_\delta$ whose initial state is physical, bound ($E < 0$) and not radial ($|L|^2 > 0$); T is given by its defining equation, which is Kepler's third law.

Conclusion

The motion repeats with period T at every time. No omniscience principle and no choice principle is used, although the perihelion direction of every such state requires LLPO (Theorem 25.13).

Identification

The coordinates are a body's position, velocity, inverse distance and GM (classical).

Not claimed

That a motion exists for every initial state. For the states of Theorem 25.11 one does (the ellipse in the time $t(\eta)$); the certified solver of Section 25.9 provides motions on bounded intervals of time.

Now the polar angle. Along a planar motion with $I = 1$ the unit vector $(x_0 u, x_1 u)$ turns at the rate Lu^2. Classically the polar angle is $\theta_0 + \int Lu^2\, dt$; on $\mathbb{R}_\delta$ a time integral of a given motion is a limit that in general costs countable choice (Proposition 12.85), and θ_0 itself is the direction of a vector. Both are avoided: the angle is built from finitely many small angles, each an arcsine.

Lemma 25.22 (A polar angle without choice)**.**

(a) Every unit vector (c, s) has an angle: a point θ with $\cos\theta = c$ and $\sin\theta = s$.

(b) Let $e = (e_0, e_1)$ be a map from $\mathbb{R}_\delta$ to the unit circle, and $h = 1/(p+1)$ with p natural, such that $|e_i(\alpha) - e_i(\beta)| \le \frac{1}{8}$ for $i = 0, 1$ whenever $|\alpha - \beta| \le h$. Let θ_0 be an angle of $e(0)$. There is a map θ with $\cos\theta(t) = e_0(t)$ and $\sin\theta(t) = e_1(t)$ for every t.

(c) If moreover, on $[-T, T]$, e_0 has the uniform derivative $-\omega e_1$ and e_1 the uniform derivative ωe_0, with $|\omega|$ bounded there, then θ has the uniform derivative ω on $[-T, T]$.

No omniscience principle and no choice principle is used.

Proof. Arcsine. On $[-1, 1]$ the sine increases with rate at least $\cos 1 > \frac{1}{2}$, so for $|y| \le \sin 1$ the root of $\sin\theta = y$ in $[-1, 1]$ is computed by Theorem 12.92, with modulus of uniqueness from the rate, as in Theorem 25.10. This is a map arcsin.

Small angles. For unit vectors a, b whose coordinates differ by at most $\frac{1}{8}$, $|a \times b| \le |a_0 - b_0| + |a_1 - b_1| \le \frac{1}{4} < \sin 1$ and $a \cdot b = 1 - \frac{1}{2}|a - b|^2 > 0$. Put $\angle(a, b) = \arcsin(a \times b)$. Then $\sin\angle(a, b) = a \times b$ and $\cos\angle(a, b) = a \cdot b$, since both cosines are positive with the same square. For three such vectors, the addition formula gives $\sin\big(\angle(a, b) + \angle(b, c)\big) = (a \times b)(b \cdot c) + (a \cdot b)(b \times c) = a \times c$, an identity for unit plane vectors, and so $\angle(a, b) + \angle(b, c) = \angle(a, c)$, the sine being injective on $[-1, 1]$.

(a) Note $\sin 1 \ge 1 - \frac{1}{6} > \frac{4}{5}$, and $\cos\vartheta > 0$ for $|\vartheta| \le 1$. By cotransitivity $|s| < \frac{7}{10}$ or $|s| > \frac{6}{10}$. In the first case $c^2 > \frac{1}{2}$, so c is apart from 0: if $c > 0$ take $\theta = \arcsin s$, whose cosine is the positive root of $1 - s^2$, namely c; if $c < 0$ take $\theta = 2q - \arcsin s$. In the second case s is apart from 0 and $|c| < \frac{4}{5}$: take $\theta = q - \arcsin c$ if $s > 0$ and $\theta = \arcsin c - q$ if $s < 0$. The addition formulas with $\cos q = 0$, $\sin q = 1$ check each case.

(b) For t read a natural N with $Nh \ge |t|$ off a regular sequence naming t, let $c_n(t)$ be t clamped into $[-nh, nh]$, and put

$$\theta(t) = \theta_0 + \sum_{n<N} \angle\big(e(c_n(t)), e(c_{n+1}(t))\big).$$

Consecutive clamped times are within h, so each angle is defined. Once $Nh \ge |t|$ the further terms are $\angle(e(t), e(t)) = 0$, so $\theta(t)$ does not depend on N or on the sequence: θ is a map. By additivity, $\theta(t) - \theta(s) = \angle(e(s), e(t))$ when $|t - s| \le h$, and induction on the number of steps from 0, with the addition formulas, gives $\cos\theta = e_0$ and $\sin\theta = e_1$.

(c) For $|t-s| \le h$, $\sin\big(\theta(t)-\theta(s)\big) = e(s)\times e(t) = \omega(s)(t-s) + o(|t-s|)$ uniformly on $[-T,T]$, and $|\alpha - \sin\alpha| \le |\alpha|^3$ for $|\alpha| \le 1$. □

Theorem 25.23 (Binet's clock: the motion is the oscillator)**.** *Let ϵ be a ratio and $y = (x, v, u, m)$ a motion of the field c^ϵ of Definition 25.18 such that for every t: $I = 1$, $u > 0$, $x_2 = v_2 = 0$, and $x_0 v_1 - x_1 v_0 = L$, a constant apart from 0 with $\epsilon < L^2$. Then there are a map θ and points w_0, A, B, κ with $\kappa > 0$ such that:*

(a) θ has derivative $Lu(t)^2$ at every t, and $x_0(t)u(t) = \cos\theta(t)$, $x_1(t)u(t) = \sin\theta(t)$ for every t;

(b) $\kappa^2 L^2 = L^2 - \epsilon$ and $w_0\kappa^2 L^2 = m$;

(c) $u(t) = w_0 + A\cos\kappa\theta(t) + B\sin\kappa\theta(t)$ for every t;

(d) whenever $\theta(t') = \theta(t) + 4q/\kappa$, $u(t') = u(t)$; whenever $\theta(t') = \theta(t) + 2q/\kappa$, $u(t') - w_0 = -(u(t) - w_0)$;

(e) $4q/\kappa = 4q$ if and only if $\epsilon = 0$.

No omniscience principle and no choice principle is used.

Proof. The polar angle. The energy $E_\epsilon = \frac{1}{2}|v|^2 - mu - \frac{1}{2}\epsilon u^2$ is constant (Theorem 25.19(a) and Proposition 25.5). On a planar state with $I = 1$, Lagrange's identity gives $|v|^2 = L^2u^2 + P^2 \ge L^2u^2$, so $(L^2-\epsilon)u^2 \le 2E_\epsilon + 2mu$, and since $L^2 - \epsilon > 0$ this bounds u by a natural number U for all times. The unit vector $e = (x_0u, x_1u)$ satisfies $\dot e_0 = -Lu^2 e_1$ and $\dot e_1 = Lu^2 e_0$: for instance $\frac{d}{dt}(x_0u) = u\big(v_0 - x_0u^2(x\cdot v)\big) = -u^3x_1(x_0v_1 - x_1v_0)$, using $u^2(x_0^2 + x_1^2) = 1$. Its speed is at most $|L|U^2$, so one step h serves for all times in Lemma 25.22(b). Part (a) of the lemma gives an angle of $e(0)$, parts (b) and (c) give θ with (a).

The oscillator. Put $\kappa = \sqrt{(L^2-\epsilon)/L^2} > 0$ and $w_0 = m/(L^2-\epsilon)$, so (b) holds. Let $\theta_0 = \theta(0)$, $C_0 = \cos\kappa\theta_0$, $S_0 = \sin\kappa\theta_0$, $u_0 = u(0)$, $P_0 = P(0)$, $D = P_0/(L\kappa)$, and

$$A = (u_0 - w_0)C_0 - DS_0, \qquad B = (u_0 - w_0)S_0 + DC_0.$$

Then $U(\vartheta) = w_0 + A\cos\kappa\vartheta + B\sin\kappa\vartheta$ has $U(\theta_0) = u_0$ and $U'(\theta_0) = \kappa(BC_0 - AS_0) = \kappa D = P_0/L$: the two conditions solved in closed form, with no search.

The motion is the oscillator. Put $\tilde u(t) = U(\theta(t))$ and $\tilde P(t) = LU'(\theta(t))$. The computation of Theorem 25.19(b) applies term by term to $B\sin\kappa\vartheta$, whose second derivative is $-\kappa^2 B\sin\kappa\vartheta$, so $U'' = -\kappa^2(U - w_0)$ and $L^2U'' = m - (L^2-\epsilon)U$. By the chain rule, $\dot{\tilde u} = u^2\tilde P$ and $\dot{\tilde P} = L^2u^2U''(\theta) = u^2\big(m - (L^2-\epsilon)\tilde u\big)$, while Theorem 25.19(a) gives $\dot u = u^2P$ and $\dot P = u^2(m - (L^2-\epsilon)u)$. So

$$Q = (L^2-\epsilon)(u-\tilde u)^2 + (P-\tilde P)^2$$

has derivative $2(L^2-\epsilon)(u-\tilde u)u^2(P-\tilde P) - 2(P-\tilde P)u^2(L^2-\epsilon)(u-\tilde u) = 0$ on every bounded interval and vanishes at 0; by Lemma 25.4(c) it vanishes everywhere, and $u = \tilde u$. That is (c).

(d) The cosine and sine have period $4q$ and change sign under a shift by $2q$. (e) $4q/\kappa = 4q$ exactly when $\kappa = 1$, which is $\epsilon = 0$ by Theorem 25.19(d). □

Conditions and scope.

Base and premises
A planar motion of Newton's field with an inverse-cube correction ϵ/r^3, ϵ a ratio below L^2, on all of $\mathbb{R}_\delta$.

Conclusion
In its own polar angle, the inverse radius of the motion is an oscillator of frequency $\kappa = \sqrt{1-\epsilon/L^2}$; the pericentre advances by $2\pi_\delta(1/\kappa - 1)$ per radial period, and the orbit closes after one turn exactly when $\epsilon = 0$.

Not claimed
The value of ϵ for any physical system. Section 25.6 treats the relativistic correction, whose orbit equation is not linear.

Two choices of form keep the proof free. The angle is a sum of small arcsines, not an integral. And the oscillator is written as $w_0 + A\cos\kappa\theta + B\sin\kappa\theta$, not as $w_0 + C\cos\kappa(\theta - \theta_1)$: the phase form needs the direction of the vector (A, B), which near a circular orbit requires LLPO, exactly as the perihelion direction does.

25.5 The N-body problem

With several bodies every pair needs its own inverse distance. Carrying one coordinate per pair keeps the field polynomial for every N.

Definition 25.24 (The N-body field)**.** For a natural number N, a state consists of positions and velocities $x_a, v_a \in \mathbb{R}^3_\delta$ and masses m_a for $a < N$, one slot u_{ab} for each pair $a < b$ (we write $u_{ba} = u_{ab}$), and a clock τ. The *N-body field* is

$$\dot{x}_a = v_a, \quad \dot{v}_a = \sum_{b \neq a} m_b\, u_{ab}^3\,(x_b - x_a), \quad \dot{u}_{ab} = -u_{ab}^3\,(x_a - x_b)\cdot(v_a - v_b), \quad \dot{m}_a = 0, \quad \dot{\tau} = 1.$$

Its quantities are the energy $\sum_a m_a|v_a|^2/2 - \sum_{a<b} m_a m_b u_{ab}$, the momentum $\mathscr{P} = \sum_a m_a v_a$, the angular momentum $\sum_a m_a\, x_a \times v_a$, the centre-of-mass integral $\sum_a m_a x_a - \tau\mathscr{P}$, and the pair defects $D_{ab} = u_{ab}^2|x_a - x_b|^2 - 1$. A state is *physical* when every $D_{ab} = 0$ and every $u_{ab} > 0$; then $u_{ab} = 1/|x_a - x_b|$ and the field is Newton's, with $m_a = GM_a$.

Theorem 25.25 (Ten integrals for every N)**.** *For every N:*

(a) the energy and the three components each of the momentum, the angular momentum and the centre-of-mass integral are first integrals of the N-body field;

(b) for every pair $a < b$, $\mathscr{L}\, D_{ab} = -2u_{ab}^2\,(x_a - x_b)\cdot(v_a - v_b)\, D_{ab}$.

These are identities in every commutative ring with the ratios, in particular on $\mathbb{R}_\delta$. No omniscience principle and no choice principle is used.

Proof. A double sum over $a \neq b$ is the sum over $a < b$ of the two terms (a,b) and (b,a), and three cancellations follow. *Momentum:* $\sum_a m_a \dot{v}_a = \sum_{a \neq b} m_a m_b u_{ab}^3 (x_b - x_a) = 0$, the summand being antisymmetric in (a,b); this is Newton's third law. *Centre of mass:* its derivative is $\mathscr{P} - \mathscr{P} - \tau \dot{\mathscr{P}} = 0$. *Angular momentum:* $\sum_a m_a\, x_a \times \dot{v}_a = \sum_{a<b} m_a m_b u_{ab}^3 \big(x_a \times (x_b - x_a) + x_b \times (x_a - x_b) \big) = \sum_{a<b} m_a m_b u_{ab}^3 (x_a - x_b) \times (x_b - x_a) = 0$. *Energy:* $\sum_a m_a v_a \cdot \dot{v}_a = \sum_{a<b} m_a m_b u_{ab}^3 (x_b - x_a) \cdot (v_a - v_b)$, while $-\sum_{a<b} m_a m_b \dot{u}_{ab} = \sum_{a<b} m_a m_b u_{ab}^3 (x_a - x_b) \cdot (v_a - v_b)$; the two cancel. (b) involves one pair: with $\Delta x = x_a - x_b$ and $\Delta v = v_a - v_b$, $\frac{d}{dt}(u^2 |\Delta x|^2 - 1) = 2u\dot{u}|\Delta x|^2 + 2u^2\, \Delta x \cdot \Delta v = -2u^2 (\Delta x \cdot \Delta v)(u^2 |\Delta x|^2 - 1)$. □

The ten classical integrals are due to Newton, Euler and Lagrange; Wintner's treatise collects them [170]. The identities are polynomial, so they hold for every N at once and need no physical state; the pair invariant says that the field keeps every D_{ab} at zero at a rate proportional to D_{ab} itself.

Theorem 25.26 (Two bodies are a Kepler problem with a drifting centre)**.**

(a) For $N = 2$ the relative coordinates $(x_0 - x_1,\ v_0 - v_1,\ u_{01},\ m_0 + m_1)$ have Lie derivatives equal to the Kepler field evaluated at them. So along every motion of the two-body field the relative coordinates form a motion of the Kepler field.

(b) Let x be a motion of the two-body field whose relative state at time 0 has $I = 1$, $u_{01} > 0$, relative energy $E = \frac{1}{2}|v_0 - v_1|^2 - (m_0 + m_1) u_{01} < 0$ and $|L|^2 > 0$ for $L = (x_0 - x_1) \times (v_0 - v_1)$, and let $T \geq 0$ satisfy $T^2 (-2E)^3 = (2\pi_\delta)^2 (m_0 + m_1)^2$. Then the relative position, the relative velocity and u_{01} at time T equal their values at 0, and for each body a,

$$(m_0 + m_1)\big(x_a(T) - x_a(0)\big) = T\big(m_0 v_0(0) + m_1 v_1(0)\big).$$

No omniscience principle and no choice principle is used.

Proof. (a) The relative accelerations are $\dot{v}_0 - \dot{v}_1 = m_1 u^3 (x_1 - x_0) - m_0 u^3 (x_0 - x_1) = -(m_0 + m_1) u^3 (x_0 - x_1)$, and $\dot{u}_{01} = -u^3 (x_0 - x_1) \cdot (v_0 - v_1)$: the Kepler field with $m = m_0 + m_1$. If polynomials F_j satisfy $\mathscr{L}_p F_j = q_j(F)$, then by Lemma 25.4(b) the curve $F(x(t))$ has the uniform derivative $q(F(x(t)))$, so it is a motion of q.

(b) By (a) and Theorem 25.20, the relative state returns at T. The masses are constant, $\tau(T) = \tau(0) + T$ by Lemma 25.4(c), and the momentum and the centre-of-mass integral are constant (Theorem 25.25 and Proposition 25.5). So $\sum_b m_b x_b(T) - \sum_b m_b x_b(0) = T\mathscr{P}(0)$. Since $x_0 - x_1$ has returned, both bodies are displaced by the same vector d, and $(m_0 + m_1) d = T\mathscr{P}(0)$. □

Lagrange found that three bodies at the corners of an equilateral triangle, rotating rigidly at the right rate, move exactly so for all time [162]. On $\mathbb{R}_\delta$ the rotation is constructed with no choice, and uniqueness says more: nothing else starts there.

Theorem 25.27 (Lagrange's triangle rotates rigidly, and no other motion starts there)**.**

(a) Let N be a natural number, $c_a \in \mathbb{R}^2_\delta$ and m_a points for $a < N$, and u, ω points with $\sum_a m_a c_a = 0$ and $\omega^2 = \big(\sum_a m_a\big)u^3$. Let $\ell(t)$ be the rotating state*: $x_a(t)$ is c_a turned by the angle ωt in the plane $x_2 = 0$, $v_a(t)$ is ω times $x_a(t)$ turned by a further quarter turn, every pair slot equals u, the masses are m_a and $\tau = t$. Then ℓ is a motion of the N-body field, and every motion x of the field with $x(0) = \ell(0)$ satisfies $x(t) = \ell(t)$ for every t.*

(b) Let m_0, m_1, m_2 have sum $M > 0$ and let $\ell_0 > 0$. There are planar points c_a and points u, ω with $u\ell_0 = 1$, $\omega \geq 0$ and $\omega^2 \ell_0^3 = M$, such that the rotating state is a motion of the three-body field, every pair defect vanishes along it, and every coordinate other than the clock returns after any T with $T\omega = 2\pi_\delta$.

No omniscience principle and no choice principle is used.

Proof. (a) Each coordinate of ℓ is of the form $\alpha \cos \omega t + \beta \sin \omega t$, a constant, or t, so it has uniform derivatives on bounded intervals. With every pair slot equal to u and $\sum_b m_b x_b = 0$, the field at the rotating state is

$$\dot{v}_a = u^3\Big(\sum_b m_b x_b - \big(\textstyle\sum_b m_b\big)x_a\Big) = -\Big(\textstyle\sum_b m_b\Big)u^3 x_a = -\omega^2 x_a,$$

which is the acceleration of the rotation, and $\dot{u}_{ab} = -u^3(x_a - x_b)\cdot(v_a - v_b) = 0$ because $v_a - v_b$ is $x_a - x_b$ turned by a quarter turn. So ℓ is a motion, and Theorem 25.3(b) gives the uniqueness.

(b) Put the corners at $(0,0)$, $(\ell_0, 0)$ and $(\ell_0/2, \ell_0\sqrt{3}/2)$ and subtract their centre of mass $\sum_a m_a c_a / M$; take $u = 1/\ell_0$ and $\omega = \sqrt{Mu^3}$, all by free square roots and inverses of points apart from 0. Each side has length ℓ_0, so every pair defect vanishes at time 0, and rotation keeps the distances. The return follows from the period $4q = 2\pi_\delta$ of the cosine and sine. □

For $N \geq 4$ no planar points are pairwise equidistant, so part (a) has its Newtonian reading, all pair slots equal to the inverse distance, only for $N \leq 3$. As a statement about the polynomial field it holds for every N.

A physical state needs $u_{ab} = 1/|x_a - x_b|$, and forming it from the knowledge that two bodies do not collide has a price.

Theorem 25.28 (The physical state from non-collision costs Markov's principle)**.** *The following are equivalent, with no choice principle: for every N, every state and every pair $a < b$ with $\neg\big(|x_a - x_b|^2 = 0\big)$ there is a point u with $u^2|x_a - x_b|^2 = 1$; and* $\mathrm{MP}_{\mathrm{bin}}$*. The same holds for a single pair of points of $\mathbb{R}^3_\delta$.*

Proof. If $u^2 r^2 = 1$ then r^2 is invertible, hence apart from 0, and being nonnegative it is positive. Every nonnegative point z is such an r^2, for $x_a = (\sqrt{z}, 0, 0)$ and $x_b = 0$. So the assertion gives $z > 0$ from $\neg(z = 0)$ for every $z \geq 0$, which applied to $|z|$ is $\mathrm{MP}_{\mathrm{bin}}$ (Corollary 12.18). Conversely $\mathrm{MP}_{\mathrm{bin}}$ gives $r^2 > 0$ and then $u = 1/\sqrt{r^2}$. □

For ratio data the test $|x_a - x_b|^2 = 0$ is decided by arithmetic, and the price does not arise. That matters for the certified solver of Chapter 15, which starts from ratio data. Its runs keep the physical manifold exactly, and more generally keep every family of polynomials that the field maps into combinations of itself. The proof follows the solver's trajectories of Chapter 15, and one more estimate about them comes first: they stay close to each other at every step, not only at the end, and close to the monomial surface on which the quadratized field moves as the original one does.

Lemma 25.29 (Level trajectories stay together and near the monomial surface)**.** *Let the quadratized problem of a polynomial field p on d coordinates, with initial point $x_0 \in \mathbb{Q}_\delta^d$, radius M and end time T, satisfy the region certificate at level n_0, with trajectories $y_k^{(n)}$, S steps and orders $N(n)$ (Definition 15.18). For a state y of the quadratized field write $x(y)$ for its coordinates w_{e_i} and $z(y) = W(x(y))$, where $W(x) = (x^\alpha)_{|\alpha|\le D}$: the point of the monomial surface $w_\alpha = x^\alpha$ above y.*

(a) If $n_0 \le n \le n'$, then $\|y_k^{(n')} - y_k^{(n)}\| \le 2^{-(n+2)}$ for every $k \le S$.

(b) There are natural numbers A and $n_1 \ge n_0$ such that $\|y_k^{(n)} - z(y_k^{(n)})\| \le A\,2^{-n}$ and $\|z(y_k^{(n)})\| \le M$ for every $n \ge n_1$ and every $k \le S$.

No omniscience principle and no choice principle is used.

Proof. (a) This is the proof of Theorem 15.21, read at every step: both trajectories stay in the ball of radius M (Theorem 15.20), the second form of Theorem 15.9, with $\Phi = \Phi_{N(n)}$ and the defects of Lemma 15.17(b), bounds their distance at step k by E_k, and Lemma 15.19 gives $E_k \le 2^{-(n+2)}$ for every $k \le S$.

(b) Write $h = 2^{-\sigma}$ for the step, so that $8Rh \le 1$.

A margin. For the levels $n_0 \le n$, the proof of Lemma 15.19 bounds E_k by $G(1+2S(M+2))2^{-N(n_0)}$, and the integer $G(1 + 2S(M + 2))$ is less than $2^{\kappa-2}$. So $E_k \le 2^{-(n_0+2)} - 2^{-N(n_0)}$, and the proof of Theorem 15.20 gives $\|y_k^{(n)}\| \le M - 2^{-N(n_0)}$ for $k \le S$.

Products of series. If power series $X_1, \dots, X_d$ have $|X_{i,j}| \le MR^j$, the coefficient of order j of a product of $r \ge 1$ of them is at most $M^r\binom{j+r-1}{r-1}R^j \le M^r 2^{j+r-1}R^j$. For a polynomial F write $|F|(c)$ for the value at $(c, \dots, c)$ of F with every coefficient replaced by its absolute value. Summing over the terms of F,

$$|F(X)_j|\,(4s)^j \le |F|(2M) \quad \text{when } 8Rs \le 1, \tag{25.2}$$

and at a point $\tau \ge 0$ with $8R\tau \le 1$ the terms of $F(X)$ of order at least $K \ge 1$ add up to at most $|F|(2M)\,2^{-K}$ in absolute value. Both hold as well when each X_i is cut off at some order.

One step from the surface. Let z be a point of the monomial surface with $\|z\| \le M$, and let X be the formal solution of $x' = p(x)$ from $x(z)$: $X_0 = x(z)$, and $(j+1)X_{j+1}$ is the coefficient of order j of $p(X)$. The identity $B(W(x), W(x))_\alpha = \sum_i \alpha_i x^{\alpha - e_i} p_i(x)$ in the proof of Corollary 15.30 holds for power series, so $W(X)$ solves the quadratized equation formally from z, and its coefficients are the exact coefficient table at z; by Lemma 15.14, $|X_{i,j}| \le MR^j$. Coordinate α of $\Phi_N(z)$ is the part of X^α of order at most N, at h; that of $z(\Phi_N(z))$ is the product of the parts

of order at most N of the X_i, at h. They differ by the terms of order above N of the latter, so $\|\Phi_N(z) - z(\Phi_N(z))\| \le (2M)^D 2^{-(N+1)}$; and $\|\Phi_N(z)\| \le M \sum_j 8^{-j} \le 2M$.

The induction. On points whose coordinates are at most $2M$, the monomials of degree at most D are Lipschitz with constant $K_W = D(2M)^{D-1}$. Put $a_0 = 1 + K_W$ and $a_{k+1} = (1 + K_W)(M + 2 + 2a_k) + (2M)^D$, an increasing sequence, and $n_1 = n_0 + a_S$. Let $n \ge n_1$, $N = N(n)$ and $y_k = y_k^{(n)}$; we show $\|y_k - z(y_k)\| \le a_k 2^{-N}$ for $k \le S$. At $k = 0$, $y_0 = [W(x_0)]_N$ and $z(y_0) = W([x_0]_N)$, which are within $(1 + K_W)2^{-(N+1)}$. If the bound holds at k, then $a_k 2^{-N} \le 2^{-N(n_0)}$, and the margin gives $\|z(y_k)\| \le M$. With $Y = \Phi_N(z(y_k))$, Lemma 15.17(a) and (b), with the bound $(M+2)2^{-N}$ of the proof of Lemma 15.19, give $\|y_{k+1} - Y\| \le (M + 2 + 2a_k)2^{-N}$; the step from the surface gives $\|Y - z(Y)\| \le (2M)^D 2^{-N}$; and $\|z(Y) - z(y_{k+1})\| \le K_W \|Y - y_{k+1}\|$. The three add up to $a_{k+1}2^{-N}$. So $A = a_S$ will do, and $\|z(y_k)\| \le M$ was shown on the way. □

Theorem 25.30 (Invariant families are kept exactly by certified runs)**.** *Let p be a polynomial field on d coordinates, $D_0, \dots, D_{n-1}$ polynomials and G_{ij} polynomials with $\mathscr{L}_p D_i = \sum_j G_{ij} D_j$ at every point of $\mathbb{R}^d_\delta$. Let $x_0 \in \mathbb{Q}^d_\delta$ have $D_j(x_0) = 0$ for every j, and let the quadratized problem of p with initial point x_0, radius M and end time T satisfy the region certificate at some level, so that the polynomial solution $\bar{x}$ at time T is defined (Definition 15.27). Then $D_i(\bar{x}) = 0$ for every i. In particular:*

- *(a) if $\mathscr{L}_p D = G D$, $\mathscr{L}_p F = A D$ and $D(x_0) = 0$, then $F(\bar{x}) = F(x_0)$: conservation modulo an invariant;*
- *(b) a certified run of the Kepler field started with $I = 1$ ends with $I = 1$ and with its initial vector A;*
- *(c) a certified run of the N-body field started with every pair defect 0 ends with every pair defect 0.*

No omniscience principle and no choice principle is used.

Proof. The idea is linearity: along a formal solution the family obeys a linear equation, which keeps it at 0 when it starts at 0 and multiplies a small start by a fixed factor. Let A, n_1, $x(y)$, $z(y)$ and $|F|(c)$ be as in Lemma 25.29 and its proof, $h = 2^{-\sigma}$ the step of the schedule, and μ a natural number with $|G_{ij}|(2M) \le \mu$ for all i, j.

A linear formal equation. Let power series $E_0, \dots, E_{n-1}$ and g_{ij} satisfy $E_i' = \sum_j g_{ij} E_j$, with $|g_{ij,l}|(4s)^l \le \mu$ for every l, $|E_{i,0}| \le \varepsilon$, and $sn\mu \le \Gamma$ for a natural number Γ. Then $\sum_{k \le K} |E_{i,k}| s^k \le 2\varepsilon(4\Gamma + 2)^{4\Gamma}$ for every K. Indeed, the weights $a_{i,k} = |E_{i,k}|(2s)^k$ satisfy $(k+1)a_{i,k+1} \le 2s\mu \sum_j \sum_{l \le k} 2^{-l} a_{j,k-l}$. With $P = 4\Gamma + 2$, induction on k gives $a_{i,k} \le \varepsilon P^k$, since the right side is then at most $4\Gamma\varepsilon P^k$. In particular $a_{i,k} \le \varepsilon P^{4\Gamma}$ for $k \le 4\Gamma$, and for $k \ge 4\Gamma$ the same inequality gives $a_{i,k+1} \le 4\Gamma\varepsilon P^{4\Gamma}/(k+1) \le \varepsilon P^{4\Gamma}$. So $\sum_k a_{i,k} 2^{-k} \le 2\varepsilon P^{4\Gamma}$.

One step. Fix a level $\ell \ge n_1$ of the solver (we keep n for the number of the D_i), and write $N = N(\ell)$ and $y_k = y_k^{(\ell)}$. Let z be a point of the monomial surface with $\|z\| \le M$ and $|D_i(x(z))| \le \varepsilon$ for every i. As in the proof of Lemma 25.29, $x(\Phi_N(z)) = \sum_{j \le N} h^j X_j$, where X is the formal solution of $x' = p(x)$ from $x(z)$ and $|X_{i,j}| \le MR^j$. The rate law holds at every ratio

point, so it is an identity of polynomials, and the series $E_i = D_i(X)$ satisfy $E_i' = \sum_j g_{ij}E_j$ with $g_{ij} = G_{ij}(X)$. By (25.2) with $s = h$, $|g_{ij,l}|(4h)^l \le \mu$; so for a natural number $\Gamma \ge hn\mu$ the linear equation gives $\sum_{k\le N}|E_{i,k}|h^k \le L\varepsilon$, with $L = 2(4\Gamma+2)^{4\Gamma}$. Now $D_i(x(\Phi_N(z)))$ is $\sum_{k\le N}E_{i,k}h^k$ plus the terms of order above N of D_i at the cut-off series, which add up to at most $|D_i|(2M)2^{-(N+1)}$. So $|D_i(x(\Phi_N(z)))| \le L\varepsilon + C2^{-(N+1)}$, with $C = \max_i |D_i|(2M)$.

Along the trajectory. Let L_D be a natural number with $|D_i(x) - D_i(x')| \le L_D\|x - x'\|$ for all i on points whose coordinates are at most $2M$, and put $b_0 = L_D$ and $b_{k+1} = Lb_k + C + L_D(M+2+2A)$. Then $|D_i(x(y_k))| \le b_k 2^{-\ell}$ for $k \le S$. At $k = 0$, $x(y_0) = [x_0]_N$ is within $2^{-(N+1)}$ of x_0, where every D_i vanishes. For the step take $z = z(y_k)$: $x(z) = x(y_k)$ and $\|z\| \le M$ by Lemma 25.29(b), so $|D_i(x(\Phi_N(z)))| \le (Lb_k + C)2^{-\ell}$; and, as in the proof of that lemma, $\|y_{k+1} - \Phi_N(z)\| \le (M + 2 + 2A)2^{-\ell}$. The numbers b_k depend only on p, the family, M and T, which fix the schedule.

The limit. By Theorem 15.22 the coordinates $x(y_S^{(\ell)})$ converge to $\bar{x}$ as ℓ grows, so $|D_i(\bar{x})| \le b_S 2^{-\ell} + L_D\|\bar{x} - x(y_S^{(\ell)})\|$ tends to 0, and $D_i(\bar{x}) = 0$. No uniqueness of solutions is used. For (a) take the family $(D, F - F(x_0))$ with rates (GD, AD); for (b) use Theorem 25.9(b); for (c) Theorem 25.25(b). □

The same argument bounds the family by a fixed multiple of its initial size when the initial point is near the invariant set rather than on it, and it holds at every certified dyadic time of the run. The conserved quantities of Theorem 25.25 are a family with all rates zero, so a certified N-body run keeps all ten integrals exactly (Theorem 25.52 below).

25.6 The relativistic orbit equation

In the polar angle θ the inverse radius $u = 1/r$ of a Newtonian orbit obeys $u'' + u = m/L^2$, whose solutions are the conics. The first relativistic correction adds a term $3mu^2$: for a test body about a mass m in Schwarzschild's geometry the orbit equation is $u'' + u = m/L^2 + 3mu^2$ [155, 168]. This section treats the equation with general coefficients, $u'' + u = a + bu^2$, in the polar angle; the next one treats the motion in proper time.

Definition 25.31 (The orbit equation)**.** For ratios a and b, the *orbit field* acts on pairs (u, w) of points:

$$u' = w, \qquad w' = a - u + bu^2,$$

the derivative being in the clock θ, read as the polar angle. Its *energy* is $\mathscr{E}(u, w) = w^2 + u^2 - 2au - \frac{2}{3}bu^3$. The *window* is $a > 0$, $b \ge 0$ and $100ab \le 1$; an initial value u_0 is in the window when $a \le 2u_0 \le 3a$.

For the Schwarzschild orbit $a = m/L^2$, the Newtonian inverse semi-latus rectum, and $b = 3m$; then $100ab \le 1$ reads $300m^2 \le L^2$. The window on u_0 allows the starting radius $1/u_0$ to lie between $\frac{2}{3}$ and 2 times the Newtonian semi-latus rectum $1/a$.

Proposition 25.32 (The energy of the orbit equation)**.** *For all ratios a, b, $\mathscr{E}$ is a first integral of the orbit field. No omniscience principle and no choice principle is used.*

Proof. $\partial_u \mathscr{E} \cdot w + \partial_w \mathscr{E} \cdot (a - u + bu^2) = (2u - 2a - 2bu^2)w + 2w(a - u + bu^2) = 0$. □

The classical way to the orbit is to integrate $d\theta = du/\sqrt{\cdots}$, a cubic under a square root. We avoid both the integral and the square root of a polynomial that changes sign. The idea is to look for the motion in the form of a harmonic oscillator run on an auxiliary clock ψ:

$$u = U_c + A_t \cos\psi, \qquad w = -A_t\, G(\psi) \sin\psi, \qquad \frac{d\psi}{d\theta} = G(\psi) = \sqrt{D - \alpha\cos\psi}. \tag{25.3}$$

Then $du/d\theta = -A_t \sin\psi \cdot G = w$, and since $GG' = \tfrac{1}{2}\alpha \sin\psi$,

$$\frac{dw}{d\theta} = -A_t\big(G' \sin\psi + G\cos\psi\big)G = -A_t\Big(\frac{\alpha}{2} + D\cos\psi - \frac{3\alpha}{2}\cos^2\psi\Big).$$

This equals $a - u + bu^2$ identically in ψ exactly when the coefficients of $\cos^2\psi$, $\cos\psi$ and 1 agree:

$$\alpha = \tfrac{2}{3}\, bA_t, \qquad D = 1 - 2bU_c, \qquad U_c - a - bU_c^2 = \tfrac{1}{3}\, bA_t^2. \tag{25.4}$$

Four parameters, three equations, and the fourth condition is the initial value $U_c + A_t = u_0$. No equation divides by a or b, so $b = 0$ is allowed.

Lemma 25.33 (The centre, the half-amplitude and the clock constants)**.** *Let a, b, u_0 be points with $a > 0$, $b \geq 0$, $100ab \leq 1$ and $a \leq 2u_0 \leq 3a$, and put $\epsilon = ab$. There are points U_c, A_t, D_m, r with*

$$U_c + A_t = u_0, \qquad U_c - a - bU_c^2 = \tfrac{1}{3} bA_t^2,$$
$$D_m^2(1 - 2bU_c) = 1, \qquad D_m \geq 0, \qquad r(1 - 2bU_c) = \tfrac{2}{3} bA_t,$$
$$|r| \leq \tfrac{1}{250}, \quad 1 + |r| \leq D_m^2, \quad D_m \leq 1 + \tfrac{1}{50}, \quad 1 + \epsilon \leq D_m, \quad D_m(1 + 2r^2) \leq 1 + \epsilon + 4\epsilon^2,$$

and moreover $a \leq U_c$, $25A_t^2 \leq 9U_c^2$ and $250\, bU_c \leq 3$. No omniscience principle and no choice principle is used.

Proof. Put $W_0 = u_0 - a \in [-\tfrac{a}{2}, \tfrac{a}{2}]$, $\beta = 1 - 2\epsilon + \tfrac{2}{3} bW_0$ and $Q_2 = a^2 + \tfrac{1}{3} W_0^2$, so that $\beta \geq 1 - \tfrac{7}{3}\epsilon \geq \tfrac{293}{300}$ and $a^2 \leq Q_2 \leq \tfrac{13}{12} a^2$. With $U_c = a + V$ and $A_t = W_0 - V$, the centre relation becomes $\tfrac{4}{3} bV^2 - \beta V + bQ_2 = 0$. Put $\gamma = \tfrac{16}{3} b^2 Q_2$, so $\gamma \leq \tfrac{52}{9}\epsilon^2 \leq \beta^2$, and take the small root written without dividing by b:

$$V = \frac{2bQ_2}{\beta + \sqrt{\beta^2 - \gamma}}.$$

Substituting, $\tfrac{4}{3} bV^2 - \beta V + bQ_2 = bQ_2(\gamma - \beta^2 + (\beta^2 - \gamma))/(\beta + \sqrt{\beta^2 - \gamma})^2 = 0$. Since $\sqrt{\beta^2 - \gamma} \geq \beta - \gamma/\beta$ (square both sides, using $\gamma \leq \beta^2$), $0 \leq V \leq bQ_2/(\beta - \gamma/(2\beta)) \leq \tfrac{10}{9}\epsilon a$; so $a \leq U_c \leq \tfrac{91}{90} a$, $bV \leq \tfrac{10}{9}\epsilon^2$, and $|A_t| \leq \tfrac{a}{2} + V \leq \tfrac{23}{45} a$.

Let $D = 1 - 2bU_c = 1 - 2\epsilon - 2bV \geq 1 - 2\epsilon - \frac{20}{9}\epsilon^2 > \frac{97}{100}$, and put $D_m = 1/\sqrt{D}$ and $r = \frac{2}{3}bA_t/D$. Then $|r| \leq \frac{2}{3}\cdot\frac{23}{45}\epsilon\cdot\frac{100}{97} \leq \frac{9}{25}\epsilon \leq \frac{1}{250}$. Next, $1+|r| \leq D_m^2$ means $|r|D \leq 2bU_c$, that is $\frac{2}{3}b|A_t| \leq 2bU_c$, which holds since $|A_t| \leq 3U_c$. The bound $1+\epsilon \leq D_m$ means $(1+\epsilon)^2 D \leq 1$, and $(1+\epsilon)^2(1-2\epsilon-2bV) = 1 - 3\epsilon^2 - 2\epsilon^3 - 2bV(1+\epsilon)^2 \leq 1$. For the upper bound,

$$(1+\epsilon+3\epsilon^2)^2\Big(1-2\epsilon-\tfrac{20}{9}\epsilon^2\Big) - 1 = \tfrac{7}{9}\epsilon^2 - \tfrac{112}{9}\epsilon^3 - \tfrac{167}{9}\epsilon^4 - \tfrac{94}{3}\epsilon^5 - 20\epsilon^6 \geq 0$$

for $0 \leq \epsilon \leq \frac{1}{100}$, so $D_m \leq 1+\epsilon+3\epsilon^2 \leq 1+\frac{1}{50}$, and with $2r^2 \leq \frac{162}{625}\epsilon^2$, $D_m(1+2r^2) \leq (1+\epsilon+3\epsilon^2)(1+\frac{162}{625}\epsilon^2) \leq 1+\epsilon+4\epsilon^2$. Finally $3U_c - 5A_t \geq 3a - \frac{5}{2}a > 0$ and $3U_c + 5A_t \geq 3a + 3V - \frac{5}{2}a - 5V = \frac{a}{2} - 2V > 0$, so $25A_t^2 \leq 9U_c^2$; and $bU_c \leq \frac{91}{90}\epsilon \leq \frac{3}{250}$. □

The polar angle as a function of the clock is $\vartheta(\psi) = \int_0^\psi d\psi'/G(\psi') = D_m\int_0^\psi (1-r\cos\psi')^{-1/2}\,d\psi'$, an elliptic integral. We build it as a series whose terms are elementary.

Lemma 25.34 (The clock). *Let D_m, r be points with $|r| \leq \frac{1}{250}$, $1+|r| \leq D_m^2$ and $0 \leq D_m \leq 1+\frac{1}{50}$, and put $h(\psi) = D_m(1-r\cos\psi)^{-1/2}$. There is a map ϑ with $\vartheta(0)=0$ such that:*

(a) ϑ has the uniform derivative h on every bounded interval, and $1 \leq h \leq 2$;

(b) $\psi' - \psi \leq \vartheta(\psi') - \vartheta(\psi) \leq 2(\psi'-\psi)$ whenever $\psi \leq \psi'$;

(c) $\vartheta(\psi+4q) = \vartheta(\psi) + \vartheta(4q)$ for every ψ;

(d) $4qD_m \leq \vartheta(4q) \leq 4qD_m(1+2r^2)$ and $4qD_m\big(1+\frac{3}{16}r^2\big) \leq \vartheta(4q) \leq 4qD_m\big(1+\frac{3}{16}r^2+2r^4\big)$.

No omniscience principle and no choice principle is used.

Proof. The binomial coefficients. Let $\beta_n = \binom{2n}{n}/4^n$. Then $\beta_0 = 1$, $\beta_1 = \frac{1}{2}$ and $\beta_{n+1} = \frac{2n+1}{2n+2}\beta_n \leq \beta_n$, so $0 \leq \beta_n \leq \frac{1}{2}$ for $n \geq 1$. The formal series $B(t) = \sum \binom{2n}{n}t^n$ satisfies $(1-4t)B' = 2B$, because $(n+1)\binom{2n+2}{n+1} = 2(2n+1)\binom{2n}{n}$. Hence $(1-4t)(B^2)' = 4B^2$, which is also the equation of $\sum 4^n t^n$, and both start with 1; so $\sum_{i+j=n}\beta_i\beta_j = 1$ for every n. For $|z| \leq \frac{1}{2}$ the series $K(z) = \sum \beta_n z^n$ therefore satisfies $K(z)^2(1-z) = 1$, and $K(z) \geq 1 - \frac{1}{2}\sum_{n\geq 1}|z|^n \geq 1 - \frac{1}{2} = \frac{1}{2}$, since $\sum_{n\geq 1}|z|^n = |z|/(1-|z|) \leq 1$; so $K(z) = (1-z)^{-1/2}$. Its tail after N terms is at most $\sum_{n\geq N}|z|^n \leq 2|z|^N$.

Powers of the cosine. Put $J_0(\psi) = \psi$, $J_1(\psi) = \sin\psi$ and

$$J_{n+2}(\psi) = \frac{\cos^{n+1}\psi\,\sin\psi}{n+2} + \frac{n+1}{n+2}J_n(\psi).$$

Differentiating, $J_n' = \cos^n$, since $\frac{d}{d\psi}\cos^{n+1}\psi\sin\psi = (n+2)\cos^{n+2}\psi - (n+1)\cos^n\psi$. With the Wallis numbers $\varpi_0 = 1$, $\varpi_1 = 0$, $\varpi_{n+2} = \frac{n+1}{n+2}\varpi_n$, induction gives $J_n(\psi+4q) = J_n(\psi) + 4q\varpi_n$ and $|J_n(\psi) - \varpi_n\psi| \leq 1$, because the first term of the recursion has period $4q$ and absolute value at most $\frac{1}{n+2}$.

The series. Put $\vartheta(\psi) = D_m\sum_n \beta_n r^n J_n(\psi)$. The terms are at most $|r|^n(|\psi|+1)$, so the partial sums, computed from a regular sequence naming ψ, converge at a geometric rate on every

bounded interval; Theorem 12.25 gives the limit with no choice, and limits are unique, so ϑ is a map. The derivatives of the partial sums are $D_m \sum_{n<N} \beta_n (r\cos\psi)^n$, within $2D_m|r|^N$ of $h(\psi) = D_m K(r\cos\psi)$, and they are Lipschitz with one constant for all N; so the limit has the uniform derivative h. From $1+|r| \le D_m^2$ and $(1-r\cos\psi)^{-1/2} \ge (1+|r|)^{-1/2}$ we get $h \ge 1$, and $h \le \frac{51}{50}(1-\frac{1}{250})^{-1/2} < 2$. That is (a), and (b) follows from Proposition 12.82, item 2, applied to $\vartheta(\psi)-\psi$ and to $2\psi-\vartheta(\psi)$ after rescaling the interval.

(c) and (d). The shift adds $D_m \sum_n \beta_n r^n\, 4q\varpi_n = \vartheta(4q)$. Odd Wallis numbers vanish, so $\vartheta(4q) = 4qD_m \sum_k \beta_{2k}\varpi_{2k} r^{2k}$, a series of nonnegative terms that starts $1+\beta_2\varpi_2 r^2 = 1+\frac{3}{8}\cdot\frac{1}{2}r^2$, and whose terms after the first, and after the first two, add up to at most $\sum_{n\ge2}|r|^n \le 2r^2$ and $\sum_{n\ge4}|r|^n \le 2r^4$. □

Theorem 25.35 (The relativistic orbit: motion, period and closure)**.** *Let a, b be ratios in the window: $a > 0$, $b \ge 0$, $100ab \le 1$.*

(a) For every point u_0 with $a \le 2u_0 \le 3a$ there is a motion $X = (u, w)$ of the orbit field on all of $\mathbb{R}_\delta$ with $X(0) = (u_0, 0)$.

(b) Every motion X of the orbit field with $w(0) = 0$ and $a \le 2u(0) \le 3a$ has an apsidal period: a point Φ with $X(\theta+\Phi) = X(\theta)$ for every θ and

$$2\pi_\delta ab \;\le\; \Phi - 2\pi_\delta \;\le\; 2\pi_\delta ab + 8\pi_\delta (ab)^2.$$

(c) In (b), $\Phi = 2\pi_\delta$ if and only if $b = 0$.

No omniscience principle and no choice principle is used.

Proof. (a) Take the constants of Lemma 25.33, put $D = 1-2bU_c$ and $\alpha = rD = \frac{2}{3}bA_t$, so the relations (25.4) hold, and let ϑ be the clock of Lemma 25.34. Since $1/D_m = \sqrt{D}$, $1/h(\psi) = \sqrt{D(1-r\cos\psi)} = G(\psi)$. By (b) of that lemma and Lemma 25.7 with $K = 2$, ϑ has an inverse map ψ with the uniform derivative $G(\psi(\theta))$, and $|G| \le 1$. Put $X(\theta) = \big(U_c + A_t\cos\psi(\theta),\ -A_t G(\psi(\theta))\sin\psi(\theta)\big)$. The computation before (25.4), with the chain rule and the free square root of the positive $D-\alpha\cos\psi$, shows that X is a motion of the orbit field, and $X(0) = (U_c + A_t, 0) = (u_0, 0)$.

(b) Let X be such a motion, with $u_0 = u(0)$, and let $\tilde{X}$ be the motion of (a). By Theorem 25.3(b), $X = \tilde{X}$. Put $\Phi = \vartheta(4q)$. By Lemma 25.34(c), $\vartheta(\psi(\theta)+4q) = \theta+\Phi = \vartheta(\psi(\theta+\Phi))$, and ϑ is injective, so $\psi(\theta+\Phi) = \psi(\theta)+4q$; the cosine, the sine and G have period $4q$, so $X(\theta+\Phi) = X(\theta)$. By Lemma 25.34(d) and the bounds $1+ab \le D_m$ and $D_m(1+2r^2) \le 1+ab+4(ab)^2$, $4q(1+ab) \le \Phi \le 4q(1+ab+4(ab)^2)$, and $4q = 2\pi_\delta$.

(c) If $\Phi = 2\pi_\delta$, then $2\pi_\delta ab \le 0$ with $2\pi_\delta > 0$ and $a > 0$, so $b \le 0$, and $b = 0$. If $b = 0$, the bounds squeeze $\Phi = 2\pi_\delta$. □

The circular orbit, $A_t = 0$, needs no separate case, and nothing is decided about whether the motion starts at the pericentre or at the apocentre: A_t is signed. For a ratio b the test $b = 0$ is arithmetic; for a coefficient given only as a point, deciding "no advance" would cost WLPO, as in Theorem 25.19(e).

The same clock gives the second order. Its bounds are stated in the orbit's own constants, and for the Schwarzschild coefficient $b = 3m$ they become the classical expansion, with a remainder.

Lemma 25.36 (Dominant lowest terms)**.** *Let $F(x,v)$ be a polynomial with ratio coefficients c_{ij} (the coefficient of $x^i v^j$), let n_0 be the least total degree of its terms, and let $f(s) = \sum_{i+j=n_0} c_{ij} s^j$. If $f(s) \geq \mu$ for $0 \leq s \leq \frac{3}{5}$ and*

$$\tau = \sum_{i+j>n_0} |c_{ij}| \Big(\frac{3}{5}\Big)^j \Big(\frac{1}{250}\Big)^{i+j-n_0} < \mu,$$

then $F(x,v) \geq 0$ whenever $0 \leq x \leq \frac{1}{250}$ and $0 \leq v \leq \frac{3}{5}x$.

Proof. For $0 \leq v \leq \frac{3}{5}x$ a term of degree $i+j$ has absolute value at most $|c_{ij}|(\frac{3}{5})^j x^{i+j}$, and for $i+j > n_0$ this is at most $|c_{ij}|(\frac{3}{5})^j(\frac{1}{250})^{i+j-n_0} x^{n_0}$; so the terms of degree above n_0 add up to at least $-\tau x^{n_0}$. If $n_0 = 0$, the term of degree 0 is $c_{00} = f(0) \geq \mu$, and $F(x,v) \geq \mu - \tau \geq 0$. Let $n_0 \geq 1$. Since $F \geq 0$ is the negation of $F < 0$, suppose $F(x,v) < 0$. The same bound for all terms gives $0 < -F(x,v) \leq C x^{n_0}$ with the ratio $C = \sum_{i+j \geq n_0} |c_{ij}|(\frac{3}{5})^j(\frac{1}{250})^{i+j-n_0}$, so $x^{n_0} > 0$; and $x \geq x^{n_0}$ because $0 \leq x \leq 1$, so $x > 0$. Now x is apart from 0, $s = v/x$ is a point with $0 \leq s \leq \frac{3}{5}$, and the terms of degree n_0 add up to $x^{n_0} f(s) \geq \mu x^{n_0}$. So $F(x,v) \geq (\mu - \tau) x^{n_0} \geq 0$, which contradicts $F(x,v) < 0$. □

Theorem 25.37 (The second order in the orbit's own apsides)**.** *Let a, b be ratios in the window and $X = (u, w)$ a motion of the orbit field with $w(0) = 0$ and $a \leq 2u(0) \leq 3a$.*

(a) There are points $\Phi, U_c, A_t, D_m, r, \theta_1$ with $X(\theta + \Phi) = X(\theta)$ for every θ, $u(0) = U_c + A_t$, $u(\theta_1) = U_c - A_t$, $w(\theta_1) = 0$, $D_m^2(1 - 2bU_c) = 1$, $D_m \geq 0$, $r(1 - 2bU_c) = \frac{2}{3}bA_t$ and

$$2\pi_\delta D_m\Big(1 + \tfrac{3}{16}r^2\Big) \leq \Phi \leq 2\pi_\delta D_m\Big(1 + \tfrac{3}{16}r^2 + 2r^4\Big).$$

(b) If $b = 3m$ for a ratio $m \geq 0$, there are points Φ, U_c, A_t, θ_1 with $X(\theta + \Phi) = X(\theta)$ for every θ, $u(0) = U_c + A_t$, $u(\theta_1) = U_c - A_t$, $w(\theta_1) = 0$, $(u(\theta) - U_c)^2 \leq A_t^2$ for every θ, and, with $x = mU_c$ and $y = mA_t$,

$$2\pi_\delta S \leq \Phi \leq 2\pi_\delta (S + 80x^3), \qquad S = 1 + 3x + \tfrac{27}{2}x^2 + \tfrac{3}{4}y^2.$$

No omniscience principle and no choice principle is used.

Proof. (a) Take the constants and the motion of Theorem 25.35, with $\Phi = \vartheta(4q)$, and $\theta_1 = \vartheta(2q)$: there $\psi = 2q$, $\cos\psi = -1$ and $\sin\psi = 0$. The bounds are Lemma 25.34(d).

(b) Now $1 - 2bU_c = 1 - 6x$ and $r(1 - 6x) = 2y$, so $r = 2yD_m^2$. By Lemma 25.33, $0 \leq x \leq \frac{1}{250}$ and $25y^2 \leq 9x^2$; and $D_m \geq 1$ since $D_m^2(1 - 6x) = 1$. The range is $(u - U_c)^2 = A_t^2 \cos^2\psi \leq A_t^2$.

Lower bound. Expanding, $(2 + 6x + 27x^2)^2(1 - 6x) = 4 - 540x^3 - 1215x^4 - 4374x^5 \leq 4 = 4D_m^2(1 - 6x)$, so $2D_m \geq 2 + 6x + 27x^2$. Using $3D_m r^2 = 12y^2 D_m^5$,

$$16\Big(D_m\big(1 + \tfrac{3}{16}r^2\big) - S\Big) = 8\big(2D_m - 2 - 6x - 27x^2\big) + 12y^2(D_m^5 - 1) \geq 0.$$

Upper bound. With $Q = 2 + 6x + 27x^2 + 140x^3$,

$$Q^2(1-6x) - 4 = x^3\big(20 - 2895x - 6894x^2 - 25760x^3 - 117600x^4\big),$$

which is positive for $0 \le x \le \frac{1}{250}$ (the bracket exceeds $20 - 11.58 - 0.12 - 0.01 > 8$). So $2D_m \le Q$, $D_m \le 1 + 3x + \frac{27}{2}x^2 + 70x^3 \le 1 + 4x$, and with $e = D_m - 1 \le \frac{4}{250}$, $D_m^5 - 1 \le 6e \le 24x$ and $D_m^9 \le 2$. Since $D_m \frac{3}{16} r^2 = \frac{3}{4} y^2 D_m^5$ and $2D_m r^4 = 32 y^4 D_m^9$,

$$S + 80x^3 - D_m\big(1 + \tfrac{3}{16}r^2 + 2r^4\big) = \big(1 + 3x + \tfrac{27}{2}x^2 + 80x^3 - D_m\big) - \tfrac{3}{4}y^2(D_m^5 - 1) - 32y^4 D_m^9,$$

and the three parts are at least $10x^3$, at least $-\frac{3}{4}\cdot\frac{9}{25}x^2\cdot 24x = -6.48x^3$ and at least $-64\cdot\frac{81}{625}x^4 \ge -0.04x^3$. So the difference is positive, and (a) gives the bounds. □

Conditions and scope.

Base and premises

The orbit equation $u'' + u = a + bu^2$ with ratio coefficients in the window $100ab \le 1$, started at an apside whose radius $1/u_0$ lies between $2/(3a)$ and $2/a$.

Conclusion

With $p = 1/U_c$ and $e = A_t/U_c$, so that $u = (1 \pm e)/p$ at the two apsides, and $b = 3m$,

$$\Phi - 2\pi_\delta = 6\pi_\delta\frac{m}{p} + \frac{3\pi_\delta}{2}(18 + e^2)\Big(\frac{m}{p}\Big)^2 + R, \qquad 0 \le R \le 160\,\pi_\delta\Big(\frac{m}{p}\Big)^3,$$

since $2\pi_\delta(S-1) = 6\pi_\delta x + 27\pi_\delta x^2 + \frac{3}{2}\pi_\delta y^2$ with $x = m/p$ and $y = ex$.

Identification

That a test body's inverse radius obeys this equation in its polar angle, with $a = m/L^2$ and $b = 3m$, is the Schwarzschild orbit equation of general relativity, taken here as given.

Not claimed

Anything about the motion in time; that is Section 25.7.

The first-order term $6\pi_\delta m/p$ is Einstein's perihelion advance [155]. The expansion of $(1-6x)^{-1/2}$ is $1 + 3x + \frac{27}{2}x^2 + \frac{135}{2}x^3 + \cdots$, so the constant 80 in the remainder is close to the truth.

25.7 The perihelion advance of actual orbits

Section 25.6 studied a function of the polar angle. A planet moves in time. This section takes the motion in proper time as the object: a polynomial field on the eight Kepler coordinates whose motions, in the identification of general relativity, are the orbits of a test body about a mass m. We prove that every such motion in the weak-field window returns in its radial coordinates after a radial period, that its polar angle advances by the amount of Theorem 25.37 in the orbit's own apsides and in its conserved quantities, and we compute the period in proper time and in the time of a distant observer.

Definition 25.38 (The proper-time field and its window orbits)**.** For a ratio Λ, the *proper-time field* G_Λ acts on Kepler states (x, v, u, m) by

$$\dot{x} = v, \qquad \dot{v} = -m\,u^3 x - 3\Lambda m\,u^5 x, \qquad \dot{u} = -u^3(x \cdot v), \qquad \dot{m} = 0,$$

with energy $H_\Lambda = \frac{1}{2}|v|^2 - mu - \Lambda m u^3$. A *window orbit* is a motion y of G_Λ such that for every t: $I = 1$, $u > 0$, $x_2 = v_2 = 0$ and $x_0 v_1 - x_1 v_0 = L$, a constant apart from 0 with $L^2 = \Lambda$; such that $x(0) \cdot v(0) = 0$; and for which there is a point a with $a\Lambda = m$, $a > 0$, $300\,am \le 1$ and $a \le 2u(0) \le 3a$. We write $H = H_\Lambda(y(0))$, $Z = -2H$ and $\lambda = am$. A *radial period* of y is a triple (θ, T, Φ) of a map θ and two points such that θ has derivative $Lu(t)^2$ at every t, $x_0 u = \cos\theta$ and $x_1 u = \sin\theta$ at every t, $LT > 0$, and for every t

$$\theta(t+T) = \theta(t) + \Phi, \qquad u(t+T) = u(t), \qquad (x \cdot v)(t+T) = (x \cdot v)(t).$$

So Λ is the squared angular momentum per unit mass, L^2, which enters the field as a ratio coefficient; $a = m/L^2$ is the Newtonian inverse semi-latus rectum, and $\lambda = m^2/L^2$. Every statement below holds for every ratio value of Λ. The orbit starts at an apside, $x \cdot v = 0$, and the condition $300am \le 1$ is the weak-field window $300m^2 \le L^2$. The sign of L is the sense of rotation and is never decided; $LT > 0$ says that T is a positive time when the body turns counterclockwise and a negative one otherwise. After a radial period the inverse radius and the radial velocity are back, and the polar angle has advanced by Φ: the pericentre turns by $\Phi - 2\pi_\delta$ per radial period.

The identification is classical. A timelike geodesic of Schwarzschild's geometry, written in the flat chart with $u = 1/r$ for the areal radius r, proper time as the clock, $m = GM/c^2$ and L the angular momentum per unit mass, is a motion of G_Λ, and $H = (\mathscr{E}^2 - 1)/2$ with $\mathscr{E}$ the conserved Killing energy: the radial equation $\ddot{r} = -m/r^2 + L^2/r^3 - 3mL^2/r^4$ is the radial part of $\dot{v} = -(m/r^3 + 3mL^2/r^5)x$. The theorems do not supply this identification; they are statements about motions of the polynomial field.

Proposition 25.39 (Energy and rates)**.** *For every ratio Λ, H_Λ and the three components of $x \times v$ are first integrals of G_Λ. On states with $I = 1$, the radial rate $P = -(x \cdot v)u$ satisfies $\mathscr{L}u = u^2 P$ and $\mathscr{L}P = u^2\big(m - |x \times v|^2 u + 3\Lambda m u^2\big)$. No omniscience principle and no choice principle is used.*

Proof. $\frac{d}{dt}\frac{1}{2}|v|^2 = -(mu^3 + 3\Lambda m u^5)(x \cdot v)$ and $\frac{d}{dt}(-mu - \Lambda m u^3) = -(m + 3\Lambda m u^2)\dot{u} = (mu^3 + 3\Lambda m u^5)(x \cdot v)$. The force is central, so $x \times v$ is constant. The rates follow as in Theorem 25.19(a), with $|v|^2 = |L|^2 u^2 + (x \cdot v)^2 u^2$ on $I = 1$. □

Theorem 25.40 (Confinement, the polar angle and motions on the whole line)**.** *Let Λ be a ratio.*

(a) Every window orbit has a map θ with derivative $Lu(t)^2$ at every t and $(x_0 u, x_1 u) = (\cos\theta, \sin\theta)$, and a point Φ with

$$2\pi_\delta(1 + 3am) \le \Phi \le 2\pi_\delta(1 + 3am) + 8\pi_\delta(3am)^2$$

such that $u(t') = u(t)$ and $(x \cdot v)(t') = (x \cdot v)(t)$ whenever $\theta(t') = \theta(t) + \Phi$.

(b) For every state s with $I = 1$, $u > 0$, $x_2 = v_2 = 0$, $L = x_0v_1 - x_1v_0$ apart from 0, $L^2 = \Lambda$ and $x \cdot v = 0$, and every point a with $a\Lambda = m$, $a > 0$, $300am \le 1$ and $a \le 2u \le 3a$, there is a motion y of G_Λ with $y(0) = s$.

No omniscience principle and no choice principle is used.

Proof. (a) *The energy relation.* On a planar state with $I = 1$, $|v|^2 = \Lambda u^2 + P^2$, so $2H_\Lambda = P^2 + \Lambda u^2 - 2mu - 2\Lambda m u^3$, which is constant. With $m = a\Lambda$ this reads $P^2 = \Lambda\big(\mathcal{G}(u_0) - \mathcal{G}(u)\big)$, where $\mathcal{G}(z) = z^2 - 2az - 2mz^3$, $u_0 = u(0)$ and $P(0) = 0$.

The band $2a \le u \le 3a$ is empty. The identity

$$4\big(\mathcal{G}(u) - \mathcal{G}(u_0)\big) - 2a^2 = 4u(u-2a) + (3a-2u_0)(2u_0-a) + 8mu_0^3 \\ + 8m(3a-u)(9a^2+3au+u^2) + a^2\big((1-300am)+84am\big)$$

shows $\mathcal{G}(u) - \mathcal{G}(u_0) \ge a^2/2 > 0$ whenever $2a \le u \le 3a$, and then $P^2 < 0$, which is impossible.

So $u \le 2a$ at every time. On $[-T', T']$ the motion is Lipschitz with some constant C. Take a step h with $Ch \le a/2$. If $u(t) \le 2a$, then at a time t' within h of t, $u(t') \le \frac{5}{2}a \le 3a$, and since the band is empty, $\neg(2a \le u(t'))$, which gives $u(t') \le 2a$. Starting from $u(0) \le \frac{3}{2}a$, induction on the number of steps covers $[-T', T']$. Every statement in this induction is an inequality, so no case is decided.

The polar angle. With $u \le 2a$ the rate Lu^2 is bounded, and Lemma 25.22 gives θ, exactly as in Theorem 25.23.

The orbit equation along the motion. Take the constants of Lemma 25.33 for a and $b = 3m$, which are points here, and the motion (U, W) of the orbit equation $U'' + U = a + 3mU^2$ built in the proof of Theorem 25.35(a); that construction uses only that a and b are points in the window. On the five coordinates $(u, P, m, \tilde{U}, \tilde{Q})$ consider the polynomial field with ratio coefficients

$$\dot{u} = u^2P, \quad \dot{P} = mu^2 - \Lambda u^3 + 3\Lambda m u^4, \quad \dot{m} = 0, \quad \dot{\tilde{U}} = u^2\tilde{Q}, \quad \dot{\tilde{Q}} = mu^2 - \Lambda u^2\tilde{U} + 3\Lambda m u^2\tilde{U}^2.$$

Both (u, P, m, u, P) and $(u, P, m, U(\theta-\theta(0)),\ LW(\theta-\theta(0)))$ are motions of it, by Proposition 25.39 and the chain rule with $\dot{\theta} = Lu^2$ and $\Lambda a = m$; they agree at time 0. By Theorem 25.3, $u(t) = U(\theta(t) - \theta(0))$ and $P(t) = LW(\theta(t) - \theta(0))$. Theorem 25.35(b), whose proof also uses only points, gives the period Φ of (U, W) with the stated bounds for $ab = 3am$, and $x \cdot v = -P/u$.

(b) Take the motion (U, W) and its clock as in (a), and let $\sigma(\psi) = \int_0^\psi h/U^2$ by Lemma 25.6; the integrand is Lipschitz because $U \ge U_c - |A_t| \ge \frac{2}{5}U_c > 0$. The proper time along the orbit is σ/L, a map with slopes bounded above and below, so Lemma 25.7 (after scaling) gives the clock ψ as a function of time, and with θ_0 an angle of (x_0u, x_1u) from Lemma 25.22(a) and $\phi = \theta_0 + \vartheta(\psi)$ put

$$x = \frac{(\cos\phi, \sin\phi, 0)}{U}, \qquad v = -W(\cos\phi, \sin\phi, 0) + LU(-\sin\phi, \cos\phi, 0),$$

with $u = U$ and m constant. The eight coordinates have uniform derivatives on bounded intervals, and $\dot{v} = -(\dot{W} + \Lambda U^3)(\cos\phi, \sin\phi, 0) = -(mu^3 + 3\Lambda m u^5)x$ by the orbit equation. At an apside $v = Lu^2(-x_1, x_0, 0)$, so the motion starts at s. □

For the remaining theorems the motion is read in the clock ψ of Section 25.6, with $U(\psi) = U_c + A_t\cos\psi$. The proper time per unit of clock is $h/(LU^2)$, so we put

$$\sigma(\psi) = \int_0^{\psi} \frac{h(\psi')}{U(\psi')^2}\,d\psi',$$

the antiderivative of Lemma 25.6.

Proposition 25.41 (The window orbit in its clock)**.** *Let y be a window orbit, and let U_c, A_t, D_m, r be the constants of Lemma 25.33 for a, $b = 3m$ and $u_0 = u(0)$, with the clock ϑ of Lemma 25.34 and its inverse. There are a polar angle θ and points T, t_1 such that, with $\psi(t)$ the clock value of $\theta(t) - \theta(0)$:*

(a) $u(t) = U(\psi(t))$ and $\sigma(\psi(t)) = Lt$ for every t;

(b) $(\theta, T, \vartheta(4q))$ is a radial period, $LT = \sigma(4q)$, and $\psi(t+T) = \psi(t) + 4q$ for every t;

(c) $u(t_1) = U_c - A_t$ and $(x\cdot v)(t_1) = 0$;

(d) $(u(t) - U_c)^2 \le A_t^2$ for every t.

No omniscience principle and no choice principle is used.

Proof. (a) The first identity is the proof of Theorem 25.40(a). For the second, ψ has derivative $G(\psi)\,Lu^2$, so $\sigma(\psi(t))$ has derivative $\frac{h}{U^2}\cdot G\cdot Lu^2 = L$, because $hG = 1$ and $u = U(\psi)$; it vanishes at 0, and Lemma 25.4(c) applies.

(b) The integrand h/U^2 has period $4q$, so $\sigma(\psi + 4q) - \sigma(\psi)$ has derivative 0 and equals $\sigma(4q)$; and σ increases at a rate bounded below, so $\sigma(4q) > 0$ and σ is injective. Put $T = \sigma(4q)/L$, using the inverse of L. Then $\sigma(\psi(t+T)) = Lt + \sigma(4q) = \sigma(\psi(t) + 4q)$, so $\psi(t+T) = \psi(t) + 4q$, and $\theta(t+T) - \theta(t) = \vartheta(\psi(t)+4q) - \vartheta(\psi(t)) = \vartheta(4q)$. The inverse radius $U(\psi)$ and $P = LW$, hence $x\cdot v = -P/u$, have period $4q$ in ψ.

(c) Put $t_1 = \sigma(2q)/L$; then $\psi(t_1) = 2q$, where $\cos\psi = -1$ and $\sin\psi = 0$.

(d) $(U - U_c)^2 = A_t^2\cos^2\psi \le A_t^2$. □

Each theorem below asserts a radial period with further properties; in every proof the radial period of Proposition 25.41 serves. We use the abbreviations

$$x = mU_c, \qquad y = mA_t, \qquad k = \sqrt{U_c^2 - A_t^2},$$

so that $0 \le x \le \frac{1}{250}$, $25y^2 \le 9x^2$, $k > 0$, $D_m^2(1-6x) = 1$ and $r = 2yD_m^2$ (Lemma 25.33). The centre relation times m and the energy at the starting apside give the two exact relations

$$\lambda = x - 3x^2 - y^2, \qquad 2H\lambda = -(x^2 - y^2)(1 - 4x). \tag{25.5}$$

For the first, $m(U_c - a) = m(3mU_c^2 + mA_t^2)$. For the second, at the apside $|v| = |L|u_0$, so $2H = \Lambda u_0^2 - 2mu_0 - 2\Lambda m u_0^3$; multiplying by $\lambda = m^2/\Lambda$ and writing $mu_0 = x + y$ gives $(x+y)^2 - 2(x+y)\lambda - 2(x+y)^3$, which expands to the right side.

Theorem 25.42 (The advance in the orbit's own apsides)**.** *Let Λ be a ratio and y a window orbit. There are a radial period (θ, T, Φ) and a time t_1 with $(x\cdot v)(t_1) = 0$ such that, with $u_0 = u(0)$ and $u_1 = u(t_1)$:*

(a) $\big(2u(t) - (u_0 + u_1)\big)^2 \le (u_0 - u_1)^2$ *for every t: the inverse radius stays between its two apsidal values;*

(b) with p and e defined by $p(u_0 + u_1) = 2$ and $e(u_0 + u_1) = u_0 - u_1$,

$$6\pi_\delta\frac{m}{p} + \frac{3\pi_\delta}{2}(18 + e^2)\Big(\frac{m}{p}\Big)^2 \;\le\; \Phi - 2\pi_\delta \;\le\; 6\pi_\delta\frac{m}{p} + \frac{3\pi_\delta}{2}(18 + e^2)\Big(\frac{m}{p}\Big)^2 + 160\,\pi_\delta\Big(\frac{m}{p}\Big)^3.$$

No omniscience principle and no choice principle is used.

Proof. Take the radial period and t_1 of Proposition 25.41. Then $u_0 = U_c + A_t$ and $u_1 = U_c - A_t$, so (a) is part (d) of the proposition, and $p = 1/U_c$, $e = A_t/U_c$ and $m/p = x$, $em/p = y$. The proof of Theorem 25.37(b), which uses only Lemmas 25.33 and 25.34, gives $2\pi_\delta S \le \Phi \le 2\pi_\delta(S + 80x^3)$ with $S = 1 + 3x + \frac{27}{2}x^2 + \frac{3}{4}y^2$, and $2\pi_\delta(S - 1) = 6\pi_\delta x + 27\pi_\delta x^2 + \frac{3}{2}\pi_\delta e^2 x^2$. □

Here p is the semi-latus rectum and e the eccentricity, read off the two apsidal radii $1/u_0$ and $1/u_1$ of the actual orbit; e is negative when the orbit starts at its apocentre, and only e^2 enters. Binary-pulsar timing states the advance in conserved quantities instead [154], and so can we.

Theorem 25.43 (The advance in the conserved energy and angular momentum)**.** *Let Λ be a ratio and y a window orbit, with $H = H_\Lambda(y(0))$ and $\lambda = am = m^2/L^2$. There is a radial period (θ, T, Φ) with*

$$2\pi_\delta F \le \Phi \le 2\pi_\delta\big(F + 300\lambda^3\big), \qquad F = 1 + 3\lambda + \tfrac{105}{4}\lambda^2 + \tfrac{15}{2}H\lambda.$$

So $\Phi - 2\pi_\delta = 2\pi_\delta\big(3\lambda + \frac{105}{4}\lambda^2 + \frac{15}{2}H\lambda\big) + R$ with $0 \le R \le 600\pi_\delta\lambda^3$. No omniscience principle and no choice principle is used.

Proof. By (25.5), F is a polynomial in x and y^2, and

$$\begin{aligned} 4(S - F) &= 570x^3 + 270xy^2 - 945x^4 - 630x^2y^2 - 105y^4, \\ 4(F + 300\lambda^3 - S - 80x^3) &= 310x^3 - 270xy^2 + 105y^4 - 9855x^4 + 21600x^3y^2 \\ &\quad - 2970x^2y^2 + 3600xy^4 + 32400x^5 - 32400x^6 \\ &\quad - 32400x^4y^2 - 10800x^2y^4 - 1200y^6. \end{aligned}$$

Both are nonnegative on the window by Lemma 25.36 with $v = |y|$ (Table 25.2): their lowest terms are $30(19 + 9s^2)x^3$ and $10(31 - 27s^2)x^3$ with $s = v/x$. So $2\pi_\delta F \le 2\pi_\delta S \le \Phi \le 2\pi_\delta(S + 80x^3) \le 2\pi_\delta(F + 300\lambda^3)$. □

The proper-time period needs the time integral $\sigma(4q) = \int_0^{4q} h/U^2$. In Newton's theory $h = 1$ and the integral is done classically by the eccentric anomaly, an angle defined by its cosine and sine. On $\mathbb{R}_\delta$ we construct the eccentric anomaly as an antiderivative and pin its winding by inequalities, without deciding the sign of anything.

Table 25.2. The polynomial inequalities of Section 25.7, each checked by Lemma 25.36 on $0 \le x \le \frac{1}{250}$, $0 \le v \le \frac{3}{5}x$, with $s = v/x$. The column $f(s)$ is the lowest-degree form, μ a lower bound for it on $[0, \frac{3}{5}]$, and τ the tail sum of the lemma, rounded up.

Inequality	Used in	n_0	$f(s)$	μ	τ
$4(S - F) \ge 0$	Thm. 25.43	3	$570 + 270s^2$	570	4.75
$4(F + 300\lambda^3 - S - 80x^3) \ge 0$	Thm. 25.43	3	$310 - 270s^2$	212	44.5
$q_1 \ge 0$	Thm. 25.46(a)	4	$3 - 6s^2$	0.84	0.32
$q_2 \ge 0$	Thm. 25.46(a)	4	$1 + 2s^2$	1	0.50
$q_3 \ge 0$	Thm. 25.46(b)	5	$18 - 24s^2 + 3s^4$	9.36	1.25
$q_4 \ge 0$	Thm. 25.46(b)	5	$2 + 20s^2 - 3s^4$	2	1.48
(F1)	Thm. 25.47	5	$\frac{1}{2}(1 - s^2)(5 - 4s + 3s^2)$	1.17	0.34
(F2)	Thm. 25.47	5	$\frac{1}{25}(73 - 50s + 52s^2 + 50s^3 + 25s^4)$	2.43	0.68
(F3)	Thm. 25.47	5	$7 - 4s - 10s^2 + 4s^3$	1.86	1.12
(F4)	Thm. 25.47	5	$\frac{1}{25}(121 - 100s + 354s^2 + 100s^3)$	4.55	2.99

Lemma 25.44 (The eccentric anomaly, pinned to one turn)**.** *Let U_c, A_t be points with $U_c > 0$ and $25A_t^2 \le 9U_c^2$, put $U(\psi) = U_c + A_t\cos\psi$ and $k = \sqrt{U_c^2 - A_t^2}$. There is a map η with $\eta(0) = 0$ and the uniform derivative k/U on every bounded interval such that:*

(a) $U(\psi)\cos\eta(\psi) = A_t + U_c\cos\psi$ and $U(\psi)\sin\eta(\psi) = k\sin\psi$ for every ψ;

(b) $\frac{1}{2} \le k/U(\psi) \le 2$ for every ψ;

(c) if $\cos d = 1$ and $-q \le d \le 2q$, then $d = 0$;

(d) $\eta(2q) = 2q$ and $\eta(4q) = 4q$;

(e) the functions $(U_c\eta - A_t\sin\eta)/k^3$ and $(U_c\sin\eta - A_t\eta)/k^3$ have the uniform derivatives $1/U^2$ and $\cos\psi/U^2$ on every bounded interval.

No omniscience principle and no choice principle is used.

Proof. $U \ge U_c - |A_t| \ge \frac{2}{5}U_c > 0$ and $k^2 \ge \frac{16}{25}U_c^2$, so k/U is a Lipschitz map, and Lemma 25.6 gives η.

(a) Put $N = \cos\eta\,(A_t + U_c\cos\psi) + \sin\eta\,k\sin\psi$. Differentiating with $\eta' = k/U$ and collecting,

$$N' = -\frac{A_t\sin\psi}{U}\Big(k\sin\eta\sin\psi + \cos\eta\,(A_t + U_c\cos\psi)\Big) = \frac{U'}{U}\,N,$$

using $k^2 = U_c^2 - A_t^2$; so $(N/U)' = 0$, and N/U equals its value 1 at 0. Then

$$\begin{aligned}\big(U\cos\eta - (A_t + U_c\cos\psi)\big)^2 + \big(U\sin\eta - k\sin\psi\big)^2 &= U^2 - 2UN + (A_t + U_c\cos\psi)^2 + k^2\sin^2\psi \\ &= -U^2 + U^2 = 0,\end{aligned}$$

and a sum of two squares that vanishes has both terms 0.

(b) $U \le 2k$ and $k \le 2U$ follow from the identities

$$4k^2 - U^2 = \tfrac{1}{2}\big[(1-c)(3U_c + 5A_t)(U_c - A_t) + (1+c)(3U_c - 5A_t)(U_c + A_t)\big] + A_t^2(1-c)(1+c),$$
$$4U^2 - k^2 = 4\big(U^2 - (U_c - |A_t|)^2\big) + (U_c - |A_t|)(3U_c - 5|A_t|),$$

with $c = \cos\psi$, in which every term is nonnegative.

(c) $1 - \cos d = 2\sin^2(d/2)$, so $\sin(d/2) = 0$, with $-\frac{q}{2} \le d/2 \le q < 2$. The sine is positive on $(0, 2]$ (Section 21.2), which refutes $d/2 > 0$ and, by oddness, $d/2 < 0$. So $d/2 = 0$.

(d) By (b) and Proposition 12.82, $q \le \eta(2q) \le 4q$, so $d = \eta(2q) - 2q \in [-q, 2q]$. At $\psi = 2q$, (a) gives $(U_c - A_t)\cos\eta(2q) = -(U_c - A_t)$ and $(U_c - A_t)\sin\eta(2q) = 0$, with $U_c - A_t > 0$; so $\cos\eta(2q) = -1$, $\sin\eta(2q) = 0$, and $\cos d = \cos\eta(2q)\cos 2q + \sin\eta(2q)\sin 2q = 1$. By (c), $\eta(2q) = 2q$. The same argument on $[2q, 4q]$ gives $\eta(4q) = 4q$.

(e) By (a), $U_c - A_t\cos\eta = k^2/U$ and $U_c\cos\eta - A_t = k^2\cos\psi/U$, so the derivatives are $(U_c - A_t\cos\eta)\eta'/k^3 = 1/U^2$ and $(U_c\cos\eta - A_t)\eta'/k^3 = \cos\psi/U^2$. □

At $\psi = 4q$, where $\eta = 4q$ and $\sin\eta = 0$, part (e) gives the two Newtonian time integrals

$$\int_0^{4q} \frac{d\psi}{U^2} = \frac{4qU_c}{k^3}, \qquad \int_0^{4q} \frac{\cos\psi\, d\psi}{U^2} = -\frac{4qA_t}{k^3}, \tag{25.6}$$

in the sense that the antiderivatives of (e) take these values. The relativistic factor h is compared with them.

Lemma 25.45 (The relativistic factor). *If $|z| \le \frac{1}{250}$, $v \ge 0$ and $v^2(1-z) = 1$, then $1 + \frac{z}{2} \le v \le 1 + \frac{z}{2} + \frac{z^2}{2}$. Consequently, in the window,*

$$4qD_m\Big(U_c - \frac{r}{2}A_t\Big) \le k^3\sigma(4q) \le 4qD_m\Big(\big(1 + \tfrac{r^2}{2}\big)U_c - \frac{r}{2}A_t\Big).$$

No omniscience principle and no choice principle is used.

Proof. $(1+\frac{z}{2})^2(1-z) = 1 - \frac{3}{4}z^2 - \frac{1}{4}z^3 \le 1$ and $(1+\frac{z}{2}+\frac{z^2}{2})^2(1-z) = 1 + \frac{1}{4}z^2(1 - 3z - z^2 - z^3) \ge 1$; compare squares of nonnegative points. With $z = r\cos\psi$ and $z^2 \le r^2$, $D_m(1 + \frac{r}{2}\cos\psi) \le h(\psi) \le D_m(1 + \frac{r}{2}\cos\psi + \frac{r^2}{2})$. So $\sigma - D_m\big(F_0 + \frac{r}{2}F_1\big)$ and $D_m\big((1 + \frac{r^2}{2})F_0 + \frac{r}{2}F_1\big) - \sigma$, with F_0, F_1 the antiderivatives of Lemma 25.44(e), have nonnegative derivatives and vanish at 0; at $4q$ use (25.6). □

Theorem 25.46 (Kepler's third law in proper time). *Let Λ be a ratio and y a window orbit, with $Z = -2H$ and $\lambda = am$.*

(a) There is a radial period (θ, T, Φ) with

$$(2\pi_\delta)^2 m^2 \le T^2 Z^3 \le (2\pi_\delta)^2 m^2 (1 + 4\lambda^2).$$

(b) There are a radial period (θ, T, Φ), a time t_1 with $(x \cdot v)(t_1) = 0$, and points a_K and w with $2a_K u_0 u_1 = u_0 + u_1$ and $2w = m(u_0 + u_1)$, where $u_0 = u(0)$ and $u_1 = u(t_1)$, such that

$$(2\pi_\delta)^2 a_K^3 (a_K + 3m) \le T^2 m\, a_K \le (2\pi_\delta)^2 a_K^3 \big(a_K + 3m + 20 a_K w^2\big).$$

No omniscience principle and no choice principle is used.

Proof. (a) Take the radial period of Proposition 25.41, so $LT = \sigma(4q)$. From $L^2 = \Lambda = m^2/\lambda$ and (25.5), $Z\lambda = (x^2 - y^2)(1 - 4x) = m^2k^2(1 - 4x)$, so

$$T^2Z^3(\lambda m)^2 = \sigma^2\frac{\lambda}{m^2}Z^3\lambda^2m^2 = \sigma^2(Z\lambda)^3 = \big(mk^3\sigma(4q)\big)^2m^4(1 - 4x)^3.$$

By Lemma 25.45, with $4q = 2\pi_\delta$, $r = 2yD_m^2$ and $w_2 = y^2$,

$$2\pi_\delta D_m\,\ell_- \le mk^3\sigma(4q) \le 2\pi_\delta D_m\,\ell_+, \qquad \ell_- = x - w_2D_m^2, \quad \ell_+ = x + 2xw_2D_m^4 - w_2D_m^2,$$

and $\ell_- > 0$. Dividing by $m^4 > 0$, the claim follows from $D_m^2\ell_-^2(1-4x)^3 \ge \lambda^2$ and $D_m^2\ell_+^2(1-4x)^3 \le \lambda^2(1+4\lambda^2)$. Since $D_m^2(1-6x) = 1$, $\ell_- = D_m^2\big(x(1-6x) - w_2\big)$ and $\ell_+ = D_m^4\big(x(1-6x)^2 + 2xw_2 - w_2(1-6x)\big)$; multiplying the two inequalities by $(1-6x)^3$ and $(1-6x)^5$ turns them into $q_1 \ge 0$ and $q_2 \ge 0$, where

$$\begin{aligned} q_1 &= \big(x(1-6x) - w_2\big)^2(1-4x)^3 - (1-6x)^3\lambda^2,\\ q_2 &= (1-6x)^5(1+4\lambda^2)\lambda^2 - \big(x(1-6x)^2 + 2xw_2 - w_2(1-6x)\big)^2(1-4x)^3, \end{aligned}$$

with $\lambda = x - 3x^2 - w_2$. Table 25.2 checks both.

(b) Take the same radial period and t_1; then $u_0u_1 = k^2$ and $u_0 + u_1 = 2U_c$, so $a_K = U_c/k^2$ and $w = x$. Multiplying by m^4k^8, the claim becomes $(2\pi_\delta)^2x^3(x + 3x^2 - 3w_2) \le (mk^3\sigma)^2\lambda x \le (2\pi_\delta)^2x^3(x + 3x^2 - 3w_2 + 20x^3)$, and the same substitution reduces it to $q_3 \ge 0$ and $q_4 \ge 0$:

$$\begin{aligned} q_3 &= \lambda\big(x(1-6x) - w_2\big)^2 - (1-6x)^3x^2(x + 3x^2 - 3w_2),\\ q_4 &= (1-6x)^5x^2(x + 3x^2 - 3w_2 + 20x^3) - \lambda\big(x(1-6x)^2 + 2xw_2 - w_2(1-6x)\big)^2. \end{aligned}$$

Table 25.2 checks both. □

The first-order terms cancel in (a): both $D_m^2\ell_-^2(1-4x)^3$ and λ^2 begin $x^2\big(1-6x-2xe^2+\cdots\big)$ with $e = y/x$, which is why the lowest terms of q_1 and q_2 have degree 4. So in proper time Kepler's energy–period law has *no* first-order relativistic correction: $T^2(-2H)^3 = (2\pi_\delta)^2m^2(1+R)$ with $0 \le R \le 4\lambda^2$. In the semi-major axis a_K, the mean of the two apsidal radii, the same period carries the first-order term $3m/a_K$: $T^2m = (2\pi_\delta)^2a_K^3(1 + 3m/a_K + R')$ with $0 \le R' \le 20(m/p)^2$.

A distant observer does not measure proper time. In Schwarzschild's geometry the coordinate time t_c of a geodesic runs at $dt_c/d\tau = \mathscr{E}/(1 - 2m/r)$. We construct this clock along the motion and compute its period.

Theorem 25.47 (The far observer's period)**.** *Let Λ be a ratio and y a window orbit, with $Z = -2H$ and $\lambda = am$. There are a radial period (θ, T, Φ), a point $E_\infty \ge 0$ with $E_\infty^2 = 1 + 2H$, maps t_c and D with $t_c(0) = 0$ such that t_c has derivative $D(\tau)$ at every τ and $D(\tau)\big(1 - 2m\,u(\tau)\big) = E_\infty$, and a point T_c with $t_c(\tau + T) = t_c(\tau) + T_c$ for every τ, such that*

$$\begin{gathered} (2\pi_\delta)^2m^2(1 + 3Z) \le T_c^2Z^3 \le (2\pi_\delta)^2m^2(1 + 3Z + 20\lambda^2),\\ LT\big(1 + \tfrac{3}{2}Z - 2\lambda^2\big) \le LT_c \le LT\big(1 + \tfrac{3}{2}Z + 8\lambda^2\big). \end{gathered}$$

No omniscience principle and no choice principle is used.

Proof. The clock. By (25.5), $Z = (x^2 - y^2)(1 - 4x)/\lambda$ lies in $[0, \frac{1}{2}]$, so $E_\infty = \sqrt{1 - Z}$ exists, and squaring shows $1 - \frac{Z}{2} - \frac{Z^2}{2} \le E_\infty \le 1 - \frac{Z}{2}$. Along the orbit $2mU \le 2(x + |y|) \le \frac{1}{51}$, so $1 - 2mU \ge \frac{1}{2}$. Let $\sigma_c(\psi) = \int_0^\psi h/(U^2(1 - 2mU))$ by Lemma 25.6, and put $t_c(\tau) = E_\infty \sigma_c(\psi(\tau))/L$ and $T_c = E_\infty \sigma_c(4q)/L$. As in Proposition 25.41, t_c has derivative

$$E_\infty \frac{h}{U^2(1 - 2mU)}\, G\, \frac{Lu^2}{L} = \frac{E_\infty}{1 - 2mu},$$

and the integrand has period $4q$, so $t_c(\tau + T) = t_c(\tau) + T_c$.

The integral. For $0 \le z \le \frac{1}{51}$, $1 + z \le 1/(1 - z) \le 1 + z + \frac{51}{50} z^2$, because

$$(1 + z + cz^2)(1 - z) = 1 + z^2(c - 1 - cz).$$

Apply it with $z = 2mU$, the bounds on h from Lemma 25.45, $\int_0^{4q} d\psi/U = \eta(4q)/k = 4q/k$, $\int_0^{4q} \cos\psi\, d\psi = 0$ and $|\int_0^{4q} \cos\psi\, d\psi/U| \le 4q/k$, each in the sense of antiderivatives. With $mk = \sqrt{x^2 - y^2} \le x$ this gives

$$2\pi_\delta D_m\, c_- \le mk^3 \sigma_c(4q) \le 2\pi_\delta D_m\, c_+,$$

where, with $v = |y|$ and $w_2 = y^2$,

$$\begin{aligned} c_- &= \ell_- + (x^2 - w_2)(2 - 2vD_m^2), \\ c_+ &= \ell_+ + (x^2 - w_2)(2 + 4w_2 D_m^4 + 2vD_m^2) + \tfrac{204}{50}\, x(x^2 - w_2)(1 + 2w_2 D_m^4). \end{aligned}$$

The bridge. $LT = \sigma(4q)$ and $LT_c = E_\infty \sigma_c(4q)$. With the bounds on σ from the proof of Theorem 25.46 and on E_∞ above, the bridge follows from

$$\begin{aligned} &\text{(F1)} \quad \big(1 - \tfrac{Z}{2} - \tfrac{Z^2}{2}\big) c_- \ge \big(1 + \tfrac{3}{2} Z - 2\lambda^2\big) \ell_+, \\ &\text{(F2)} \quad \big(1 - \tfrac{Z}{2}\big) c_+ \le \big(1 + \tfrac{3}{2} Z + 8\lambda^2\big) \ell_-. \end{aligned}$$

The period law. As in Theorem 25.46, with $B = Z\lambda$, $T_c^2 Z^3 (\lambda m)^2 \lambda = (mk^3 \sigma_c)^2 m^4 (1 - 4x)^3 (\lambda - B)$, and $\lambda - B \ge 0$. The period law follows from

$$\begin{aligned} &\text{(F3)} \quad D_m^2 c_-^2 (1 - 4x)^3 (\lambda - B) \ge \lambda^2 (\lambda + 3B), \\ &\text{(F4)} \quad D_m^2 c_+^2 (1 - 4x)^3 (\lambda - B) \le \lambda^2 (\lambda + 3B + 20\lambda^3). \end{aligned}$$

The certificates. Multiply (F1) and (F2) by $\lambda^2 (1 - 6x)^2$, (F3) by $(1 - 6x)^3$ and (F4) by $(1 - 6x)^5$. Each becomes a polynomial inequality in x and v, with 52, 50, 62 and 107 terms, whose terms of lowest degree, 5, and tails are listed in Table 25.2; Lemma 25.36 proves each. □

Conditions and scope.

Base and premises

The proper-time field G_Λ for a ratio value of the squared angular momentum $\Lambda = L^2$, and a motion in the weak-field window $300m^2 \le L^2$ starting at an apside whose radius lies between $\frac{2}{3}$ and 2 times L^2/m.

Conclusion

The advance per radial period is $6\pi_\delta m/p + \frac{3}{2}\pi_\delta(18 + e^2)(m/p)^2$ up to a remainder in $[0, 160\pi_\delta(m/p)^3]$, and $2\pi_\delta(3\lambda + \frac{105}{4}\lambda^2 + \frac{15}{2}H\lambda)$ up to one in $[0, 600\pi_\delta\lambda^3]$. The radial period obeys Kepler's third law in proper time with no first-order correction; the far observer's period is $T_c = T(1 + \frac{3}{2}Z + R)$ with $-2\lambda^2 \le R \le 8\lambda^2$. So the first-order relativistic term in the period belongs to the far observer's clock alone. Everything is free: no omniscience principle and no choice principle.

Identification

That these motions are Schwarzschild geodesics in proper time, that t_c is the Schwarzschild coordinate time, and that a planet is such a test body, is classical general relativity, not derived here. Section 25.8 says which parts of it Recognition Science derives.

Not claimed

Any planetary data; no measured value enters any proof. The window is a hypothesis: Mercury, with m/p near $3 \cdot 10^{-8}$, lies deep inside it.

Reading off a period from a motion is itself a claim: could the inverse radius return earlier than T? Not unless the orbit is circular.

Theorem 25.48 (The radial period is the least one)**.** *Let Λ be a ratio and y a window orbit. There is a radial period (θ, T, Φ) such that, if $u(t_0)$ is apart from $u(0)$ for some t_0, then every T' with $LT' > 0$ and $u(T') = u(0)$ satisfies $LT \le LT'$. No omniscience principle and no choice principle is used.*

Proof. Take the radial period of Proposition 25.41. There $u(t) - u(0) = A_t\big(\cos\psi(t) - 1\big)$, so A_t is apart from 0, and $u(T') = u(0)$ forces $\cos\psi(T') = 1$. The goal $LT \le LT'$ is the negation of $LT' < LT$, so assume $0 < LT' < LT$. Since σ is increasing with $\sigma(0) = 0$, $\sigma(4q) = LT$ and $\sigma(\psi(T')) = LT'$, we get $0 \le \psi(T') \le 4q$. If $\psi(T') \le 2q$, then Lemma 25.44(c) with $d = \psi(T')$ gives $\psi(T') = 0$ and $LT' = 0$. If $\psi(T') \ge 2q$, the same with $d = 4q - \psi(T')$ gives $\psi(T') = 4q$ and $LT' = LT$. Both contradict the assumption. The two cases need not be decided: each is refuted, so neither $\psi(T') > 2q$ nor $\psi(T') < 2q$ can fail, and the assumption yields a contradiction. □

On a circular orbit every T' returns the inverse radius, so the hypothesis is needed. It is stated positively, as apartness, so that the theorem asks for no decision whether a given orbit is circular.

25.8 What Recognition Science supplies here

Every theorem of this chapter is a statement about a polynomial field on $\mathbb{R}_\delta$. Reading it as the motion of a planet or of a spinning body brings in physics: the dimension of space, the force law, the coefficient of the relativistic term, the rule that test bodies follow geodesics, and the values of constants. Recognition Science derives some of these inputs, reaches the relativistic coefficient through the lattice's own field equations, with Regge's action read as the gravitational action and $\kappa = 1$, and leaves the rest as identifications and calibrations. Table 25.3 lists them.

Table 25.3. The physical inputs of this chapter and where each comes from.

Input	Used in	Status
Space has three dimensions; angular momentum is a vector	Sections 25.2–25.5, 25.10	Recognition Science's chain, from its premise that a kept pair of records is a linked closed curve (Chapter 23); $D(D-1)/2 = D$ only for $D = 0$ and $D = 3$
Newton's inverse square; $m = GM$	Sections 25.2–25.5	Identification; the value of G is a calibration
$b = 3m$, that is $\beta = \gamma = 1$	Sections 25.6 and 25.7, first order	The $\kappa = 1$ lattice's own equations: long-wave solutions with point-mass jets have $\gamma = \beta = 1$ (Theorem 25.70); open: the existence of that profile
A post-Newtonian β of J's own cost	none	Does not exist, with Chapter 27's static test field read as that of a static mass: on the Euclidean lattice J's cost is Regge's action plus a gauge-breaking cubic vertex
The second-order coefficients	Theorems 25.37, 25.42, 25.43	Schwarzschild's; the lattice's own equations give $\chi_2 = 1$ (Theorem 25.70), and at the jet level the jet equations built from the quartic cell form give $\chi_3 = 1$ and the whole coefficient, the cell form's reading as the action's fourth derivative resting on two named premises on third-order rates (Theorem 25.71)
Geodesic motion; proper time and a far observer's time	Section 25.7	Classical general relativity (identification)
Masses, moments of inertia, initial states	throughout	Calibration and observation; no measured value enters a proof

Three dimensions and vectors. Recognition Science identifies a kept pair of records with a closed curve linked in space. That identification is a premise of its forcing chain, and from it the chain derives that space has three dimensions, the dimension in which a closed curve can be linked; Chapter 23 proves that topological step. This chapter uses the number in one more way. A rotation in D dimensions turns $D(D-1)/2$ independent planes, and $D(D-1)/2 = D$ holds exactly for $D = 0$ and $D = 3$. So in three dimensions, and only there, each plane of rotation is named by the axis perpendicular to it: angular momentum is the vector $L = x \times v$ of Kepler's problem, and Euler's equations of Section 25.10 are equations for a vector.

The inverse square. The Kepler field is Newton's law, and this chapter takes it as given. Reading it as gravity, and m as GM, is an identification; the value of G is a calibration, and no theorem needs it.

The first relativistic order. For a static, spherically symmetric field with the Eddington–Robertson parameters β and γ, the advance of the pericentre per revolution at the first post-Newtonian order is $\frac{1}{3}(2 + 2\gamma - \beta)$ times Einstein's $6\pi m/p$ [169]. An orbit equation $u'' + u = a + bu^2$ with $a = 1/p$ advances by $2\pi_\delta ab$ per revolution at first order (Theorem 25.35), which is that advance exactly when $b = (2 + 2\gamma - \beta)\, m$; only this first-order advance is compared here, not the orbit equations themselves. In the areal radius, Schwarzschild's field, with $\beta = \gamma = 1$, gives exactly $u'' + u = m/L^2 + 3mu^2$ [155, 168], and the fields of Sections 25.6 and 25.7 have its coefficient $b = 3m$. The lattice's own field equations reach $\beta = \gamma = 1$, and with them $b = 3m$: on $\mathbb{R}_\delta$, with no omniscience principle and no choice, Regge's cubic vertex on the $\kappa = 1$ lattice is Einstein–Hilbert's at long wavelength, and a long-wave solution of the lattice's equations whose jets at a point are those of a point mass has $\gamma = \beta = 1$ (Section 25.12). What remains open is the existence of that long-wave profile; Regge's action read as the gravitational action, and $\kappa = 1$, are identifications. The first-order advance $6\pi_\delta m/p$ of Theorem 25.42, and the absence of a first-order term in the proper-time Kepler law of Theorem 25.46, rest on these.

The cost J has no post-Newtonian β of its own. The clause cannot be traded for the cost J. Chapter 27 proves, in classical mathematics, that the cost built from J on the Kuhn lattice at its Euclidean background equals Regge's action plus an extra cubic vertex, that this extra vertex breaks gauge invariance on shell, and that for one static first-order field, the field of two null static waves, the second-order equation of the J-built cost has no solution. Read as the Euclidean form of the linearized field of a static mass, the identification named there, this is the field from which β is read, so J's cost has no β of its own: the question has no answer, not an unknown one. The theorem concerns that field; it does not treat the point-mass field itself. With Regge's action read as the gravitational action, an identification also named in Chapter 27, nonlinear gravity on that lattice is Regge's action, and there the J-built cost agrees with it to second order in the perturbation and no further; the Lorentzian lattice is not treated. The relativistic coefficient of this chapter comes from Regge's action on the lattice (Section 25.12), not from J.

The second order. The second-order coefficients, $27\pi_\delta(m/p)^2$ and $\frac{3}{2}\pi_\delta e^2(m/p)^2$ in Theorem 25.42 and $\frac{105}{4}$ and $\frac{15}{2}$ in Theorem 25.43, are properties of Schwarzschild's orbit equation and of the proper-time field G_Λ. A static metric that agrees with Schwarzschild's only to the first post-Newtonian order changes them through its higher-order terms, so the second order tests more of a field equation than the first does. Write the temporal and the spatial part of a static, spherically symmetric metric in isotropic coordinates as power series in $U = m/\bar{r}$, with $\bar{r}$ the

isotropic radius. At $\beta = \gamma = 1$ they read

$$-g_{00} = 1 - 2U + 2U^2 - \tfrac{3}{2}\chi_3 U^3 + O(U^4), \qquad g_{ij} = \big(1 + 2U + \tfrac{3}{2}\chi_2 U^2 + O(U^3)\big)\delta_{ij}, \tag{25.7}$$

which names the coefficient of U^2 in the spatial part and that of U^3 in the temporal part. Schwarzschild's field, with $-g_{00} = \big((1-U/2)/(1+U/2)\big)^2$ and $g_{ij} = (1+U/2)^4\delta_{ij}$, has $\chi_2 = \chi_3 = 1$. At $\beta = \gamma = 1$ the second-order advance depends on exactly these two further coefficients, and on no other (Proposition 25.49). The lattice's own equations also give the spatial coefficient its Schwarzschild value, $\chi_2 = 1$ (Theorem 25.70), and at the jet level the third-order jet equation, built from the quartic cell form, gives $\chi_3 = 1$ (Theorem 25.71), with the quartic cell form read as the fourth derivative of δ's action, a reading that rests on two named premises about third-order rates. The proposition is the classical expansion of the apsidal angle of a timelike geodesic, carried to second order. It concerns the geodesics of a given metric and is proved in classical mathematics, over the classical real numbers.

Proposition 25.49 (The second-order advance depends on two more coefficients)**.** *Let a static, spherically symmetric metric satisfy* (25.7)*, with $-g_{00}$ and the spatial factor smooth functions of U near* 0*. Take a timelike geodesic whose areal radius r moves between two distinct apsidal values, with inverses u_0 and u_1; define p and e by $p(u_0+u_1) = 2$ and $e(u_0+u_1) = u_0 - u_1$, as in Theorem 25.42, and put $x = m/p$. Let Φ be the polar angle swept in one radial period, E and L the conserved energy and angular momentum per unit mass, $H = (E^2-1)/2$ and $\lambda = m^2/L^2$. There are numbers C and $x_* > 0$, depending only on the metric, such that for every such geodesic with $x \le x_*$*

$$\Big|\frac{\Phi}{2\pi} - 1 - 3x - \Big(\frac{15}{2} + \frac{15}{4}\chi_2 + \frac{9}{4}\chi_3 + \frac{3}{4}\chi_2\, e^2\Big)x^2\Big| \le Cx^3$$

and

$$\Big|\frac{\Phi}{2\pi} - 1 - 3\lambda - \Big(\frac{39}{2} + \frac{9}{2}\chi_2 + \frac{9}{4}\chi_3\Big)\lambda^2 - \Big(6 + \frac{3}{2}\chi_2\Big)H\lambda\Big| \le C\lambda^3.$$

Proof. Every $O(\cdot)$ below is bounded by a constant that depends only on the metric, for $x \le x_*$ with x_* small.

The areal form. Let $\Psi(U) = 1 + 2U + \frac{3}{2}\chi_2 U^2 + O(U^3)$ be the spatial factor. The areal radius is $r = \bar r\,\Psi^{1/2}$, so $v = m/r$ satisfies $v = U\Psi^{-1/2} = U - U^2 + (\frac{3}{2} - \frac{3}{4}\chi_2)U^3 + O(U^4)$, which inverts to $U = v + v^2 + (\frac{1}{2} + \frac{3}{4}\chi_2)v^3 + O(v^4)$. Since $dr/d\bar r = \Psi^{1/2}\big(1 - U\Psi'/(2\Psi)\big)$ with $\Psi' = d\Psi/dU$, and $U\Psi'/(2\Psi) = v + (\frac{3}{2}\chi_2 - 1)v^2 + O(v^3)$, the metric is $-A\,dt^2 + B\,dr^2 + r^2 d\Omega^2$ with

$$A = 1 - 2v + a_3 v^3 + O(v^4), \qquad B = \Psi\Big(\frac{d\bar r}{dr}\Big)^2 = \Big(1 - \frac{U\Psi'}{2\Psi}\Big)^{-2} = 1 + 2v + b_2 v^2 + O(v^3),$$

where $a_3 = 3 - \frac{3}{2}(\chi_2 + \chi_3)$ and $b_2 = 1 + 3\chi_2$. Schwarzschild's $A = 1 - 2v$, $B = 1/(1-2v)$ has $a_3 = 0$ and $b_2 = 4$.

The orbit equation. Along the geodesic, in its proper time, $E = A\dot t$ and $L = r^2\dot\phi$ are constant and $-A\dot t^2 + B\dot r^2 + r^2\dot\phi^2 = -1$. With $u = 1/r$ as a function of the polar angle, $\dot r = -Lu'$, so

$$B\,u'^2 = N(u), \qquad N(u) = \frac{1}{L^2}\Big(\frac{E^2}{A} - 1\Big) - u^2. \tag{25.8}$$

For Schwarzschild's metric this reads $u'^2 = (E^2 - 1 + 2mu)/L^2 - u^2 + 2mu^3$, whose derivative is the orbit equation $u'' + u = m/L^2 + 3mu^2$ of Section 25.6. The apsides are the zeros u_0, u_1 of N, which is positive between them, and $\Phi = 2\bigl|\int_{u_1}^{u_0} \sqrt{B/N}\,du\bigr|$.

The factor between the apsides. Write $1/A = 1 + 2v + 4v^2 + (8 - a_3)v^3 + v^4\rho(v)$ with ρ smooth near 0. Then

$$N(u) = n_0 + n_1 u - (1 - n_2)u^2 + n_3 u^3 + N_4(u),$$

with $n_0 = (E^2 - 1)/L^2$, $n_1 = 2mE^2/L^2$, $n_2 = 2mn_1$, $n_3 = \tfrac{1}{2}(8 - a_3)m^2 n_1$ and $N_4(u) = \tfrac{1}{2}n_1 m^3 u^4 \rho(mu)$. As N vanishes at u_0 and u_1, $N(u) = (u_0 - u)(u - u_1)R(u)$, where $R(u)$ is minus the second divided difference of N at u_0, u_1, u. Divided differences are linear. That of a constant or of u is 0, that of u^2 is 1, and that of u^3 is $u_0 + u_1 + u$, because $u^3 - (u - u_0)(u - u_1)(u + u_0 + u_1)$ is linear in u; that of N_4 is $\tfrac{1}{2}N_4''(\xi)$ for some ξ between the least and the greatest of u_0, u_1, u. So

$$R(u) = 1 - n_2 - n_3(u_0 + u_1 + u) - \tfrac{1}{2}N_4''(\xi),$$

and the first divided difference, $(N(u_0) - N(u_1))/(u_0 - u_1) = 0$, gives in the same way

$$n_1 = (1 - n_2)(u_0 + u_1) - n_3(u_0^2 + u_0 u_1 + u_1^2) - N_4'(\xi')$$

for some ξ' between u_0 and u_1. Here $u_0 + u_1 = 2/p$, and on $0 \le u \le 2/p$ we have $mu \le 2x$, so $N_4' = n_1 O(x^3)$ and $N_4'' = n_1 m\, O(x^2)$. The last relation is therefore linear in n_1: $n_1\bigl(1 + 4x + O(x^2)\bigr) = 2/p$, so $n_1 = (2/p)\bigl(1 - 4x + O(x^2)\bigr)$. Hence $n_2 = 4x - 16x^2 + O(x^3)$, $N_4'' = O(x^3)$, and at $u = (1 + e\cos\psi)/p$

$$R = 1 - 4x + 16x^2 - (8 - a_3)x^2(3 + e\cos\psi) + O(x^3).$$

The integral. Put $u = (1 + e\cos\psi)/p$ for $0 \le \psi \le \pi$. Then $(u_0 - u)(u - u_1) = (e/p)^2 \sin^2\psi$ and $du = -(e/p)\sin\psi\,d\psi$, so

$$\Phi = 2\int_0^{\pi} \sqrt{B/R}\,d\psi = \int_0^{2\pi} \sqrt{B/R}\,d\psi.$$

With $v = x(1 + e\cos\psi)$ and $(1 - \delta)^{-1/2} = 1 + \tfrac{1}{2}\delta + \tfrac{3}{8}\delta^2 + O(\delta^3)$,

$$\sqrt{B} = 1 + x(1 + e\cos\psi) + \tfrac{3}{2}\chi_2 x^2(1 + e\cos\psi)^2 + O(x^3),$$
$$1/\sqrt{R} = 1 + 2x - 2x^2 + \tfrac{1}{2}(8 - a_3)x^2(3 + e\cos\psi) + O(x^3).$$

Multiply and average over ψ, where $\cos\psi$ has mean 0 and $\cos^2\psi$ mean $\tfrac{1}{2}$:

$$\frac{\Phi}{2\pi} = 1 + 3x + \Bigl(2 - 2 + \tfrac{3}{2}(8 - a_3) + \tfrac{3}{2}\chi_2\bigl(1 + \tfrac{1}{2}e^2\bigr)\Bigr)x^2 + O(x^3),$$

and $8 - a_3 = 5 + \tfrac{3}{2}(\chi_2 + \chi_3)$ gives the first estimate.

The conserved quantities. Multiplied by $m/2$, the value of n_1 reads $E^2\lambda = x - 4x^2 + O(x^3)$. At $u = 0$ the factorization gives $n_0 = -u_0 u_1 R(0)$, and multiplied by m^2 this reads $2H\lambda = -(1 - e^2)x^2\bigl(1 + O(x)\bigr)$. So $H = O(x)$, $e^2x^2 = x^2 + 2H\lambda + O(x^3)$, and with $E^2 = 1 + 2H$, $\lambda = x - 4x^2 - 2Hx + O(x^3)$, that is, $x = \lambda + 4\lambda^2 + 2H\lambda + O(\lambda^3)$, and $\lambda \ge x/2$. Substituting into the first estimate, $3x$ gives $3\lambda + 12\lambda^2 + 6H\lambda$, the term in x^2 without e gives $(\tfrac{15}{2} + \tfrac{15}{4}\chi_2 + \tfrac{9}{4}\chi_3)\lambda^2$, and $\tfrac{3}{4}\chi_2 e^2 x^2$ gives $\tfrac{3}{4}\chi_2(\lambda^2 + 2H\lambda)$; their sum is the second estimate. $\square$

At $\chi_2 = \chi_3 = 1$ the proposition gives $\frac{3\pi}{2}(18 + e^2)(m/p)^2$ and the coefficients $\frac{105}{4}$ and $\frac{15}{2}$, those of Theorems 25.42 and 25.43, which prove more for Schwarzschild's orbit equation: explicit remainders, for motions on $\mathbb{R}_\delta$ with no choice. So, through that clause, the eccentricity term $\frac{3}{2}\pi_\delta e^2(m/p)^2$ of Theorem 25.42 and the coefficient $\frac{15}{2}$ of Theorem 25.43 are reached; the temporal coefficient contributes $\frac{9}{2}\pi_\delta(m/p)^2$ of the term $27\pi_\delta(m/p)^2$ and $\frac{9}{4}$ of the coefficient $\frac{105}{4}$, and these shares are not. The theorems hold for the Schwarzschild orbit equation, whatever field equation produces it.

Geodesics, clocks and data. That a test body moves on a timelike geodesic in its proper time, that a planet is such a body, and that t_c is the time of a distant observer, is classical general relativity; the theorems of Section 25.7 name these identifications and prove nothing about them. No measured number enters a proof. Masses, the moments of inertia of Section 25.10 and initial states are calibrations and observations, and the weak-field window is a hypothesis about the orbit.

25.9 Certified runs are motions

Chapter 15 built a solver for polynomial fields. From ratio data and one finite check, the region certificate, it computes a point of $\mathbb{R}_\delta$ to every precision (Theorem 15.22). What that point is was settled there classically: Theorem 15.29 identifies it with the value of a classical solution that is assumed to exist and to stay in the ball of the certificate. On $\mathbb{R}_\delta$ we can now say what the point is without assuming anything. It is the value at time T of a motion in the sense of Definition 25.2, built from the solver's own Taylor steps.

Throughout this section p is a polynomial field on d coordinates, $x_0 \in \mathbb{Q}_\delta^d$, $M \geq 1$ and T are natural numbers, the quadratized problem of p with initial point x_0, radius M and end time T satisfies the region certificate at a level n_0, and $\bar{x} \in \mathbb{R}_\delta^d$ is its polynomial solution at time T (Definition 15.27). We call this a *passing run.* For a natural number j, $2^{-j}p$ is the field whose coefficients are those of p times 2^{-j}; its motions are those of p run 2^j times more slowly. We write $\|y\|$ for the largest absolute value of a coordinate of y.

Theorem 25.50 (The certified point is the end of a motion)**.** *For a passing run there is a motion X of p on $[0, T]$ with $X(0) = x_0$ and $X(T) = \bar{x}$. Moreover, let a and j be natural numbers with $a \leq T2^j$, and suppose that the quadratized problem of $2^{-j}p$ with initial point x_0, end time a and some radius $M_1 \geq 1$ satisfies the region certificate at some level. Then $X(a/2^j)$ is the polynomial solution of that problem. No omniscience principle and no choice principle is used.*

Proof. Cells. The schedule has the step $h = 2^{-\sigma}$, with $8R \leq 2^\sigma$, and $S = T2^\sigma$ steps (Definition 15.18); let A, n_1, $x(y)$ and $z(y)$ be as in Lemma 25.29. For $k < S$, a level $n \geq n_1$ and $0 \leq \tau \leq h$, let $V_k^{(n)}(\tau) = \sum_{j \leq N(n)} \tau^j X_j$, where X is the formal solution of $x' = p(x)$ from $x(y_k^{(n)})$. By the proof of Lemma 25.29, the monomials of X form the exact coefficient table at the surface point $z(y_k^{(n)})$, so $V_k^{(n)}$ is a Taylor polynomial of the quadratized field, read in the coordinates w_{e_i}. For $n_1 \leq n \leq n'$ the surface points of the two levels at step k lie in the ball of radius M and within $(2A + 1)2^{-n}$ of each other, by Lemma 25.29. By Lemmas 15.14 and 15.15 their coefficients are at most MR^j and differ by at most $(2A + 1)2^{-n}(j + 1)R^j$. The proof of Lemma 15.16 bounds the sums term by term, so the lemma holds with h replaced by

any $\tau \in [0,h]$: the sums of order $N(n)$ differ by at most $2(2A+1)2^{-n}$, and raising the order from $N(n)$ to $N(n')$ changes a sum by at most $M2^{-N(n)}$. Hence

$$\left|V_k^{(n')}(\tau) - V_k^{(n)}(\tau)\right| \le (M + 4A + 2)\, 2^{-n}$$

for all k and τ, and Theorem 12.25 gives the limit $V_k(\tau)$ with no choice.

The equation inside a cell. The polynomial $V = V_k^{(n)}$ has coefficients at most MR^j, and $8R\tau \le 1$ on the cell. For $0 \le \tau_1 \le \tau_2 \le h$ and $\Delta = \tau_2 - \tau_1$, the binomial theorem gives $0 \le \tau_2^j - \tau_1^j - j\tau_1^{j-1}\Delta \le \binom{j}{2}\tau_2^{j-2}\Delta^2$, and $\sum_j \binom{j}{2}8^{2-j} = (8/7)^3 < \frac{3}{2}$; so $|V(\tau_2) - V(\tau_1) - V'(\tau_1)\Delta| \le \frac{3}{2}MR^2\Delta^2$. Since $X' = p(X)$ as power series, $V'(\tau) - p(V(\tau))$ consists of the terms of order at least $N(n)$ of p at the cut-off series, which add up to at most $|p|(2M)2^{-N(n)}$, with $|p|(2M)$ the largest of the $|p_i|(2M)$, by the bound on products in the proof of Lemma 25.29. So $|V(\tau_2) - V(\tau_1) - p(V(\tau_1))\Delta| \le \frac{3}{2}MR^2\Delta^2 + |p|(2M)2^{-n}\Delta$, and in the limit, with $C_1 = \frac{3}{2}MR^2$,

$$\left|V_k(\tau_2) - V_k(\tau_1) - p\big(V_k(\tau_1)\big)(\tau_2 - \tau_1)\right| \le C_1(\tau_2 - \tau_1)^2 \qquad (0 \le \tau_1 \le \tau_2 \le 2^{-\sigma}).$$

At level n the end $V_k^{(n)}(h)$ of cell k is $x\big(\Phi_{N(n)}(z(y_k^{(n)}))\big)$, which is within $(M+2+2A)2^{-n}$ of $x(y_{k+1}^{(n)})$, the start of cell $k+1$, as in the proof of Lemma 25.29. So neighbouring cells meet in the limit; cell 0 starts at $[x_0]_{N(n)}$, hence at x_0 in the limit, and the last cell ends at $\bar{x}$ by Theorem 15.22.

A motion. At a ratio time $t \in [0,T]$ put $X(t) = V_k(t - k2^{-\sigma})$ with $k = \lfloor t2^\sigma \rfloor$, and $k = S - 1$ at $t = T$. Two ratio times closer than one step lie in one cell or in neighbouring ones, so the cell estimate chains to $|X(t) - X(s) - p(X(s))(t-s)| \le C_2(t-s)^2$ for all ratio times in $[0,T]$, with C_2 fixed, and X is Lipschitz there. A Lipschitz family indexed by ratio times extends to every point of $[0,T]$ with no choice: the value at t is the limit of the values at the stages of a regular sequence naming t, which converge at the rate the Lipschitz constant gives, and two names of t give the same limit. The estimate passes to the limit, so X is a motion of p on $[0,T]$, with the modulus $j = (k+1)C_2$ at tolerance k.

Dyadic times. The curve $s \mapsto X(s/2^j)$ is a motion of $2^{-j}p$ on $[0,a]$, by the chain rule. The first part, applied to the problem of $2^{-j}p$, gives a second motion of $2^{-j}p$ on $[0,a]$ from x_0, ending at that problem's polynomial solution. Theorem 25.3(a) makes the two agree at $s = a$. □

Corollary 25.51 (The certified value does not depend on the schedule)**.**

(a) *If the quadratized problems of p with initial point x_0, end time T and radii M and M' satisfy the region certificate at levels n_0 and n_0', their polynomial solutions at time T are equal.*

(b) *Let the coefficients of p be polynomials in parameters $\vartheta_0, \dots, \vartheta_{l-1}$, and let $\vartheta \in \mathbb{Q}_\delta^l$. If the run of the field that carries the parameters as l further coordinates of rate 0, from (x_0, ϑ), and the run of the field with ϑ substituted, from x_0, both pass with the same radius and end time, then the first d coordinates of the first polynomial solution equal the second.*

(c) *For a passing run and every first integral F of p, the motion X of Theorem 25.50 has $F(X(t)) = F(x_0)$ for every $t \in [0,T]$, in particular at every certified dyadic time.*

No omniscience principle and no choice principle is used.

Proof. (a) Theorem 25.50 gives two motions of p on $[0,T]$ from x_0, ending at the two values, and Theorem 25.3(a) makes them equal at T. (b) Along the motion of the first run the parameter coordinates are constant (Lemma 25.4(c)), so its first d coordinates form a motion of the substituted field from x_0; uniqueness compares it with the motion of the second run. (c) is Proposition 25.5. □

Exact conservation follows at once, and it is the statement a physicist checks first: a certified run of a conservative system does not drift.

Theorem 25.52 (Exact conservation along certified runs)**.** *For a passing run and a first integral F of p, $F(\bar{x}) = F(x_0)$. Moreover, let K be a natural number with $|F(y)-F(z)| \le K\|y-z\|$ whenever $\|y\|, \|z\| \le M+3$, and let $a^{(m)}$ be the vector of the answers at precision m of the coordinates of $\bar{x}$. Then $|F(a^{(m)}/2^m) - F(x_0)| \le 2K2^{-m}$ for every m. No omniscience principle and no choice principle is used.*

Proof. Corollary 25.51(c) at $t = T$ gives $F(\bar{x}) = F(x_0)$. By Theorem 15.20 the coordinates of $\bar{x}$ have absolute value at most M, and by Theorem 15.22 each answer satisfies $|a_i^{(m)}/2^m - \bar{x}_i| \le 2^{-m}$; so both points lie in the box of radius $M+3$, and $|F(a^{(m)}/2^m) - F(\bar{x})| \le K2^{-m} \le 2K2^{-m}$. □

A second proof never builds the motion. Along one exact Taylor step the values of F form a power series whose derivative is the series of $\mathscr{L}_p F = 0$, so a step moves F by at most a fixed multiple of its truncation error; the steps are counted by the schedule, and the limit over the levels removes the error. It is the argument of Theorem 25.30 with all rates zero, and it is how the ten integrals of N bodies (Theorem 25.25) are kept exactly by every certified N-body run.

Proposition 25.53 (Hamilton's energy is a first integral)**.** *Let H be a polynomial with ratio coefficients in $q_0, \dots, q_{n-1}, p_0, \dots, p_{n-1}$, and let Hamilton's field be $\dot{q}_i = \partial H/\partial p_i$, $\dot{p}_i = -\partial H/\partial q_i$. Then H is a first integral, and every passing run of Hamilton's field keeps H exactly. No omniscience principle and no choice principle is used.*

Proof. $\mathscr{L}H = \sum_i \big(\partial_{q_i} H \partial_{p_i} H - \partial_{p_i} H \partial_{q_i} H\big) = 0$, an identity of polynomials. Theorem 25.52 does the rest. □

Whether a polynomial is a first integral at all is a question about every point of $\mathbb{R}_\delta^d$, and it might look as if answering it needed an omniscience principle. It does not.

Theorem 25.54 (First integrals are decided on a finite grid)**.** *Let p and F have ratio coefficients, and let E be the largest exponent of a single variable in a term of the polynomial $\mathscr{L}_p F$. Then F is a first integral of p if and only if $\mathscr{L}_p F$ vanishes at every point of the grid $\{0,1,\dots,E\}^d$. Consequently, for all p and F, either F is a first integral of p or it is not, and when it is not, a grid point where $\mathscr{L}_p F \neq 0$ is a witness. No omniscience principle and no choice principle is used.*

Proof. One direction is immediate. For the other, a polynomial P in d variables, of degree at most E in each and vanishing on the grid, is the zero polynomial. For $d = 1$, P has the $E+1$ distinct ratio roots $0, \dots, E$, so its coefficients solve a Vandermonde system with nonzero

determinant and vanish. For $d+1$ variables write $P = \sum_{j \le E} P_j\, x_d^j$; at each grid point of the first d coordinates the one-variable case makes every P_j vanish, and induction makes every P_j zero. A zero polynomial vanishes at every point of $\mathbb{R}_\delta^d$. The grid values are ratios, and each test whether one of them is 0 is integer arithmetic, so the finite conjunction is decided. □

The certified value depends on the initial point with a Lipschitz constant that the solver itself supplies, and that makes predictions from measured states possible. A measured state is a point of $\mathbb{R}_\delta^d$, not a ratio.

Theorem 25.55 (Lipschitz dependence and measured initial states)**.** *Fix a passing run at x_0, with radius M and end time T.*

(a) There is a natural number Λ, which depends only on p, M and T, such that for all ratio points y_1, y_2 whose runs with radius M and end time T pass at some levels, and every ratio r with $\|y_1 - y_2\| \le r$, the polynomial solutions satisfy $\|\bar y_1 - \bar y_2\| \le \Lambda r$.

(b) There are natural numbers ρ, Λ and a map S from $\mathbb{R}_\delta^d$ to $\mathbb{R}_\delta^d$ such that $S(y) = \bar y$ for every ratio point y with $\|y - x_0\| \le 2^{-\rho}$ whose run passes at some level, and $\|S(x) - S(x')\| \le \Lambda \|x - x'\|$ for all points x, x' within $2^{-\rho}$ of x_0.

No omniscience principle and no choice principle is used.

Proof. (a) Take $\Lambda = GK_D$, with G the growth bound of the schedule, which does not depend on the initial point (Definition 15.18), and K_D a Lipschitz constant of the monomials of degree at most D on the ball of radius $M+1$. Run both trajectories at a common level n. By the region certificates both stay in the ball of radius M (Theorem 15.20), where the Taylor step is Lipschitz; the initial points lie in it too, so no closeness to x_0 is needed. The error recursion of Theorem 15.9 is affine in the initial gap: after S steps the trajectories are within $G(K_D r + 2^{-N(n)+1})$ plus defects of order $2^{-N(n)}$. Letting n grow gives $\|\bar y_1 - \bar y_2\| \le \Lambda r$.

(b) One region check at x_0 certifies every ratio point y near it: at the level n_0 plus a fixed number of extra bits, the level trajectories from y stay within the margin $2^{-(n_0+2)}$ of the checked trajectory from x_0 once $\|y - x_0\| \le 2^{-\rho}$, for ρ fixed by n_0 and Λ. Clamping each coordinate into $[x_{0,i} - 2^{-\rho}, x_{0,i} + 2^{-\rho}]$ makes $y \mapsto \bar y$ a Λ-Lipschitz map on all ratio points, and it extends to points through their regular sequences as in the proof of Theorem 25.50, with no choice and the same constant. □

Does every problem have a passing run? For short enough times, yes, and the certificate can be written down.

Theorem 25.56 (Every polynomial field is certified for a short time)**.** *Let p be a polynomial field with $D = \max(1, \deg p)$, t_p terms in all, and c the largest of the numbers $\lceil |q| \rceil$ over its coefficients q, and let $x_0 \in \mathbb{Q}_\delta^d$. Put*

$$X = \max\bigl(1, \max_i \lceil |x_{0,i}| \rceil\bigr), \quad M = X^D + 1, \quad s = \mathrm{bl}(M) + 3, \quad k = \mathrm{bl}\bigl(t_p D c M 2^s\bigr).$$

Then the quadratized problem of $2^{-k}p$ with initial point x_0, radius M and end time 1 satisfies the region certificate at every level. So the solver computes the state at time 2^{-k} of every polynomial field from every ratio initial point. No omniscience principle and no choice principle is used.

Proof. The initial monomials have absolute value at most $X^D = M - 1$. Every coefficient of the quadratized field of $2^{-k}p$ has the form $\alpha_i q 2^{-k}$ with $\alpha_i \le D$, and each variable receives at most t_p of them, so on the ball of radius M the field has size at most $t_p D c M^2/2^k \le M/2^s \le \frac{1}{8}$, and its exact Taylor coefficients of order j are at most $M2^{-sj}$ by the majorants of Lemma 15.14. Over one time unit the exact steps move each coordinate by at most $\frac{1}{4}$, truncation and rounding add at most $\frac{1}{4}$ at every level, and so every level trajectory stays within $M - \frac{1}{2}$, inside the bound $M - 2^{-(n+2)}$ the certificate asks. □

For $\dot{x} = x^2$ from $x_0 = 1$, whose solution $1/(1-t)$ blows up at $t = 1$, the recipe gives $D = 2$, $X = 1$, $M = 2$, $s = 5$ and $k = \mathrm{bl}(128) = 8$: the run to time 2^{-8} passes at every level.

What is classical. Two kinds of statement in this section and the next are proved with classical logic. The first is any claim that one fixed tape machine does all the computing in polynomial time: part (b) of the next proposition, clause (3) of Theorem 25.59, and with it the rigid body's certificate (Corollary 25.67). The second is the completeness of the region check, part (a) of the next proposition, which reasons about a classical solution. Everything else in both sections is free.

Proposition 25.57 (Completeness of the check, and one machine for short times)**.** *The following are proved in classical mathematics over the classical real numbers.*

(a) Let $x : [0,T] \to \mathbb{R}^d$ be continuous with $x(0) = x_0$ and the right derivative $p(x(t))$ at every $t \in [0,T)$, and suppose that for some $\varepsilon > 0$ every monomial $x(t)^\alpha$ with $|\alpha| \le D$ has absolute value at most $M - \varepsilon$ on $[0,T]$, as it does when $\|x(t)\| \le r$ with $r \ge 1$ and $r^D \le M - \varepsilon$. Then the quadratized problem with radius M and end time T satisfies the region certificate at some level.

(b) There are one control table H and one natural number λ such that for every polynomial field p and every $x_0 \in \mathbb{Q}_\delta^d$ there is a constant K with: for every $i < d$ and every m, H runs from the request of the problem of Theorem 25.56 for coordinate i at precision m to the answer word of its answer within $K(m+1)^\lambda$ steps.

Proof. (a) The estimate behind Theorem 15.29 does not use the certificate step by step: on each step the true solution and the Taylor polynomial of the computed point differ by a residual of order 8^{-N}, and the classical Gronwall inequality, applied while the trajectory stays in the ball, bounds the level-n trajectory's distance from the true monomials by $C2^{-n}$. For n with $C2^{-n} + 2^{-(n+2)} \le \varepsilon$ the trajectory stays within $M - 2^{-(n+2)}$, which is the certificate at level n. (b) Combine Theorem 25.56 with Theorem 15.33. □

Everything so far concerns one field. The certificates combine into one statement for a physical theory, which we now make precise.

Definition 25.58 (Polynomial theories and sign laws)**.** A *polynomial theory* consists of a polynomial field p on d coordinates, a finite list of polynomials, the *candidate integrals*, and a finite list of *sign laws* on polynomial observables. A sign law on an observable F is built from the atoms $F < c$, $F \le c$, $F = c$, $c < F$ and $c \le F$, with ratio thresholds c, by "and" and "or". *Kepler's theory* has the Kepler field, the candidates E, L_0, L_1, L_2, and the sign law $E < 0 \vee E = 0 \vee E > 0$ on the energy.

Theorem 25.59 (One conversion certificate for a whole polynomial theory)**.** *Every polynomial theory satisfies the following five clauses.*

(1) Each candidate integral is a first integral of p, or it is not.

(2) Each candidate that is a first integral keeps exactly its initial value at the polynomial solution of every passing run, and at the answers at precision m it is within $C2^{-m}$ of that value, for a natural number C.

(3) There are one control table H and one natural number λ such that, for every passing run, H computes the answer at precision m of each coordinate of the polynomial solution from its request within $K(m+1)^{\lambda}$ steps, with K depending on the run only.

(4) For each sign law on an observable that takes every value of $\mathbb{R}_\delta$ as the state varies, "the law holds at every state" is equivalent to one of: false, true, LLPO, LPO; and "at every state the law holds or fails" is equivalent to one of: true, LLPO, WLPO, LPO. Which one is computed by a finite rule, given in the proof, from where the law's atoms hold near each threshold.

(5) Every passing run has the extension S of Theorem 25.55(b) to measured initial states.

Clauses (1), (2), (4) and (5) are proved with no omniscience principle and no choice principle. Clause (3) is proved in classical mathematics, and so is the certificate as a whole. For Kepler's theory all four candidates are first integrals, and deciding the sign law at every state costs exactly LPO.

Proof. (1) is Theorem 25.54, (2) is Theorem 25.52, (3) is Theorem 15.33 for polynomial fields, and (5) is Theorem 25.55.

(4) *Germs.* Since the observable takes every value, a statement about all states is the same statement about all points z of the line, as in Theorem 25.14. Write the law as a disjunction of conjunctions of atoms; distributing "and" over "or" does not change where it holds. The *window* of a threshold c is the open interval between its neighbouring thresholds, unbounded where there is none. On the window every atom with another threshold is constantly true or constantly false, and comparing ratios tells which. So on the window each disjunct says that z lies in one of seven sets, its *germ* at c: the empty set, $\{c\}$, the open sides $z < c$ and $z > c$, the closed sides $z \le c$ and $z \ge c$, or the whole window. The law's six readings at c record whether some germ contains the point c, the open side $z < c$, the open side $z > c$, the closed side $z \le c$, the closed side $z \ge c$, or the whole window, the *neighbourhood reading*. A germ that contains one point of an open side contains that whole side.

From a threshold to all points. Cotransitivity, applied once at each gap between consecutive thresholds, places every point in some window. And for a ratio $\epsilon > 0$ less than the distance

from c to its neighbours, the point $w = c + \max(-\epsilon, \min(\epsilon, z - c))$ lies in the window of c, and $w < c$, $w = c$ or $w > c$ exactly when $z < c$, $z = c$ or $z > c$. So a comparison with c that holds at every point of the window holds at every point, and by translation it is the same comparison with 0; its price is in Corollary 12.18 and Table 12.2: $z \le c \vee z \ge c$ costs LLPO, $z = c \vee z \ne c$ and $z \le c \vee \neg(z \le c)$ cost WLPO, and $z < c \vee \neg(z < c)$ and $z < c \vee z = c \vee z > c$ cost LPO. Recall that $\neg(z < c)$ is $c \le z$.

Holds at every point. If at some threshold c the point c, or one of the open sides, is in no germ, then the law fails at c, or at a ratio point of the window beside c, and "holds at every point" is false. Otherwise the upper bounds: with the neighbourhood reading at every threshold the law holds wherever a point is placed, with no principle; with both closed sides in germs at every threshold, LLPO places a point of a window at $z \le c$ or $z \ge c$; and LPO places it at $z < c$, $z = c$ or $z > c$, each in a germ. The lower bounds: suppose the law holds at every point. If the neighbourhood reading fails at c, every germ lies in $z \le c$ or in $z \ge c$, so at each point of the window the disjunct that holds gives $z \le c \vee z \ge c$, which is LLPO. If moreover the closed side $z \le c$, say, is in no germ, every nonempty germ lies in $z < c$ or in $z \ge c$, and the law gives $z < c \vee \neg(z < c)$, which is LPO.

Holds or fails at every point. At a threshold c the rule gives a local price, and the upper bound places a point of the window as finely as the price allows. It is free when the neighbourhood reading holds, or when no germ contains c or an open side, for then the law fails on the whole window. When c and both open sides are in germs, the law cannot fail on the window, since $\neg(z < c)$ and $\neg(z > c)$ give $z = c$; so "holds or fails" is "holds", and the price is LLPO if both closed sides are in germs and LPO otherwise, as above. When c is in no germ but an open side, say $z < c$, is, the law holds only apart from c and fails only where $\neg(z < c)$, so deciding it decides $z < c \vee \neg(z < c)$: LPO. When c is in a germ and neither open side is, the law holds exactly at c, and deciding it is $z = c \vee z \ne c$: WLPO. When c and exactly one open side, say $z < c$, are in germs, the law holds only where $z \le c$: if the closed side $z \le c$ is in a germ, the law holds exactly where $z \le c$ and costs WLPO; otherwise it holds exactly where $z < c$ or $z = c$, and deciding it decides $z < c \vee \neg(z < c)$: LPO. The law's price is the strongest local price, in the order free, LLPO, WLPO, LPO: each local price is a lower bound, by the clamping above, and the strongest principle gives the weaker ones and decides the law at every point, since every point is placed in some window.

Section 26.8 states the same rule as a table of these bits (Definition 26.51 and Table 26.3) and proves it again, with two further classes, as Theorem 26.52.

For Kepler's theory, the four candidates are first integrals by Theorem 25.9(a). The germs of the sign law at its threshold 0 are $E < 0$, $\{0\}$ and $E > 0$: the point and both open sides are covered, but no closed side and not the whole window. So both its truth at every state and its decision at every state cost LPO, as Theorem 25.14 found directly. □

The certificate is the bridge from a physical theory written as polynomials to computations one can trust: its integrals are found by a finite check, conserved exactly by every certified run, its laws priced, and its predictions extended to measured states. In the certificate, only the claim that one fixed tape machine does all the computing in polynomial time is classical, and the reason is the construction of that machine in Chapter 15, not anything about mechanics.

25.10 The rigid body

Toss a book into the air, spinning it about each of its three principal axes in turn. About the axis of largest moment of inertia and about the axis of smallest it keeps turning steadily; about the middle axis it flips over and over. Euler's equations explain both behaviours, and the classical explanation is short: two first integrals and one inequality for the stable axes, and Chetaev's growth argument for the middle one [153, 163]. On $\mathbb{R}_\delta$ all of it holds as it stands, with no omniscience and no choice. The one omniscience principle this section meets sits where classical mechanics makes a decision without comment: which axis of a given body is the one of largest moment.

Definition 25.60 (Euler's field)**.** *Euler's field* is the polynomial field on the six coordinates $L_0, L_1, L_2, J_0, J_1, J_2$ given by

$$\dot{L}_k = (J_{k+2} - J_{k+1})\, L_{k+1} L_{k+2}, \qquad \dot{J}_k = 0 \qquad (k = 0, 1, 2),$$

with indices read modulo 3. Its *energy* and its *Casimir* are the polynomials

$$E = \tfrac{1}{2}\big(J_0 L_0^2 + J_1 L_1^2 + J_2 L_2^2\big), \qquad C = L_0^2 + L_1^2 + L_2^2.$$

Read L as the angular momentum of a body in its principal frame and J_k as the inverse $1/I_k$ of its principal moments of inertia. Then $\omega_k = J_k L_k$ is the angular velocity, and the field is Euler's equation $\dot{L} = L \times \omega$ for a body on which no torque acts [156]. That reading is an identification; so is the restriction to the body frame, which drops the orientation of the body from the state, and the moments are calibrations. Three dimensions enter as in Section 25.8: only for $D = 3$, among positive dimensions, is the number $D(D-1)/2$ of planes of rotation equal to D, so only there is angular momentum a vector with one component per axis. The moments have rate 0, so each J_k is a first integral and stays at its initial value along every motion (Proposition 25.5).

Theorem 25.61 (The energy and the Casimir are first integrals)**.** *E and C are first integrals of Euler's field. No omniscience principle and no choice principle is used.*

Proof. Both Lie derivatives carry the factor $L_0 L_1 L_2$, and the remaining sums cancel cyclically:

$$\begin{aligned} \mathscr{L}C &= \sum_k 2L_k(J_{k+2} - J_{k+1})L_{k+1}L_{k+2} = 2L_0L_1L_2 \sum_k (J_{k+2} - J_{k+1}) = 0, \\ \mathscr{L}E &= \sum_k J_k L_k (J_{k+2} - J_{k+1}) L_{k+1} L_{k+2} \\ &= L_0 L_1 L_2 \,(J_0 J_2 - J_0 J_1 + J_1 J_0 - J_1 J_2 + J_2 J_1 - J_2 J_0) = 0. \end{aligned}$$

□

So every motion keeps E and C (Proposition 25.5), every passing run keeps them exactly (Theorem 25.52), and L moves on the intersection of the sphere $C = \text{const}$ with the ellipsoid $E = \text{const}$. Here is the stability argument in a small case. Take $J = (1, 2, 3)$, so that axis 0 has the largest moment, and start at $L = (1, \frac{1}{10}, \frac{1}{10})$. The combination $2E - J_0 C$ removes L_0:

$$2E - J_0 C = (J_1 - J_0)L_1^2 + (J_2 - J_0)L_2^2 = L_1^2 + 2L_2^2,$$

and its value $\frac{3}{100}$ is kept along the motion. So $L_1^2 + L_2^2 \le \frac{3}{100}$ at every time: the part of L transverse to axis 0, of length $\sqrt{2}/10$ at the start, never exceeds $\sqrt{3}/10$. The argument uses only that $J_1 - J_0$ and $J_2 - J_0$ are positive.

Theorem 25.62 (Rotation about an extreme axis is stable, with an explicit bound)**.** *Let s and s' be states of Euler's field with the same moments, the same energy and the same Casimir. Write J_k and L_k for the coordinates of s, L'_k for those of s', and put $\alpha = J_1 - J_0$, $\beta = J_2 - J_0$.*

(a) At every state and for each k, $2E - J_k C = (J_{k+1} - J_k)L_{k+1}^2 + (J_{k+2} - J_k)L_{k+2}^2$.

(b) If $J_0 < J_1$ and $J_0 < J_2$, then

$$\min(\alpha, \beta)\,(L_1'^2 + L_2'^2) \le \max(\alpha, \beta)\,(L_1^2 + L_2^2),$$

and $L_1'^2 + L_2'^2 \le \varepsilon$ for every ε with $\max(\alpha, \beta)(L_1^2 + L_2^2) \le \min(\alpha, \beta)\,\varepsilon$.

(c) If $J_1 < J_0$ and $J_2 < J_0$, then $L_1'^2 + L_2'^2 \le \varepsilon$ for every ε with

$$\max(J_0 - J_1, J_0 - J_2)\,(L_1^2 + L_2^2) \le \min(J_0 - J_1, J_0 - J_2)\,\varepsilon.$$

(d) With no order among the moments assumed, the first inequality of (b) holds between $X(t)$ and $X(0)$ for every motion X of Euler's field on $[0, 1]$ and every $t \in [0, 1]$, with α and β read at $X(0)$; between the polynomial solution of every passing run of Euler's field and its initial point; and, for every passing run at x_0, between $S(x)$ and x for every point x within $2^{-\rho}$ of x_0 in each coordinate, for some natural number ρ and a map S that agrees with the polynomial solution at every ratio point of that cube whose run passes.

No omniscience principle and no choice principle is used.

Proof. (a) $2E - J_k C = \sum_i (J_i - J_k)L_i^2$, and the term $i = k$ vanishes.

(b) By (a), $Q = \alpha L_1^2 + \beta L_2^2$ equals $2E - J_0 C$, so it has the same value at s and at s'. Put $\underline{m} = \min(\alpha, \beta)$ and $\overline{m} = \max(\alpha, \beta)$, which exist on the line with no choice (Proposition 12.19). Since $\alpha - \underline{m}$, $\beta - \underline{m}$, $\overline{m} - \alpha$, $\overline{m} - \beta$ and all squares are nonnegative, and products of nonnegative points are nonnegative,

$$\underline{m}\,(L_1'^2 + L_2'^2) \le Q(s') = Q(s) \le \overline{m}\,(L_1^2 + L_2^2).$$

This uses no order among the moments. When $\alpha > 0$ and $\beta > 0$ are witnessed, so is $\underline{m} > 0$, and $\underline{m}$ is invertible (Proposition 12.19); multiplying $\underline{m}(L_1'^2 + L_2'^2) \le \overline{m}(L_1^2 + L_2^2) \le \underline{m}\,\varepsilon$ by its inverse gives $L_1'^2 + L_2'^2 \le \varepsilon$.

(c) The same argument applied to $J_0 C - 2E = (J_0 - J_1)L_1^2 + (J_0 - J_2)L_2^2$.

(d) Along a motion on $[0, 1]$ the moments, E and C are constant (Theorem 25.61 and Proposition 25.5), and the displayed inequality of (b), whose proof used no order, applies to $s = X(0)$ and $s' = X(t)$. For a passing run, Theorem 25.52 keeps E, C and each J_k exactly at the polynomial solution. For measured states, take ρ and S from Theorem 25.55(b); its proof shows that every ratio point of the cube has a passing run. At such a point y, $S(y)$ is the polynomial solution and keeps every first integral F by Theorem 25.52. A point x of the cube is the limit of ratio points y_n of the cube, S is Lipschitz and F is a polynomial, so $F(S(x)) = \lim F(S(y_n)) = \lim F(y_n) = F(x)$, and the inequality holds between $S(x)$ and x. □

Conditions and scope.

Base and premises
Euler's field on $\mathbb{R}^6_\delta$; two states on the same level sets of the moments, the energy and the Casimir; for the bound, the strict inequalities $J_0 < J_1$ and $J_0 < J_2$ given with their witnesses.

Conclusion
Rotation about the axis of largest moment, and likewise about the axis of smallest, is stable with an explicit modulus: a transverse part with $L_1^2 + L_2^2 \le \varepsilon \min(\alpha, \beta)/\max(\alpha, \beta)$ at the start stays at most ε, along every motion, at every certified end state and for measured initial states.

Identification
L is a free body's angular momentum in its principal frame and J_k its inverse principal moments; three dimensions are those Recognition Science derives from its premise that a kept pair of records is a linked closed curve (Section 25.8). The moments are calibrations.

Not claimed
That axis 0 of a given body is its axis of largest moment: deciding that for all bodies costs LPO (Theorem 25.63).

The hypothesis of (b) is a decision about three points of the line: since $J_k = 1/I_k$, it says that axis 0 is the axis of largest moment. For a body whose moments are ratios it is integer arithmetic. For moments that are general points it is not free, and the price is exact.

Theorem 25.63 (Deciding the axis of largest moment costs exactly LPO)**.** *The statement "for all $J_0, J_1, J_2 \in \mathbb{R}_\delta$, either $J_0 < J_1$ and $J_0 < J_2$, or not" is equivalent to LPO. No choice principle is used.*

Proof. Assume LPO. By Corollary 12.18 every point x satisfies $x > 0 \vee x = 0 \vee x < 0$. Applied to $J_1 - J_0$ this decides $J_0 < J_1$, since the last two cases refute it; likewise $J_0 < J_2$, and a conjunction of two decided statements is decided.

Conversely, take $J_1 = J_2 = 0$. The statement then decides $J_0 < 0$ for every point J_0, so with $J_0 = -z$ it decides $z > 0$ for every z. Given a binary sequence α, let z be the point of its positive jump presentation (Chapter 12). If $z > 0$, a stage of the presentation exceeds its tolerance, and the bounded search at that stage finds an n with $\alpha(n) = 1$, as in the proof of Theorem 12.7. If not $z > 0$, then $z \le 0$; every numerator of the presentation is nonnegative, so $z \ge 0$, hence $z = 0$, and Lemma 12.5 makes every $\alpha(k)$ zero. This is LPO. The values can be taken positive, as the inverse moments of a body are: the point z of a jump presentation satisfies $0 \le z \le 1$, so $J_1 = J_2 = 2$ and $J_0 = 2 - z$ are all positive, and $J_0 < J_1$ exactly when $z > 0$. So deciding, for all bodies, whether axis 0 is the axis of largest moment already gives LPO. □

The middle axis behaves differently, and the difference is visible in the energy–Casimir form of axis 1: $2E - J_1 C = (J_2 - J_1)L_2^2 + (J_0 - J_1)L_0^2$ has one coefficient of each sign when $J_2 < J_1 < J_0$, so its level set through the axis is not a small closed curve but a pair of crossing curves. For $J = (3, 2, 1)$, the states $(0, \sqrt{2}, 0)$ and $(1, 0, 1)$ both have $C = 2$ and $2E = 4$: one turns exactly about the middle axis, the other is perpendicular to it, and the integrals cannot tell them apart.

Theorem 25.64 (Chetaev: rotation about the middle axis is unstable). *Let $J_0, J_1, J_2 \in \mathbb{R}_\delta$ with $J_2 < J_1 < J_0$, so that axis 1 has the middle moment.*

(a) For every $K \in \mathbb{R}_\delta$ there are states s and s' with moments J_0, J_1, J_2, the same energy and the same Casimir, such that $K(L_0^2 + L_2^2) < L_0'^2 + L_2'^2$. So no bound $L_0'^2 + L_2'^2 \le K(L_0^2 + L_2^2)$ holds on the level sets of the integrals.

(b) Let X be a motion of Euler's field on $[0,1]$ whose moments at time 0 are J_0, J_1, J_2, let $W(t) = -L_0(t)L_2(t)$, and let $\mu \ge 0$ with $\mu^2 \le (J_0 - J_1)(J_1 - J_2)$ and $\ell_1 \ge 0$. If $L_1(t) \ge \ell_1$ for every $t \in [0,1]$, then

$$2W(0)\,(1 + 2\mu\ell_1 h)^n \le C\big(X(0)\big)$$

for every $h \ge 0$ and every natural number n with $nh \le 1$. Consequently, if $C(X(0)) < 2W(0)(1 + 2\mu\ell_1 h)^n$ for one such h and n, then it is not the case that $L_1(t) \ge \ell_1$ for every $t \in [0,1]$.

No omniscience principle and no choice principle is used.

Proof. (a) The points $J_0 - J_2$, $J_1 - J_2$ and $J_0 - J_1$ are positive, so they have square roots with no choice (Proposition 12.19). Let s have $L = (0, \sqrt{J_0 - J_2}, 0)$ and s' have $L' = (\sqrt{J_1 - J_2}, 0, \sqrt{J_0 - J_1})$, both with moments J_0, J_1, J_2. Both have $C = J_0 - J_2$, and

$$2E(s') = J_0(J_1 - J_2) + J_2(J_0 - J_1) = J_1(J_0 - J_2) = 2E(s).$$

At s, $K(L_0^2 + L_2^2) = 0$, while $L_0'^2 + L_2'^2 = J_0 - J_2 > 0$.

(b) Suppose $L_1(t) \ge \ell_1$ on $[0,1]$. The moments are constant along X. With $a = J_0 - J_1 > 0$ and $b = J_1 - J_2 > 0$, Lemma 25.4(b) gives W the uniform derivative

$$\dot W = -(\dot L_0 L_2 + L_0 \dot L_2) = L_1\big(bL_2^2 + aL_0^2\big)$$

on $[0,1]$. Completing the square,

$$a\big(bL_2^2 + aL_0^2 + 2\mu L_0 L_2\big) = (aL_0 + \mu L_2)^2 + (ab - \mu^2)L_2^2 \ge 0,$$

and $a > 0$, so $bL_2^2 + aL_0^2 \ge 2\mu W$. With $L_1 \ge \ell_1 \ge 0$ and $bL_2^2 + aL_0^2 \ge 0$ this gives, on $[0,1]$,

$$\dot W \ge 0 \qquad \text{and} \qquad \dot W \ge \ell_1\big(bL_2^2 + aL_0^2\big) \ge 2\mu\ell_1 W.$$

A function f whose uniform derivative is at least γ on $[s,t]$ satisfies $f(t) - f(s) \ge \gamma(t-s)$: divide $[s,t]$ into steps within the modulus at tolerance k, so that each step gains at least $(\gamma - \frac{1}{k+1})$ times its length, and let k grow. With $\gamma = 0$, W is nondecreasing on $[0,1]$. On $[s, s+h]$ the derivative is then at least $2\mu\ell_1 W(s)$, since $2\mu\ell_1 \ge 0$, so $W(s+h) \ge W(s)(1 + 2\mu\ell_1 h)$. Induction, multiplying by the positive factor $1 + 2\mu\ell_1 h$ at each step, gives $W(nh) \ge W(0)(1 + 2\mu\ell_1 h)^n$ whenever $nh \le 1$. Finally $2W = -2L_0L_2 \le L_0^2 + L_2^2 \le C$, because $(L_0 + L_2)^2 \ge 0$, and C is constant along X. So $2W(0)(1 + 2\mu\ell_1 h)^n \le 2W(nh) \le C(X(nh)) = C(X(0))$. Under the strict inequality of the last sentence, the assumption would give $C(X(0)) < C(X(0))$. □

The best choice is $\mu = \sqrt{(J_0 - J_1)(J_1 - J_2)}$, which the line supplies with no choice. As $h = t/n$ shrinks, $(1 + 2\mu\ell_1 t/n)^n$ increases to $e^{2\mu\ell_1 t}$, so while $L_1 \geq \ell_1$ the product W grows at least like $e^{2\lambda t}$ with $\lambda = \ell_1\sqrt{(J_0 - J_1)(J_1 - J_2)}$, the growth rate of the classical linearization about steady rotation with $L_1 = \ell_1$. A state near the middle axis with $W(0) > 0$ must therefore leave the region $L_1 \geq \ell_1$ within a time of about $(2\lambda)^{-1}\log\bigl(C/2W(0)\bigr)$. The window $[0, 1]$ loses nothing: Euler's field is homogeneous of degree two in L, so $t \mapsto (cL(ct), J)$ is a motion on $[0, 1]$ whenever (L, J) is one on $[0, c]$. The escape is not only forced; it can be written down.

Proposition 25.65 (The separatrix leaves the middle axis at the rate λ)**.** *Let $u, v, h, r, J_1, \tau \in \mathbb{R}_\delta$ with $h^2 = u^2 + v^2$, put $\lambda = uvhr$ and $\phi(t) = \lambda(t - \tau)$, and write $\operatorname{sech} z = 2/(e^z + e^{-z})$ and $\tanh z = (e^z - e^{-z})/(e^z + e^{-z})$. The curve with moments $(J_1 + v^2, J_1, J_1 - u^2)$ and*

$$L(t) = \bigl(-ur\operatorname{sech}\phi(t),\ -hr\tanh\phi(t),\ vr\operatorname{sech}\phi(t)\bigr)$$

is a motion of Euler's field on $[0, 1]$. Along it $C = h^2r^2$ at every time; at time 0, $L_0^2 + L_2^2 \leq h^2r^2\,(2e^{-\lambda\tau})^2$; at time τ the state is $L = (-ur, 0, vr)$ with the same moments; and

$$\lambda^2 = (J_0 - J_1)(J_1 - J_2)\,C.$$

No omniscience principle and no choice principle is used.

Proof. For every z, $e^z + e^{-z} - 2 = (e^{z/2} - e^{-z/2})^2 \geq 0$, so the denominator is apart from 0 and has an inverse with no choice (Proposition 12.19); the exponential is free (Proposition 12.88). By the chain and quotient rules, $S = \operatorname{sech}\phi$ and $T = \tanh\phi$ have the uniform derivatives $-\lambda ST$ and λS^2 on $[0, 1]$, and $S^2 + T^2 = 1$ is algebra. The moments give $J_0 - J_1 = v^2$, $J_1 - J_2 = u^2$ and $J_0 - J_2 = h^2$, and

$$\begin{aligned}
\dot L_0 &= ur\lambda ST = u^2vhr^2ST = (-u^2)(-hrT)(vrS) = (J_2 - J_1)L_1L_2,\\
\dot L_1 &= -hr\lambda S^2 = -uvh^2r^2S^2 = h^2(vrS)(-urS) = (J_0 - J_2)L_2L_0,\\
\dot L_2 &= -vr\lambda ST = -uv^2hr^2ST = (-v^2)(-urS)(-hrT) = (J_1 - J_0)L_0L_1.
\end{aligned}$$

Further $C = (u^2 + v^2)r^2S^2 + h^2r^2T^2 = h^2r^2$. At $t = 0$, $\phi = -\lambda\tau$, and sech is even with $0 < \operatorname{sech} z \leq 2e^{-z}$, because $e^z + e^{-z} \geq e^z$; so $L_0^2 + L_2^2 = h^2r^2\operatorname{sech}^2(\lambda\tau) \leq h^2r^2(2e^{-\lambda\tau})^2$. At $t = \tau$, $S = 1$ and $T = 0$. Finally $\lambda^2 = u^2v^2 \cdot h^2r^2 = (J_1 - J_2)(J_0 - J_1)\,C$. □

For $u, v, r > 0$ and $0 \leq \tau \leq 1$ the body starts with its angular momentum within relative distance $2e^{-\lambda\tau}$ of the middle axis, in the sense $\sqrt{L_0^2 + L_2^2} \leq 2e^{-\lambda\tau}\sqrt{C}$, and at time τ it has turned fully across it, $L_1 = 0$. The rate λ is the one of Theorem 25.64 with $\ell_1 = \sqrt{C}$. The time to flip grows only like the logarithm of the initial distance: this is the tumbling of a tennis racket or a wing nut thrown about its middle axis [151].

When two moments are equal the body is a symmetric top, and the motion is a uniform turning.

Proposition 25.66 (The free symmetric top precesses)**.** *Let $A, B, c, j, \kappa \in \mathbb{R}_\delta$ and $\Omega = (\kappa - j)c$. The curve with moments (j, j, κ) and*

$$L(t) = \big(A\cos\Omega t + B\sin\Omega t,\ B\cos\Omega t - A\sin\Omega t,\ c\big)$$

is a motion of Euler's field on $[0,1]$. No omniscience principle and no choice principle is used.

Proof. With $J_0 = J_1 = j$ and $J_2 = \kappa$ the field reads $\dot L_0 = (\kappa - j)L_1L_2$, $\dot L_1 = (j - \kappa)L_2L_0$, $\dot L_2 = 0$; at $L_2 = c$ these are $\dot L_0 = \Omega L_1$ and $\dot L_1 = -\Omega L_0$, which the curve satisfies because the derivatives of cos and sin are $-$ sin and cos. What needs checking on $\mathbb{R}_\delta$ is a uniform derivative at every speed Ω, with one modulus on $[0,1]$. For $|y - x| \le 1$ the addition formulas of Section 21.2 give, with $h = y - x$,

$$\cos y - \cos x + h\sin x = \cos x(\cos h - 1) - \sin x(\sin h - h),$$

and the alternating series give $|\cos h - 1| \le h^2/2$ and $|\sin h - h| \le |h|^3/6$; with $|\cos x|, |\sin x| \le 1$ the right side is at most h^2. The same holds for $\sin y - \sin x - h\cos x = \sin x(\cos h - 1) + \cos x(\sin h - h)$. Hence $|g(\Omega t) - g(\Omega s) - \Omega g'(\Omega s)(t - s)| \le \Omega^2(t - s)^2$ whenever $|\Omega(t - s)| \le 1$, for $g = \cos$ and $g = \sin$, which is a uniform derivative with an explicit modulus; sums and constant multiples keep it. □

In the body frame the angular momentum of a free symmetric top turns about the symmetry axis at the constant rate $\Omega = (1/I_2 - 1/I_0)L_2$, on a circle of the sphere $C = A^2 + B^2 + c^2$. The reading is again an identification.

Last, the certificate of Section 25.9 applies to the rigid body as a whole.

Corollary 25.67 (The rigid body's conversion certificate)**.** *The polynomial theory with Euler's field, the candidate integrals E and C and no sign laws satisfies the five clauses of Theorem 25.59, and both candidates are first integrals. Clauses (1), (2), (4) and (5) are proved with no omniscience principle and no choice principle; clause (3), and with it the corollary as stated, is proved in classical mathematics.*

Proof. Theorem 25.59 applies to every polynomial theory, and Theorem 25.61 shows that both candidates are first integrals. Clause (4) has no laws to price. □

So a certified simulation of a tumbling body cannot drift off its level sets: its energy and its $|L|^2$ are exactly those of the initial state at every certified time, and by Theorem 25.62(d) its transverse part obeys the stability bound there too.

25.11 Pulsar timing in far time

A binary pulsar is timed, not watched. The pulsar is a clock riding the orbit; it ticks in its own proper time τ. A timer far away reads the time t_c of a distant observer, and from the arrival times fits three things: the rate at which the pericentre turns per unit of far time, the delay of the pulsar's clock against its mean rate inside each period, and how these two are related. Section 25.7 proved that every window orbit returns after a radial period T of proper time, that its pericentre turns by $\Phi - 2\pi_\delta$, and that the far clock advances by the same T_c in every

period (Theorem 25.47). Here we turn those statements into the three observables. Each is proved on $\mathbb{R}_\delta$ with no omniscience principle and no choice principle, for every ratio Λ; as in Definition 25.38, Λ is the squared angular momentum per unit mass, L^2, entering the field as a coefficient, not a cosmological term.

Fix a window orbit y of G_Λ. A *timing frame* of y consists of a radial period (θ, T, Φ); a time t_1 with $(x\cdot v)(t_1)=0$ such that $\big(2u(t)-(u_0+u_1)\big)^2 \le (u_0-u_1)^2$ for every t, where $u_0=u(0)$ and $u_1=u(t_1)$; the eccentricity e and the ratio $w=m/p$ read off the apsides, $e(u_0+u_1)=u_0-u_1$ and $2w=m(u_0+u_1)$; a point $E_\infty\ge 0$ with $E_\infty^2=1+2H_\Lambda(y(0))$; and a far clock t_c with $t_c(0)=0$, derivative $D(\tau)$ at every τ with $D(\tau)\big(1-2m\,u(\tau)\big)=E_\infty$, and a point T_c with $t_c(\tau+T)=t_c(\tau)+T_c$ for every τ. The *Einstein delay* is $\Delta(\tau)=t_c(\tau)-(T_c/T)\,\tau$: the far clock minus its mean rate. The *Kepler radius* a_K is read off the apsides by $2a_Ku_0u_1=u_0+u_1$.

Theorem 25.68 (The advance as a rate, the Einstein delay, and a mass-free identity)**.** *Let Λ be a ratio and y a window orbit. Then y has a timing frame, and:*

(a) The advance as a rate. *With $K=(6\pi_\delta)^3(2\pi_\delta)^2m^2$,*

$$K\Big(1+\big(\tfrac{39}{2}-\tfrac{21}{4}e^2\big)w\Big)\ \le\ T_c^2\big((\Phi-2\pi_\delta)(1-e^2)\big)^3\ \le\ K\Big(1+\big(\tfrac{39}{2}-\tfrac{21}{4}e^2\big)w+300\,w^2\Big).$$

(b) The Einstein delay. *With $k\ge 0$, $k^2=u_0u_1$ and $\lambda=am$, for every proper time τ*

$$\big|kL\big(Tt_c(\tau)-T_c\,\tau-2T(x\cdot v)(\tau)\big)\big|\ \le\ 20\,m\,|e|\,\lambda\,|T|,$$

that is, $|\Delta(\tau)-2(x\cdot v)(\tau)|\le 20|e|\lambda m/(|L|k)$.

(c) The timing identity. *At every proper time τ at which $u(\tau)a_K=1$,*

$$2|e|(1-e^2)(\Phi-2\pi_\delta)\,TT_c\,(1-3w)\ \le\ 3(2\pi_\delta)^2\big|Tt_c(\tau)-T_c\,\tau\big|\ \le\ 2|e|(1-e^2)(\Phi-2\pi_\delta)\,TT_c\,(1-2w),$$

that is, $3(2\pi_\delta)^2|\Delta| = 2|e|(1-e^2)(\Phi-2\pi_\delta)T_c(1+R)$ with $-3w\le R\le -2w$.

No omniscience principle and no choice principle is used.

Conditions and scope.

Base and premises

Motions of the polynomial field G_Λ in the weak-field window of Definition 25.38, for every ratio Λ, on $\mathbb{R}_\delta$. Free.

Conclusion

(a) is the binary-pulsar law $\dot\omega(1-e^2)=3\,n^{5/3}m^{2/3}\big(1+(\tfrac{13}{2}-\tfrac{7}{4}e^2)\tfrac{m}{p}\big)$, with $\dot\omega=(\Phi-2\pi_\delta)/T_c$ and $n=2\pi_\delta/T_c$, in the test-mass limit, cubed so that no fractional power appears, with its first relativistic correction exact and a proved remainder. (b) says that the clock part of the timing record is a polynomial in the orbit's state, $2\,x\cdot v=2r\,dr/d\tau$, up to a remainder proportional to e; it vanishes on circular orbits. (c) ties the clock record to the advance with no mass in it; its leading 3 is the post-Newtonian combination $2+2\gamma-\beta$.

Identification

A timelike Schwarzschild geodesic as a motion of G_Λ, E_∞ as its Killing energy and t_c as the static observer's time at infinity (Section 25.7); the pulsar's proper time as its clock.

Not claimed

Light-travel delays (Roemer and Shapiro), the companion's mass, and any timing data.

Proof. All three parts start from one construction on one orbit. As in Proposition 25.41, write the orbit in its clock angle ψ: $u = U_c + A_t \cos\psi$, $L\,d\tau = (h/U^2)\,d\psi$, with the radial period $T = L^{-1}\sigma(4q)$ and the far clock $t_c = E_\infty \sigma_c(\psi)/L$, $T_c = E_\infty \sigma_c(4q)/L$ (Theorem 25.47); $p = 1/U_c$, $e = A_t/U_c$ and $w = x = mU_c$.

(a) Theorem 25.42 gives $6\pi_\delta w + \frac{3\pi_\delta}{2}(18+e^2)w^2 \le \Phi - 2\pi_\delta \le \text{that} + 160\pi_\delta w^3$, and Theorem 25.47 gives $(2\pi_\delta)^2 m^2(1+3Z) \le T_c^2 Z^3 \le (2\pi_\delta)^2 m^2(1+3Z+20\lambda^2)$, with $Z = -2H$. The conserved quantities satisfy $\lambda = x(1-(3+e^2)x)$ and $Z\lambda = x^2(1-e^2)(1-4x)$. Multiplying each target bound by $8(Z\lambda)^3$ and cancelling with explicit inverses leaves two polynomial inequalities $x^2 q(x, e^2) \ge 0$ on the window $x \le \frac{1}{250}$, $e^2 \le \frac{9}{25}$, which exact Bernstein certificates close (degrees $(5,6)$ and $(8,6)$, with nonnegative rational coefficients over 3^{12}). The second-order margin of the lower inequality is at least 127.5, and the upper one stays at least 25.4 below the allowed 300.

(b) With $\tau = L^{-1}\sigma(\psi)$, $T t_c - T_c \tau = L^{-2} E_\infty W(\psi)$ for the Wronskian of the two clocks

$$W(s) = \sigma(4q)\big(\sigma_c - \sigma\big)(s) - \big(\sigma_c - \sigma\big)(4q)\,\sigma(s),$$

which has period $4q$ exactly; and $T t_c - T_c \tau - 2T(x \cdot \nu) = L^{-2} F(\psi)$ with F explicit. In the Wronskian the secular parts of the two clocks cancel identically, which a bound on the mean rate T_c/T alone could not achieve, since its remainder is not proportional to e. By periodicity, extended constructively by cotransitivity and induction, it suffices to bound kF for $|\psi| \le 3q$. There the clock integrals are exact through the eccentric anomaly, $\int d\psi/U = \text{ecc}/k$, $\int d\psi/U^2 = (U_c\,\text{ecc} - A_t \sin\text{ecc})/k^3$, with second-order expansions of $(1 - r\cos\psi)^{-1/2}$ and $1/(1-2mU)$ and explicit third-order remainders. The one transcendental function left, the true minus the eccentric anomaly, is bounded by its three-term series $2\beta\sin\psi - \beta^2 \sin 2\psi + \frac{2}{3}\beta^3 \sin 3\psi$, $\beta = e/(1+\sqrt{1-e^2})$, through an exact identity and the bounds $|T_3|, |T_4| \le 1$ on Chebyshev polynomials. The coefficients are then bounded by exact certificates; at the edge of the window the normalized bound is 18.5 against the allowed 19.9.

(c) At a crossing $u\,a_K = 1$ the eccentric anomaly is a quarter turn: $A_t \cos(\text{ecc}\,\psi) = 0$, so $|\sin(\psi - \text{ecc}\,\psi)| = |e|$ with no case split on e. The angle $A = |\psi - \text{ecc}\,\psi|$ lies between $|e|$ and $|e| + |e|^3/6 + |e|^5/8$, by the sine bounds and the anomaly bound of (b). Inserting this into the Wronskian of (b), with the bounds of (a) on $\Phi - 2\pi_\delta$ and on the two periods, leaves after clearing $(1-6x)^5(1-2x)^2$ two polynomial inequalities in x and $k/U_c \in [\frac{17}{20}, 1]$, closed by exact certificates; the upper margin at the worst corner $x = \frac{1}{250}$ is 0.13 w in R. □

25.12 The point mass from the lattice's own field equations

The relativistic coefficient $b = 3m$ of this chapter, that is $\beta = \gamma = 1$, rests on the cubic vertex of the gravitational action being Einstein–Hilbert's. Chapter 27 proves this for Regge's action on the Kuhn lattice in classical mathematics (Theorem 27.61). This section proves it on $\mathbb{R}_\delta$ at the Lorentzian background $\kappa = 1$, with no omniscience principle and no choice principle, and goes further: a long-wave solution of the lattice's own field equations whose jets at a point are those of a point mass has $\gamma = 1$, $\beta = 1$ and $\chi_2 = 1$, and at the next order the jet equations built from the quartic cell form give the third-order coefficient its Schwarzschild value.

Use the periodic $\kappa = 1$ Kuhn lattice of side N, the flat field ℓ, and δ's own Regge action S, flat Hessian $\mathscr{H}$ and cubic cell form C of Section 27.7. For integer wave vectors m_0, m_1 and $m_2 = -m_0 - m_1$ and symmetric H_0, H_1, H_2, let $C_N(H; m)$ be the cubic form of S on the three plane waves u_{m_i, H_i} (Theorem 27.30), and let

$$V_{\mathrm{EH}}(H; m_0, m_1) = \sqrt{3}\, W_3\big(g_1^{-1}; H_0, H_1, H_2; m_0, m_1, m_2\big),$$

with W_3 Einstein's cubic vertex of Definition 27.58 at the integer momenta m_i. By Theorem 27.60 this is the cubic vertex of the Einstein–Hilbert density $\sqrt{-\det g}\, R$ around g_1, since $\sqrt{|\det g_1|} = \sqrt{3}$. It is $\sqrt{3}$ times a polynomial with rational coefficients in the entries of the H_i and m_0, m_1.

Theorem 25.69 (Regge's cubic vertex is Einstein–Hilbert's on $\mathbb{R}_\delta$)**.** *On $\mathbb{R}_\delta$, free: for all integer m_0, m_1 and symmetric H_0, H_1, H_2 over $\mathbb{R}_\delta$ there is a constant B such that for every side N such that each of $2m_0$, $2m_1$ and $2m_2$ has a component that N does not divide,*

$$\Big| N^{-2} C_N(H; m) - \tfrac{\pi_\delta^2}{2}\, V_{\mathrm{EH}}(H; m_0, m_1) \Big| \le B\, N^{-2}.$$

The order-zero part of the cubic cell form vanishes on all constant strains, over every commutative ring.

Proof. On the torus the cubic form of three plane waves is N^4 times a cell symbol, a finite sum over the 24 simplices of the integer cubic table of $1152\sqrt{3}\, C$ with cosine phases, as in Theorem 27.30(c); the condition on N removes the aliased frequencies. Expanding the cosines to second order with the remainder bounds of δ's cosine, the order-zero part vanishes (an identity over every commutative ring), the order-one part vanishes by the central inversion, and the order-two part satisfies the integer polynomial identity $3S_2 + 64\, C_{\mathrm{EH}} = 0$, where S_2 is the second-order symbol and $C_{\mathrm{EH}} = 8\, W_3(3g_1^{-1}) = 648\, W_3(g_1^{-1})$ is eight times Einstein's cubic vertex on $3g_1^{-1} = 3 \cdot 1 - \mathbf{1}\mathbf{1}^{\mathsf{T}}$ (W_3 is homogeneous of degree four in the inverse metric), so that $V_{\mathrm{EH}} = \sqrt{3}\, C_{\mathrm{EH}}/648$; this identity is checked coefficient by coefficient. The remainder is bounded by the fourth-order cosine bound, which gives B. □

To pass from vertices to solutions we need what it means for a smooth field to solve the lattice's equations at long wavelength. A *δ-native C^3 jet field* near a point x_0 assigns to each point a symmetric matrix with first and second derivatives, all bounded by a constant K on a ball of radius r, each Lipschitz with constant K, and with second-order Taylor remainder

at most $K|y-z|^3$. Sample such a field h on the edges of the side-N lattice at edge midpoints, $s(e) = d_e^{\mathsf{T}} h(x_e) d_e$. For a symmetric test tensor W and a box of lattice cells around a rational point x_0, the *weak sums* of order one and two are $\mathscr{H}(s_1, t_{\mathrm{box}})$ and $\mathscr{H}(s_2, t_{\mathrm{box}}) + \frac{1}{2}C(s_1, s_1, t_{\mathrm{box}})$, with t_{box} the box test field of W. Let $x_0 = p/q$. A pair (h_1, h_2) is a *long-wave solution* at x_0 when there is a radius $\rho_0 > 0$, with the box of half-width ρ_0 around x_0 inside the unit torus and $8\rho_0$ at most the radius r of the jet bounds, such that for every box of lattice cells of half-width $\rho = 1/(j+1) \le \rho_0$ centred at x_0, every symmetric W and every rational tolerance $\varepsilon > 0$, both weak sums on that box are at most εN^2 on all lattices of large enough side N among the multiples of $q(j+1)$. The box is fixed before the tolerance: on each fixed box the weak sums are $o(N^2)$. The *jet equations* of order one and two at x_0 are the edge-local long-wave equations of $\mathscr{H}$ and of $\mathscr{H} + \frac{1}{2}C$ on the jets of h_1 and (h_1, h_2) at x_0. The *isotropic point-mass jets* with mass m are $h_1 = 2mU(\mathrm{tt} + \gamma S)$ and $h_2 = m^2U^2(\alpha\,\mathrm{tt} + \beta_2 S)$ and, at third order, $h_3 = m^3U^3(\zeta\,\mathrm{tt} + \xi S)$, with tt the time-time strain, S the spatial identity and U the potential, taken at a point x_0 of the lattice's chart. In the notation of (25.7), $\alpha = -2\beta_{\mathrm{PN}}$, $\beta_2 = \frac{3}{2}\chi_2$ and $\zeta = \frac{3}{2}\chi_3$.

Theorem 25.70 (Long-wave solutions with point-mass jets have $\gamma = \beta = \chi_2 = 1$)**.** *On $\mathbb{R}_\delta$, free, at the background $\kappa = 1$:*

(a) at every field within $1/40000000$ of the flat field, cell by cell, and along every direction with entries in $[-1, 1]$, δ's Regge action has the first variation $\sum_t A_t'(\theta_t - \theta_t^{\mathrm{flat}})$ with a uniform quadratic modulus, and along $\ell + xw$ this first variation is $x\,\mathscr{H}(w, \cdot) + \frac{1}{2}x^2 C(w, w, \cdot)$ up to $O(N^4|x|^3)$; the weak sums are the coefficients of this first variation along the box test field;

(b) if (h_1, h_2) is a long-wave solution at a rational point x_0 of two δ-native C^3 jet fields, then the jets at x_0 satisfy the jet equations of order one and two, for every symmetric test tensor;

(c) if moreover those jets are the isotropic point-mass jets with $m \neq 0$, then $\gamma = 1$, $\alpha = -2$ and $\beta_2 = \frac{3}{2}$: the post-Newtonian γ and β, and the second-order spatial coefficient χ_2, are 1.

Conditions and scope.

Base and premises

δ's Regge action on the $\kappa = 1$ Kuhn lattice on $\mathbb{R}_\delta$; the hypothesis of (b), a long-wave solution in weak form with box test profiles. Free; the mass need only be not equal to zero, and no step inverts it, so Markov's principle is not used.

Conclusion

$\beta = \gamma = 1$, and with them $b = 3m$ and the first-order advance $6\pi_\delta m/p$, follow from the lattice's own equations, not from a clause about the action; so does $\chi_2 = 1$.

Not claimed

That an actual lattice solution of a mass is such a C^3 family solving the weak equations: the existence and regularity of the long-wave profile, open in classical mathematics too.

Identification

Regge's action on the $\kappa = 1$ lattice as the gravitational action, $\kappa = 1$ itself (Section 27.8), and the jets as those of a static point mass.

Proof. (a) Schläfli's identity (Theorem 27.32(b)) gives the first variation in Schläfli form at every nondegenerate field, and the moduli of Theorem 27.32(a) make the quadratic remainder uniform; expanding the angle differences to second order around the flat field gives $\mathscr{H}$ and $\frac{1}{2}C$, the cubic term by polarizing the per-hinge cubic identity. Evaluating at the sampled field $\ell + ms_1 + m^2 s_2$ along the box test field identifies the weak sums. (b) The argument is integral-free. Expand each edge's equation for a sampled field: the stencils of $\mathscr{H}$ and of C at $\kappa = 1$ have vanishing zero and first moments, integer identities checked exactly, so the equation at an edge is N^{-2} times an edge-local jet operator plus $O(N^{-3})$; the sum of these operators against the strains of W is exactly the jet operator of the jet equations. The box test sums reorganize, on the torus, by test edges; the N^4 interior edges carry the jet operator at nearby points, the boundary only $O(N)$ terms, and the non-metric part of an actual solution enters only on the boundary. Shrinking the boxes and using the Lipschitz bounds of the jets squeezes the jet equations at x_0 out of the $o(N^2)$ hypothesis. (c) The first-order jet equation with the test tensor $E_{01} + E_{10}$ forces $\gamma = 1$; the second-order equations with E_{00} and $E_{01} + E_{10}$ are two linear equations in α and β_2 with a nonzero determinant once $m \neq 0$, solved by $\alpha = -2$, $\beta_2 = \frac{3}{2}$; the Schwarzschild jets solve all of them. □

At third order the lattice's equation needs the quartic vertex. Its table is the quartic cell form of the $\kappa = 1$ lattice, which is $\sqrt{3}$ times the continued table $-T_4(1)$ of Chapter 27 (Proposition 27.63); that this cell form is the fourth derivative of δ's own Regge action on $\mathbb{R}_\delta$ rests on two named premises about its third-order rates of areas and angles, which are open. The *third-order jet equation* at x_0 with test tensor W is the edge-local equation $\mathscr{H}(s_3) + C(s_1, s_2) + \frac{1}{6}D(s_1, s_1, s_1) = 0$ on the jets, D the quartic cell form.

Theorem 25.71 (The third order at the jet level). *On $\mathbb{R}_\delta$, free, at the background $\kappa = 1$, with the isotropic point-mass jets and $m \neq 0$:*

(a) with the lower-order jets at $\gamma = 1$, $\alpha = -2$ and $\beta_2 = \frac{3}{2}$, the values Theorem 25.70(c) forces, the third-order jet equations with the test tensors E_{00} and $E_{01} + E_{10}$ hold exactly when $\zeta = \frac{3}{2}$ and $\xi = \frac{1}{2}$; so $\chi_3 = \frac{2}{3}\zeta = 1$, and the spatial U^3 coefficient is $\frac{1}{2}$, as in isotropic Schwarzschild, $(1 + U/2)^4$; the Schwarzschild jets solve both equations;

(b) every solution of the first-, second- and third-order jet equations with $m \neq 0$ has $4\varphi_2 = 54 + 3e^2$, where

$$4\varphi_2 = 40 + 48\gamma - 40\beta_{\mathrm{PN}} - 2\beta_{\mathrm{PN}}^2 - 24\beta_{\mathrm{PN}}\gamma + 8\gamma^2 + 15\chi_2 + 9\chi_3 + 3\chi_2 e^2$$

is the second-order perihelion coefficient in the coefficients read off the jets, $\beta_{\mathrm{PN}} = -\alpha/2$, $\chi_2 = \frac{2}{3}\beta_2$, $\chi_3 = \frac{2}{3}\zeta$; at $\beta_{\mathrm{PN}} = \gamma = 1$ it is four times the coefficient $\frac{15}{2} + \frac{15}{4}\chi_2 + \frac{9}{4}\chi_3 + \frac{3}{4}\chi_2 e^2$ of Proposition 25.49. So $\varphi_2 = \frac{27}{2} + \frac{3}{4}e^2$, the coefficient of Theorem 25.42: $\Phi - 2\pi_\delta = 2\pi_\delta(3w + \varphi_2 w^2)$ up to the third order.

Conditions and scope.

Base and premises

The jet equations of the $\kappa = 1$ lattice at one point, with the quartic cell form; free.

Conclusion

At the jet level the jet equations built from the quartic cell form give the Schwarzschild point mass through third order, and with it the whole second-order perihelion advance of Theorem 25.42, $27\pi_\delta(m/p)^2 + \frac{3}{2}\pi_\delta e^2(m/p)^2$.

Not claimed

The bridge of Theorem 25.70(b) at third order; and that the quartic cell form is the fourth derivative of δ's action, which rests on the two named premises on third-order rates.

Proof. On tensor-product jets the third-order operator has closed forms: the quadratic part on the U^3 jet, the mixed cubic part on the first- and second-order jets, and the quartic part with the integer quartic table of the cell form. With $\gamma = 1$, $\alpha = -2$ and $\beta_2 = \frac{3}{2}$ substituted, in units of $\sqrt{3}/3456$ the two components are $432\,m^3(6\zeta - 4\xi - 7)$ and $864\,m^3(-12\zeta + 14\xi + 11)$. With $m \neq 0$, so $m^3 \neq 0$, the two linear equations $6\zeta - 4\xi = 7$ and $12\zeta - 14\xi = 11$ have the unique solution $\zeta = \frac{3}{2}$, $\xi = \frac{1}{2}$. For (b), the first- and second-order jet equations alone give $\gamma = 1$, $\alpha = -2$ and $\beta_2 = \frac{3}{2}$, as in the proof of Theorem 25.70(c), and part (a) gives $\zeta = \frac{3}{2}$; substituting $\beta_{\mathrm{PN}} = \gamma = \chi_2 = \chi_3 = 1$ gives $30 + 15 + 9 + 3e^2$. □

So the chain from the lattice to the planet's orbit has one link left open at first and second order, the existence of the long-wave profile, and at third order two more, the bridge and the third-order rates. Everything else is a theorem on $\mathbb{R}_\delta$ at no price.

Questions to work through

1. (Theorem 25.9.) Take $m = 1$, $u = 1$, $x = (1, 0, 0)$ and $v = (0, \frac{1}{2}, \frac{1}{2})$. Compute E, L and A, check $|A|^2 = 2E|L|^2 + m^2$, and show that the body is at its aphelion. Find the perihelion distance, the eccentricity $|A|/m$, the semi-major axis and the period given by Kepler's third law.

2. (Theorem 25.54.) For the field $\dot{x}_0 = x_1$, $\dot{x}_1 = -x_0 + x_0^2$, show by the grid test that $F = x_0^2 + x_1^2$ is not a first integral, and name a grid point that witnesses it. Then show that $G = \frac{1}{2}x_1^2 + \frac{1}{2}x_0^2 - \frac{1}{3}x_0^3$ is one.

3. (Theorem 25.56.) For $\dot{x} = x^2$ from $x_0 = 3$, compute X, M, s and k. Up to which time does the theorem certify the solution, and how does that compare with the time $\frac{1}{3}$ at which $3/(1 - 3t)$ blows up? Why can no region certificate reach time $\frac{1}{3}$?

4. (Theorem 25.42.) Take as an identification that Mercury's orbit is a window orbit with $m/p = 2.66 \times 10^{-8}$ and $e = 0.206$, completing 415.2 revolutions per century. Compute the advance of its perihelion per century from the first-order term, in seconds of arc, and bound the contributions of the second-order term and of the remainder.

5. (Theorem 25.62.) For $J = (1, 3, 4)$ and $L(0) = (2, \frac{1}{5}, \frac{1}{5})$, give the bound the theorem places on $L_1^2 + L_2^2$ along the motion, and compare it with the exact range allowed by the energy–Casimir form. How small must $L_1^2 + L_2^2$ be at the start to keep it at most 10^{-2}? What goes wrong for $J = (2, 1, 3)$?

6. (Theorem 25.64 and Proposition 25.65.) For $J = (3, 2, 1)$ and $|L|^2 = 200$, find u, v, h, r and λ of the separatrix with $\tau = 1$. How close to the middle axis does it start, and where is it at time 1? Compare λ with the rate of Theorem 25.64(b).

Chapter 26

Fields, heat and the second law

A voxel of Recognition Science keeps three axis accounts. At each tick of its history it either holds or posts one unit to one side of one account, and over many ticks its imbalance spreads out and settles toward a fixed law. Does the voxel's relative cost ever rise under Recognition Science's own tick, and how fast does it fall?

Relative cost is a definite number. For a law p on finitely many configurations with every $p_i > 0$, and a distribution q, the *relative entropy* is

$$D(q \,\|\, p) = \sum_{q_i > 0} q_i \log \frac{q_i}{p_i}.$$

It is 0 at $q = p$ and positive at every other distribution. When p is the Gibbs law of costs c_i at temperature T, that is $p_i = e^{-c_i/T}/Z$, the free energy $F(q) = \sum_i q_i c_i - TS(q)$, with $S(q) = -\sum_i q_i \log q_i$, exceeds its least value by exactly $TD(q \,\|\, p)$ (Section 26.1). Recognition Science takes the costs to be values of the comparison cost J of Chapter 21 and reads this excess as the relative J-cost. A step of a Markov chain with transition matrix K moves q to qK.

Here is a small case. Take three configurations $0, 1, 2$ and the chain

$$K = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} :$$

from 0 it holds or moves to 1 with equal chances, from 1 it moves to 2, and from 2 back to 0. The law $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ is stationary, since $pK = (\frac{1}{4} + \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) = p$. Start from $q_0 = (\frac{2}{3}, 0, \frac{1}{3})$ and step:

tick t	q_t	$D(q_t \,\|\, p)$	value
0	$(\frac{2}{3}, 0, \frac{1}{3})$	$\log \frac{4}{3}$	0.2877
1	$(\frac{2}{3}, \frac{1}{3}, 0)$	$\log \frac{4}{3}$	0.2877
2	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$\frac{1}{3} \log \frac{2}{3} + \frac{2}{3} \log \frac{4}{3}$	0.0566
3	$(\frac{1}{2}, \frac{1}{6}, \frac{1}{3})$	$\frac{1}{6} \log \frac{2}{3} + \frac{1}{3} \log \frac{4}{3}$	0.0283
4	$(\frac{7}{12}, \frac{1}{4}, \frac{1}{6})$		0.0223

The relative entropy never rises. Yet the first tick produces nothing, although q_0 is not the stationary law and every entry of K^4 is positive. The numbers show why. Configuration 0 is entered from 0 and from 2, and q_0 puts the same multiple $\frac{4}{3}$ of p on both, so the step merges two configurations that it cannot tell apart, and loses nothing. At the second tick configuration 0 is entered from 0, with ratio $q_1(0)/p(0) = \frac{4}{3}$, and from 2, with ratio 0; the merge destroys information and the relative entropy falls by a factor of five.

This chapter answers the opening question by proving what the example suggests. Relative entropy never rises under a stochastic step that keeps p (Theorem 26.10). One step loses nothing exactly when it merges no two configurations with different ratios q_i/p_i (Theorem 26.11), so a strict fall at every tick fails even when some power of K is positive, and holds over every window N with K^N positive (Theorem 26.12). Recognition Science's tick exists and is unique (Theorem 26.21), and, walled into any finite set of states that is closed under inward posts, it lowers relative entropy strictly at every tick (Theorem 26.23). Its numbers on all of $\mathbb{Z}^3$ rest on five numerical properties of its entries, (T1)–(T5) of Section 26.4, which we state and do not prove. Given them, from a start a the tick keeps $D_n(a) \le \log\bigl(1 + 0.9922^{2n}(\varphi^{9+|a|_1} - 1)\bigr)$ after n ticks (Corollary 26.29); its exact pace per tick lies between 0.7726 and 0.98447 (Corollary 26.37); and a spine of k independent voxels settles sharply at $\log k/(-\log r)$ ticks, where r is that pace (Theorem 26.42). Two voxels that each move by Recognition Science's tick and are coupled by the ledger's double entry stay at least 6.247 nats from the law of two independent voxels at equilibrium, at every tick, and, given (T1) and (T2), settle to their own equilibrium at one voxel's pace (Proposition 26.43). The information between them approaches its limit at the square root of that pace, provided two statements about one voxel's walk hold that we name but do not prove (Proposition 26.44).

The chapter has three further subjects. Section 26.1 builds entropy, the Gibbs principle and the H-theorem on $\mathbb{R}_\delta$ and prices the decisions about them. Section 26.7 puts Maxwell's and Schrödinger's equations on a lattice with δ's numbers: an exact Maxwell flow, standing light with its dispersion error, a light cone for lattice propagators and a continuum limit with error of order h^2. Section 26.8 prices physical laws: sign and classification laws, linear laws of two quantities, laws on a half-line, and the lattice premises Recognition Science uses for an inverse-square field.

Table 26.1 records the scope of each part. A result is *free* when its proof uses no omniscience principle and no choice principle, and *proved in classical mathematics* when it is a theorem about the classical real numbers (Section 21.1); we say which before each statement. Every theorem about Recognition Science's tick is proved for every kernel of a stated class. The tick enters through a short list of its properties, which we state where they are used: its existence and uniqueness, proved here, and the five numerical inputs (T1)–(T5), which are not. Every number we give for the tick itself is conditional on the inputs it names.

26.1 Entropy and the Gibbs principle on the line

Thermodynamics needs $-x\log x$ at $x \ge 0$. On $\mathbb{R}_\delta$ the case definition "$-x\log x$ if $x > 0$, and 0 if $x = 0$" is not a map: deciding $x = 0 \vee x > 0$ for every $x \ge 0$ is LPO. Applied to $|x|$ it gives $x > 0 \vee x = 0 \vee x < 0$ for every point, which is LPO by Corollary 12.18, and LPO gives it back; even the weaker $x = 0 \vee x \ne 0$ costs WLPO. We build the function instead, then use it for the Gibbs principle and the H-theorem. The logarithm is that of Definition 21.4. From Section 21.2 we use that every positive point y is $e^{\log y}$, and three order facts of the exponential, free for all points a, b, u:

$$1 + u \le e^u, \qquad e^b(a - b) \le e^a - e^b, \qquad a < b \text{ exactly when } e^a < e^b. \tag{26.1}$$

The second is the *tangent bound*. A function F on $\mathbb{R}_\delta$ is *continuous at* x when for every k there is a j with $|F(y) - F(x)| \le 1/(k+1)$ whenever $|y - x| \le 1/(j+1)$, and *continuous* when it is

Table 26.1. What each part of the chapter proves, and in which setting.

Section	Content	Setting
26.1 Heat on the line	entropy function, Gibbs principle, H-theorem, the prices of the Gibbs decisions	free on $\mathbb{R}_\delta$ (prices exact with no choice); one classical reading
26.2 Markov steps	the second law, its equality case, windows, heat baths	classical; the head theorem's form on $\mathbb{R}_\delta$ (Theorem 26.7) free
26.3 Recognition Science's dynamics	zero production, one-step strictness, Doeblin's rate	classical; Doeblin's rate and lazy connected strictness also free on $\mathbb{R}_\delta$
26.4 The tick	existence and uniqueness, walls, Doeblin windows, the rate from a point start; numbers for the tick from (T1)–(T5)	classical
26.5 Gap and start	Metropolis tick with a rate	free on $\mathbb{R}_\delta$
	a gap for every ball; start dependence	classical
26.6 The spine	pace, additivity, couplings, cutoff, the pair law forced by double entry; the pair's sharp rate, reduced to two named statements; numbers for the tick from (T1)–(T5)	classical
26.7 Lattice fields	Maxwell flow, light, Schrödinger	free on $\mathbb{R}_\delta$
26.8 Laws and premises	sign, plane and range laws	prices exact with no choice
	inverse-square lattice premises	classical; integer prices with no choice

continuous at every point.

Definition 26.1 (Entropy function and relative entropy on the line)**.** An *entropy function* is a continuous map η from $\mathbb{R}_\delta$ to $\mathbb{R}_\delta$ with $\eta(x) = -xy$ whenever $x > 0$ and $e^y = x$, and $\eta(0) = 0$. For points $q_1, \dots, q_n$ put $S(q) = \sum_i \eta(q_i)$. For *levels* $\ell_1, \dots, \ell_n$, standing for the weights $p_i = e^{\ell_i}$, the *relative entropy* on the line is

$$D_\eta(q \,\|\, \ell) = \sum_i \bigl(-\eta(q_i) - q_i \ell_i\bigr).$$

For $q_i > 0$ the ith term is $q_i \log q_i - q_i \log p_i$, the classical term. At $q_i = 0$ it is 0, with no decision about which case holds.

Lemma 26.2 (Closed conditions pass to the boundary)**.** *Let F be continuous at a point $q \geq 0$, and let $F(x) \geq 0$ for every $x > 0$. Then $F(q) \geq 0$. The proof is free.*

Proof. On $\mathbb{R}_\delta$, $F(q) \geq 0$ means that $F(q) < 0$ is impossible. Suppose $F(q) < 0$, and take k with $F(q) < -1/(k+1)$. Continuity gives j with $|F(y) - F(q)| \leq 1/(2k+2)$ for $|y - q| \leq 1/(j+1)$. The point $y = q + 1/(j+1)$ is positive, and $F(y) \leq F(q) + 1/(2k+2) < 0$, which contradicts $F(y) \geq 0$. □

Theorem 26.3 (The entropy function exists and is unique)**.** *There is an entropy function. Any two entropy functions agree at every $x \geq 0$. Both statements are free.*

Proof. Construction. Write $x^+ = \max(x, 0)$ and, for each n,

$$\eta_n(x) = -x^+ \log\Bigl(x^+ + \frac{1}{n+1}\Bigr).$$

The logarithm's argument is at least $1/(n+1)$, so each η_n is a map with no case split. For $0 < a \leq b$, the tangent bound (26.1) at the points $\log b$ and $\log a$ gives $a(\log b - \log a) \leq b - a$. With $y = x^+$, $m \geq n$, $a = y + 1/(m+1)$ and $b = y + 1/(n+1)$,

$$0 \leq \eta_m(x) - \eta_n(x) = y(\log b - \log a) \leq \frac{y}{a}(b - a) \leq \frac{1}{n+1}.$$

So the η_n converge uniformly with the rate $1/(n+1)$. The diagonal limit of Theorem 12.25, applied on a regular sequence naming x, defines $\eta(x)$, and limits are unique, so η is a map on the points. Each η_n is continuous: $x \mapsto x^+$ moves by at most $|x - x'|$, and on $[1/(n+1), \infty)$ the logarithm moves by at most $(n+1)$ times the change of its argument. A uniform limit with a rate of continuous functions is continuous, so η is. At 0 every η_n vanishes. For $x > 0$, $|\log(x + 1/(n+1)) - \log x| \leq 1/((n+1)x)$ by the same inequality, so $\eta_n(x) \to -x \log x$.

Uniqueness. Let η and η' be entropy functions. Every $x > 0$ is $e^{\log x}$, so $\eta(x) = -x \log x = \eta'(x)$ for $x > 0$. The function $F = -|\eta - \eta'|$ is continuous and vanishes at every positive point, so Lemma 26.2 gives $F(x) \geq 0$, that is $\eta(x) = \eta'(x)$, at every $x \geq 0$. □

From now on η is the entropy function, and we write $x \log x$ for $-\eta(x)$ at $x \geq 0$. The next lemma is the single inequality behind the Gibbs principle, the H-theorem and the native second laws later in the chapter.

Lemma 26.4 (The Gibbs term). *For points ℓ and $q \geq 0$ put $g(\ell, q) = e^{\ell} - q - \eta(q) - q\ell$. Then*

1. *$g(\ell, q) \geq 0$;*
2. *$g(\ell, q) > 0$ exactly when $q \# e^{\ell}$;*
3. *$g(\ell, q) = 0$ exactly when $q = e^{\ell}$.*

The proofs are free.

Proof. 1. For $q = e^m > 0$ we have $\eta(q) = -me^m$, so $g(\ell, q) = e^{\ell} - e^m - e^m(\ell - m) \geq 0$ by the tangent bound (26.1). Every $q > 0$ has this form, and $g(\ell, \cdot)$ is continuous, so Lemma 26.2 gives $g(\ell, q) \geq 0$ for every $q \geq 0$.

2. Suppose $q \# e^{\ell}$. Item 1 at a point m in place of ℓ says $-\eta(a) \geq a + am - e^m$ for every $a \geq 0$. Let $c = (q + e^{\ell})/2 > 0$ and $m = \log c$; adding the inequality at $a = q$ and at $a = e^{\ell}$ gives $-\eta(q) - \eta(e^{\ell}) \geq 2cm = -2\eta(c)$. The terms of g that are linear in q then give $g(\ell, c) \leq \frac{1}{2}\big(g(\ell, q) + g(\ell, e^{\ell})\big) = \frac{1}{2}g(\ell, q)$, since $g(\ell, e^{\ell}) = 0$. Now $c \# e^{\ell}$, so $u = \ell - m \# 0$ by the third fact in (26.1), and

$$g(\ell, c) = e^m\big(e^u - 1 - u\big) = e^m\Big(2\big(e^{u/2} - 1 - \tfrac{u}{2}\big) + \big(e^{u/2} - 1\big)^2\Big) \geq e^m\big(e^{u/2} - 1\big)^2 > 0,$$

by the first fact in (26.1) at $u/2$, and the square is positive because $e^{u/2} \# 1$. So $g(\ell, q) > 0$. Conversely let $g(\ell, q) > 0$, and take k with $g(\ell, q) > 1/(k+1)$. Continuity of $g(\ell, \cdot)$ at e^{ℓ}, where it vanishes, gives j with $g(\ell, y) \leq 1/(k+1)$ whenever $|y - e^{\ell}| \leq 1/(j+1)$. Either $|q - e^{\ell}| > 1/(j+2)$ or $|q - e^{\ell}| < 1/(j+1)$; the second would give $g(\ell, q) \leq 1/(k+1)$, so the first holds and $q \# e^{\ell}$.

3. If $g(\ell, q) = 0$, item 2 excludes $q \# e^{\ell}$, and on $\mathbb{R}_{\delta}$ two points that are not apart are equal. Conversely $g(\ell, e^{\ell}) = e^{\ell} - e^{\ell} + \ell e^{\ell} - \ell e^{\ell} = 0$. □

Theorem 26.5 (The Gibbs variational principle on the line). *Let $E_1, \dots, E_n$ and β be points, let y be a point with $e^y = Z = \sum_i e^{-\beta E_i}$, and let $q_1, \dots, q_n \geq 0$ with $\sum_i q_i = 1$. Put $F(q) = \sum_i q_i \beta E_i - S(q)$. Then*

1. *$F(q) \geq -y$;*
2. *$F(q) = -y$ exactly when $q_i Z = e^{-\beta E_i}$ for every i, that is, when q is the Gibbs state;*
3. *$F(q) > -y$ exactly when $q_i Z \# e^{-\beta E_i}$ for some i.*

The proofs are free.

Proof. Put $\ell_i = -\beta E_i - y$, so that $\sum_i e^{\ell_i} = Ze^{-y} = 1$. Then

$$\sum_i g(\ell_i, q_i) = \sum_i e^{\ell_i} - \sum_i q_i - S(q) + \sum_i q_i(\beta E_i + y) = F(q) + y.$$

Every term on the left is nonnegative (Lemma 26.4). A finite sum of nonnegative points is 0 exactly when each term is, and is positive exactly when some term is, since a positive finite sum has a positive term. Item 3 of the lemma turns "each term is 0" into $q_i = e^{\ell_i}$, which is $q_i Z = e^{-\beta E_i}$, and item 2 turns "some term is positive" into an apartness. □

Conditions and scope.

Base and premises
Finitely many levels E_i, an inverse temperature β, a logarithm y of the partition function, a distribution q; all points of $\mathbb{R}_\delta$.

Conclusion
The free energy is bounded below by $-\log Z$, the Gibbs state is its only minimiser, and "strictly above the minimum" is the same statement as "apart from the Gibbs state", with no decision about q.

Not claimed
A temperature scale or a physical calibration of β.

Gibbs stated the principle in his treatise on statistical mechanics [181]. What the line adds is the third item: it identifies an observable deviation from equilibrium with an excess of free energy, and the next theorem shows that the classical alternatives about Gibbs states each carry an exact price.

Theorem 26.6 (The prices of the Gibbs decisions)**.** *Quantify over all data of Theorem 26.5 (all n, levels, β, y and distributions q). Then, with no choice principle:*

1. *"q is the Gibbs state or it is not" holds for all data exactly when WLPO holds;*
2. *"if q is not the Gibbs state, it is apart from it" holds for all data exactly when $\mathrm{MP}_{\mathrm{bin}}$ holds;*
3. *"q is the Gibbs state, or its free energy is strictly above $-y$" holds for all data exactly when LPO holds;*
4. *"the free energy is strictly above $-y$" and "q is apart from the Gibbs state" are equivalent for all data, at no price.*

Proof. Item 4 is Theorem 26.5, item 3.

Lower bounds. Let a be any point and $A \geq |a|$ a natural number. Take $n = 2$, $E_1 = E_2 = 0$, $\beta = 0$, $Z = 2$, $y = \log 2$, and $q = (\frac{1}{2} + c, \frac{1}{2} - c)$ with $c = a/(2A + 1)$, so $q \geq 0$ and $\sum q = 1$. Then q is the Gibbs state exactly when $2q_1 = 1$ and $2q_2 = 1$, that is $a = 0$, and q is apart from it exactly when $a \# 0$. So item 1 gives $a = 0 \vee a \neq 0$ for every a, which is WLPO by Corollary 12.18; item 2 gives $a \neq 0 \to a \# 0$, which is $\mathrm{MP}_{\mathrm{bin}}$; and item 3, read through item 4, gives $a = 0 \vee a \# 0$, which is $a = 0 \vee a > 0 \vee a < 0$, LPO.

Upper bounds. With WLPO, each of the n equations $q_i Z = e^{-\beta E_i}$ is decided by a zero test, and a finite conjunction of decided statements is decided. With $\mathrm{MP}_{\mathrm{bin}}$, let q not be the Gibbs state and put $s = \sum_i |q_i Z - e^{-\beta E_i}|$. Then $s \neq 0$, so $s > 0$ by Corollary 12.18, and a positive finite sum has a positive term: some coordinate is apart. With LPO, each difference $q_i Z - e^{-\beta E_i}$ is 0 or apart from 0; if all are 0, q is the Gibbs state, and otherwise item 4 gives the strict excess. □

The second law in its oldest form, Boltzmann's H-theorem [176], says that a relaxation never raises the free energy. On the line it is a statement about a stochastic matrix and the relative entropy of Definition 26.1.

Theorem 26.7 (The H-theorem on the line). *Let K be an $n \times n$ matrix of points with $K_{ij} \geq 0$ and $\sum_j K_{ij} = 1$, and let $\ell_1, \dots, \ell_n$ be levels whose weights are stationary: $\sum_i e^{\ell_i} K_{ij} = e^{\ell_j}$ for every j. Then for every q with all $q_i \geq 0$,*

$$D_\eta(qK \,\|\, \ell) \leq D_\eta(q \,\|\, \ell), \qquad (qK)_j = \sum_i q_i K_{ij}.$$

In particular, if the Boltzmann weights $e^{-\beta E_i}$ are stationary, the free energy F of Theorem 26.5 never increases under $q \mapsto qK$, and for a distribution q it stays at least $-\log Z$. The proofs are free.

Proof. Column bound. Fix a column j, put $r = (qK)_j$, and let μ be any point. Lemma 26.4, item 1, at $\ell = \mu + \ell_i$ reads $-\eta(q_i) - q_i \ell_i \geq q_i \mu + q_i - e^{\mu} e^{\ell_i}$. Multiply by $K_{ij} \geq 0$, sum over i, and use stationarity:

$$\sum_i K_{ij}\big(-\eta(q_i) - q_i \ell_i\big) \geq \mu r + r - e^{\mu} e^{\ell_j}. \tag{26.2}$$

If $r > 0$, the choice $\mu = \log r - \ell_j$ makes the right side $r \log r - r\ell_j = -\eta(r) - r\ell_j$, the jth term of $D_\eta(qK \,\|\, \ell)$. Summing over j and using $\sum_j K_{ij} = 1$ on the left gives the theorem when every column of qK is positive.

Columns of unknown sign. For a point $\varepsilon > 0$ the vector $q + \varepsilon e^{\ell}$ is mapped to $qK + \varepsilon e^{\ell}$, whose columns are positive. So $G(\varepsilon) = D_\eta(q + \varepsilon e^{\ell} \,\|\, \ell) - D_\eta(qK + \varepsilon e^{\ell} \,\|\, \ell)$ is nonnegative at every $\varepsilon > 0$. It is continuous, because η is, and Lemma 26.2 gives $G(0) \geq 0$. No decision whether a column vanishes is made.

Free energy. With $\ell_i = -\beta E_i$ the relative entropy is $\sum_i (-\eta(q_i) + q_i \beta E_i) = F(q)$, and $F \geq -\log Z$ by Theorem 26.5. □

The weights in the theorem need not sum to 1; only stationarity is used. The theorem reads back into the classical setting through the unique reading of $\mathbb{R}_\delta$ in the classical real numbers (Section 21.2).

Proposition 26.8 (The classical reading is Recognition Science's thermodynamics). *The following is proved in classical mathematics. Let ρ be the reading of $\mathbb{R}_\delta$. Then $\rho(\eta(x)) = -\rho(x) \log \rho(x)$ for $x \geq 0$ (with $0 \log 0 = 0$), and $\rho(S(q))$ is the recognition entropy $-\sum_i \rho(q_i) \log \rho(q_i)$ of Recognition Science. Let a recognition system have temperature $T_R > 0$, let X_i be positive ratios, and suppose the calibration $\rho(\beta E_i) = J(X_i)/T_R$. Then q is the Gibbs state of Theorem 26.5 exactly when $\rho(q_i) = e^{-J(X_i)/T_R}/Z$ for every i, and the theorem reads as Recognition Science's free-energy principle: $F_R(q) = \sum_i \rho(q_i) J(X_i) - T_R\, S(\rho(q)) \geq -T_R \log Z$, with equality exactly at the Gibbs measure.*

Proof. The reading preserves sums, products, limits with a modulus and the logarithm (Section 21.2). At $x > 0$ it sends $-x \log x$ to $-\rho(x) \log \rho(x)$; at $x = 0$ both sides are 0; and both sides are continuous in $x \geq 0$, so they agree for all $x \geq 0$. The Gibbs condition $q_i Z = e^{-\beta E_i}$ reads as $\rho(q_i) Z = e^{-J(X_i)/T_R}$, and $T_R\, \rho(F(q)) = F_R(q)$. □

The calibration $\rho(\beta E_i) = J(X_i)/T_R$ is a stated hypothesis. The proposition does not identify a temperature, and nothing in this section uses an observation.

26.2 The second law for a Markov step

This section and the next work in classical mathematics: distributions, stochastic matrices and relative entropies are classical real objects. The head theorem below has a form on $\mathbb{R}_\delta$, Theorem 26.7, which is free; reading that form back to the classical real numbers through the reading ρ of Proposition 26.8 is a classical step.

Definition 26.9 (Markov steps and J-descent operators)**.** Let Ω be a finite set. A *distribution* is a vector $q \geq 0$ on Ω with $\sum_i q_i = 1$. A *stochastic matrix* K has $K_{ij} \geq 0$ and $\sum_j K_{ij} = 1$; it moves q to qK, $(qK)_j = \sum_i q_i K_{ij}$. A vector p with all $p_i > 0$ is *stationary* when $pK = p$. For $q \geq 0$ the relative entropy is $D(q \,\|\, p) = \sum_{q_i>0} q_i \log(q_i/p_i)$; with $f(x) = x\log x$ and $f(0) = 0$, it is $\sum_i p_i f(q_i/p_i)$. A *$\mathcal{J}$-descent operator* with equilibrium p, in Recognition Science's term, is a map R on distributions with $R(p) = p$ and $D(R(q) \,\|\, p) \leq D(q \,\|\, p)$ for every q.

The function f is convex on $[0,\infty)$, and strictly convex: equality in Jensen's inequality $f(\sum_i w_i x_i) \leq \sum_i w_i f(x_i)$, for weights $w_i > 0$ summing to 1, holds only when all x_i are equal.

Theorem 26.10 (Relative entropy never rises under a stochastic step)**.** *Let K be a stochastic matrix on a finite set and p a stationary vector with every $p_i > 0$. Then for every vector $q \geq 0$,*

$$D(qK \,\|\, p) \leq D(q \,\|\, p).$$

So $q \mapsto qK$ is a $\mathcal{J}$-descent operator with equilibrium p. If p is the Gibbs law $p_i = e^{-c_i/T}/Z$ of costs c_i at temperature $T > 0$, the free energy $F(q) = \sum_i q_i c_i - TS(q)$ of a distribution q satisfies $F(q) - F(p) = TD(q \,\|\, p)$ and never increases under the step. The theorem is proved in classical mathematics, and its matrix form on $\mathbb{R}_\delta$ is free (Theorem 26.7).

Proof. Fix a column j and put $w_{ij} = p_i K_{ij}/p_j \geq 0$. Stationarity gives $\sum_i w_{ij} = 1$, and $(qK)_j/p_j = \sum_i w_{ij}\, q_i/p_i$. By Jensen's inequality,

$$f\Big(\frac{(qK)_j}{p_j}\Big) \leq \sum_i w_{ij}\, f\Big(\frac{q_i}{p_i}\Big).$$

Multiply by p_j and sum over j: $D(qK \,\|\, p) \leq \sum_i \sum_j p_i K_{ij} f(q_i/p_i) = \sum_i p_i f(q_i/p_i) = D(q \,\|\, p)$, since the rows of K sum to 1. For the Gibbs law and a distribution q, $\log p_i = -c_i/T - \log Z$ and $\sum_i q_i = 1$, so $TD(q \,\|\, p) = -TS(q) + \sum_i q_i c_i + T\log Z = F(q) + T\log Z$; at $q = p$ this gives $F(p) = -T\log Z$, hence $F(q) - F(p) = TD(q \,\|\, p)$. □

Conditions and scope.

Base and premises

A finite configuration space, a stochastic matrix, a stationary vector with every entry positive.

Conclusion

Relative entropy to the stationary law, and with it the free energy of a Gibbs law, never rises. Recognition Science's second-law notes assume this monotonicity for its relaxation step; for every Markov step with a positive stationary law it is a theorem.

Not claimed

That a given recognition tick acts on a finite configuration space as a stochastic matrix with the Gibbs law stationary. That is an identification; Section 26.4 treats Recognition Science's own tick directly.

Positivity of p matters. Recognition Science writes relative entropy as a sum over the configurations where both q and p are positive. For the stochastic matrix with rows $(1,0)$ and $(1,0)$, the stationary law $(1,0)$ and $q = (\frac{1}{2}, \frac{1}{2})$, that sum is $-\frac{1}{2}\log 2$ at q and 0 at $qK = (1,0)$: it rises. The theorem is the Markov form of Boltzmann's H-theorem, and in information theory it is the data-processing inequality for relative entropy [172, 177]. The proof also locates exactly where the arrow acts.

Theorem 26.11 (Where a step loses relative entropy)**.** *Let K and p be as in Theorem 26.10 and let $q \geq 0$.*

(a) $D(q \,\|\, p) - D(qK \,\|\, p) = \sum_j \Gamma_j$ *with* $\Gamma_j = \sum_i p_i K_{ij} f(q_i/p_i) - p_j f((qK)_j/p_j) \geq 0$.

(b) $D(qK \,\|\, p) = D(q \,\|\, p)$ *exactly when* $q_i/p_i = q_{i'}/p_{i'}$ *whenever* $K_{ij} \neq 0$ *and* $K_{i'j} \neq 0$*: the step merges no two configurations with different ratios.*

(c) This holds exactly when the time reversal $\hat{K}_{ji} = p_i K_{ij}/p_j$ *restores q:* $qK\hat{K} = q$.

(d) This holds exactly when some stochastic L with $pL = p$ restores q: $qKL = q$.

(e) If every $K_{ij} > 0$, $\sum_i p_i = 1$, *and $q \neq p$ is a distribution, then* $D(qK \,\|\, p) < D(q \,\|\, p)$.

These results are proved in classical mathematics.

Proof. (a) The two double sums in the proof of Theorem 26.10 are $D(q \,\|\, p)$ and $D(qK \,\|\, p)$, and $\Gamma_j \geq 0$ is Jensen's inequality for column j.

(b) By strict convexity, $\Gamma_j = 0$ exactly when q_i/p_i is constant on $\{i : w_{ij} > 0\} = \{i : K_{ij} \neq 0\}$, the *predecessor set* of j; and a sum of nonnegative terms vanishes exactly when each does.

(c) $\hat{K}$ is stochastic, $\sum_i \hat{K}_{ji} = 1$, and $p\hat{K} = p$. Suppose the ratios are constant, equal to c_j, on the predecessor set of each j. Then $(qK)_j = c_j p_j$, and $(qK\hat{K})_i = \sum_j c_j p_i K_{ij} = p_i \sum_j K_{ij} q_i/p_i = q_i$, because $c_j = q_i/p_i$ whenever $K_{ij} \neq 0$. So (b) gives (c), and (c) gives (d) with $L = \hat{K}$.

(d) If L restores q, Theorem 26.10 for K and for L gives $D(q \,\|\, p) = D(qKL \,\|\, p) \leq D(qK \,\|\, p) \leq D(q \,\|\, p)$.

(e) Every predecessor set is all of Ω, so equality would make q/p a constant c; summing, $c = 1$ and $q = p$. □

The recovery map $\hat{K}$ is the classical case of Petz's recovery map for quantum relative entropy [187]. Statement (b) explains the opening example: in the step from q_0 the only predecessor set with two members is $\{0, 2\}$, and q_0 has the ratio $\frac{4}{3}$ at both.

Theorem 26.12 (Strict decrease: not per tick, but per window)**.** *The following are proved in classical mathematics.*

(a) A strict decrease at every tick fails for primitive chains: the chain K of the opening, with rational entries, has $K^4 > 0$ entrywise and the positive stationary law $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, and the distribution $q_0 = (\frac{2}{3}, 0, \frac{1}{3}) \neq p$ has $D(q_0K \,\|\, p) = D(q_0 \,\|\, p) = \log \frac{4}{3}$.

(b) For every stochastic K with positive stationary distribution p and every N with $K^N > 0$ entrywise, $D(qK^N \,\|\, p) < D(q \,\|\, p)$ for every distribution $q \neq p$.

(c) For every stochastic K with positive stationary distribution p and every distribution q, one step lowers $D(q \,\|\, p)$ exactly when some predecessor set carries two different ratios q_i/p_i.

Proof. (a) The rows of K^2 are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$, $(1, 0, 0)$ and $(\frac{1}{2}, \frac{1}{2}, 0)$, and those of $K^4 = (K^2)^2$ are $(\frac{9}{16}, \frac{5}{16}, \frac{1}{8})$, $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $(\frac{5}{8}, \frac{1}{8}, \frac{1}{4})$. The values of D are in the opening table: $D(q_0 \,\|\, p) = \frac{2}{3} \log \frac{4}{3} + \frac{1}{3} \log \frac{4}{3}$ and $D(q_0K \,\|\, p) = \frac{2}{3} \log \frac{4}{3} + \frac{1}{3} \log \frac{1/3}{1/4}$, both $\log \frac{4}{3}$.

(b) K^N is stochastic with $pK^N = p$ and all entries positive; apply Theorem 26.11(e) to it.

(c) is Theorem 26.11(b) together with Theorem 26.10. □

Heat baths are the commonest physical steps of this kind. Let r be a *record* of the configurations, a map from Ω to some set, and let π be a positive law. The *heat bath* for r keeps the record and resamples the configuration from π conditioned on it: $K(x, y) = \pi(y)/\pi(r^{-1}(r(x)))$ when $r(y) = r(x)$, and 0 otherwise.

Proposition 26.13 (Heat baths, Gibbs resampling and Glauber dynamics)**.** *The following is proved in classical mathematics. Let π be the Gibbs law of a recognition system for costs X.*

(a) For every record r, the heat bath is a $\mathcal{J}$-descent operator with equilibrium π, and it leaves $D(q \,\|\, \pi)$ unchanged exactly when q/π is constant on every fibre $\{r(x) = r(y)\}$.

(b) Gibbs resampling, $K(x, y) = \pi(y)$, lowers $D(q \,\|\, \pi)$ strictly from every distribution $q \neq \pi$.

(c) Every mixture $\sum_k w_k K_k$ of heat baths, with $w_k \geq 0$ and $\sum_k w_k = 1$ (random scan), and in particular Glauber dynamics of a spin system on $L+1$ sites (choose a site uniformly, resample its spin from the conditional Gibbs law given the other spins), is a $\mathcal{J}$-descent operator with equilibrium π.

Proof. A heat bath is stochastic, and $\pi(x)K(x, y) = \pi(x)\pi(y)/\pi(\text{fibre})$ is symmetric in x and y; summing over x gives $\pi K = \pi$. The predecessor set of y is the fibre of y, so Theorem 26.11(b) gives the criterion in (a). Gibbs resampling is the heat bath of a constant record, with all entries positive; apply Theorem 26.11(e). A mixture of stochastic matrices that keep π is stochastic and keeps π, and Glauber dynamics is the uniform mixture of the heat baths whose record is the configuration with one site erased. □

The Metropolis rule [186] and Glauber's heat bath [182] were introduced to sample Gibbs laws; the proposition says that each of them is a second-law step for the law it samples.

26.3 What the second law gives Recognition Science's dynamics

A J-descent operator never raises relative entropy. Three questions remain: which steps produce no entropy at all, which produce some at every tick, and how fast the production drives a distribution to equilibrium. All three have exact answers for Markov steps. Throughout, K is a stochastic matrix on a finite set with a stationary distribution p, every $p_i > 0$; the results are proved in classical mathematics. Theorem 26.17 gives the forms on $\mathbb{R}_\delta$ of two of them: Doeblin's rate and one-step strictness for lazy connected steps.

Theorem 26.14 (Zero production means a permutation of the costs)**.** *$D(qK \,\|\, p) = D(q \,\|\, p)$ for every distribution q exactly when K is the permutation matrix of a permutation σ with $p(\sigma(i)) = p(i)$ for every i. For the Gibbs law $p_i \propto e^{-J(X_i)/T}$ of Recognition Science, this holds exactly when K relabels the configurations by a permutation that preserves every cost: $J(X_{\sigma(i)}) = J(X_i)$. The point masses alone decide it.*

Proof. If K permutes by a p-preserving σ, then $D(qK \,\|\, p)$ is the same sum as $D(q \,\|\, p)$ with its terms relabelled. Conversely, suppose there is no production from any point mass δ_i. The ratio δ_i / p is $1/p_i$ at i and 0 elsewhere, so by Theorem 26.11(b) no column j can have $K_{ij} \neq 0$ and $K_{i'j} \neq 0$ with $i' \neq i$: every column has at most one nonzero entry. Each row has a nonzero entry, since it sums to 1; let $\sigma(i)$ be its column. The map σ is injective, hence a permutation of the finite set, and each row's single nonzero entry is 1. Stationarity gives $p_{\sigma(i)} = p_i$. For the Gibbs law, $p_{\sigma(i)} = p_i$ exactly when $e^{-J(X_{\sigma(i)})/T} = e^{-J(X_i)/T}$, that is, when the costs agree. □

So among Markov steps, the reversible relabellings are exactly the steps that produce nothing. The cyclic shift of eight configurations of equal cost is one; it has the form of the eight-tick clock of Recognition Science, a match of formulas rather than an identification. Every other step produces entropy from some point mass.

Theorem 26.15 (One-step strictness: one merging class)**.** *Say that i and i'* feed a common configuration *when $K_{ij} \neq 0$ and $K_{i'j} \neq 0$ for some j, and call the classes of the equivalence relation this generates the* merging classes *of K.*

(a) $D(qK \,\|\, p) = D(q \,\|\, p)$ exactly when q/p is constant on every merging class.

(b) One step lowers $D(q \,\|\, p)$ strictly from every distribution $q \neq p$ exactly when K has a single merging class.

(c) If K is lazy, $K_{ii} > 0$ for every i, and connected, any two configurations joined by a path of steps $a \to b$ with $K_{ab} > 0$ or $K_{ba} > 0$, then K has a single merging class.

(d) Connectivity is necessary for (b), and laziness is not: the matrix on three configurations with $K_{ij} = \frac{1}{2}$ for $i \neq j$ and $K_{ii} = 0$, with uniform p, lowers D strictly at every step. The swap of two configurations is connected and produces nothing from $(1, 0)$; the identity on two configurations is lazy and produces nothing from $(1, 0)$.

Proof. (a) By Theorem 26.11(b), no production means q/p is constant on each predecessor set, which is the same as constant on each merging class.

(b) With one class, no production forces $q = cp$, and $c = 1$ because both sum to 1. With a class C that misses some configuration, the distribution $q = p\,1_C/p(C)$ differs from p and q/p is constant on every class, so it produces nothing.

(c) If $K_{ab} > 0$, then a and b both feed b, because $K_{bb} > 0$. So the two ends of every edge lie in one class, and connectivity gives one class.

(d) If i and i' feed a common j, then i, j, i' is a path; so one merging class makes K connected. In the three-configuration matrix any two configurations feed the third, so there is one class. The swap and the identity have two classes each. □

Theorem 26.16 (Doeblin's condition gives a rate)**.** *Suppose $(K^N)_{ij} \geq \delta\, p_j$ for all i, j, for some window N and some $\delta \geq 0$. Then $\delta \leq 1$, and for every distribution q and every m,*

$$D(qK^{Nm} \,\|\, p) \leq (1-\delta)^m D(q \,\|\, p), \qquad D(qK^n \,\|\, p) \leq (1-\delta)^{\lfloor n/N \rfloor} D(q \,\|\, p) \text{ for all } n.$$

If $K^N > 0$ entrywise, some $\delta > 0$ works, so $D(qK^n \,\|\, p) \to 0$ exponentially. For a Gibbs law the free-energy excess $F(q) - F(p)$ decays at the same rate.

Proof. Step 1. Summing the hypothesis over j in any row gives $\delta \leq \sum_j (K^N)_{ij} = 1$.

Step 2: one window, $\delta < 1$. Put $M = K^N$, stochastic with $pM = p$, and $R_{ij} = (M_{ij} - \delta p_j)/(1-\delta)$. Then $R \geq 0$ by hypothesis, its rows sum to $(1-\delta)/(1-\delta) = 1$, and $pR = (p - \delta p)/(1-\delta) = p$. So $M = \delta\Pi + (1-\delta)R$ with $\Pi_{ij} = p_j$, and for a distribution q, $qM = \delta p + (1-\delta)\,qR$. Convexity of f with $f(1) = 0$ gives $p_i f\big(\delta + (1-\delta)x_i\big) \leq (1-\delta)p_i f(x_i)$ with $x_i = (qR)_i/p_i$, so $D(qM \,\|\, p) \leq (1-\delta)D(qR \,\|\, p)$, and Theorem 26.10 for R gives $D(qR \,\|\, p) \leq D(q \,\|\, p)$.

Step 3: $\delta = 1$. Each row of M is at least p and sums to 1, so $M = \Pi$, $qM = p$ and $D(qM \,\|\, p) = 0$.

Step 4. Nm steps of K are m steps of M, so the bound iterates. For general n use $\lfloor n/N \rfloor$ full windows and Theorem 26.10 on the remaining steps. If $K^N > 0$, take $\delta = \min_{i,j} (K^N)_{ij}/p_j > 0$. The free-energy statement is $F(q) - F(p) = TD(q \,\|\, p)$. □

Conditions and scope.

Base and premises
A stochastic matrix on a finite set, a positive stationary distribution, and a Doeblin minorisation at one window.

Conclusion
A geometric rate for relative entropy and free energy.

Not claimed
A rate for J-descent operators in general: Proposition 26.18 shows that none exists.

Doeblin introduced the minorisation that bears his name for the ergodic theory of Markov chains [180]; its use for mixing is standard [184]. Two of the results above also hold on $\mathbb{R}_\delta$, with relative entropy computed by the entropy function, and there they are free: Doeblin's

rate, and one-step strictness for lazy connected steps. The one point that needs care is the strict statement, where "$q \neq p$" must become "q is apart from p". The zero-production criterion and the merging-class criterion of Theorems 26.14 and 26.15 are proved here only in classical mathematics.

Theorem 26.17 (Doeblin's rate and lazy strictness are free on the line)**.** *Let K be an $n \times n$ matrix of points of $\mathbb{R}_\delta$ with $K_{ij} \geq 0$ and rows summing to 1, and let ℓ be levels with $\sum_i e^{\ell_i} = 1$ and $\sum_i e^{\ell_i} K_{ij} = e^{\ell_j}$. Write $p_i = e^{\ell_i}$ and $D = D_\eta(\cdot \,\|\, \ell)$.*

(a) If $d \geq 0$ and $(K^N)_{ij} \geq d\, p_j$ for all i, j, then for every $q \geq 0$ with $\sum_i q_i = 1$ and every m, $D(qK^{Nm}) \leq (1-d)^m D(q)$.

(b) If $K_{ii} > 0$ for every i and any two indices are joined by a path of positive entries, then $D(qK) < D(q)$ for every $q \geq 0$ with $\sum_i q_i = 1$ and some $q_i \# p_i$.

Both proofs are free.

Proof. (a) Write $M = K^N$ and $S = M - d\,\Pi$, where $\Pi_{ij} = p_j$; then $S \geq 0$, with no division by $1 - d$ and so no decision whether $d = 1$. Summing the hypothesis over j gives $d \leq 1$. Fix a column j, put $r = (qM)_j$ and $Q = \sum_i q_i$, and let μ be a point. Inequality (26.2), applied to the nonnegative weights S_{ij}, for which $\sum_i p_i S_{ij} = (1-d)p_j$, and the same Gibbs-term bound at the single atom Qp_j give

$$\sum_i S_{ij}\big(-\eta(q_i) - q_i \ell_i\big) + d\big(-\eta(Qp_j) - Qp_j \ell_j\big) \geq \mu r + r - e^{\mu} p_j,$$

since $\sum_i S_{ij} q_i + dQp_j = r$. As in Theorem 26.7, the choice $\mu = \log r - \ell_j$ for positive columns and the shift $q + \varepsilon p$ make the right side the jth term of $D(qM)$. Summing over j, with $\sum_j S_{ij} = 1 - d$, gives $D(qM) \leq (1-d)D(q) + d\,D(Qp)$, and here $Q = 1$ and $D(p) = 0$. Since $1 - d \geq 0$, the bound iterates over m windows.

(b) The exact form of (26.2) at a column j with $r_j > 0$ and $\mu_j = \log r_j - \ell_j$ is the identity

$$\sum_i K_{ij}\big(-\eta(q_i) - q_i \ell_i\big) - \big(-\eta(r_j) - r_j \ell_j\big) = \sum_i K_{ij}\, g(\mu_j + \ell_i, q_i),$$

whose terms are nonnegative by Lemma 26.4. Summed over j, the left sides give $D(q) - D(qK)$, and every column contributes at least 0: for $r_j > 0$ this is (26.2) at μ_j, and in general it follows by applying that bound to $q + \varepsilon p$ and letting $\varepsilon \to 0$ (Lemma 26.2), as in the proof of Theorem 26.7. So it suffices to find one column b with $r_b > 0$ and one positive term. Since $q_i \# p_i$ for some i and both vectors sum to 1, some i has $q_i > p_i$ and some k has $q_k < p_k$ (a negative finite sum has a negative term), so $q_i/p_i \# q_k/p_k$. Along a path of positive entries from i to k, cotransitivity of $<$ finds consecutive indices a, b with $q_a p_b \# q_b p_a$, and we may orient them so that $K_{ab} > 0$. One of q_a, q_b is positive, so $r_b \geq K_{ab} q_a + K_{bb} q_b > 0$. If both terms $g(\mu_b + \ell_a, q_a)$ and $g(\mu_b + \ell_b, q_b)$ vanished, Lemma 26.4 would give $q_a = r_b p_a / p_b$ and $q_b = r_b$, so $q_a p_b = q_b p_a$. By cotransitivity applied to the apartness, one of $q_a \# r_b p_a / p_b$ and $q_b \# r_b$ holds, and the corresponding term, multiplied by $K_{ab} > 0$ or $K_{bb} > 0$, is positive. $\square$

Recognition Science's own notes on the second law make two statements about J-descent operators that these theorems correct. The second-law notes describe the Lyapunov function $F_R(q_n) - F_R(p)$ as providing a quantitative convergence rate for the dynamics. The notes on the boundary second law list three inclusions as strict: cost-preserving permutations inside π-preserving permutations, inside the maps of the simplex that keep relative entropy, inside the injective maps of the simplex.

Proposition 26.18 (Two corrections to Recognition Science's notes)**.** *The following is proved in classical mathematics.*

(a) No rate holds for all $\mathcal{J}$-descent operators. On two configurations of equal cost the Gibbs law is $(\frac{1}{2}, \frac{1}{2})$, the swap is a $\mathcal{J}$-descent operator, and from $(1, 0)$ it keeps $D = \log 2$ at every tick; so no $\delta > 0$ satisfies $D(R(q) \,\|\, p) \le (1 - \delta) D(q \,\|\, p)$ for every $\mathcal{J}$-descent operator R and every q. A rate holds under a Doeblin condition (Theorem 26.16), and for Recognition Science's tick it is Corollary 26.29.

(b) For the Gibbs law of a cost, the first inclusion is an equality: a permutation preserves the Gibbs law exactly when it preserves every cost. Among Markov steps the second is an equality as well: a stochastic matrix that keeps relative entropy from every distribution is the matrix of a p-preserving permutation (Theorem 26.14). The third stays strict among Markov steps: the step $q \mapsto (q + p)/2$, with matrix $(\delta_{ij} + p_j)/2$, is injective and lowers $D(q \,\|\, p)$ strictly from every distribution $q \neq p$.

Proof. (a) For equal costs the Gibbs weights are equal. The swap is stochastic and keeps $(\frac{1}{2}, \frac{1}{2})$, so it is a J-descent operator by Theorem 26.10; it maps $(1, 0)$ to $(0, 1)$ and back, and $D = \log 2$ at both. (b) The first equality is the last step of the proof of Theorem 26.14. For the third, $(q + p)/2$ determines q, and Step 2 of the proof of Theorem 26.16, with $\delta = \frac{1}{2}$ and R the identity, gives $D((q + p)/2 \,\|\, p) \le \frac{1}{2} D(q \,\|\, p) < D(q \,\|\, p)$, since $D(q \,\|\, p) > 0$ for $q \neq p$. □

26.4 Recognition Science's tick: walls, windows and a rate

We now turn to the voxel of the opening. Its state space is infinite, so the finite theorems of the last two sections apply to it in two ways: on finite sets of states with a wall, and through a spectral gap on all of $\mathbb{Z}^3$. This section and the next two are proved in classical mathematics, except Theorem 26.31, which is free.

Definition 26.19 (Voxel states, the T9 law and voxel ticks)**.** A *voxel state* is an imbalance vector $s \in \mathbb{Z}^3$, the imbalances of a voxel's three axis accounts. A *move* adds $\pm e_i$ to it, one unit posted to one side of one account, so each state has six neighbours. The *imbalance cost* is $|s|_1 = |s_1| + |s_2| + |s_3|$. Let $\rho = \varphi^{-1} = \varphi - 1$; from here on ρ stands for this number, not for the reading of Section 26.1. The *weight* is $w(s) = \rho^{|s|_1}$ and the *T9 law* is $\pi(s) = \rho^{|s|_1}/\varphi^9$. A *voxel tick* is a matrix P on $\mathbb{Z}^3$ with entries $P(s, s') \ge 0$ that vanish unless $s' = s$ or s' is a neighbour of s, with $P(s, s) + \sum_m P(s, s + m) = 1$ (the sum over the six moves), with detailed balance $w(s)P(s, s') = w(s')P(s', s)$, and with the six entries $P(0, m)$ equal. $P(s, s)$ is the *hold* at s and $P(s, s + m)$ the *post* along m. We write P^n for n ticks and $D_n(a) = D(P^n(a, \cdot) \,\|\, \pi)$ for the relative entropy after n ticks from the state a.

A move changes $|s|_1$ by exactly ± 1. Call it *inward* when it lowers $|s|_1$ and *outward* otherwise. Along each axis at most one of the two moves is inward, every nonzero state has an inward move, and the *vacuum* 0 has none. The label T9 names this law only: it comes from a numbering of Recognition Science's derivations other than the one restated in Chapter 22, and the law is not step C9 there. The T9 law is a probability law: since $\sum_{n\in\mathbb{Z}} \rho^{|n|} = (1+\rho)/(1-\rho)$, and $1+\rho = \varphi$, $1-\rho = \rho^2$ and $\rho^{-1} = \varphi$, the total weight is $\big((1+\rho)/(1-\rho)\big)^3 = (\varphi^3)^3 = \varphi^9$. So $\pi(0) = \varphi^{-9} = 17\sqrt{5} - 38 \approx 0.013156$, and the vacuum carries $-\log \pi(0) = 9 \log \varphi \approx 4.331$ nats. After n ticks from a the law $P^n(a, \cdot)$ lives on the ball $B_{|a|_1+n}$, where $B_N = \{s : |s|_1 \le N\}$, so $D_n(a)$ is a finite sum. Detailed balance passes to powers: $\pi(a)P^n(a,s) = \pi(s)P^n(s,a)$.

Definition 26.20 (Recognition Science's tick and the Metropolis tick)**.** *Recognition Science's tick* is the voxel tick whose holds are all positive and which satisfies the hold-return balance

$$P(s,s)\,P(s',s') = P(s,s')\,P(s',s) \qquad \text{for every neighbouring pair } s, s'.$$

Exactly one voxel tick has these two properties (Theorem 26.21). The *Metropolis tick* proposes each of the six moves with probability $\frac{1}{6}$ and accepts the move to s' with probability $\min(1, w(s')/w(s))$: an inward post has probability $\frac{1}{6}$, an outward post $\rho/6$, and the rest of the row is the hold. At the vacuum it holds with probability $1-\rho$.

The tick is well defined: the two conditions turn the kernel into one equation for one positive function on the states, and that equation has exactly one solution.

Theorem 26.21 (Recognition Science's tick exists and is unique)**.** *In classical mathematics: exactly one voxel tick has all holds positive and satisfies hold-return balance. Its entries are*

$$P(s,s) = \frac{g(s)^2}{w(s)}, \qquad P(s, s+m) = \frac{g(s)\,g(s+m)}{w(s)},$$

and 0 *elsewhere, where g is the only positive function on $\mathbb{Z}^3$ that takes one value on the six unit vectors and satisfies*

$$g(s)\Big(g(s) + \sum_m g(s+m)\Big) = w(s) \qquad \textit{for every state } s, \tag{26.3}$$

the sum running over the six moves.

Proof. Step 1: such a tick gives a solution. Let P be a voxel tick with positive holds and hold-return balance, and put $g(s) = \sqrt{w(s)P(s,s)} > 0$. For neighbours s, s', detailed balance and hold-return balance give

$$\big(w(s)P(s,s')\big)^2 = w(s)P(s,s')\,w(s')P(s',s) = w(s)w(s')\,P(s,s)P(s',s') = g(s)^2 g(s')^2,$$

and both $w(s)P(s,s')$ and $g(s)g(s')$ are nonnegative, so $w(s)P(s,s') = g(s)g(s')$. Multiplying the row sum $P(s,s) + \sum_m P(s, s+m) = 1$ by $w(s)$ gives (26.3). The six posts $P(0,m) = g(0)g(m)$ at the vacuum are equal, so g takes one value on the six unit vectors. So P has the displayed form, with g a positive solution of (26.3).

Step 2: a solution gives such a tick. Let g be a positive solution of (26.3) that takes one value on the six unit vectors, and define P by the display. Its entries are nonnegative and vanish off

the diagonal and the neighbours; its rows sum to 1 by (26.3); detailed balance holds because $w(s)P(s,s') = g(s)g(s')$ is symmetric; the holds are positive; the six posts $P(0,m) = g(0)g(m)$ at the vacuum are equal, since $w(0) = 1$; and $P(s,s)P(s',s') = g(s)^2 g(s')^2/(w(s)w(s')) = P(s,s')P(s',s)$.

Step 3: at most one solution. Let g be a positive solution. The neighbour sum in (26.3) is positive, so $g(s)^2 < w(s)$. A move changes $|s|_1$ by one, so $w(s+m) \le \varphi\, w(s) < 4w(s)$ and $g(s+m) < \sqrt{w(s+m)} < 2\sqrt{w(s)}$. Then (26.3) gives $w(s) < g(s)\big(\sqrt{w(s)} + 12\sqrt{w(s)}\big)$, that is, $\sqrt{w(s)} < 13\, g(s)$, and so $p(s) = g(s)^2/w(s) > 1/169$. Put $q(s,m) = g(s)g(s+m)/w(s)$; by (26.3), $p(s) + \sum_m q(s,m) = 1$. Now let h be a second positive solution and $r(s) = h(s)/g(s)$. Both $r(s)$ and $1/r(s)$ are below $\sqrt{w(s)}\big/\big(\sqrt{w(s)}/13\big) = 13$, so

$$B = \sup_s \max\big(r(s), 1/r(s)\big)$$

is finite, and $B \ge 1$. Dividing (26.3) for h by $w(s)$ gives

$$r(s)\Big(p(s)\, r(s) + \sum_m q(s,m)\, r(s+m)\Big) = 1.$$

Suppose $r(s) \ge 1$. Every $r(s+m)$ is at least $1/B$, so

$$1 \ge r(s)\Big(p(s) + \frac{1-p(s)}{B}\Big) = r(s)\,\frac{1 + p(s)(B-1)}{B} \ge r(s)\,\frac{B+168}{169\,B},$$

because $p(s) > 1/169$ and $B \ge 1$. So $r(s) \le 169B/(B+168)$, and this also holds when $r(s) < 1$, since $169B/(B+168) \ge 1$. Exchanging g and h bounds $1/r(s)$ in the same way. Taking the supremum over s, $B \le 169B/(B+168)$, that is $B \le 1$. Hence $r(s) = 1$ for every s, and $h = g$.

Step 4: a solution exists. Write $g(s) = \sqrt{w(s)}\, x(s)$ and $c(s,m) = \sqrt{w(s+m)/w(s)}$, which is $\rho^{1/2}$ or $\varphi^{1/2}$, so $0 < c(s,m) < 2$. Then (26.3) reads $x(s)\big(x(s) + S_x(s)\big) = 1$ with $S_x(s) = \sum_m c(s,m)\, x(s+m)$. For $S \ge 0$ the positive root of $y(y+S) = 1$ is $R(S) = 2/\big(S + \sqrt{S^2+4}\big)$, which decreases in S, with $R(0) = 1$ and $R(12) = 2/(12 + \sqrt{148}) > 1/13$. Let X be the set of functions x from $\mathbb{Z}^3$ to $[1/13, 1]$. For $x \in X$, $0 \le S_x(s) \le 12$, so $U(x)(s) = R(S_x(s))$ defines a map U of X into itself; and $x \le y$ pointwise gives $U(x) \ge U(y)$. So $U^2 = U \circ U$ preserves the pointwise order. Start from the constant $x_0 = 1/13$ and put $x_{k+1} = U^2(x_k)$. Then $x_0 \le x_1$, and by induction $x_k \le x_{k+1}$; so at each state x_k increases to a limit $x_\infty(s) \le 1$. The value $U^2(x)(s)$ is a continuous function of the values of x at the states within distance 2 of s, so $U^2(x_\infty) = \lim_k x_{k+1} = x_\infty$. Then $U(x_\infty)$ is a fixed point of U^2 as well. The value $U(x)(s)$ depends only on x at the neighbours of s, whose cost has the other parity. Put $x(s) = U(x_\infty)(s)$ when $|s|_1$ is odd and $x(s) = x_\infty(s)$ when it is even. At an odd state the neighbours are even, so $U(x)(s) = U(x_\infty)(s) = x(s)$; at an even state they are odd, so $U(x)(s) = U^2(x_\infty)(s) = x_\infty(s) = x(s)$. So $U(x) = x$, and $g = \sqrt{w}\,x$ is a positive solution of (26.3).

Step 5: symmetry. A signed permutation σ of the three axes keeps $|s|_1$, so it keeps w, and it carries the neighbours of s to the neighbours of $\sigma(s)$. So $g \circ \sigma$ is again a positive solution, and $g \circ \sigma = g$ by Step 3. The six unit vectors are carried to one another by such permutations, so g takes one value on them. By Step 2 the solution gives a voxel tick with positive holds and hold-return balance, and by Steps 1 and 3 every such tick is this one. □

By the theorem every post of Recognition Science's tick is positive, and by Step 3 of its proof every hold is at least $1/169$. That this kernel is the tick of Recognition Science's history

dynamics rests on the premises from which Recognition Science derives it: imbalance states, the T9 weight, detailed balance and hold-return balance. The theorems below concern the kernel. Its entries are not known in closed form: they come from the solution of (26.3), the maximum-entropy equation of that dynamics. We use five numerical properties of them. This book does not prove them; they are stated inputs, obtained by interval computations on the kernel's entries that we do not reproduce:

(T1) every hold exceeds 0.135;

(T2) it has the Poincaré gap $\gamma = 39/5000$ on $\mathbb{Z}^3$ (Definition 26.27);

(T3) $\varphi^9 P^2(0,0) - 1 < 15.26$ and $\varphi^9 P^8(0,0) - 1 < 3.151$;

(T4) $P^{10}(0,0) > 0.045166$ and $e_5/e_4 \geq 0.879^2$, for the even-return excess e_n of Section 26.6;

(T5) the hold at the vacuum lies between 0.18 and 0.19.

Every theorem in this section and the next two is proved for every voxel tick with the stated properties. The five facts enter only the statements named for Recognition Science's tick; each such statement names the facts it uses, and is proved here from them.

Definition 26.22 (Walled ticks and inward-closed sets)**.** Let S be a finite set of states containing the vacuum. The *walled tick* P_S on S moves as P inside S and turns every post that would leave S into a hold:

$$P_S(x,y) = P(x,y)\ (x \neq y), \qquad P_S(x,x) = P(x,x) + \sum_{m\,:\,x+m\notin S} P(x,x+m).$$

Its law is $\pi_S(x) = w(x)/\sum_{s\in S} w(s)$. The set S is *inward-closed* when every inward move from a state of S stays in S. Every ball B_N is inward-closed and contains the vacuum.

Theorem 26.23 (The walled tick is a second-law step, strict on inward-closed sets)**.** *Let P be a voxel tick and S a finite set of states containing the vacuum.*

(a) P_S is stochastic and satisfies $w(x)P_S(x,y) = w(y)P_S(y,x)$ on S; so $\pi_S P_S = \pi_S$, and $q \mapsto qP_S$ is a $\mathcal{J}$-descent operator with equilibrium π_S.

(b) If S is inward-closed and every post of P is positive, then $D(qP_S \,\|\, \pi_S) < D(q \,\|\, \pi_S)$ for every distribution $q \neq \pi_S$ on S.

No condition on the holds of P is needed.

Proof. (a) The six posts from x land on six distinct states, none equal to x. In the row of P_S at x the posts that stay in S keep their place and the others are added to the hold, so the row sums to 1. Off the diagonal $P_S = P$, so detailed balance is inherited, and summing it over x gives stationarity. Theorem 26.10 gives the J-descent operator.

(b) By Theorem 26.15(b) it suffices to show that P_S has one merging class. (i) Let $x \in S$ have $|x|_1 \geq 2$, let x' be x after an inward move and x'' be x' after an inward move; both lie in S.

Then x feeds x' by its inward post, and x'' feeds x' by the reverse, outward, post, which stays in S. So x and x'' lie in one class. Repeating, every state of even cost lies in the class of the vacuum, and every state of odd cost in the class of a unit vector. (ii) Every unit vector in S feeds the vacuum, so all of them lie in one class. (iii) Let x^* be a state of largest cost in S, and x' its inward neighbour. An outward move from x^* leaves S, so $P_S(x^*, x^*) > 0$, and x' feeds x^* by an outward post. So x^* and x', whose costs differ by one, lie in one class, which joins the even and the odd states. □

Theorem 26.24 (Doeblin's condition at window $2R$)**.** *Let P be a voxel tick with every hold at least $h \geq 0$ and every post at least $e \geq 0$, and let S be a finite inward-closed set containing the vacuum with $|s|_1 \leq R$ on S. Then for all $x, y \in S$,*

$$P_S^{2R}(x, y) \geq \min(h, e)^{2R}\, \pi_S(y),$$

and therefore $D(qP_S^{2Rm} \,\|\, \pi_S) \leq \big(1 - \min(h, e)^{2R}\big)^m D(q \,\|\, \pi_S)$ for every distribution q on S and every m.

Proof. An entry of a power of a nonnegative matrix is at least the product along any one path. Take the path that makes $|x|_1$ inward moves from x to the vacuum, staying in S by inward closure; holds at the vacuum $2R - |x|_1 - |y|_1$ times, each at least $P(0, 0) \geq h$; and then the inward path of y in reverse, $|y|_1$ outward moves inside S. It has $2R$ steps, each with probability at least $\min(h, e)$, and $\pi_S(y) \leq 1$. The rate is Theorem 26.16 with $N = 2R$ and $\delta = \min(h, e)^{2R}$. □

For Recognition Science's tick both bounds can be made explicit, and the window cannot be shortened.

Proposition 26.25 (Posts and Doeblin windows of Recognition Science's tick)**.**

(a) With (T1), every post of Recognition Science's tick lies in $[0.1061, 0.3345]$.

(b) With (T1), on the ball B_N the walled tick satisfies $P_{B_N}^{2N}(x, y) \geq 0.1061^{2N} \pi_{B_N}(y)$, where $0.1061^2 = 0.011257$. Every δ with $\delta\, \pi_{B_N}(y) \leq P_{B_N}^{2N}(x, y)$ for all x, y satisfies $\delta \leq \varphi^9 (0.3345^2 \varphi)^N \leq 76.02 \cdot 0.1811^N$, which is below 1 from $N = 3$ on.

(c) For every voxel tick and every N, a minorisation $\delta\, \pi_{B_N}(y) \leq P_{B_N}^{M}(x, y)$ for all $x, y \in B_N$ with $\delta > 0$ requires $M \geq 2N$.

Proof. (a) Let s, s' be neighbours. Detailed balance and hold-return balance give

$$\big(\pi(s)P(s, s')\big)^2 = \pi(s)\pi(s')\, P(s, s')P(s', s) = \pi(s)\pi(s')\, P(s, s)P(s', s') > 0.135^2\, \pi(s)\pi(s'),$$

so $P(s, s') > 0.135\, \rho^{(|s'|_1 - |s|_1)/2}$. An outward post exceeds $0.135\sqrt{\rho} > 0.135 \cdot 0.786 > 0.1061$, and an inward post exceeds $0.135/\sqrt{\rho} > 0.17$. In each row the hold exceeds 0.135 and five other posts exceed 0.1061, so every post is below $1 - 0.135 - 5 \cdot 0.1061 = 0.3345$.

(b) The lower bound is Theorem 26.24 with $R = N$, $h = 0.135$ and $e = 0.1061$. For the upper bound take $x = (N, 0, 0)$ and $y = (-N, 0, 0)$. A path of k ticks from x to $x - (k, 0, 0)$ must lower the first coordinate at every tick, since each tick moves ℓ^1-distance at most one; so only the

straight path contributes, its steps are posts of P inside B_N, and $P^{2N}_{B_N}(x,y) \le 0.3345^{2N}$. The total weight of B_N is at most φ^9, so $\pi_{B_N}(y) \ge \rho^N/\varphi^9$, and $\delta \le 0.3345^{2N}\varphi^9\rho^{-N} = \varphi^9(0.3345^2\varphi)^N$, with $0.3345^2\varphi = 0.18104$ and $\varphi^9 = 76.013$. At $N = 3$ the bound is 0.451.

(c) $P^M_{B_N}(x,y) = 0$ when $|x-y|_1 > M$, and $(N,0,0)$ and $(-N,0,0)$ are $2N$ apart while $\pi_{B_N} > 0$. □

So the walled form of Recognition Science's tick is strict on every inward-closed set (Theorem 26.23(b); its posts are positive by Theorem 26.21), and, given (T1), its Doeblin constant at window $2N$ decays exponentially in the radius. A rate that does not degrade with the radius needs a different tool, a spectral gap; it comes in Theorem 26.32. On $\mathbb{Z}^3$ itself the wall is never reached early.

Proposition 26.26 (On a ball the tick is a J-descent operator)**.** *Let P be a voxel tick, a a state and n, N natural numbers with $|a|_1 + n \le N$. Then $P^n_{B_N}(a,\cdot) = P^n(a,\cdot)$ and*

$$D\big(P^n(a,\cdot)\,\|\,\pi_{B_N}\big) = D_n(a) + \log\pi(B_N), \qquad \pi_{B_N} = \pi/\pi(B_N).$$

So on every ball the walled tick is a $\mathcal{J}$-descent operator with equilibrium π_{B_N} whose evolution from δ_a is the tick's own law for the first $N - |a|_1$ ticks.

Proof. After $k < n$ ticks the law lives on $B_{|a|_1+k} \subseteq B_{N-1}$, from which no post leaves B_N, so P_{B_N} and P act alike on it; induct on k. For the relative entropy, each term with $G(s) = P^n(a,s) > 0$ splits as $G\log(G/\pi) + G\log\pi(B_N)$, and $\sum G = 1$. □

Definition 26.27 (Poincaré gap)**.** For signed measures X, Y on $\mathbb{Z}^3$, one of them finitely supported, put $\langle X,Y\rangle_\pi = \sum_s X(s)Y(s)/\pi(s)$, and write XP for $(XP)(s') = \sum_s X(s)P(s,s')$. A voxel tick P has *Poincaré gap* γ on $\mathbb{Z}^3$ when every finitely supported X with $\sum_s X(s) = 0$ satisfies

$$\langle X,X\rangle_\pi - \langle X,XP\rangle_\pi \ge \gamma\langle X,X\rangle_\pi.$$

Theorem 26.28 (The rate from a point start)**.** *Let P be a voxel tick with every hold at least h and Poincaré gap γ, with $2h - 1 < 1 - \gamma$ and $(1-\gamma) + (2h-1) \ge 0$.*

(a) For every state a and every $n \ge 0$,

$$D_n(a) \le \log\Big(1 + (1-\gamma)^{2n}\big(\varphi^{9+|a|_1} - 1\big)\Big),$$

with equality at $n = 0$.

(b) For every finitely supported start law μ, $D(\mu P^n\,\|\,\pi) \le \log\big(1 + (1-\gamma)^{2n}\chi^2(\mu)\big)$, where $\chi^2(\mu) = \langle\mu,\mu\rangle_\pi - 1$.

(c) For all n_0 and n, $D_{n_0+n}(0) \le \log\big(1 + (1-\gamma)^{2n}(\varphi^9 P^{2n_0}(0,0) - 1)\big)$.

Proof. Step 1: relative entropy below the log of the energy. For a finitely supported law G, concavity of log, with the weights $G(s)$, gives $D(G\,\|\,\pi) = \sum_{G>0} G\log(G/\pi) \le \log\sum_{G>0} G\cdot(G/\pi) =$

$\log\langle G, G\rangle_\pi$. For $G = P^n(a,\cdot)$, detailed balance gives $\langle G, G\rangle_\pi = \sum_s P^n(a,s)P^n(s,a)/\pi(a) = P^{2n}(a,a)/\pi(a)$, which we call $E_n(a)$.

Step 2: one tick contracts the energy. Write $TX = XP$. On finitely supported measures T keeps finite support and total mass, and detailed balance makes it symmetric: $\langle X, TY\rangle_\pi = \langle TX, Y\rangle_\pi$. For every finitely supported Z, with $z = Z/\pi$,

$$\langle Z, Z\rangle_\pi + \langle Z, TZ\rangle_\pi = \frac{1}{2}\sum_{s,s'} \pi(s)P(s,s')\big(z(s) + z(s')\big)^2 \geq 2h\,\langle Z, Z\rangle_\pi,$$

keeping only the terms $s' = s$. So $\langle Z, TZ\rangle_\pi \geq a\langle Z, Z\rangle_\pi$ with $a = 2h - 1$. Let X be finitely supported with total mass 0, put $b = 1 - \gamma$, $c_k = \langle X, T^k X\rangle_\pi$, $Y = TX - aX$ and $Z = bX - TX$. Expanding with the symmetry of T gives the identity

$$(b-a)(b^2c_0 - c_2) = \big[b\langle Y, Y\rangle_\pi - \langle Y, TY\rangle_\pi\big] + \big[\langle Z, TZ\rangle_\pi - a\langle Z, Z\rangle_\pi\big] + (b^2 - a^2)(bc_0 - c_1).$$

The first bracket is nonnegative by the gap, since Y has total mass 0; the second by the hold bound; and the third term because $b > a$, $b + a \geq 0$ and $bc_0 - c_1 \geq 0$ by the gap. Since $b - a > 0$, $\langle TX, TX\rangle_\pi = c_2 \leq (1-\gamma)^2\langle X, X\rangle_\pi$, and by induction $\langle T^n X, T^n X\rangle_\pi \leq (1-\gamma)^{2n}\langle X, X\rangle_\pi$. No spectral theorem is used.

Step 3: truncation. $\delta_a - \pi$ is not finitely supported, so we truncate π. Let $Z_L = \pi(B_L)$ for $L \geq |a|_1 + 2n$, put $M = L - n$, and $X = \delta_a - \pi 1_{B_L}/Z_L$. Then $\langle X, X\rangle_\pi = 1/\pi(a) - 2/Z_L + 1/Z_L = 1/\pi(a) - 1/Z_L$. For $s \in B_M$, every state reachable from s in n ticks lies in B_L, so by detailed balance $(\pi 1_{B_L} P^n)(s) = \pi(s)\sum_x P^n(s,x) = \pi(s)$. Keeping in $\langle T^n X, T^n X\rangle_\pi$ only the terms with $s \in B_M$, where $P^n(a,\cdot)$ lives, Step 2 gives

$$E_n(a) - \frac{2}{Z_L} + \frac{Z_M}{Z_L^2} \leq (1-\gamma)^{2n}\Big(\frac{1}{\pi(a)} - \frac{1}{Z_L}\Big).$$

As $L \to \infty$, Z_L and Z_M tend to 1, so $E_n(a) - 1 \leq (1-\gamma)^{2n}\big(1/\pi(a) - 1\big)$, and $1/\pi(a) = \varphi^9\rho^{-|a|_1} = \varphi^{9+|a|_1}$.

Step 4. Combining, $D_n(a) \leq \log E_n(a) \leq \log\big(1 + (1-\gamma)^{2n}(\varphi^{9+|a|_1} - 1)\big)$. At $n = 0$ both sides are $-\log\pi(a)$.

(b) is the same argument with μ in place of δ_a: Step 1 bounds $D(\mu P^n \,\|\, \pi)$ by $\log\langle \mu P^n, \mu P^n\rangle_\pi$, and Step 3 with $X = \mu - \pi 1_{B_L}/Z_L$ gives $\langle \mu P^n, \mu P^n\rangle_\pi - 1 \leq (1-\gamma)^{2n}\chi^2(\mu)$. (c) is (b) for $\mu = P^{n_0}(0,\cdot)$, whose χ^2 is $\langle \mu, \mu\rangle_\pi - 1 = P^{2n_0}(0,0)/\pi(0) - 1$ by Step 1. □

Conditions and scope.

Base and premises

A voxel tick on $\mathbb{Z}^3$, a lower bound on its holds, a Poincaré gap, and the two inequalities linking them.

Conclusion

An explicit bound for the relative entropy to the T9 law at every tick, from every start, tight at the start.

Not claimed

That a tick is a unit of physical time, or that the T9 law is the Gibbs law of a J-cost; the latter holds only in the sense of Corollary 26.29(c).

Corollary 26.29 (The rate of Recognition Science's tick)**.** *Let P be Recognition Science's tick, and use (T1)–(T3).*

(a) For every state a and every n,
$D_n(a) \le \log\bigl(1 + (4961/5000)^{2n}(\varphi^{9+|a|_1} - 1)\bigr)$*, where* $4961/5000 = 0.9922$.

(b) $D_N(0) \le \log\bigl(1 + 15.26 \cdot (4961/5000)^{2(N-1)}\bigr)$ *for* $N \ge 1$*, and*
$D_N(0) \le \log\bigl(1 + 3.151 \cdot (4961/5000)^{2(N-4)}\bigr)$ *for* $N \ge 4$.

(c) On every ball B_N there is a $\mathcal{J}$-descent operator with equilibrium π_{B_N} whose n-fold evolution from δ_a is $P^n(a,\cdot)$ and stays within the bound of (a) for $|a|_1 + n \le N$. For a recognition system at temperature T_R, there are costs X on B_N whose Gibbs law is π_{B_N}, and the free-energy excess along this evolution is at most T_R times the bound of (a).

Proof. With $h = 0.135$ and $\gamma = 39/5000$: $2h - 1 = -0.73 < 1 - \gamma = 0.9922$ and $0.9922 - 0.73 \ge 0$, so Theorem 26.28 applies. (b) is part (c) of that theorem at $n_0 = 1$ and $n_0 = 4$, with (T3). For (c), Proposition 26.26 turns the walled evolution into the tick's law, and $D(P^n(a,\cdot) \,\|\, \pi_{B_N}) = D_n(a) + \log \pi(B_N) \le D_n(a)$. Every positive law on a finite set is a Gibbs law: take X with $J(X(x)) = -T_R \log \pi_{B_N}(x)$, possible because J takes every value in $[0,\infty)$ and $\pi_{B_N} \le 1$; then $Z = 1$, and $F_R(q) - F_R(\pi_{B_N}) = T_R\, D(q \,\|\, \pi_{B_N})$ by Theorem 26.10. □

In numbers: from the vacuum the bound of (a) starts at $9 \log \varphi \approx 4.331$ nats, the value of $D_0(0)$, and falls below one nat from tick 242 on; the first bound of (b) does so from tick 141, and the second from tick 43. These are proved worst cases, computed from (T1)–(T3), not relaxation times. Part (c) is the rate that Recognition Science's second-law notes ask of their Lyapunov function, and Proposition 26.18 showed that it fails for J-descent operators in general; for the tick itself, given (T1)–(T3), it is a theorem, on every ball, in both its relative-entropy and its free-energy form.

26.5 The Metropolis tick on the line, and a gap for every ball

The Metropolis tick has entries in $\mathbb{Q}(\varphi)$, so $\mathbb{R}_\delta$ represents it exactly, and its second law with a rate can be proved there with no omniscience and no choice. The decisions the proof needs are all about integer triples.

Definition 26.30 (Enumerated sets and the walled Metropolis tick on the line)**.** An *enumerated set* E is a finite list $e_0, \dots, e_{n-1}$ of distinct voxel states that contains the vacuum and contains every state reached from a listed state by an inward move. Whether a state is listed is found by searching the list, which compares integer triples. Let ρ be the point $\varphi - 1$ of $\mathbb{R}_\delta$, so $\rho^2 = 1 - \rho$. The *walled Metropolis tick* K_E has, for $i \ne j$, the entry $\frac{1}{6}$ if e_j is e_i after an inward move, $\rho/6$ if it is e_i after an outward move, and 0 otherwise; the diagonal entry is 1 minus the rest of the row. Its law is $\pi_E(i) = \rho^{|e_i|_1} / \sum_k \rho^{|e_k|_1}$, with levels $\ell_i = \log \pi_E(i)$.

Theorem 26.31 (A second law with a rate on the line)**.** *Let E be an enumerated set with $|e_i|_1 \le R$ for every i, and put $c = \rho^2/6 \approx 0.063661$. Write $D = D_\eta(\cdot \,\|\, \ell)$.*

(a) Every diagonal entry of K_E is at least $\rho^2/2$.

(b) K_E is stochastic, $\pi_E(i)K_E(i,j) = \pi_E(j)K_E(j,i)$, and π_E is stationary.

(c) $D(qK_E) \le D(q)$ for every $q \ge 0$.

(d) $D(qK_E) < D(q)$ for every distribution q with some $q_i \# \pi_E(i)$.

(e) $D(qK_E^{2Rm}) \le (1 - c^{2R})^m D(q)$ for every distribution q and every m.

All five statements are free. On the ball of radius one, with seven states, the factor in (e) is $1 - c^2 \approx 0.995947$ per two ticks.

Proof. (a) Which entry a move gets is decided by comparing integer costs and searching the list. Along each axis at most one move is inward, so the two posts of an axis sum to at most $(1 + \rho)/6$, the six posts to at most $(1 + \rho)/2$, and the diagonal entry is at least $1 - (1 + \rho)/2 = (1 - \rho)/2 = \rho^2/2$. No decision about which posts are kept is needed.

(b) An inward move from x and the reverse, outward, move from its target are both kept, by inward closure, and $\rho^{|x|_1} \cdot \frac{1}{6} = \rho^{|x|_1 - 1} \cdot \frac{\rho}{6}$. So detailed balance holds entry by entry, and summing it gives stationarity.

(c) is Theorem 26.7 with the stationary weights $e^{\ell_i} = \pi_E(i)$.

(d) By (a) the tick is lazy, and every index reaches the vacuum through inward moves inside E, so it is connected; Theorem 26.17(b) applies.

(e) As in Theorem 26.24, follow the path from e_i to the vacuum by inward moves, hold at the vacuum, and reach e_j by the reverse of its inward path. It has $2R$ steps, each with probability at least c: $\frac{1}{6} \ge c$, $\rho/6 \ge \rho^2/6$, and $\rho^2/2 \ge c$. Since $\pi_E(j) \le 1$, $K_E^{2R}(i,j) \ge c^{2R}\pi_E(j)$, and Theorem 26.17(a) gives the rate. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$, the golden point ρ, and a finite list of integer triples closed under inward moves.

Conclusion

Relative entropy to the walled T9 law never rises, falls strictly from every law apart from equilibrium, and falls geometrically over windows of $2R$ ticks; no omniscience and no choice.

Not claimed

The same for Recognition Science's own tick on $\mathbb{R}_\delta$: its entries solve a maximum-entropy equation and are not represented here. Under the reading, K_E is the classical walled Metropolis tick, entry for entry.

The Doeblin route gives a factor that decays with the radius. The next theorem gives a rate that does not: a Poincaré inequality for the walled tick with one constant for every inward-closed set. It is proved in classical mathematics, by canonical paths through the vacuum, a method of Jerrum and Sinclair and of Diaconis and Stroock [183, 179], here with weights along the path.

Theorem 26.32 (A spectral gap for every inward-closed set)**.** *Let P be a voxel tick whose inward posts are all at least $e > 0$, and let S be a finite inward-closed set containing the vacuum. Then every function f on S satisfies*

$$\operatorname{Var}_{\pi_S}(f) \le \frac{1030}{e}\,\mathcal{E}_S(f), \qquad \mathcal{E}_S(f) = \tfrac{1}{2}\sum_{x,y\in S} \pi_S(x)P_S(x,y)\big(f(x)-f(y)\big)^2,$$

so the walled tick has spectral gap at least $e/1030$, whatever the size of S. For Recognition Science's tick, with (T1), the constant is $10300000/1061 \approx 9707.8$ and the gap exceeds $1.03\cdot 10^{-4}$; for the Metropolis tick it is 6180, a gap of at least $1/6180 \approx 1.618\cdot 10^{-4}$; both on every ball B_N.

Proof. Let $L(u)$ be u after lowering its first nonzero coordinate by one in absolute value, and $L(0) = 0$. Inward closure keeps the path $x, Lx, L^2x, \dots, 0$ inside S; it has $|x|_1$ steps. Write w for the weight and $W = \sum_{s\in S} w(s)$.

(1) Mean. $\operatorname{Var}_{\pi_S}(f) \le \sum_x \pi_S(x)\big(f(x)-f(0)\big)^2$, since the mean minimises the mean square deviation.

(2) Weighted Cauchy–Schwarz along the path. With $g_k = f(L^kx) - f(L^{k+1}x)$ and $s = \frac{9}{10}$,

$$\big(f(x)-f(0)\big)^2 = \Big(\sum_{k<|x|_1} g_k\Big)^2 \le \Big(\sum_k s^k\Big)\Big(\sum_k \frac{g_k^2}{s^k}\Big) \le 10 \sum_{k<|x|_1} \frac{g_k^2}{s^k}.$$

(3) Exchange of sums. Put $G(u) = (f(u)-f(Lu))^2$ and $F(u) = \sum_{x\in S}\sum_{k<|x|_1} [L^kx = u]\, w(x)/s^k$. Then $\sum_x w(x)\sum_k G(L^kx)/s^k = \sum_{u\ne 0} G(u)F(u)$.

(4) The load on a state. A path through $u \ne 0$ either starts at u or passed one step earlier through a state $v \in S$ with $Lv = u$, so $F(u) \le w(u) + \frac{10}{9}\sum_{v\in S,\,Lv=u} F(v)$. The predecessors are explicit. If $u_1 \ne 0$, the only one is $u + \operatorname{sgn}(u_1)e_1$. If $u_1 = 0 \ne u_2$, there are three: $u \pm e_1$ and $u + \operatorname{sgn}(u_2)e_2$. If $u_1 = u_2 = 0 \ne u_3$, there are five: $u \pm e_1$, $u \pm e_2$ and $u + \operatorname{sgn}(u_3)e_3$. Each has cost $|u|_1 + 1$ and weight $\rho\, w(u)$. Give the three types of u the constants $C = \frac{7}{2}$, 19 and 103. With $\frac{10}{9}\rho < \frac{10}{9}\cdot 0.6181 < 0.68678$,

$$\begin{aligned} 1 + 0.68678\cdot\tfrac{7}{2} &= 3.404 \le \tfrac{7}{2},\\ 1 + 0.68678\cdot\big(\tfrac{7}{2}+\tfrac{7}{2}+19\big) &= 18.857 \le 19,\\ 1 + 0.68678\cdot\big(7+38+103\big) &= 102.64 \le 103. \end{aligned}$$

Induction from the states of largest cost in S, which have no predecessor in S, inward, gives $F(u) \le C(u)\, w(u) \le 103\, w(u)$.

(5) Edges. The maps $u \mapsto (u, Lu)$ and $u \mapsto (Lu, u)$ on $S \setminus \{0\}$ are injective with disjoint images, and detailed balance gives the two families equal weight; so $\sum_{u\ne 0} w(u)P(u,Lu)G(u) \le \frac{1}{2}\sum_{x,y\in S} w(x)P(x,y)(f(x)-f(y))^2$.

(6) Combine. $P(u, Lu) \geq e$, because Lu is an inward move. Dividing by W,

$$\mathrm{Var}_{\pi_S}(f) \leq \frac{10}{W} \sum_{u \neq 0} G(u)F(u) \leq \frac{10 \cdot 103}{eW} \sum_{u \neq 0} w(u)P(u, Lu)G(u) \leq \frac{1030}{e}\, \mathscr{E}_S(f),$$

where in $\mathscr{E}_S$ the walled tick may replace P, because the two agree off the diagonal and diagonal terms vanish. For Recognition Science's tick every post, inward ones included, is at least 0.1061 (Proposition 26.25(a)), and for the Metropolis tick the inward posts equal $\frac{1}{6}$. □

Conditions and scope.

Base and premises
A voxel tick with inward posts bounded below, and a finite inward-closed set containing the vacuum.

Conclusion
One Poincaré constant, $1030/e$, for every such set.

Not claimed
The size of the true gap: $1030/e$ is a proved worst case, not an estimate of the actual relaxation.

So the Doeblin constant of Proposition 26.25 decays like 0.18^N, while the gap does not degrade at all. Two statements remain open: a contraction of χ^2 at every tick that is uniform in the ball, and a form of the gap on $\mathbb{R}_\delta$. Some dependence on the start, however, cannot be removed.

Theorem 26.33 (No rate is uniform in the start)**.** *For every voxel tick P, every number of ticks n and every level M there is a c such that $D_n(a) \geq M$ for every start a with $|a|_1 \geq c$. Quantitatively, if $L + n < |a|_1$ then $D_n(a) \geq -\log\bigl(1 - \pi(B_L)\bigr)$. So the factor $\varphi^{9+|a|_1} - 1$ in Theorem 26.28 cannot be dropped.*

Proof. (i) A finitely supported law G carried by a set T has $D(G \,\|\, \pi) \geq -\log \pi(T)$: the inequality $\log y \geq 1 - 1/y$ at $y = G\pi(T)/\pi$ gives $G\log(G/\pi) \geq G - \pi/\pi(T) - G\log\pi(T)$ for each term with $G > 0$, and summing over T gives $D \geq 1 - 1 - \log\pi(T)$. (ii) After n ticks from a the law lives on $\{s : |s|_1 \geq |a|_1 - n\}$, since each tick changes the cost by at most one; if $L + n < |a|_1$, this set misses B_L, so its π-mass is at most $1 - \pi(B_L)$. (iii) $\pi(B_L) \to 1$, so $-\log(1 - \pi(B_L)) \to \infty$. □

26.6 The pace of the spine

A *spine* is a finite set of voxels. Started from the empty record, every voxel at the vacuum, each voxel relaxes toward the T9 law. The bound of Corollary 26.29 is a worst case. This section finds the exact pace, shows how a spine of many voxels settles, and shows what couplings between voxels can do. Everything here is proved in classical mathematics.

Definition 26.34 (The even-return excess and the exact rate)**.** For a voxel tick P the *even-return excess* is $e_n = P^{2n}(0,0) - \pi(0)$, and the *exact rate* of P is $r(P) = \sup_n e_{n+1}/e_n$.

Lemma 26.35 (The excess decays at exactly the rate $r(P)$)**.** *For every voxel tick P:*

(a) $e_n/\pi(0) = \sum_s \pi(s)\big(P^n(0,s)/\pi(s) - 1\big)^2 > 0$;

(b) $e_{n+1}^2 \le e_n e_{n+2}$, *so the ratios* e_{n+1}/e_n *are nondecreasing; they are at most* 1 *and tend to* $r(P)$, *and* $0 < r(P) \le 1$;

(c) $e_n \le e_0\, r(P)^n$, *and* $e_n \ge e_m (e_{m+1}/e_m)^{n-m}$ *for* $n \ge m$;

(d) if every hold is at least h *and* P *has Poincaré gap* γ, *with* $2h-1 < 1-\gamma$ *and* $(1-\gamma)+(2h-1) \ge 0$, *then* $r(P) \le (1-\gamma)^2$.

Proof. (a) With $G_n = P^n(0,\cdot)$, detailed balance gives $P^{2n}(0,0) = \sum_s G_n(s) P^n(s,0) = \pi(0)\sum_s G_n(s)^2/\pi(s)$, and since G_n and π both sum to 1, $\sum_s G_n^2/\pi - 1 = \sum_s \pi (G_n/\pi - 1)^2$. The terms outside B_n, where $G_n = 0$, equal $\pi(s) > 0$.

(b) With $u_n = G_n/\pi - 1$, the same computation gives $\sum_s \pi(s) u_n(s) u_{n+2}(s) = P^{2n+2}(0,0)/\pi(0) - 1 = e_{n+1}/\pi(0)$, and the Cauchy–Schwarz inequality for the weights π gives $e_{n+1}^2 \le e_n e_{n+2}$. If some ratio were $q > 1$, the later e_n would grow like q^n, but $e_n \le 1$. So the ratios increase to their supremum.

(c) Write e_n as e_0, or e_m, times a product of consecutive ratios, and bound each ratio by $r(P)$, or below by e_{m+1}/e_m.

(d) Step 3 of the proof of Theorem 26.28 at $a = 0$ gives $e_n/\pi(0) \le (1-\gamma)^{2n}(\varphi^9 - 1)$. A ratio $q > (1-\gamma)^2$ would give $e_n \ge e_m q^{n-m}$, which is too large for large n. □

Theorem 26.36 (The pace of one voxel)**.** *For every voxel tick P and every n,*

$$2e_n^2 \le D_{2n}(0) \qquad \text{and} \qquad D_n(0) \le \varphi^9 e_n.$$

Consequently $D_N(0)^{1/N} \to r(P)$ *as* $N \to \infty$, *and for* $m \le n$, $D_{2n}(0) \ge 2(e_m/q^m)^2 (q^2)^n$ *with* $q = e_{m+1}/e_m$.

Proof. Lower bound. Let $G = P^{2n}(0,\cdot)$ and $g = G(0) = e_n + \pi(0) \ge \pi(0)$. The log-sum inequality $\sum_i a_i \log(a_i/b_i) \ge A \log(A/B)$, with $A = \sum a_i$ and $B = \sum b_i$, follows by summing $a_i \log \frac{a_i B}{b_i A} \ge a_i - b_i A/B$, which is $\log y \ge 1 - 1/y$. Applied to the states other than the vacuum, whose π-mass is at most $1 - \pi(0)$, it gives

$$D_{2n}(0) \ge g \log \frac{g}{\pi(0)} + (1-g) \log \frac{1-g}{1-\pi(0)} \ge 2\big(g - \pi(0)\big)^2 = 2e_n^2.$$

The second inequality is the two-point case of Pinsker's inequality: for fixed $b \in (0,1)$ the function $\psi(a) = a \log \frac{a}{b} + (1-a)\log \frac{1-a}{1-b} - 2(a-b)^2$ has $\psi(b) = \psi'(b) = 0$ and $\psi''(a) = 1/a + 1/(1-a) - 4 \ge 0$.

Upper bound. Step 1 of the proof of Theorem 26.28 and $\log y \le y - 1$ give $D_n(0) \le P^{2n}(0,0)/\pi(0) - 1 = \varphi^9\big(e_n + \pi(0)\big) - 1 = \varphi^9 e_n$.

The pace. Write $u_N = D_N(0)$. By the upper bound and Lemma 26.35(c), $u_N \le \varphi^9 e_0\, r(P)^N$, so $\limsup u_N^{1/N} \le r(P)$. For $a < r(P)$ choose m with $q = e_{m+1}/e_m > a$. For $k \ge m$, $u_{2k} \ge 2e_k^2 \ge 2(e_m/q^m)^2 q^{2k}$. The relative entropy never rises from one tick to the next (apply Proposition 26.26 and Theorem 26.23(a) on a ball large enough), so $u_{2k+1} \ge u_{2k+2}$. Hence $u_N \ge c\,q^N$ for all $N \ge 2m$ with $c = 2(e_m/q^m)^2 q$, and $\liminf u_N^{1/N} \ge q > a$. The last statement is the lower bound with Lemma 26.35(c). □

The bound $2e_n^2 \le D_{2n}(0)$ uses the sharp constant 2 in Pinsker's inequality [188], due to Csiszár [172]. For Recognition Science's tick the bracket becomes a numerical interval. One part of it rests on a named hypothesis about a finite computation.

The return-probability hypothesis. We name one statement and use it only where it is named:

(R) a specified interval computation of the vacuum return probabilities of Recognition Science's tick returns, at tick 800, an upper end at most $h_{800} = 60670985342324649/2^{62}$ and, at tick 802, a lower end at least $\ell_{802} = 60670956941227723/2^{62}$.

The computation's enclosures contain the true values, so (R) gives $P^{800}(0,0) \le h_{800}$ and $\ell_{802} \le P^{802}(0,0)$, and these two inequalities are all we use. Statement (R) concerns the output of a finite computation; we state it as a hypothesis and do not prove it.

Corollary 26.37 (The pace of Recognition Science's tick)**.** *Let P be Recognition Science's tick and $r = r(P)$.*

(a) With (T1), (T2) and (T4), $0.879^2 \le r \le (4961/5000)^2$, that is $r \in [0.7726, 0.98447]$; in particular $r < 1$. Moreover $D_{2n}(0) \ge 2 \cdot 0.032^2 \cdot (0.879^2)^{2(n-5)}$ for every $n \ge 5$.

(b) Under hypothesis (R), $r \ge 0.9898^2 = 0.97970$, and for every $n \ge 400$

$$D_{2n}(0) \ge 2\,(1.525 \cdot 10^{-7})^2\,(0.9898^2)^{2(n-400)}.$$

(c) For every start a and every N, the entropy produced at all later ticks adds up to the relative entropy still present: $\sum_{i\ge 0}\big(D_{N+i}(a) - D_{N+i+1}(a)\big) = D_N(a)$, and with (T1), (T2) this is at most $\log\big(1 + (4961/5000)^{2N}(\varphi^{9+|a|_1} - 1)\big)$.

Proof. (a) The upper bound is Lemma 26.35(d) with $h = 0.135$ and $\gamma = 39/5000$. By (T4) and Lemma 26.35(b), $r \ge e_5/e_4 \ge 0.879^2$, and $e_6/e_5 \ge 0.879^2$ as well. Since $\pi(0) = 17\sqrt{5} - 38 < 0.013157$, (T4) gives $e_5 > 0.045166 - 0.013157 > 0.032$, so $e_n \ge 0.032\,(0.879^2)^{n-5}$ for $n \ge 5$, and Theorem 26.36 gives the floor.

(b) Hypothesis (R) gives $e_{400} \le h_{800} - \pi(0) < 3.0533 \cdot 10^{-7}$ and $e_{401} \ge \ell_{802} - \pi(0) > 2.9916 \cdot 10^{-7}$, so $e_{401}/e_{400} > 0.97979 > 0.9898^2$, and r is at least this ratio. The ratios are at most 1, so $e_{400} \ge e_{401} > 1.525 \cdot 10^{-7}$, and the floor follows as in (a).

(c) The partial sums telescope to $D_N(a) - D_{N+m}(a)$, and $D_{N+m}(a) \to 0$ by Corollary 26.29(a). □

So, given (T1), (T2) and (T4), the proved gap rate $(4961/5000)^2 = 0.9844608\ldots$ of Corollary 26.29 is the true pace to within a factor of 1.28, and to within half a percent under hypothesis (R) as well. A spine of independent voxels relaxes at the same pace, because relative entropy adds.

Theorem 26.38 (Relative entropy adds over independent voxels)**.** *Let V be a finite set of voxels and, for each v, let G_v be a law on a finite set of states. Then*

$$D\Big(\bigotimes_v G_v \,\Big\|\, \bigotimes_v \pi\Big) = \sum_v D(G_v \,\|\, \pi).$$

In particular, k voxels moving independently by a voxel tick P from starts a_v have joint relative entropy $\sum_v D_n(a_v)$ after n ticks, and from the empty record $k\,D_n(0)$. For Recognition Science's tick, with (T1)–(T4), for $N \geq 1$ and $n \geq 5$,

$$k\,D_N(0) \leq k \log\big(1 + 15.26 \cdot (4961/5000)^{2(N-1)}\big), \qquad k\,D_{2n}(0) \geq k \cdot 2 \cdot 0.032^2 \cdot (0.879^2)^{2(n-5)}.$$

Proof. At a joint state $p = (p_v)$ with $\prod_v G_v(p_v) > 0$,

$$\log \frac{\prod_v G_v(p_v)}{\prod_v \pi(p_v)} = \sum_v \log \frac{G_v(p_v)}{\pi(p_v)};$$

where some factor vanishes both sides of the term-by-term identity are 0. Summing the vth term against the product law leaves G_v, because every other factor sums to 1. The bounds are Corollary 26.29(b) and Corollary 26.37(a). □

Table 26.2 turns these bounds, and the second bound of Corollary 26.29(b), into windows for spines of Recognition Science's tick, proved from (T1)–(T4). A spine of k voxels is below 0.01 nats at every tick from the second column on: since $\log(1 + x) \leq x$, it suffices that $15.26 \cdot (4961/5000)^{2(N-1)} \leq 0.01/k$, the first bound of Corollary 26.29(b), and for $k = 10^6$ this holds once $N - 1 \geq \log(1.526 \cdot 10^9)/0.015661 \approx 1350.2$. The second bound of Corollary 26.29(b) gives the earlier ticks of the third column: it suffices that $3.151 \cdot (4961/5000)^{2(N-4)} \leq 0.01/k$, and for $k = 10^6$ this holds once $N - 4 \geq \log(3.151 \cdot 10^8)/0.015661 \approx 1249.5$. The spine is above 0.01 nats at every tick up to the fourth column: the floor of the theorem holds at the even tick $2n$, and $D_N(0) \geq D_{2n}(0)$ for $N \leq 2n$ because relative entropy never rises. Under hypothesis (R) the floor of Corollary 26.37(b) gives the fifth column; for $k = 10^6$ it adds nothing, since $10^6 \cdot 2\,(1.525 \cdot 10^{-7})^2 < 0.01$.

The second and third columns grow by about 63.85 ticks per nat of $\log k$, the fourth by about 3.9, and the fifth by about 48.8. These are the upper bound and the two lower bounds on $1/(-\log r)$ that Corollary 26.37 gives for the exact rate r, and Theorem 26.42 below shows that the true settling time grows by exactly $1/(-\log r)$ ticks per nat. The same pace governs the distance of the joint law from equilibrium in ℓ^1.

Table 26.2. Windows for a spine of k independent voxels of Recognition Science's tick from the empty record, proved from (T1)–(T4): the relative entropy $k\,D_N(0)$ is at most 0.01 nats at every tick N from the second column on, by the first bound of Corollary 26.29(b), and from the third column on, by its second bound; it is more than 0.01 nats at every tick up to the fourth column (fifth column: the same under the return-probability hypothesis (R)).

	below 0.01 from tick		above 0.01 through tick	
voxels k	first bound	second bound	proved	under (R)
10^{6}	1352	1254	56	56
10^{12}	2234	2136	110	874
10^{24}	3998	3900	218	2222
10^{80}	12232	12134	716	8510

Proposition 26.39 (The spine in ℓ^1)**.** *Let P be a voxel tick with every hold at least h and Poincaré gap γ, with $2h-1<1-\gamma$ and $(1-\gamma)+(2h-1)\geq 0$. Let voxels $v=1,\dots,k$ start from states s_v and move independently for m ticks, with joint law $Q(p)=\prod_v P^m(s_v,p_v)$, and put $\Pi(p)=\prod_v \pi(p_v)$. Then for every finite set B of joint states,*

$$\sum_{p\in B}\lvert Q(p)-\Pi(p)\rvert \leq \Big(\prod_{v=1}^{k}\big(1+(1-\gamma)^{2m}(\varphi^{9+\lvert s_v\rvert_1}-1)\big)-1\Big)^{1/2}.$$

Proof. Take $L\geq m+\max_v\lvert s_v\rvert_1$ so large that the box $\mathscr{B}=(B_L)^k$ contains B; it contains the support of Q. Write $E_m(a)=P^{2m}(a,a)/\pi(a)$. By Step 1 of the proof of Theorem 26.28, $\sum_{s\in B_L}P^m(a,s)^2/\pi(s)=E_m(a)$, and the sum over a product box of a product of functions is the product of the sums, so $\sum_{p\in\mathscr{B}}Q(p)^2/\Pi(p)=\prod_v E_m(s_v)$. Since Q sums to 1 on $\mathscr{B}$ and Π to at most 1,

$$\sum_{p\in\mathscr{B}}\frac{(Q(p)-\Pi(p))^2}{\Pi(p)}=\prod_v E_m(s_v)-2+\Pi(\mathscr{B})\leq\prod_v E_m(s_v)-1.$$

By the Cauchy–Schwarz inequality with the weights Π,

$$\sum_{p\in B}\lvert Q-\Pi\rvert=\sum_{p\in B}\frac{\lvert Q-\Pi\rvert}{\sqrt{\Pi}}\sqrt{\Pi}\leq\Big(\sum_{p\in B}\frac{(Q-\Pi)^2}{\Pi}\Big)^{1/2}\Big(\sum_{p\in B}\Pi\Big)^{1/2}\leq\Big(\prod_v E_m(s_v)-1\Big)^{1/2},$$

and Step 3 of the proof of Theorem 26.28 gives $E_m(a)\leq 1+(1-\gamma)^{2m}(\varphi^{9+\lvert a\rvert_1}-1)$. □

For k voxels of Recognition Science's tick from the empty record, with (T1) and (T2), the bound is at most $\big(\exp(75.02\,k\,(4961/5000)^{2m})-1\big)^{1/2}$, since $1+x\leq e^x$ and $\varphi^9-1<75.02$. It is below ε once $m\geq\big(\log k+\log(75.02/\log(1+\varepsilon^2))\big)/0.015661$, so in ℓ^1 as well the spine settles within 63.86 ticks per nat of $\log k$.

Independent voxels are the simplest joint law. Recognition Science also allows couplings, and they change the answer.

Definition 26.40 (Joint ticks, the target, and total correlation)**.** A *joint tick* of a spine V of k voxels assigns to each joint configuration reached from the empty record a probability law on joint moves, in which each voxel holds or makes one unit post, with the bookkeeping of Recognition Science's ledger: each post enters the journal as a reciprocal pair. It *meets Recognition Science's target* when, at every reached configuration, each voxel's own move law is the row of Recognition Science's tick at that voxel's imbalance. We write P_n for the law of the joint imbalance $p \in (\mathbb{Z}^3)^V$ after n ticks. The *product law* moves each voxel by its own row, independently. The *mirror law* of two voxels moves voxel 0 by its row and makes voxel 1 hold when voxel 0 holds and post the reverse move when voxel 0 posts. For a joint law P with marginals G_v, the *total correlation* is $\mathrm{TC}(\mathsf{P}) = \sum_p \mathsf{P}(p) \log\bigl(\mathsf{P}(p)/\prod_v G_v(p_v)\bigr)$.

Theorem 26.41 (Couplings add their total correlation)**.** *Let a joint tick of k voxels meet Recognition Science's target, and let n be a number of ticks.*

(a) Each voxel's marginal of P_n is $P^n(0,\cdot)$, and $\mathsf{P}_n(p) \le P^n(0, p_v)$ for every v.

(b) $D(\mathsf{P}_n \,\|\, \pi^{\otimes V}) = k\,D_n(0) + \mathrm{TC}(\mathsf{P}_n)$ and $\mathrm{TC}(\mathsf{P}_n) \ge 0$. So the joint relative entropy is at least $k\,D_n(0)$, and the product law attains this floor exactly.

(c) $D(\mathsf{P}_n \,\|\, \pi^{\otimes V}) \le C\,k\,D_n(0)$ exactly when $\mathrm{TC}(\mathsf{P}_n) \le (C-1)\,k\,D_n(0)$.

(d) The mirror law meets the target, carries all its mass on pairs $(s,-s)$, and has $D(\mathsf{P}_n \,\|\, \pi\otimes\pi) \ge 9\log\varphi \approx 4.331$ at every tick, while with (T1) and (T2) $2D_n(0) \to 0$. So for every constant C some tick n has $D(\mathsf{P}_n \,\|\, \pi\otimes\pi) > C \cdot 2D_n(0)$: not every joint tick that meets the target has an upper rate $C\,k\,D_n(0)$.

Proof. (a) By induction on n: at each reached configuration, voxel v moves by the row of the tick at its own imbalance, so its marginal is carried forward by P. A joint probability is at most each marginal.

(b) Where $\mathsf{P}_n(p) > 0$, every $G_v(p_v) \ge \mathsf{P}_n(p) > 0$, and $\log\frac{\mathsf{P}_n}{\prod\pi} = \log\frac{\mathsf{P}_n}{\prod G_v} + \sum_v \log\frac{G_v}{\pi}$. Summing against P_n and using (a) for each v gives the identity. For $\mathrm{TC} \ge 0$, $\log y \ge 1 - 1/y$ gives $\mathsf{P}_n \log(\mathsf{P}_n/\prod G_v) \ge \mathsf{P}_n - \prod_v G_v(p_v)$, and summing, $\mathrm{TC} \ge 1 - \sum_p \prod_v G_v(p_v) \ge 0$. Under the product law $\mathsf{P}_n = \prod_v G_v$, so $\mathrm{TC} = 0$.

(c) is (b) rearranged.

(d) Reflection $s \mapsto -s$ maps Recognition Science's tick to a voxel tick with positive holds and hold-return balance, hence to itself, since that tick is unique (Theorem 26.21); so the row at $-s$ is the reflected row at s, and voxel 1, which mirrors voxel 0, moves by its own row. Every reached configuration has $p_1 = -p_0$. A law carried by $T = \{(s,-s)\}$ has relative entropy at least $-\log(\pi\otimes\pi)(T)$, by part (i) of the proof of Theorem 26.33, and $(\pi\otimes\pi)(T) = \sum_s \pi(s)^2 \le \pi(0)\sum_s \pi(s) = \varphi^{-9}$, since $\pi(-s) = \pi(s) \le \pi(0)$. Finally $D_n(0) \to 0$ by Corollary 26.29(a), so for every C some n has $2C\,D_n(0) < 9\log\varphi$. □

Conditions and scope.

Base and premises

Recognition Science's tick, a spine of k voxels started from the empty record, and a joint tick that meets the target; (T1) and (T2) for (d).

Conclusion

The independent product law is the fastest joint law that meets the target; a coupling settles exactly as fast as its total correlation allows. The mirror law never settles to the product law: it stays at least $9\log\varphi$ from it, while it settles to its own anti-diagonal law (Proposition 26.43).

Identification

That the product law is Recognition Science's joint law rests on its joint-entropy premise. On two voxels, cross-voxel double entry together with the target forces the joint law of the mirror law (Proposition 26.43); which coupling nature uses is not claimed.

Total correlation is Watanabe's measure of multivariate dependence [189]. For independent voxels the spine settles at a sharp time. The phenomenon, a cutoff, was found by Aldous and Diaconis for card shuffling [175, 178].

Theorem 26.42 (The spine's sharp cutoff)**.** *Let P be a voxel tick with exact rate $r = r(P) < 1$, put $t^*(k) = \log k/(-\log r)$, and consider k independent voxels started from the empty record, with relative entropy $k\,D_N(0)$ after N ticks.*

(a) For every $\varepsilon > 0$, $k\,D_N(0) \le \varepsilon$ at every tick $N \ge t^*(k) + \log(\varphi^9/\varepsilon)/(-\log r)$.

(b) For every $\delta \in (0,1)$ and every level M there is K_0 such that $k\,D_N(0) \ge M$ for all $k \ge K_0$ and every tick $N \le (1-\delta)\,t^*(k)$.

(c) The settling time $T_\varepsilon(k) = \min\{N : k\,D_N(0) \le \varepsilon\}$ satisfies $T_\varepsilon(k)/\log k \to 1/(-\log r)$ as $k \to \infty$.

For Recognition Science's tick, with (T1), (T2) and (T4), $r < 1$ and $1/(-\log r)$ lies between 3.87 and 63.86 ticks per nat of $\log k$, and between 48.76 and 63.86 under hypothesis (R); for $\varepsilon = 0.01$ the window in (a) beyond $t^*(k)$ is at most 571 ticks.

Proof. (a) By Theorem 26.36 and Lemma 26.35(c), $D_N(0) \le \varphi^9 e_0\, r^N \le \varphi^9 r^N$, since $e_0 = 1-\pi(0) \le 1$. And $k\varphi^9 r^N \le \varepsilon$ exactly when $\log k + \log(\varphi^9/\varepsilon) \le N(-\log r)$.

(b) Put $a = r^{(1-\delta/2)/(1-\delta)}$, so $a < r$, since the exponent exceeds 1. The proof of Theorem 26.36 gives $c > 0$ and $q \in (a, 1]$ with $D_N(0) \ge c\,q^N$ for all large N, and, shrinking c, for all N, because every $D_N(0) \ge D_{2N}(0) \ge 2e_N^2 > 0$. For $N \le (1-\delta)t^*(k)$,

$$q^N \ge a^N \ge a^{(1-\delta)t^*(k)} = r^{(1-\delta/2)t^*(k)} = k^{-(1-\delta/2)},$$

because $r^{t^*(k)} = 1/k$. So $k\,D_N(0) \ge c\,k^{\delta/2}$, which exceeds M for all large k.

(c) By (a), $T_\varepsilon(k) \le t^*(k) + |\log(\varphi^9/\varepsilon)|/(-\log r) + 1$. By (b) with the level $M = 2\varepsilon$, $T_\varepsilon(k) > (1-\delta)t^*(k)$ for all large k. Divide by $\log k$ and let $\delta \to 0$.

For Recognition Science's tick, $0.879^2 \le r \le (4961/5000)^2$ by Corollary 26.37(a), so $-\log r \ge 0.015661$, giving $1/(-\log r) \le 63.86$ and a window of at most $\log(7602)/0.015661 \le 571$ ticks; and $1/(-\log 0.879^2) = 3.877$, $1/(-\log 0.9898^2) = 48.77$. □

Conditions and scope.

Base and premises

A voxel tick with exact rate $r < 1$, and k independent voxels from the empty record.

Conclusion

Settling at $\log k/(-\log r)$ ticks, with a window of bounded width above and a window of width $o(\log k)$ below.

Not claimed

A window of bounded width below $t^*(k)$; that would need a pure geometric lower bound $e_n \ge c r^n$ at the exact pace, which is open. A tick is not identified with a unit of physical time.

Theorem 26.41 leaves the coupling open. One coupling comes from the ledger's own double entry, read across voxels: a post at one voxel is matched, in the same tick, by the reverse post at another. A joint tick is *paired*, or keeps *cross-voxel double entry*, when at every reached configuration each joint move of positive probability matches every posting voxel with another voxel, matched back to it, that posts the reverse move: the same account, the opposite side. On two voxels a paired move is a joint hold, or a post of voxel 0 together with the reverse post of voxel 1; so the mirror law is paired. On two voxels double entry, together with the target, fixes the joint law completely, and we can compute everything about it.

Write $G_n = P^n(0,\cdot)$ for one voxel's law after n ticks, $m_n = \sum_s G_n(s)\,|s|_1$ for its mean imbalance cost, and $H(G) = -\sum_{G(s)>0} G(s)\log G(s)$ for the Shannon entropy of a law. On pairs of states (s,t) let Π_A and Π_c vanish off the anti-diagonal $t=-s$, and put

$$\Pi_A(s,-s) = \pi(s), \qquad \Pi_c(s,-s) = \frac{\pi(s)\,\pi(-s)}{Z}, \qquad Z = \sum_s \pi(s)\,\pi(-s).$$

Π_A is the *anti-diagonal T9 law*, and Π_c is the product law $\pi\otimes\pi$ conditioned on a zero total imbalance.

Proposition 26.43 (Double entry fixes the pair's law)**.** *The following is proved in classical mathematics. Let P be Recognition Science's tick with exact rate $r = r(P)$, and let a joint tick of two voxels meet the target and be paired, with joint law P_n after n ticks from the empty record.*

(a) For every n, $\mathsf{P}_n(s,t) = G_n(s)$ if $t=-s$ and 0 otherwise: voxel 1's imbalance is minus voxel 0's, and the pair's law is one voxel's law placed on the anti-diagonal. Its relative entropy to its own equilibrium is one voxel's: $D(\mathsf{P}_n\,\|\,\Pi_A) = D_n(0)$.

(b) $Z = (1+\varphi^{-2})^3/\varphi^{15}$, and for every n

$$D(\mathsf{P}_n\,\|\,\Pi_c) = D_n(0) + \log Z + (9+m_n)\log\varphi.$$

(c) $\mathrm{TC}(\mathsf{P}_n) = H(G_n) = (9+m_n)\log\varphi - D_n(0)$*, and for every* n

$$\big|15\log\varphi - \mathrm{TC}(\mathsf{P}_n)\big| \le C r^{n/2}, \qquad C = \log\varphi\sqrt{\varphi^9-1}\Big(\sum_s \pi(s)\,|s|_1^2\Big)^{1/2} + \varphi^9 - 1.$$

(d) For every n, $D(\mathsf{P}_n \,\|\, \pi\otimes\pi) = D(\mathsf{P}_n \,\|\, \Pi_c) - \log Z$*, so*

$$D(\mathsf{P}_n \,\|\, \pi\otimes\pi) \ge 15\log\varphi - 3\log(1+\varphi^{-2}) \approx 6.248.$$

(e) With (T1) and (T2), as $n\to\infty$*:* $D(\mathsf{P}_n \,\|\, \Pi_A)$ *never rises and tends to* 0*;* $D(\mathsf{P}_n \,\|\, \Pi_c)$ *tends to* $3\log(1+\varphi^{-2}) \approx 0.9705$*, not to* 0*; and* $\mathrm{TC}(\mathsf{P}_n)$ *and* $D(\mathsf{P}_n \,\|\, \pi\otimes\pi)$ *tend to* $15\log\varphi \approx 7.218$*, the entropy of the T9 law.*

(f) Let k *disjoint pairs of voxels move independently, each by the mirror law, with joint law* Q_n *on the* $2k$ *voxels. Then* $D(\mathsf{Q}_n \,\|\, \Pi_A^{\otimes k}) = k\,D_n(0)$*, the relative entropy of* k *independent voxels, so with (T1) and (T2) parts (a) and (b) of Theorem 26.42 hold for it verbatim; and* $D(\mathsf{Q}_n \,\|\, \pi^{\otimes 2k}) = k\,D(\mathsf{P}_n \,\|\, \pi\otimes\pi) \ge k\big(15\log\varphi - 3\log(1+\varphi^{-2})\big)$ *at every tick, where* P_n *is the mirror law's joint law.*

Proof. (a) A paired move of two voxels adds m to one imbalance and $-m$ to the other, or holds both, so it keeps the total imbalance p_0+p_1. The empty record has total 0, so every reached configuration has $p_1 = -p_0$, and $\mathsf{P}_n(s,t) = 0$ unless $t=-s$. By Theorem 26.41(a) voxel 0's marginal is G_n, so $G_n(s) = \sum_t \mathsf{P}_n(s,t) = \mathsf{P}_n(s,-s)$. Then $D(\mathsf{P}_n \,\|\, \Pi_A) = \sum_s G_n(s)\log\big(G_n(s)/\pi(s)\big) = D_n(0)$.

(b) The T9 law is a product over the three axes, $\pi(s) = \prod_i \rho^{|s_i|}/\varphi^3$, because $\sum_{j\in\mathbb{Z}} \rho^{|j|} = \varphi^3$ (the computation after Definition 26.19). So $\pi(-s) = \pi(s)$ and

$$Z = \sum_s \pi(s)^2 = \frac{1}{\varphi^{18}}\Big(\sum_{j\in\mathbb{Z}} \rho^{2|j|}\Big)^3 = \frac{1}{\varphi^{18}}\Big(\frac{1+\rho^2}{1-\rho^2}\Big)^3.$$

Since $\rho^2+\rho=1$, we have $1-\rho^2=\rho$, and $(1+\rho^2)/\rho = \varphi(1+\varphi^{-2})$; so $Z = \varphi^3(1+\varphi^{-2})^3/\varphi^{18} = (1+\varphi^{-2})^3/\varphi^{15}$. By (a),

$$D(\mathsf{P}_n \,\|\, \Pi_c) = \sum_s G_n(s)\log\frac{G_n(s)\,Z}{\pi(s)^2} = D_n(0) + \log Z - \sum_s G_n(s)\log\pi(s),$$

and $-\log\pi(s) = (9+|s|_1)\log\varphi$; summing against G_n, which has total mass 1, gives $(9+m_n)\log\varphi$.

(c) By (a), voxel 1's marginal at $-s$ is $\sum_{s'} \mathsf{P}_n(s',-s) = \mathsf{P}_n(s,-s) = G_n(s)$. So the term of $\mathrm{TC}(\mathsf{P}_n)$ at $(s,-s)$ is $G_n(s)\log\big(G_n(s)/(G_n(s)\,G_n(s))\big) = -G_n(s)\log G_n(s)$, and $\mathrm{TC}(\mathsf{P}_n) = H(G_n)$. The last computation of (b), with π in place of π^2/Z, gives $D_n(0) = -H(G_n) + (9+m_n)\log\varphi$.

For the bound we need the mean imbalance cost of the T9 law. Along one axis $\sum_{j\in\mathbb{Z}} |j|\rho^{|j|} = 2\rho/(1-\rho)^2$, and dividing by $\varphi^3 = (1+\rho)/(1-\rho)$ gives $2\rho/(1-\rho^2) = 2$; so $\sum_s \pi(s)|s|_1 = 6$, and the entropy of the T9 law is $\sum_s \pi(s)(9+|s|_1)\log\varphi = 15\log\varphi$. Hence

$$15\log\varphi - H(G_n) = (6-m_n)\log\varphi + D_n(0).$$

Since G_n vanishes outside B_n, $6-m_n = \sum_s \big(\pi(s)-G_n(s)\big)|s|_1$, and the Cauchy–Schwarz inequality with the weights π gives

$$|6-m_n| \le \Big(\sum_s \frac{(G_n(s)-\pi(s))^2}{\pi(s)}\Big)^{1/2}\Big(\sum_s \pi(s)|s|_1^2\Big)^{1/2} = \big(\varphi^9 e_n\big)^{1/2}\Big(\sum_s \pi(s)|s|_1^2\Big)^{1/2}$$

by Lemma 26.35(a) and $\pi(0) = \varphi^{-9}$. The last sum is finite: it has at most $4N^2 + 2$ states with $|s|_1 = N$, each of weight ρ^N/φ^9. By Lemma 26.35(c), $\varphi^9 e_n \le \varphi^9 e_0\, r^n = (\varphi^9 - 1)\, r^n$, since $e_0 = 1 - \pi(0)$; and $D_n(0) \le \varphi^9 e_n$ by Theorem 26.36. As $r \le 1$, $r^n \le r^{n/2}$, and the two bounds add up to $C r^{n/2}$.

(d) By Theorem 26.41(b) and part (c), $D(\mathsf{P}_n \,\|\, \pi \otimes \pi) = 2D_n(0) + H(G_n) = D_n(0) + (9 + m_n)\log\varphi$, which is $D(\mathsf{P}_n \,\|\, \Pi_c) - \log Z$ by (b). The masses of Π_c sum to 1, so $D(\mathsf{P}_n \,\|\, \Pi_c) \ge 0$: the inequality $\log y \ge 1 - 1/y$ gives $D(\mathsf{P}_n \,\|\, \Pi_c) \ge \sum_{\mathsf{P}_n > 0} (\mathsf{P}_n - \Pi_c) \ge 0$. Finally $-\log Z = 15\log\varphi - 3\log(1+\varphi^{-2})$ by (b).

(e) With (T1) and (T2), Lemma 26.35(d) gives $r \le (4961/5000)^2 < 1$, as in Corollary 26.37(a). So $D_n(0) \le (\varphi^9 - 1) r^n \to 0$, and $D_n(0)$ never rises from one tick to the next (the proof of Theorem 26.36); by (a) this is $D(\mathsf{P}_n \,\|\, \Pi_A)$. The bound in (c) gives $m_n \to 6$. Then (b) gives $D(\mathsf{P}_n \,\|\, \Pi_c) \to \log Z + 15\log\varphi = 3\log(1 + \varphi^{-2}) > 0$, and (c) and (d) give the limit $15\log\varphi$ of $\mathrm{TC}(\mathsf{P}_n)$ and of $D(\mathsf{P}_n \,\|\, \pi \otimes \pi)$.

(f) The mirror law meets the target (Theorem 26.41(d)) and is paired, so (a)–(e) apply to it. The k pairs move independently, so Q_n is the product over the pairs of P_n, and relative entropy adds over the factors of a product law, by the computation in the proof of Theorem 26.38, against $\Pi_A^{\otimes k}$ and against $\pi^{\otimes 2k}$ alike. So $D(\mathsf{Q}_n \,\|\, \Pi_A^{\otimes k}) = k\, D(\mathsf{P}_n \,\|\, \Pi_A) = k\, D_n(0)$ by (a), and $D(\mathsf{Q}_n \,\|\, \pi^{\otimes 2k}) = k\, D(\mathsf{P}_n \,\|\, \pi \otimes \pi)$, which (d) bounds below. The proof of Theorem 26.42(a) and (b) uses only the quantity $k\, D_N(0)$ and $r < 1$. □

Conditions and scope.

Base and premises

Recognition Science's tick; two voxels started from the empty record; a joint tick that meets the target and is paired at every reached configuration; (T1) and (T2) for the limits in (e) and the cutoff in (f).

Conclusion

Double entry and the target fix the pair's joint law: one voxel's law on the anti-diagonal. The pair settles to its own equilibrium at exactly one voxel's pace, while the information the coupling puts between the two voxels is 0 at the empty record and tends to the T9 law's entropy $15\log\varphi$, within $C r^{n/2}$ at tick n; and the pair stays at least 6.247 nats from the product T9 law at every tick. This sharpens the floor $9\log\varphi \approx 4.331$ of Theorem 26.41(d).

Identification

Cross-voxel double entry is a hypothesis of the proposition, not identified with a physical process. A tick is not identified with a unit of physical time.

Not claimed

That double entry fixes the joint law of more than two voxels; the smallest value of $D(\mathsf{P}_n \,\|\, \pi \otimes \pi)$ over the ticks, of which $-\log Z$ is only a lower bound; and that the exponent $r^{1/2}$ in (c) is sharp, which Proposition 26.44 reduces to two unproved statements about one voxel's walk.

Part (c) of Proposition 26.43 bounds the approach of the total correlation to its limit by $C r^{n/2}$. Is $r^{1/2}$ the true rate? By part (c) the gap is

$$g_n = 15\log\varphi - \mathrm{TC}(\mathsf{P}_n) = (6 - m_n)\log\varphi + D_n(0),$$

and $D_n(0)$ falls at the full pace r (Theorem 26.36), so the answer turns on how fast one voxel's mean imbalance cost m_n approaches the T9 value 6. Write $G_n(B_j) = \sum_{s\in B_j} G_n(s)$ for the mass of a cost ball after n ticks, and $\pi(B_j)$ for its equilibrium mass. We name two statements about one voxel's walk from the empty record:

(B) *ball dominance*: $G_n(B_j) \geq \pi(B_j)$ for every tick n and every radius j, that is, at every tick each cost ball keeps at least its equilibrium mass;

(M) *monotone mean*: $m_n \leq m_{n+1}$ for every n, that is, the mean imbalance never falls.

Neither statement is proved. A floating-point computation of the walk on the ball B_{80} finds (M) at every tick up to 1200, and (B) at every tick up to 1200 for every radius up to 65. This is evidence, not proof, and nothing below uses it.

Proposition 26.44 (The coupling's sharp rate, reduced to two statements)**.** *The following is proved in classical mathematics. Let P be Recognition Science's tick with exact rate $r = r(P)$.*

(a) For all $n \leq N$,

$$6 - m_n \geq \sum_{j=0}^{N-1} \big(G_n(B_j) - \pi(B_j)\big).$$

(b) If (B) holds, then $G_n(0) - \pi(0) \leq 6 - m_n$ for every $n \geq 1$.

(c) If (B) and (M) hold, then for every q with $0 < q < r^{1/2}$ there are $c > 0$ and n_0 with $c\,q^n \leq 6 - m_n$ for every $n \geq n_0$; and for every joint tick of two voxels that meets the target and is paired,

$$\big|15 \log \varphi - \mathrm{TC}(\mathsf{P}_n)\big|^{1/n} \longrightarrow r^{1/2} \qquad (n \to \infty).$$

(d) With (T5), $P^2(0,0) > P(0,0)$. So the statement that no ball ever gains mass, $G_{n+1}(B_j) \leq G_n(B_j)$ for every n and j, is false: the ball $B_0 = \{0\}$ gains mass from tick 1 to tick 2.

Proof. (a) A state s of B_N has $|s|_1$ equal to the number of $j < N$ with $j < |s|_1$, so for every law f on states

$$\sum_{s\in B_N} f(s)\,|s|_1 = \sum_{j=0}^{N-1} \big(f(B_N) - f(B_j)\big).$$

G_n lives on $B_n \subseteq B_N$, so $m_n = \sum_{j<N}\big(1 - G_n(B_j)\big)$. Every state outside B_N has $|s|_1 > N$, so, with the T9 mean 6 from the proof of Proposition 26.43(c) and the identity for $f = \pi$,

$$6 = \sum_s \pi(s)\,|s|_1 \geq \sum_{s\in B_N} \pi(s)\,|s|_1 + N\big(1 - \pi(B_N)\big) = \sum_{j<N}\big(1 - \pi(B_j)\big).$$

Subtracting the expression for m_n gives (a).

(b) Under (B) every term of the sum in (a) is nonnegative. Take $N = n \geq 1$ and keep the term $j = 0$, which is $G_n(0) - \pi(0)$.

(c) Since $q^2 < r$ and the ratios e_{k+1}/e_k increase to r (Lemma 26.35(b)), some ℓ has $\theta = e_{\ell+1}/e_\ell > q^2$. By Lemma 26.35(c), for every $k \geq \ell$, $e_k \geq e_\ell\,\theta^{k-\ell} \geq c'q^{2k}$ with $c' = e_\ell\,\theta^{-\ell} > 0$. Let $n \geq 2\ell + 2$. At an even tick $n = 2k$, part (b) gives $6 - m_{2k} \geq G_{2k}(0) - \pi(0) = e_k \geq c'q^{2k}$. At an odd tick $n = 2k + 1$, (M) gives $m_{2k+1} \leq m_{2k+2}$, so $6 - m_{2k+1} \geq e_{k+1} \geq c'q \cdot q^{2k+1}$. As $q < 1$, the constant $c = c'q$ serves at every tick $n \geq n_0 = 2\ell + 2$.

For the second claim, Proposition 26.43(c) gives $g_n = (6 - m_n)\log\varphi + D_n(0)$ and $|g_n| \leq C(r^{1/2})^n$, so $\limsup_n |g_n|^{1/n} \leq r^{1/2}$. Since $D_n(0) \geq 0$, the floor gives $g_n \geq c\log\varphi \cdot q^n > 0$ for $n \geq n_0$, so $\liminf_n |g_n|^{1/n} \geq q$. This holds for every q in $(0, r^{1/2})$, which is not empty because $r > 0$ (Lemma 26.35(b)).

(d) Let $h = P(0,0)$. Then $G_1(0) = h$ and $G_2(0) = h^2 + \sum_m P(0,m)P(m,0)$, the sum over the six moves. The six posts at the vacuum are equal, so each is $(1-h)/6$, and detailed balance $w(0)P(0,m) = w(m)P(m,0)$, with $w(0) = 1$ and $w(m) = \rho$, gives $P(m,0) = \varphi(1-h)/6$. Hence

$$G_2(0) - G_1(0) = h^2 + \frac{\varphi(1-h)^2}{6} - h = \frac{(1-h)\big(\varphi - (6+\varphi)h\big)}{6}.$$

By (T5), $h < 0.19 < 1.6/7.6 < \varphi/(6+\varphi)$, so both factors are positive. □

Part (d) closes the most natural route to (B) and (M). If no ball ever gained mass, (M) would follow from the identity $m_n = \sum_{j<N}\big(1 - G_n(B_j)\big)$ in the proof of (a), and (B) as well, since $G_n(B_j) \to \pi(B_j)$ by Pinsker's inequality once $D_n(0) \to 0$ (Corollary 26.29(a), with (T1)–(T3)). But the vacuum ball already gains mass at the second tick. With (B) and (M), part (c) and Proposition 26.43(c) together say that $r^{1/2}$, which lies between 0.879 and 0.9922 with (T1), (T2) and (T4), is the exact pace at which the coupling's information approaches $15\log\varphi$.

Conditions and scope.

Base and premises

Recognition Science's tick from the empty record; for (b) the statement (B); for (c) the statements (B) and (M), and a joint tick of two voxels that meets the target and is paired; for (d) the stated input (T5).

Conclusion

Under (B) the mean imbalance gap is at least the return excess; under (B) and (M) the information the coupling carries approaches the T9 law's entropy at exactly the square root of one voxel's pace; the stronger statement that no ball ever gains mass is false.

Not claimed

(B) and (M) themselves: both are open, and the computation reported before the proposition is evidence for them, not a proof. A tick is not identified with a unit of physical time.

26.7 Fields on the lattice

The voxels of Recognition Science form a cubic lattice. This section places the fields of classical physics on such a lattice, as functions on sites, edges and faces, and puts Maxwell's and Schrödinger's equations on it with the numbers of $\mathbb{R}_\delta$; identifying these lattice fields with Recognition Science's own fields is not claimed. Everything in it is free; the lattice

identities hold over every commutative ring. Amplitudes are pairs of points, $a = (\operatorname{Re} a, \operatorname{Im} a)$, multiplied as complex numbers, with $e^{i\theta} = (\cos\theta, \sin\theta)$ for the functions of Definition 21.6, and $\operatorname{sz}(a) = |\operatorname{Re} a| + |\operatorname{Im} a|$. A matrix U *is* $\exp(zH)$ when each entry of U is the limit of the corresponding entries of the partial sums $\sum_{k<M} z^k H^k/k!$. The quarter period q is the point with $0 \le q \le 2$ and $\cos q = 0$; by Theorem 21.8, $4q = 2\pi_\delta$.

Definition 26.45 (Lattice fields and the Maxwell generator)**.** Let $\Lambda = (\mathbb{Z}/(N+1))^3$ be the periodic cubic lattice with $N+1$ sites per side and unit vectors e_0, e_1, e_2; axis indices are read modulo 3. Over a commutative ring, a *scalar field* f is a function on sites, an *edge field* A gives $A_i(x)$ on the edge from x to $x + e_i$, and a *face field* B gives $B_k(x)$ on the face at x normal to e_k. Put

$$\begin{aligned}
(\nabla f)_i(x) &= f(x+e_i) - f(x),\\
(\operatorname{curl} A)_k(x) &= \big[A_{k+2}(x+e_{k+1}) - A_{k+2}(x)\big] - \big[A_{k+1}(x+e_{k+2}) - A_{k+1}(x)\big],\\
\operatorname{div}_F B(x) &= \textstyle\sum_i \big[B_i(x+e_i) - B_i(x)\big],\\
(\operatorname{curl}^* B)_i(x) &= \big[B_{i+1}(x-e_{i+2}) - B_{i+1}(x)\big] - \big[B_{i+2}(x-e_{i+1}) - B_{i+2}(x)\big],\\
\operatorname{div}_E E(x) &= \textstyle\sum_i \big[E_i(x-e_i) - E_i(x)\big].
\end{aligned}$$

A *Maxwell field* is a pair $\psi = (E, B)$ of an edge field and a face field with values in $\mathbb{R}_\delta$. The *Maxwell generator* is $\mathfrak{A}\psi = (\operatorname{curl}^* B, -\operatorname{curl} E)$, and the *energy* is $\|\psi\|^2 = \sum_x \sum_i \big(E_i(x)^2 + B_i(x)^2\big)$.

The evolution $\dot E = \operatorname{curl}^* B$, $\dot B = -\operatorname{curl} E$ is the vacuum Maxwell system on Yee's staggered lattice [190], with the electric field on edges and the magnetic field on faces. Summation by parts on the periodic lattice shows that curl^* is the adjoint of curl and div_E that of ∇ for the counting inner product.

Theorem 26.46 (The lattice Maxwell flow is free)**.**

(a) Over every commutative ring, $\operatorname{curl}\nabla f = 0$, $\operatorname{div}_F \operatorname{curl} A = 0$ *and* $\operatorname{div}_E \operatorname{curl}^* B = 0$.

(b) On $\mathbb{R}_\delta$*, for every point* t *there is a map* $U(t) = \exp(t\mathfrak{A})$ *on Maxwell fields, the limit of its exponential series with an explicit modulus, such that* $U(0)$ *is the identity,* $U(t+s) = U(t)U(s)$*,* $\|U(t)\psi\| = \|\psi\|$*, and the Gauss residuals* $\operatorname{div}_E E$ *and* $\operatorname{div}_F B$ *of* $U(t)\psi$ *do not depend on* t*.*

Proof. (a) Shifts along different axes commute. Expanding, $(\operatorname{curl}\nabla f)_k(x)$ is $[f(x+e_{k+1}+e_{k+2}) - f(x+e_{k+1}) - f(x+e_{k+2}) + f(x)]$ minus the same expression with the two shifts in the other order, which is 0; the other two identities expand in the same way.

(b) *The series.* In the basis of unit fields $\mathfrak{A}$ is a finite integer matrix with at most four entries ± 1 in each row, since each component of curl and of curl^* has four terms. So every entry of $\mathfrak{A}^m$ is an integer of absolute value at most 4^m. Let V be a natural number with $|t| \le V$, and let $S_M(t) = \sum_{m<M} t^m \mathfrak{A}^m/m!$. Once $M \ge 8V$ consecutive terms of $\sum_m (4V)^m/m!$ at least halve, so every entry of $S_{M'}(t) - S_M(t)$, for $M' \ge M \ge 8V$, is at most $2(4V)^M/M!$ in absolute value, which tends to 0 with an explicit rate. The diagonal limit (Theorem 12.25) computes each entry of $U(t) = \lim_M S_M(t)$ from a regular sequence naming t, and limits are unique, so $U(t)$ depends only on the point t. Then $U(t)\psi = \sum_b \psi_b U(t) e_b$ is a finite combination over the unit fields e_b, and no regular sequence of ψ is ever chosen. Clearly $U(0)$ is the identity.

The group law. For $|t|, |s| \le V$, expanding $(t+s)^n$ by the binomial theorem gives

$$S_M(t)S_M(s) - S_M(t+s) = \sum_{\substack{a,b<M \\ a+b\ge M}} \frac{t^a s^b}{a!\,b!}\,\mathfrak{A}^{a+b},$$

and since $\sum_{a+b=n} 1/(a!\,b!) = 2^n/n!$, every entry of the right side is at most $\sum_{n\ge M}(8V)^n/n! \le 2(8V)^M/M!$ once $M \ge 16V$. Products and sums of finitely many limits are limits, so $U(t)U(s) = U(t+s)$.

Energy. Summation by parts on the periodic lattice, $\sum_x f(x+e_j)g(x) = \sum_x f(x)g(x-e_j)$, shows that curl^* is the adjoint of curl, so $\langle \mathfrak{A}\phi, \chi\rangle = -\langle \phi, \mathfrak{A}\chi\rangle$ for the counting inner product. Hence $\langle \mathfrak{A}^m\phi, \chi\rangle = (-1)^m\langle \phi, \mathfrak{A}^m\chi\rangle$, and summing the series, $\langle U(t)\phi, \chi\rangle = \langle \phi, U(-t)\chi\rangle$. So $\|U(t)\psi\|^2 = \langle \psi, U(-t)U(t)\psi\rangle = \langle \psi, U(0)\psi\rangle = \|\psi\|^2$.

Gauss. A term of order $m \ge 1$ is $\mathfrak{A}(\mathfrak{A}^{m-1}\psi)$, whose electric part is a curl^* and whose magnetic part is a curl; by (a), div_E and div_F vanish on them. So the Gauss residuals of every partial sum $S_M(t)\psi$ are those of ψ, and they pass to the limit. □

Proposition 26.47 (Gauss's law: sources and prices)**.**

(a) For a static current J, an edge field, put $W(t) = \sum_m t^{m+1}\mathfrak{A}^m/(m+1)!$ and $U_J(t)\psi = U(t)\psi - W(t)(J,0)$. Then $U_J(t)\psi = \psi + W(t)\big(\mathfrak{A}\psi - (J,0)\big)$, the integrated form of $\dot E = \mathrm{curl}^*\, B - J$, $\dot B = -\mathrm{curl}\, E$; $U_J(t+s) = U_J(t)U_J(s)$; and at every site $\mathrm{div}_E E(t) = \mathrm{div}_E E(0) - t\,\mathrm{div}_E J$. A charge $Q(t)$ satisfies Gauss's law $\mathrm{div}_E E(t)(x) + Q(t) = 0$ at every t exactly when it does at $t = 0$ and $Q(t) = Q(0) + t\,\mathrm{div}_E J(x)$.

(b) On a lattice with at least two sites per side, at a fixed site x_0, with no choice principle: "every field satisfies $\mathrm{div}_E E(x_0) = 0$ or does not" is equivalent to WLPO; "every field that violates it violates it visibly, with $\mathrm{div}_E E(x_0)\ \#\ 0$" to $\mathrm{MP}_{\mathrm{bin}}$; and "every field satisfies it or violates it visibly" to LPO. The same holds for $\mathrm{div}_F B(x_0) = 0$.

The flow keeps Gauss's law at no price; deciding it carries the price.

Proof. (a) Termwise, $U(t) = 1 + W(t)\mathfrak{A}$, which gives the integrated form. The divergence kills every term of $W(t)(J,0)$ except the first, $t(J,0)$, by Theorem 26.46(a). The series identity $W(t+s) = W(t) + U(t)W(s)$ gives the flow property, and the continuity statement follows from the formula for $\mathrm{div}_E E(t)$.

(b) With two or more sites per side, $x_0 - e_0 \ne x_0$. For a point a, the edge field with $E_0(x_0) = -a$ and all other entries 0 has $\mathrm{div}_E E(x_0) = a$, so the three statements for all fields are the statements $a = 0 \vee a \ne 0$, $a \ne 0 \to a\ \#\ 0$ and $a = 0 \vee a\ \#\ 0$ for all points, which cost WLPO, $\mathrm{MP}_{\mathrm{bin}}$ and LPO by Corollary 12.18. The face field with $B_0(x_0) = -a$ does the same for div_F. □

Light on the lattice is exact, and its error against the continuum is explicit.

Theorem 26.48 (Standing light and its dispersion)**.** *Let m be a natural number, $\theta = 4qm/(N+1) = 2\pi_\delta m/(N+1)$ and $\omega = 2\sin(\theta/2)$.*

(a) The standing wave with $E_1(x) = \cos(\theta x_0)$, where $x_0 \in \{0, \dots, N\}$ is the first coordinate of x, and all other components 0, evolves under the Maxwell flow exactly as $E_1(t,x) = \cos(\theta x_0)\cos(\omega t)$, with $E_0 = E_2 = 0$ at every time.

(b) For every $\theta \geq 0$ and every point t, with $\omega = 2\sin(\theta/2)$, $|\cos(\omega t) - \cos(\theta t)| \leq |t|\theta^3/24$.

(c) For $\theta > 0$ the phase speed satisfies $1 - \theta^2/24 \leq \omega/\theta \leq 1$, and the group speed $\cos(\theta/2)$ satisfies $1 - \theta^2/8 \leq \cos(\theta/2) \leq 1$.

Proof. (a) Fields whose only components are E_1 and B_2, both depending on x_0 alone, form an invariant space: $\mathfrak{A}$ sends $E_1 = f$ to $B_2 = -(f(x_0+1) - f(x_0))$ and $B_2 = g$ to $E_1 = g(x_0 - 1) - g(x_0)$, and all other components of the image vanish because nothing depends on x_1 or x_2. So $\mathfrak{A}^2$ acts on E_1 as the periodic second difference $f(x_0+1) - 2f(x_0) + f(x_0 - 1)$. Since $\theta(N+1) = 2\pi_\delta m$, the function $\cos(\theta x_0)$ is periodic with period $N+1$, and the addition formula gives $\cos(\theta(x_0+1)) + \cos(\theta(x_0-1)) = 2\cos\theta\cos(\theta x_0)$. So $\mathfrak{A}^2$ multiplies the standing wave by $-(2 - 2\cos\theta) = -\omega^2$. The even terms of the exponential series give $E_1 = \cos(\theta x_0)\sum_k (-\omega^2 t^2)^k/(2k)! = \cos(\theta x_0)\cos(\omega t)$, and the odd terms contribute only to B_2.

(b) For $y \geq 0$, $y - y^3/6 \leq \sin y \leq y$. From the series, for $0 \leq y^2 \leq 20$ the terms of $1 - \mathrm{S}(y^2)$ pair into nonnegative brackets, and for $0 \leq y^2 \leq 42$ so do those of $\mathrm{S}(y^2) - 1 + y^2/6$; for $y \geq 1$, $\sin y \leq 1 \leq y$, and for $y^2 \geq 12$, $y - y^3/6 \leq -y < -1 \leq \sin y$. The overlapping ranges cover all $y \geq 0$. At $y = \theta/2$ this gives $\theta - \theta^3/24 \leq \omega \leq \theta$. The cosine is 1-Lipschitz, since $\cos a - \cos b = -2\sin\frac{a+b}{2}\sin\frac{a-b}{2}$ and $|\sin y| \leq |y|$; so $|\cos(\omega t) - \cos(\theta t)| \leq |t|\,|\omega - \theta| \leq |t|\theta^3/24$.

(c) The phase speed is (b) divided by θ. The group speed bound is $\cos y \geq 1 - y^2/2$ at $y = \theta/2$, proved from the series like the sine bounds. □

So lattice light travels at the continuum speed up to a relative error of $\theta^2/24$ in the phase speed, where θ is the wave number times the lattice spacing. In every direction the same computation gives plane waves of frequency Ω with $\Omega^2 = \sum_j 4\sin^2(\theta_j/2)$, and for $0 \leq \theta_j$ with $\theta_j^2 \leq 48$, $\sum_j(\theta_j^4/12 - \theta_j^6/360) \leq |\theta|^2 - \Omega^2 \leq \sum_j \theta_j^4/12$: the lattice is anisotropic at fourth order only.

Schrödinger's equation on a lattice has a light cone too, and a continuum limit with an error of order h^2. Let V be a potential with ratio values, $|V(x)| \leq K$ for a natural number K. The *box Hamiltonian* on the sites $-L, \dots, L$ is $(H_L\psi)(x) = V(x)\psi(x) - \psi(x+1) - \psi(x-1)$, with the terms outside the box omitted, and its *propagator* at time t is $\exp(-\mathrm{i}tH_L)$.

Theorem 26.49 (A light cone and convergent propagators)**.**

(a) Let U be a square matrix of amplitudes and d a function on pairs of indices with $d(x,x) = 0$ and $d(x,z) \leq d(x,y) + d(y,z)$, and suppose $U_{xy} = 0$ whenever $d(x,y) > 1$. Then $(U^k)_{xy} = 0$ whenever $k < d(x,y)$. If moreover the sizes $\mathrm{sz}(U_{rk})$ in each column of U sum to at most B, and $\mathrm{sz}(z) \leq S$ for natural numbers B, S, then every entry of $\exp(zU)$ at a pair with $d = d(x,y) \geq 2BS$ has real and imaginary part at most $4(BS)^d/d!$ in absolute value.

(b) Let $|t| \le T$ for a natural number T, put $c = (2+K)T$, let $R \le L \le L'$ and $m = L - R + 1$, and let $|x|, |y| \le R$. If $m \ge 2c$, the entries (x, y) of the propagators of H_L and $H_{L'}$ agree to within $4c^m/m!$ in real and in imaginary part. If only $c \le m$, they agree to within $4c^{2m}/(2m)!$.

(c) For $|x|, |y| \le R$ the box amplitudes converge, as $L \to \infty$, to an amplitude $A(x, y)$ within $4c^m/m!$ of every box amplitude with $m \ge 2c$. The columns of A are orthonormal, and A preserves the inner product of every pair of finitely supported states.

(d) Deciding, for all potentials, boxes and times, whether a propagator entry is exactly 0 is equivalent to WLPO, with no choice principle.

Parts (a) to (c) are free.

Proof. (a) By induction on k: $(U^{k+1})_{xy} = \sum_r (U^k)_{xr} U_{ry}$, and for each r either $k < d(x, r)$, so the first factor vanishes, or $d(r, y) > 1$ by the triangle inequality, so the second does. For the bound, the series of $\exp(zU)$ at (x, y) starts at order d, its jth term is at most $2(BS)^j/j!$ in each part, and the tail estimate of (b) below applies.

(b) Compare two exponential series whose coefficients agree below order m. The entries of H_L^k satisfy the three-term recursion of the Hamiltonian; $(H_L^k)_{xy} = 0$ for $|x - y| > k$ by (a); $|(H_L^k)_{xy}| \le (2 + K)^k$; and $(H_L^k)_{xy} = (H_{L'}^k)_{xy}$ when $|x| + k \le L$, because a site just outside the smaller box is beyond the cone of x. So the two series agree below order m, and from order $j \ge m$ each term of the difference is at most $2c^j/j!$ in real and imaginary part. Once $2c \le j + 1$ consecutive terms at least halve, so the tail is at most $2 \cdot 2c^m/m!$; bounds at every stage pass to the limits. For the sharp form, the boxes agree on every order k with $k + |x| + |y| < 2L + 2$, a round trip from x to the wall and back to y, so agreement holds below order $2m$.

(c) By (b) the box amplitudes form a Cauchy sequence in L with an explicit modulus, and the diagonal limit gives A. Each box propagator is unitary: H_L is real and symmetric, so the conjugate transpose of $\exp(-\mathrm{i}tH_L)$ is $\exp(\mathrm{i}tH_L)$, and their product is 1 by the group law. By the bound in (a), with $B = 2 + K$, the entries of every box propagator in column y decay like $4c^d/d!$ at distance $d \ge 2c$ from y, uniformly in the box; so the sums of products of columns over $|x| \le M$ converge uniformly, and orthonormality passes to the limit, and with it the preservation of inner products of finitely supported states.

(d) With WLPO, decide whether the real and the imaginary part are 0. Conversely, for $L = 1$ and $V = 0$ the entry between the sites -1 and 0 is $\mathrm{i}\sin(\sqrt{2}\,t)/\sqrt{2}$, because the odd powers of H_1 have that entry -2^j and the even powers have it 0. For $|t| \le 1$, $\sin(\sqrt{2}t) = 0$ exactly when $t = 0$, since $\mathrm{S}(y^2) \ge \frac{2}{3}$ for $y^2 \le 2$. So the decision gives $t = 0 \vee t \ne 0$ for every $|t| \le 1$, which is WLPO by Corollary 12.18 after scaling. □

Part (a) is the exact form, for range-one steps, of the finite group velocity of Lieb and Robinson [185].

Theorem 26.50 (The continuum limit of the ring, with error of order h^2)**.** *Let $n \geq 3$, let H be the matrix of $2\psi_j - \psi_{j+1} - \psi_{j-1}$ on a ring of n sites, and put $\theta_m = 4qm/n = 2\pi_\delta m/n$ for integers m. For finitely many amplitudes a_m let $\psi(j) = \sum_m a_m e^{\mathrm{i}\theta_m j}$. For every point s and every matrix $U = \exp(-\mathrm{i}sH)$, at every site j the amplitude $(U\psi)(j)$ is within*

$$\frac{|s|}{12} \sum_m \mathrm{sz}(a_m)\, \theta_m^4$$

of the continuum value $\sum_m a_m e^{-\mathrm{i}\theta_m^2 s} e^{\mathrm{i}\theta_m j}$, in real and in imaginary part. Such a matrix U exists for every s, and it is unitary. In physical units, with spacing $h = 2\pi_\delta/n$, time $t = h^2 s$ and position $x = hj$, the continuum value is $\Psi(t,x) = \sum_m a_m e^{\mathrm{i}(mx - m^2 t)}$, which solves $\mathrm{i}\,\partial_t \Psi = -\partial_x^2 \Psi$, and the bound is $|t|\, h^2 \sum_m \mathrm{sz}(a_m) m^4/12$. On the voxel lattice $(\mathbb{Z}/n)^3$ the same holds with m^4 replaced by $m_1^4 + m_2^4 + m_3^4$. All of this is free.

Proof. Eigenvectors. Since $\theta_m n = 2\pi_\delta m$, the plane wave $v(j) = e^{\mathrm{i}\theta_m j}$ is periodic around the ring, and $v(j+1) + v(j-1) = 2\cos\theta_m\, v(j)$, so $Hv = \lambda v$ with $\lambda = 2 - 2\cos\theta_m$. The partial sums of the exponential series act on v as the scalar partial sums, whose even and odd parts are the cosine and sine series; limits are unique, so $Uv = e^{-\mathrm{i}s\lambda} v$, and by linearity $(U\psi)(j) = \sum_m a_m e^{-\mathrm{i}s\lambda_m} e^{\mathrm{i}\theta_m j}$. The matrix U exists because H has integer entries, at most three in each row, so its exponential series converges entrywise with an explicit modulus, as in the proof of Theorem 26.46(b); it is unitary because H is real and symmetric, so the conjugate transpose of $\exp(-\mathrm{i}sH)$ is $\exp(\mathrm{i}sH)$, and the group law makes their product 1.

The error of one mode. Cosine and sine are 1-Lipschitz, so each part of $a\,e^{\mathrm{i}A} - a\,e^{\mathrm{i}B}$ is at most $\mathrm{sz}(a)|A - B|$, here with $|A - B| = |s|\,|\theta^2 - \lambda(\theta)|$.

The dispersion gap. $1 - \theta^2/2 \leq \cos\theta \leq 1 - \theta^2/2 + \theta^4/24$ for every θ: with $u = \theta^2$, the terms of $\mathrm{C}(u) - 1 + u/2$ pair into nonnegative brackets for $u \leq 30$, and $1 - u/2 \leq -1$ for $u \geq 4$; those of $1 - u/2 + u^2/24 - \mathrm{C}(u)$ pair into nonnegative brackets for $u \leq 56$, and $1 - u/2 + u^2/24 \geq 1$ for $u \geq 12$. So $0 \leq \theta^2 - \lambda(\theta) \leq \theta^4/12$. Summing over the modes gives the bound. With $\theta_m = mh$ and $s = t/h^2$, $|s|\theta_m^4/12 = |t|h^2 m^4/12$, and Ψ solves the equation mode by mode. On $(\mathbb{Z}/n)^3$ the Laplacian is the sum of three ring Laplacians and the gaps add. □

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$, its cosine, sine and π_δ, a lattice, and a finite trigonometric initial state.

Conclusion

The lattice evolution exists and differs from the free continuum Schrödinger evolution by at most $|t|h^2 \sum \mathrm{sz}(a_m) m^4/12$.

Not claimed

An identification of these lattice fields with Recognition Science's electromagnetism or its quantum amplitudes.

Table 26.3. The local price of deciding a sign law near a threshold p; the first row that applies gives it.

	Bits at p	Local price
1	N	free
2	not P, not L, not R	free
3	P, L and R	LLPO if LC and RC, else LPO
4	not P	LPO
5	not L and not R	WLPO
6	L	WLPO if LC, else LPO
7	otherwise	WLPO if RC, else LPO

26.8 Laws and their prices

A physical law classifies states by comparisons: an orbit is bound when its energy is negative, a field obeys Gauss's law when a divergence is zero, a displacement is causal when $E \geq |p|$. On $\mathbb{R}_\delta$ each comparison has a price (Corollary 12.18), and a law combines them. This section computes the exact price of whole classes of laws by programs that read only the thresholds, and then prices the lattice premises that Recognition Science uses for an inverse-square field. The prices are proved with no choice principle. As equivalences with omniscience principles they are trivial in classical mathematics, so their content is that each price is exact without choice. Theorem 25.59(4) of Chapter 25 lists the prices a sign law on an observable that takes every value can have; Theorem 26.52 below computes, by an explicit rule on the thresholds, which of them each such law has.

Definition 26.51 (Sign laws and their germs)**.** A *sign law* in one quantity x is built from *atoms* $x < c$, $x \leq c$, $x = c$, $c < x$, $c \leq x$, with ratio thresholds c, by "and" and "or". It *holds* at a point as written. It is *visibly refuted* at x when its failure is witnessed by strict comparisons: $x < c$ and $x \leq c$ are refuted when $c < x$, $c < x$ and $c \leq x$ when $x < c$, and $x = c$ when $x \# c$; "and" is refuted when one part is, "or" when both are. Write the law in disjunctive normal form; its *cells* are conjunctions of atoms. At a threshold p of the law record six bits, each saying whether some cell holds on the whole of a germ at p: the point p itself (P), an open interval just left of p (L), just right of p (R), an interval $(p - \varepsilon, p]$ (LC), an interval $[p, p + \varepsilon)$ (RC), and an open interval around p (N), for all small $\varepsilon > 0$. The *local decision price* at p is the first row of Table 26.3 that applies. The *decision class* of the law is the join of its local prices in the order free $<$ LLPO $<$ WLPO $<$ LPO. Its *holds class* is *refuted* if P, L or R fails at some threshold, and otherwise free if N holds at every threshold, LLPO if LC and RC hold at every threshold, and LPO in the remaining case. With a seventh bit F, whether the law is visibly refuted at p itself, its *refutation class* is false if some threshold has neither P nor F, otherwise free if every threshold has N, F, or both L and R, and $\mathrm{MP}_{\mathrm{bin}}$ in the remaining case; its *settling class* is false in the same case, and otherwise the join over the thresholds of: free if N or F, LLPO if L, R, LC and RC, and LPO otherwise.

For example, Kepler's trichotomy of an energy, $E < 0 \vee E = 0 \vee E > 0$, has holds class LPO; the closed dichotomy $E \leq 0 \vee 0 \leq E$ has holds class and decision class LLPO; the overlapping cover $x < 1 \vee -1 < x$ has holds class free; deciding $E < 0$ costs LPO, and deciding $E = 0$ or $E \leq 0$ costs WLPO.

Theorem 26.52 (One program prices every sign law)**.** *For every sign law φ, with no choice principle:*

(a) "φ holds at every point" is equivalent to the meaning of its holds class: false, true, LLPO or LPO;

(b) "φ holds or fails at every point" is equivalent to the meaning of its decision class: true, LLPO, WLPO or LPO;

(c) "wherever φ fails it is visibly refuted" is equivalent to the meaning of its refutation class: false, true or $\mathrm{MP}_{\mathrm{bin}}$; and "$\varphi$ holds or is visibly refuted at every point" to the meaning of its settling class: false, true, LLPO or LPO;

(d) (a)–(c) hold as well for sign laws whose thresholds lie in $\mathbb{Z}[\varphi]$.

Proof. Windows. Between two consecutive thresholds every atom is constant. List the thresholds in order, which is decided by comparing ratios, and apply cotransitivity of $<$ once at each gap between consecutive values: every point x lies in the *window* of some threshold p, on the same strict side as p of every other threshold, and for each p every point lies in its window or strictly on one side of p. In the window of p the law at x agrees with its germ at p: if N, it holds at x; if LC and $x \le p$, it holds; if L and $x < p$, it holds; if P and $x = p$, it holds. Each bit that fails confines where it holds: without R to $x \le p$, without P to points apart from p, without N to $x \le p$ or $p \le x$, without LC to $x < p$ or $p \le x$ (symmetrically for L and RC). Visible refutation is read atom by atom, then through "and" and "or": an atom whose threshold is not p is visibly refuted at x exactly when it is at p, and one with threshold p never is at p but holds or is visibly refuted at each $x < p$ and $x > p$. So the law is visibly refuted throughout the window if F, at each $x < p$ there if L fails (each $x > p$ if R fails), and, if F fails, nowhere else.

Upper bounds. With the principle named by a row, the position of x relative to p is decided as finely as the row's bits require: LLPO gives $x \le p \vee p \le x$; WLPO decides $x = p$ and $x \le p$; LPO decides $x < p$, $x = p$ or $x > p$. In the rows marked free, either a cell holds on a whole neighbourhood of p, or the law fails at p and on both sides, and then $\neg\varphi(x)$ follows by refuting each way it could hold. A point in a different window is strictly on one side of p, so a local decision spreads to the whole line, and the join of the local principles decides the law. For (c): if some threshold has neither P nor F, the law there neither holds nor is visibly refuted, and both statements fail. Otherwise take x in the window of p. With F the law is visibly refuted at x, with N it holds, and with P, L and R it cannot fail, since $\neg(x < p)$ and $\neg(p < x)$ give $x = p$. Otherwise P holds, and a failure at x gives $x \neq p$, which $\mathrm{MP}_{\mathrm{bin}}$ turns into $x < p$ or $p < x$ (Table 12.2), a side whose bit fails, where the law is visibly refuted. For settling, LLPO places x where LC or RC makes the law hold, and LPO places x below, at or above p.

Lower bounds. At a threshold p whose row names a principle, the confinements turn a decision of the law at each x of the window into one comparison with p: $x = p \vee x \neq p$ (row 5), $x \le p \vee \neg(x \le p)$ (row 6 with LC; mirrored, row 7 with RC), $x \le p \vee p \le x$ (row 3 with LC and RC), and $x < p \vee \neg(x < p)$ or its mirror image in the other rows; for instance, in row 4 with L the law holds only apart from p and fails only where $\neg(x < p)$. Points outside the window are strictly on one side of p. As $x < y$ exactly when $x - y + p < p$, and likewise

for $\leq$ and $=$, the comparison with p at every point is the same comparison for all pairs of points, priced WLPO, LLPO or LPO in Table 12.2. For (a), a law whose holds class is refuted fails at a threshold or at a point just beside one, found inside its window by density of the ratios; otherwise it holds everywhere under the principle named, and a threshold without N (or without one of LC, RC) is in row 3 and yields LLPO (or LPO). For (c), where refutation is not free, P holds, N and F fail, and so does L or R. If both do, the law fails at each $x \neq p$ of the window, so it is visibly refuted, at a point $x < p$ or $p < x$; if only L does, it fails where $\neg(p \leq x)$ and is visibly refuted, at a point $x < p$. This is $x \neq p \to x < p \vee p < x$, or $\neg(p \leq x) \to x < p$, at every point, which costs $\mathrm{MP}_{\mathrm{bin}}$. For settling, with P and neither N nor F: if L fails, the law holds only where $p \leq x$ and is visibly refuted only where $x < p$, or $p < x$ with R failing, so settling decides $x < p$, which is LPO; if L and R hold, it is never visibly refuted in the window, and settling is holding, priced as in (a). For (d), the proof uses of the thresholds only that their strict and weak order is decided by a computation, and in $\mathbb{Z}[\varphi]$ the sign of $a + b\varphi$ is decided by integer arithmetic: $a + b\varphi > 0$ exactly when $2a + b > -b\sqrt{5}$, which compares the integers $(2a + b)^2$ and $5b^2$ once the signs of $2a + b$ and b are known. □

Laws in two quantities are priced by the same idea, after a reduction to lines through finitely many points.

Theorem 26.53 (Linear laws of two quantities)**.** *Consider the laws built by "and" and "or" from atoms $ax + by < c$, $ax + by \leq c$ and $ax + by = c$ in two quantities x, y, with integers a, b and ratio thresholds c. There are programs, reading the coefficients, that assign to each law two classes such that, with no choice principle, "the law holds at every point of the plane" is equivalent to the meaning of the first, which is false, true, LLPO or LPO, and "the law holds or fails at every point" to the meaning of the second, which is true, LLPO, WLPO or LPO. The class is the join, over finitely many* centres*, of the class of the law's* fan *at the centre, and the class of a fan is the join of the one-quantity classes of Theorem 26.52 along finitely many lines through its vertex. The* critical heights *are 0, the ratio c/b of each atom with $b \neq 0$, and the heights where two non-parallel lines that are not horizontal cross; the centres at a critical height h are $(0, h)$ and the points where the non-horizontal lines meet the height h. They include every crossing point of the law's lines. For laws in $k \geq 3$ quantities the corresponding statement is open.*

Proof. Fans. At a centre v keep the atoms whose lines pass through v; near v the others are constant. Such a law depends only on the direction of $d = (d_1, d_2) = (x, y) - v$. Its lines through v, with the horizontal line through v, cut the plane into open sectors and open rays, each with an integer sample point, and on each piece every atom's form has the sign it has at the sample (strict or zero on the open piece, weak on its closure). So the fan law holds on an open piece exactly when it holds at the sample, and a cell holding at v and at the sample holds on the closed piece. Along the line through v in a sample direction the fan law is a one-quantity sign law at the threshold 0, priced by Table 26.3; the lower bounds come from these lines. For the upper bounds the table gives, on each sample line: a price at most WLPO forces, without P, neither L nor R (rows 2, 4), and with P and R, RC (rows 1, 3, 7); at most LLPO with P forces N, or R and RC (rows 1, 3); free with P forces N (row 1). Without P the law fails at v and on every open piece, so at every d, which cannot lie outside all pieces. With P: if the class is free, a cell holds at w, 0 and $-w$ for the sample w of an open sector, where every non-constant form is nonzero, with opposite signs at $\pm w$; so the cell's atoms are constants true at v, and it holds everywhere. If LLPO, every closed piece carries a cell holding

on it, and LLPO places d in a closed sector, by the weak sign of d_2 and then of each offset $d_1 - \kappa d_2$ from a line $d_1 = \kappa d_2$ of the law, in increasing order of κ. If WLPO, zero tests place d at v, or in a closed ray or sector C with d not excluded from its open part; if the law holds at the sample of C, RC gives a cell on C, and otherwise it fails on the open part, hence at d. LPO decides every atom. For holding everywhere, a fan law that fails at v or at a sample is refuted, and one that holds at both cannot fail at any d, so it holds wherever it is decided. Three closed half-planes meeting at a point and covering the plane cost exactly LLPO, since a line through the vertex in a generic direction already carries the dichotomy.

Centres. The critical heights are ratios, and they include the heights of the horizontal lines. For a point (x, y), the one-quantity window argument places y strictly between the critical heights next to some critical height h, so (x, y) is strictly on the side of h of every horizontal line at another height. If the law has no non-horizontal line, the centre is $v = (0, h)$. Otherwise, in that band two lines that meet the height h at different points do not cross, so their order at y is their order at h, and cotransitivity places x next to one group of lines, whose point at height h is a centre v. Either way (x, y) lies in the window of v, strictly on v's side of every line not through v, and there the law is its fan law at $(x, y) - v$. Every direction from a centre enters its window after scaling, and fan laws are invariant under scaling, so "holds everywhere" and "decided everywhere" for the law are the conjunctions of the same statements for its fans. □

For example, the light-cone law $E \geq |p|$, in units with $c = 1$, is decided everywhere exactly under WLPO, and the strict cone $E > |p|$ exactly under LPO. The program never needs a principle outside the chain free, LLPO, WLPO, LPO.

Theorem 26.54 (Laws on a half-line, and the oscillator's energy)**.**

(a) For every sign law φ and every ratio c there is a sign law φ_c with $\varphi_c(x)$ exactly when $\varphi(\max(x, c))$, for every point x.

(b) Let a quantity $\mathcal{O}$ on a set of states have range exactly $[c, \infty)$: $\mathcal{O}(s) \geq c$ for every state, and every point $a \geq c$ is $\mathcal{O}(s)$ for some state. Then "$\varphi(\mathcal{O}(s))$ at every state" costs exactly the holds class of φ_c, and "$\varphi(\mathcal{O}(s))$ is decided at every state" exactly its decision class.

(c) The energy of the harmonic oscillator, a positive quadratic form in position and momentum, has range exactly $[0, \infty)$. So $E \leq 0 \vee E \geq 0$ holds and is decided at every state with no principle; Kepler's trichotomy $E < 0 \vee E = 0 \vee E > 0$ holds at every state exactly under LPO and is decided at every state exactly under LPO; and deciding whether the oscillator is at rest, $E = 0$, costs exactly WLPO.

(d) Clamping does not carry visible refutation: for no ratio c is there a sign law ψ such that, at every point y, ψ holds exactly when the atom $\max(y, c) < c$ holds and ψ is visibly refuted exactly when that atom is visibly refuted.

All of this is proved with no choice principle.

Proof. (a) Replace each atom by its clamped form, decided by comparing q with c: $\max(x, c) < q$ is false if $q \leq c$ and is $x < q$ if $c < q$; $\max(x, c) \leq q$ is false if $q < c$ and $x \leq q$ otherwise; $\max(x, c) = q$ is false if $q < c$, is $x \leq c$ if $q = c$, and is $x = q$ if $c < q$; $q < \max(x, c)$ is true if

$q < c$ and $q < x$ otherwise; $q \le \max(x, c)$ is true if $q \le c$ and $q \le x$ otherwise. Constants are sign laws, such as $x < 1 \vee -1 < x$ and $x < -1 \wedge 1 < x$. Each equivalence holds on $\mathbb{R}_\delta$ with no case split on x; for instance, if $\max(x, c) = q > c$ then $x < q$ is impossible, so $x \ge q$, and $x \le \max(x, c) = q$.

(b) The values $\mathcal{O}(s)$ and the values $\max(x, c)$ are the same set of points, so the statements about states are the statements of Theorem 26.52 for φ_c.

(c) The energy is at least 0, and every $a \ge 0$ is attained with zero momentum at the position given by the square root on the line. Then apply (b) with $c = 0$: the clamped dichotomy holds at every point, the clamped trichotomy is $x \le 0 \vee 0 < x$, with holds and decision class LPO, and the clamped rest law is $x \le 0$, with decision class WLPO.

(d) Suppose ψ existed, and take a ratio y below c and below every threshold of ψ. At y every atom of ψ holds or is visibly refuted: $y < q$ makes $x < q$ and $x \le q$ hold, and refutes $x = q$, $q < x$ and $q \le x$ with the strict witness $y < q$. "And" and "or" keep this property, so ψ holds or is visibly refuted at y. But $\max(y, c) = c$, and the atom $c < c$ neither holds nor is visibly refuted, since a refutation would need $c < c$. □

So on a half-line the program prices holding and deciding a law, and the refutation prices of Theorem 26.52(c) do not transfer by clamping.

Recognition Science's lattice premises

Recognition Science obtains an inverse-square field on its lattice from a current with three properties. On $\mathbb{Z}^D$ a *current* assigns a rate $j(x, i)$ to the edge from x to $x + e_i$, and its divergence at x is $\operatorname{div} j(x) = \sum_i \big(j(x, i) - j(x - e_i, i)\big)$.

Definition 26.55 (Balanced, shell-uniform and sourced currents)**.** A current is *balanced off the origin* when $\operatorname{div} j(x) = 0$ for every $x \ne 0$, and *sourced* when $\operatorname{div} j(0) \ne 0$. It is *shell-uniform* with rates $f(0), f(1), \ldots$ when, for every radius r and axis i, every edge leaving the box $\|x\|_\infty \le r$ through its face $x_i = r$ carries $f(r)$, $j(x, i) = f(r)$, and every edge leaving through the face $x_i = -r$ carries the outward flow $f(r)$, $j(x - e_i, i) = -f(r)$.

Shell uniformity is Recognition Science's isotropy premise: the flux leaves every box equally through every boundary edge. The box of radius r has $2D(2r + 1)^{D-1}$ outward edges, and summing the divergence over the box, all interior edges cancel; so for a balanced shell-uniform current

$$2D\,(2r + 1)^{D-1} f(r) = \operatorname{div} j(0) \qquad \text{for every } r, \tag{26.4}$$

which is the lattice form of Gauss's law with $f(r)(2r + 1)^{D-1} = f(0)$.

Theorem 26.56 (The inverse-square premises are satisfiable in every dimension)**.** *The following is proved in classical mathematics. Let $D \ge 1$.*

(a) There is a current on $\mathbb{Z}^D$ with rational rates that is balanced off the origin and shell-uniform with $f(r) = 1/(2r + 1)^{D-1}$; in three dimensions the rates are the inverse squares $1/(2r + 1)^2$.

(b) *A profile f admits a balanced, shell-uniform current exactly when $f(r)(2r+1)^{D-1} = f(0)$ for every r.*

(c) *If $D \geq 2$, a current of potential form, $j(x,i) = \phi(x) - \phi(x+e_i)$ for a function ϕ on sites, that is balanced off the origin and shell-uniform with rates f has $f(r) = 0$ for every r, and so source 0.*

Proof. (b) Necessity is (26.4); sufficiency follows from (a), since a multiple of a current is a current.

(a) Give every outward edge of every box the rate it must carry, $1/(2r+1)^{D-1}$ outward on the boundary of the box of radius r. The edges from the origin carry 1, so the source is $2D$. The sites of the shell $\|x\|_\infty = r+1$ receive $2D$ in total through the boundary of the box of radius r and send $2D(2r+3)^{D-1}/(2r+3)^{D-1} = 2D$ out through the boundary of the next box, and every edge between two shells is one of these. So the net supply on each shell sums to 0. For $D \geq 2$ the shell is connected by its own edges, and a spanning tree routes any supply of total 0 with rational rates; for $D = 1$ each shell site has one edge in and one out, both carrying 1. This current has $f(r) = 1/(2r+1)^{D-1}$, and $f(0)$ times it has any profile with $f(r)(2r+1)^{D-1} = f(0)$.

(c) For a current of potential form, $\operatorname{div} j(x) = \sum_i \big(2\phi(x) - \phi(x+e_i) - \phi(x-e_i)\big)$. Put $a = e_0$ and $b = 2e_0$. The sites 0, e_0 and $2e_0$ lie on the faces $x_0 = 0, 1, 2$ of the boxes of radii $0, 1, 2$, so $\phi(0) - \phi(e_0) = f(0)$, $\phi(e_0) - \phi(2e_0) = f(1)$ and $\phi(2e_0) - \phi(3e_0) = f(2)$; the term of axis 0 in $\operatorname{div} j$ is $f(1) - f(0)$ at a and $f(2) - f(1)$ at b. For an axis $k \neq 0$, the sites $e_0 \pm e_k$ lie on the face $x_0 = 1$ of the box of radius 1, so $\phi(2e_0 \pm e_k) = \phi(e_0 \pm e_k) - f(1)$, and also $\phi(2e_0) = \phi(e_0) - f(1)$; so the term of axis k, $2\phi(x) - \phi(x+e_k) - \phi(x-e_k)$, is the same at b as at a. Both divergences vanish, and their difference is $f(0) - 2f(1) + f(2) = 0$. By (26.4), $f(1) = f(0)/A$ and $f(2) = f(0)/B$ with $A = 3^{D-1}$ and $B = 5^{D-1}$, so $f(0)(AB - 2B + A) = 0$. For $D \geq 2$, $A \geq 3$ and $AB - 2B + A = B(A-2) + A > 0$, so $f(0) = 0$, and then $f(r) = f(0)/(2r+1)^{D-1} = 0$ for every r. □

For $D = 1$ part (c) has no force: there $A = B = 1$, and the potential $\phi(x) = -|x|$ gives the balanced shell-uniform current with every rate 1. In two or more dimensions a lattice potential obeys the boundary-averaged law behind (26.4), that the flux out of every box equals the source, but never the shell-uniform premise with a source.

Theorem 26.57 (No integer current is balanced, shell-uniform and sourced)**.** *On $\mathbb{Z}^3$:*

(a) *every current with integer rates that is balanced off the origin and shell-uniform with integer rates has source 0; so no integer current is balanced, shell-uniform and sourced;*

(b) *on $\mathbb{Z}^D$, for every $D \geq 1$, every current with real rates that vanishes outside a box and is balanced off the origin and shell-uniform has source 0, and a balanced shell-uniform current with a nonzero source carries a nonzero rate on the axis edge leaving $(R+1)e_i$, for every R and every axis i, so it has infinitely many nonzero rates;*

(c) *deciding, for every integer current, whether it is balanced off the origin costs exactly WLPO, while for currents vanishing outside a given box it is decided with no principle;*

(d) at a state where every compared quantity lies in $\mathbb{Z}[\varphi]$, every exact law is decided, and a failing law without strict atoms is visibly refuted, with no principle.

Parts (a) and (b) concern integer and real currents in classical mathematics; (c) and (d) are proved with no choice principle.

Proof. (a) By (26.4), $6(2r+1)^2 f(r)$ is the source σ for every r. Take $r = |\sigma|$: then $6(2r+1)^2 > |\sigma|$, so the integer $f(r)$ has absolute value below 1, hence is 0, and $\sigma = 0$.

(b) A current that vanishes outside a box has $f(r) = 0$ for large r, so its source $2D(2r+1)^{D-1} f(r)$ is 0 by (26.4). If the source is not 0, then $f(r) \neq 0$ for every r, and the axis edge leaving $(R+1)e_i$ is an outward edge of the box of radius $R+1$, so it carries $f(R+1) \neq 0$.

(c) Balance off the origin is a conjunction, over all sites, of decidable integer equations. With WLPO, enumerate the sites $x_n \neq 0$ and decide whether the sequence $[\operatorname{div} j(x_n) \neq 0]$ is all zero. Conversely, for a binary sequence α let j carry the rate $\alpha(n)$ on the edge from $(2n+2)e_0$ to $(2n+3)e_0$ and 0 elsewhere; then j is balanced off the origin exactly when α is all zero. For a current vanishing outside a box, only finitely many sites can have nonzero divergence.

(d) $\mathbb{Z}[\varphi]$ has decidable equality and sign: $a+b\varphi$ with integers a, b is compared with 0 by integer arithmetic. So each atom is decided, hence each law; and a failing atom $x \leq y$ or $x = y$ fails strictly, which is a visible refutation. □

Conditions and scope.

Base and premises

Recognition Science's balance and isotropy premises for a lattice current.

Conclusion

The premises are consistent in every dimension, with rates fixed exactly to $f(0)/(2r+1)^{D-1}$. On $\mathbb{Z}^3$ they cannot be met with integer rates (on $\mathbb{Z}$ they can: the current of the potential $-|x|$ has every rate 1), and with a source they need infinitely many nonzero rates, an infinite posting history.

Not claimed

That a potential field meets them: in two or more dimensions Theorem 26.56(c) excludes it, and a pointwise inverse-square decay for a lattice potential is an open statement.

26.9 Thermal light on the lattice

Fill the cubic lattice of Section 26.7 with light at temperature T. How much energy does it hold? In the continuum the answer is the Stefan–Boltzmann law: energy density proportional to T^4, with the constant $\pi^2/15$ in natural units. On the lattice the photon's energy is not proportional to its wave number, so the law can hold only as $T \to 0$, and the first correction is of order T^2. This section proves the lattice law, its correction with exact coefficients, and the facts about the lattice photon behind them, on $\mathbb{R}_\delta$ with no omniscience principle and no choice principle. Three statements fail as first posed, for a reason worth seeing: they quantify over every zero of the cosine at which the sine is 1, rather than over the quarter period alone.

As in Section 26.7, q is the quarter period, $\pi = 2q$, and the *zone* is $[-\pi, \pi]^3$. For a function F on the zone, its *midpoint averages* are the averages of F over the $(N+1)^3$ centres of the cells of the grid with $N+1$ cells per axis, and v is its *zone average*, the continuum's $(2\pi)^{-3}\int F d^3k$, when the midpoint averages converge to v. Write $S(k) = \sum_j (1 - \cos k_j)$ and $|k|^2 = \sum_j k_j^2$.

Definition 26.58 (The lattice photon and the cold datum)**.** Let $r > 0$. The *Bose kernel* is the function K with $K(s)(e^s - 1) = s$ for every point s and $K(0) = 1$; there is exactly one. The *lattice photon energy* is the function E with $E(k) \geq 0$ and

$$\cosh E(k) - 1 = r\,S(k)$$

at every wave vector k, the shell equation of the lattice; its root is unique. A *cold datum* is $r > 0$, its square root $\sqrt{r}$, the functions E and K, a family $a(T)$, $T > 0$, of zone averages of $k \mapsto K(E(k)/T)$, and the Stefan–Boltzmann value $L = \pi^2/(15\sqrt{r}^3)$. The thermal energy density divided by T^4 is then $u(T)/T^4 = 2a(T)/T^3$, two polarizations times the average energy $TK(E/T)$ of a mode.

The parameter r sets the speed of long waves: $E(k)^2 = r|k|^2$ to leading order, as Proposition 26.59 shows.

Proposition 26.59 (The lattice photon)**.** *Let $r \geq 0$, and in the limits of parts 3 and 4 and the slope, $r > 0$.*

1. *$E(k)^2 \leq r|k|^2$ for every k; and if $|k_j| \leq 4$ for each j, then $r|k|^2 \leq 16(1+6r)E(k)^2$ and $r|k|^2 - E(k)^2 \leq C_4(r)|k|^4$, for a constant $C_4(r)$ given explicitly by the series of* cosh.
2. *At the zone edge $k = (\pi, 0, 0)$: $E^2 \leq 16/5$ when $r = 1$, and $E^2 \geq 19/5$ when $r = 4/3$.*
3. Long-wave limits. *Along a ray $k = \kappa n$ with $|n| = 1$, as $\kappa \to 0$: $E^2/\kappa^2 \to r$; with $v^2 = \sum_j (r \sin k_j / \sinh E)^2$ the squared group speed, $(1 - v^2/r)/\kappa^2 \to (\sum_j n_j^4 + r)/4$; and with $w = \sqrt{v^2/r}$, $(1-w)/E^2 \to (\sum_j n_j^4 + r)/(8r)$.*
4. Real time. *If $rS(k) \leq 2$, the real-time shell $1 - \cos\omega = rS(k)$ has a root ω with $0 \leq \omega \leq \pi$. Along a ray, for real-time roots in $[0, \pi]$, the limits of part 3 hold with $\sum_j n_j^4 - r$ in place of $\sum_j n_j^4 + r$ and ω in place of E. If $\omega(k)$ is such a root at every k with $rS(k) < 2$, then at every k with $S(k) > 0$, $rS(k) < 2$ and $\sin\omega(k)$ apart from 0, its derivative in k_j is $r \sin k_j / \sin\omega(k)$.*

Proof. (1) The series of cosh gives $E^2/2 \leq \cosh E - 1$, hence $E^2 \leq 2rS(k) \leq r|k|^2$, since $1 - \cos y \leq y^2/2$. The two other bounds compare both sides of the shell equation with their leading terms, using a lower bound for $1 - \cos y$ on $|y| \leq 4$ and the tail of the series of cosh beyond its quadratic and quartic terms on the range of E allowed by the first bound.

(2) At $k = (\pi, 0, 0)$, $S = 2$, so $\cosh E = 1 + 2r$. At $r = 1$, $\cosh E = 3$, and $E^2 \leq 16/5$ follows from $\cosh(\sqrt{16/5}) > 3$, checked on the series; at $r = 4/3$, $\cosh E = 11/3$, and $E^2 \geq 19/5$ from $\cosh(\sqrt{19/5}) < 11/3$. Each comparison is a finite computation with ratio bounds on the tail of the series.

(3) On the shell, $\sinh^2 E = (\cosh E - 1)(\cosh E + 1) = rS(2 + rS)$, and $\sum_j \sin^2 k_j = 2S - \sum_j (1 - \cos k_j)^2$, so $v^2 S(2 + rS) = r(2S - \sum_j (1 - \cos k_j)^2)$ exactly. Expanding $S = \kappa^2/2 - \kappa^4 \sum n_j^4/24 + \cdots$

along the ray, with the remainders of Proposition 12.98, gives each limit with an explicit modulus.

(4) The root is $\omega = \arccos(1 - rS(k))$, with the line's arccosine, which takes values in $[0, \pi]$ on $[-1, 1]$. On the real-time shell $\sin^2 \omega = rS(2 - rS)$, which gives the limits as in (3). The slope comes from an implicit-function argument with explicit constants: $\sin\omega\, d\omega = r \sin k_j\, dk_j$ on the shell. □

Theorem 26.60 (The lattice Stefan–Boltzmann law and its correction)**.** *Let $r > 0$. The zone averages $a(T)$ of Definition 26.58 exist for every $T > 0$, as one function of T. On every cold datum:*

1. *$u(T)/T^4 \to L = \pi^2/(15\sqrt{r}^3)$ as $T \to 0^+$;*
2. *for every $\varepsilon > 0$, eventually $L - \varepsilon T^2 \le u(T)/T^4$; for some $c > 0$, eventually $L + cT^2 \le u(T)/T^4$; and for some C, $u(T)/T^4 \le L + CT^2$ for every $T > 0$;*
3. *the coefficient $\big(u(T)/T^4 - L\big)/T^2$ converges as $T \to 0^+$, to*

$$\kappa\Big(\frac{\alpha}{r} + \beta\Big)L, \qquad \kappa = \frac{15}{\pi^2} \cdot \frac{2}{(2\pi)^3}, \quad 63\alpha = 4\pi^7, \quad 189\beta = 20\pi^7,$$

that is, to $\frac{5\pi^2}{21}\big(\frac{1}{r} + \frac{5}{3}\big)L$.

No omniscience principle and no choice principle is used.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; the quarter period; $r > 0$; the Bose kernel, the lattice photon energy and the zone averages of Definition 26.58.

Conclusion

Light on the cubic lattice obeys the Stefan–Boltzmann law as $T \to 0$, and its first correction is positive, of order T^2, with the exact coefficient $\frac{5\pi^2}{21}(\frac{1}{r} + \frac{5}{3})$ relative to the law, for the discrete-time shell $\cosh E - 1 = rS$ of Definition 26.58; the term $\frac{25\pi^2}{63}$ comes from the time step of that shell.

Not claimed

The identification of r, T or the lattice spacing with Recognition Science's own constants.

Proof. The zone averages. The integrand $K(E(k)/T)$ has a modulus of uniform continuity on the zone computed from T, and its midpoint averages form a sequence with an explicit Cauchy modulus. Such a presented sequence converges with no completeness and no choice, and the limit is a function of T because the modulus is computed, not chosen.

The law. Scale $k = Ty$. The midpoint average at temperature T becomes $T^3/(2\pi)^3$ times a box sum of $K(E(Ty)/T)$ over $[-\pi/T, \pi/T]^3$. By Proposition 26.59(1), $E(Ty)^2/T^2$ is within $C_4 T^2 |y|^4$ of $r|y|^2$ and at least $r|y|^2/(16(1 + 6r))$ on the scaled zone, so the integrand is dominated by

a separable product of functions $\operatorname{sech}^2(ay_j/2)$, whose box sums are bounded by telescoping, uniformly in the box and the grid. Every comparison is between two integrands on one box. As $T \to 0$ the box sums of $K(E(Ty)/T)$ approach the space value of $K(\sqrt{r}|y|)$. That value is $128q^5/(15\sqrt{r}^3) = 4\pi^5/(15\sqrt{r}^3)$: expand $K(s) = \sum_{n<M} se^{-(n+1)s} + K(s)e^{-Ms}$; each mode is the elementary integrand $|y|e^{-|y|}$ rescaled, whose space value is 24π (Theorem 12.96); the tail is bounded uniformly in the grid; and $\zeta(4) = \pi^4/90$ (Theorem 12.97) sums the modes. Multiplying by $2/(2\pi)^3$ gives L.

The T^2 coefficient. Write $b = \sqrt{r}|y|$ for the continuum photon energy at the scaled wave vector and $\sigma = E(Ty)/T$ for the lattice one. The shell equation, expanded to fourth order with the remainders of Proposition 12.98, gives

$$\sigma^2 = b^2 - \frac{T^2}{12}\Big(r\sum_j y_j^4 + r^2|y|^4\Big) + O(T^4|y|^6),$$

with an explicit constant, so the lattice photon lies just below light. Expand the kernel into its modes, $K(s) = \sum_{m\ge1} se^{-ms}$. For one mode $f(s) = se^{-ms}$, whose first two derivatives are explicit, the dispersion gives

$$f(\sigma) - f(b) = T^2\,\frac{\sqrt{r}}{24}\,(mb-1)e^{-mb}\Big[\big(\tfrac{3}{5}+r\big)|y|^3 + \frac{P_4(y)}{|y|}\Big]$$

up to explicit integrable majorants of order T^4, where $P_4(y) = \sum_j y_j^4 - \frac{3}{5}|y|^4$. The quartic P_4 is harmonic, so $P_4(y)W(|y|)/|y|$ is the divergence of $k(|y|)\nabla P_4$ with $k' = W/4$, and its box sums are fluxes through the faces of the box, which vanish as the box grows: this is how the angular average $\frac{1}{5}$ of n_1^4 enters, with no rotation symmetry of the cubic box. The radial part integrates exactly by Theorem 12.96, through $\int_0^\infty t^2(at-1)e^{-at}t^3\,dt = 600/a^6$, so mode m contributes $\sqrt{r}^{-3}(120q/r + 200q)/m^6$. Summing over the modes, with a tail bound uniform in T and $\zeta(6) = \pi^6/945$ (Theorem 12.97), the excess divided by T^2 tends to $\sqrt{r}^{-3}(\alpha/r + \beta)$ with $\alpha = 120q\,(2q)^6/945$, that is $63\alpha = 4\pi^7$, and $\beta = 200q\,(2q)^6/945$, that is $189\beta = 20\pi^7$. Multiplying by $2/(2\pi)^3$ and writing $\sqrt{r}^{-3} = 15L/\pi^2$ gives part 3; then $\kappa\alpha = \frac{20}{21}q^2$ and $\kappa\beta = \frac{100}{63}q^2$, which is the stated relative coefficient. The Bose value

$$\int_0^\infty s^2K(cs)\,ds = \frac{16q^4}{15c^3} = \frac{\pi^4}{15c^3},$$

proved on midpoint grids by the same splitting into modes, each a rescaled $\int t^3e^{-t}\,dt = 6$ (Proposition 12.98), with a tail uniform in the grid, is the one-dimensional form of the law's constant.

Part 2. The coefficient of part 3 is positive. So, eventually, $u(T)/T^4$ lies above $L + cT^2$ for c half the coefficient, and above $L - \varepsilon T^2$ for every $\varepsilon > 0$. The upper bound holds at every temperature, not only small ones: the kernel decreases, with $K(a) - K(b) \le (b-a)(1+b)K(a)$ for $0 \le a \le b$, which gives the pointwise bound $K(\sigma) \le K(b) + T^2m_1(y)$ with m_1 below an explicit separable majorant; and the box sums of $K(b)$ never exceed its space value by more than any given ε on fine grids, since the integrand is nonnegative, so a box holds no more than all of space. □

Corollary 26.61 (The correction depends on the parameter r). *The relative T^2 coefficient times the squared photon energy at the zone edge is strictly smaller at $r = 1$ than at $r = 4/3$: $\kappa(\alpha+\beta)E_1^2 < \kappa(\frac{3}{4}\alpha+\beta)E_{4/3}^2$, where E_r is the lattice photon energy at $(\pi,0,0)$.*

Proof. Since $0 < \alpha \le \beta$ (indeed $\beta = \frac{5}{3}\alpha$), $\kappa(\alpha+\beta)E_1^2 \le \frac{16}{5}\kappa(\alpha+\beta)$ and $\kappa(\frac{3}{4}\alpha+\beta)E_{4/3}^2 \ge \frac{19}{5}\kappa(\frac{3}{4}\alpha+\beta)$ by Proposition 26.59(2), and $16(\alpha+\beta) < 19(\frac{3}{4}\alpha+\beta)$ because $\frac{7}{4}\alpha < 3\beta$. □

So the size of the correction, measured against the photon energy at the zone edge, is not the same at $r = 1$ and at $r = 4/3$: it depends on the parameter r, not on the lattice alone.

Modes of a box, and two statements that needed the quarter period

On a periodic box with $N+1$ sites per side the field's own modes have wave numbers $2\pi m/(N+1)$. Theorem 12.101(3) shows that for a periodic integrand with a modulus of uniform continuity the averages over these modes converge to the zone average, with no principle; applied to $K(E(k)/T)$, the thermal energy summed over the box's own modes, divided by the number of modes, converges to $a(T)$. The characters of the box are then the plane waves.

Proposition 26.62 (Roots of unity and plane waves). *Let z be a pair of points, multiplied as a complex number, with $z^{N+1} = 1$. Then $z = e^{\mathrm{i}\,2\pi m/(N+1)}$ for exactly one natural number $m \le N$. Consequently every character ψ of the periodic box of side $N+1$, a map to the complex pairs with $\psi(0) = 1$, $\psi(x+y) = \psi(x)\psi(y)$ and period $N+1$ in each direction, is a plane wave $\psi(x) = e^{\mathrm{i}k\cdot x}$ with $k = 2\pi m/(N+1)$ for some $m \in \{0,\dots,N\}^3$. The first statement is false when $\pi = 2q$ is read with q any point having $q > 0$, $\cos q = 0$ and $\sin q = 1$: for $q' = 5q$, $N = 4$ and $z = e^{\mathrm{i}\,2\pi/5}$, no $m \le 4$ has $z = e^{\mathrm{i}\,4q'm/5}$.*

Proof. The points $\zeta_m = e^{\mathrm{i}2\pi m/(N+1)}$, $m \le N$, are roots of $z^{N+1} - 1$, pairwise at distance at least $2\sin(\pi/(N+1))$, so $z^{N+1} - 1 = \prod_m (z - \zeta_m)$. Given z with $z^{N+1} = 1$, $|z| = 1$. Let d be the least distance. Since $\prod_m |z - \zeta_m| = 0$, a finite comparison finds an m with $|z - \zeta_m| < d/2$: for each m decide, by approximation, whether $|z - \zeta_m| < d/2$ or $|z - \zeta_m| > d/4$; they cannot all exceed $d/4$, since then the product would be at least $(d/4)^{N+1} > 0$. This singles out one ζ_m near z. The product of the other factors is bounded away from 0 near ζ_m, and the whole product is 0, so $z = \zeta_m$. Uniqueness of m follows from the distances. A character with $\psi(0) = 1$ is determined by its values on the three unit steps, $\psi(x) = \prod_j \psi(e_j)^{x_j}$, each a root of unity of order $N+1$, since $\psi(e_j)^{N+1} = \psi((N+1)e_j) = \psi(0) = 1$. Without $\psi(0) = 1$ the zero map is a counterexample. For $q' = 5q$, $4q'm/5 = 4qm$, so $e^{\mathrm{i}4q'm/5} = 1$ for every m, while $e^{\mathrm{i}2\pi/5} \ne 1$. □

The same reading, every zero of the cosine with sine 1, also falsified two statements about real time, and the corrected forms are those of Proposition 26.59(4). As posed, the real-time ray limits and the real-time slope asked for them at roots in $[0, 2q']$ for every such q'. For $q' = 5q$ the root $4q - \kappa$ lies in $[0, 2q']$, satisfies the real-time shell, and has $\omega^2/\kappa^2 \ge 4/\kappa^2$, so the limit of ω^2/κ^2 fails; and a root function that jumps from ω_0 to $4q - \omega_0$ off one coordinate plane has no derivative there, which, asked for at every such function, would decide whether a point equals 0. With $q' \le 2$, that is with the quarter period, both statements hold.

26.10 Kepler's motion for all time

The rest of the chapter leaves the lattice of light: a polynomial field with a coercive first integral, the Kepler field, has a motion on the whole time line, and the lattice gauge expansions of Remark 26.64 state what is still missing below countable choice.

Theorem 26.63 (Kepler on the whole line)**.** *Let s be a physical state of the Kepler field (Definition 25.8) whose eight coordinates are ratios, with negative energy $E < 0$ and $|L|^2 > 0$. Then there is a motion $X : \mathbb{R}_\delta \to \mathbb{R}_\delta^8$ of the field, uniformly differentiable on bounded intervals, with $X(0) = s$. No omniscience principle and no choice principle is used.*

Proof. The motion is built explicitly, with no case split: no decision between a circle and an ellipse, no perihelion direction, no polar form of a vector. Write $\Lambda = |L|^2$ and $\ell = \sqrt{\Lambda}$. In the polar angle θ measured from the initial position the inverse radius is $U(\theta) = U_c + C\cos\theta + S'\sin\theta$, the solution of Binet's equation $U'' + U = m/\Lambda$, with $U_c = m/\Lambda$, $C = u - U_c$ and $S' = -u(x\cdot v)/\ell$ read from the state. The exact identity

$$2(mU + E) = \Lambda\big((U_c + C\cos\theta + S'\sin\theta)^2 + (C\sin\theta - S'\cos\theta)^2\big)$$

gives $mU \geq -E > 0$, so U has a positive lower bound. The polar angle as a function of time solves $\dot\Theta = \ell U(\Theta)^2$, $\Theta(0) = 0$; its rate is bounded above and below, and the clock construction of Theorem 25.40 (Lemma 25.22) gives Θ on the whole line. With the orthonormal pair $e_1 = ux$, $e_2 = u(L\times x)/\ell$ spanning the orbital plane, put $A = \cos\Theta\, e_1 + \sin\Theta\, e_2$ and $B = \cos\Theta\, e_2 - \sin\Theta\, e_1$, and

$$x(t) = A/U(\Theta), \quad v(t) = -\ell U'(\Theta)A + \ell U(\Theta)B, \quad u(t) = U(\Theta), \quad m(t) = m.$$

Differentiating, this is a motion of the Kepler field on the whole line, because $\Lambda U_c = m$, and at $t = 0$ it is the state s. □

Remark 26.64 (What the lattice expansions still need)**.** Six expansions of lattice gauge quantities hold given countable choice of natural numbers and are not known with no choice: the rate of a face phase, $|\Phi(h) - ch^2(g_y(0) - g_x(0))| \leq 2|c|M|h|^3$, and the cubic bounds for a face defect, for a plaquette and for a covariant difference, together with their two rates. Each follows from Taylor's step of Proposition 12.99, whose price with pointwise derivatives is open (Remark 12.104); with uniform derivatives each holds with no choice.

Questions to work through

1. (Theorem 26.11 and Theorem 26.15.) For the chain K of the opening, list the predecessor sets of the three configurations and the merging classes of K and of K^2. Find every distribution q with $D(qK \,\|\, p) = D(q \,\|\, p)$, and check that q_0 is one of them. Why does K^2 lower D strictly from every $q \neq p$, although K^2 has a zero entry?

2. (Theorem 26.16.) For the same chain, compute the largest δ with $(K^4)_{ij} \geq \delta\, p_j$ for all i, j, and the bound the theorem gives for $D(q_0K^8 \,\|\, p)$. Compare it with the true value.

3. (Theorem 26.23 and Theorem 26.24.) Let S be the ball of radius one, the vacuum and the six unit vectors, and P the Metropolis tick. Write down the walled tick P_S, check that it keeps π_S, and find its merging classes. Which Doeblin constant does Theorem 26.24 give at window 2, and what is the true largest one?

4. (Lemma 26.4 and Theorem 26.6.) Show that $g(\ell, q)$ is convex in q, and use the two-level system of the proof of Theorem 26.6 to show directly that "the free energy is $-\log Z$ or strictly above it" implies LPO.

5. (Proposition 26.43.) For two voxels under double entry, compute $D(\mathsf{P}_0 \,\|\, \Pi_c)$, $\mathrm{TC}(\mathsf{P}_0)$ and $D(\mathsf{P}_0 \,\|\, \pi \otimes \pi)$ at the empty record exactly, and check part (d) there. Show that the limit in part (e) is $D(\Pi_A \,\|\, \Pi_c)$, and write it with $\log 5$ and $\log \varphi$. By how much does the floor of part (d) exceed the floor $9 \log \varphi$ of Theorem 26.41(d)?

6. (Theorem 26.57.) Construct, for $D = 2$, a current with rational rates that is balanced off the origin and shell-uniform with $f(r) = 1/(2r + 1)$, by routing the flow around the shell $\|x\|_\infty = 1$ explicitly. Why can no integer multiple of it have integer rates?

Chapter 27

Gravity: the long-wave form of the cost and Regge's action

A geometry close to flat space is a metric $g = 1 + \varepsilon h$, with 1 the flat metric, h a symmetric strain and ε small. A ledger that compares the two geometries compares them through the matrix ratio $1^{-1}g = 1 + \varepsilon h$, whose eigenvalues are the stretches along the principal axes. When the ledger charges the cost $J(x) = \frac{1}{2}(x + x^{-1}) - 1$ of Chapter 21 for each stretch, what does it charge for the geometry, and is that charge the graviton?

Here is one wave. Use four coordinates, the zeroth read as time, take the wave covector $K = e_0$, and take the strain $h = e_1 e_2^{\mathsf{T}} + e_2 e_1^{\mathsf{T}}$, which shears the plane of the first two space axes. The matrix $1 + \varepsilon h$ has eigenvalues $1 + \varepsilon$, $1 - \varepsilon$, 1, 1, so the ledger charges

$$J(1+\varepsilon) + J(1-\varepsilon) = \frac{\varepsilon^2}{2(1+\varepsilon)} + \frac{\varepsilon^2}{2(1-\varepsilon)} = \frac{\varepsilon^2}{1-\varepsilon^2}.$$

Divided by ε^2 this tends to 1, which is $\frac{1}{2}\operatorname{tr}(h^2)$: the long-wave form of J. Linearized gravity charges a strain h at the wave K the Fierz–Pauli form [87]

$$\mathrm{FP}(h, K) = |K|^2 \operatorname{tr}(h^2) - 2|hK|^2 + 2 \operatorname{tr} h \, (K^{\mathsf{T}} h K) - |K|^2 (\operatorname{tr} h)^2,$$

which here is $2 - 0 + 0 - 0 = 2$. That is $|K|^2$ times twice the long-wave form of J. Two other strains at the same wave show where the agreement ends:

strain at $K = e_0$	$\frac{1}{2}\operatorname{tr}(h^2)$	$\mathrm{FP}(h, K)$	what it is
$e_1 e_2^{\mathsf{T}} + e_2 e_1^{\mathsf{T}}$	1	2	transverse and traceless
$e_0 e_1^{\mathsf{T}} + e_1 e_0^{\mathsf{T}}$	1	0	a change of coordinates
$\operatorname{diag}(0, 1, 1, 1)$	$\frac{3}{2}$	-6	the conformal mode

The ledger charges a mere change of coordinates, which gravity does not, and it charges the conformal mode positively, where gravity's form is negative. So the ledger's charge is not the gravitational action. On the strains that are transverse and traceless, however, the two agree exactly, at every wave, and the factor between them is the dispersion $|K|^2$.

This chapter proves that agreement and finds its scope. On transverse-traceless strains the graviton is J's long-wave form, in $D \geq 3$ space dimensions only, and J's calibration fixes the factor (Section 27.1); the identity holds over every commutative ring, hence on $\mathbb{R}_\delta$ with no choice, where the decisions about a wave have exact prices (Section 27.2). J allows every trace completion, locality selects Fierz–Pauli, locality itself rests on a depth bound for records or on taking the long-wave limit at a fixed stage, and J fixes neither the background nor Newton's constant (Section 27.3). Curvature, lattice deficits and Regge's table see exactly the non-gauge part, and no nonzero local two-derivative invariant form is nonnegative (Section 27.4).

Regge's action and J's odd hinge cost agree to second order only: J's cubic vertex breaks gauge invariance, so for a static field built from two harmonic modes the second-order equation has no solution and J's post-Newtonian parameter β_{PN} does not exist; a hinge cost linear in the deficit matches Regge's derivatives exactly when its area factor is linear to second order (Section 27.5). Schläfli's identity holds for every nondegenerate Euclidean 4-simplex, and Regge's Hessian is $\sqrt{|\det g_\kappa|}$ times one rational family of tables, with true second derivatives for every $\kappa < \frac{1}{4}$ and at $\kappa = 1$ (Sections 27.6 and 27.7); the near-flat chain from Regge's action to Fierz–Pauli, and its cubic term, run on $\mathbb{R}_\delta$, free, and at $\kappa = 1$ a wave with no deficits is a vertex displacement or, at zero wave, a constant metric strain (Section 27.7). The background $\kappa = 1$ is equivalent to the purity of one posting's lattice record and is not derived (Section 27.8); the locality clause can be a world's rule and is a named premise (Section 27.9); and reading J's comparison as the graviton's action needs two numbers at one wave, each resting on a named premise whose exact content we state (Section 27.10). Sections 27.11–27.16 go beyond linear gravity: Regge's long-wave vertices are one half of Einstein's through quartic order, with Regge's own derivatives on every Euclidean background g_κ, $\kappa < \frac{1}{4}$, and for the continued tables on every background off the poles; among hinge charges built from one function of the area, gauge consistency selects Regge's action through cubic order, and for shape costs it fixes Regge's action up to its coupling; J's cubic vertex is one quarter of Einstein's plus a part Δ that no redefinition removes; hinge shape costs cannot complete a changed coupling at quartic order; four gravitons scatter at tree level and long wavelength as in general relativity on every flat background g_κ off the poles; at order K^4 the graviton's two branch values are known on every Lorentzian background other than $\kappa = \frac{1}{2}$, and every background with $\kappa > \frac{1}{2}$, $\kappa = 1$ included, is birefringent; and each named premise of reading the cost as the graviton's action is separated from the foundation or reduced to one named reading.

Polynomial identities hold over every commutative ring and are free on $\mathbb{R}_\delta$; the native chain is proved on $\mathbb{R}_\delta$ directly, with δ's own π, cosine and square root, and its decisions carry the prices of Chapter 12. Statements about real matrices, limits, complex exponentials and records are proved in classical mathematics, and we claim no price on $\mathbb{R}_\delta$ for them; each section and scope box says which kind it states. Reading a strain as a spacetime metric, the Kuhn lattice as discretized spacetime and Regge's action as the gravitational action are identifications, named where they enter. The full treatment of linear gravity on the ledger is the preprint [95] (Zenodo, 2026); the vertices through quartic order are treated in [196], and the gravity program as a whole in [197]. Each result stated here has its proof here, in one of three forms that the proof makes plain: complete; resting on a long finite computation that the proof describes, with its inputs, without reproducing it; or, for Theorems 27.32 and 27.33, a route whose steps and constants the proof names without carrying the estimates out.

27.1 The long-wave form of J is the Fierz–Pauli form on transverse strains

We first fix the objects. Indices of four-dimensional vectors and matrices run over 0, 1, 2, 3; e_a is the ath unit vector and $E_{ab} = e_a e_b^{\mathsf{T}}$.

Definition 27.1 (Matrix cost, long-wave form, Fierz–Pauli form)**.**

1. For an invertible real $n \times n$ matrix A, the *cost of the ratio* A is $J(A) = \frac{1}{2}(\operatorname{tr} A + \operatorname{tr} A^{-1}) - n$. For a real $n \times n$ matrix X, *J's long-wave form* is $J_2(X) = \frac{1}{2}\operatorname{tr}(X^2)$.

2. A *strain* is a real symmetric 4×4 matrix h, and a *wave* is a covector $K \in \mathbb{R}^4$ with $|K|^2 = K \cdot K$. The *Fierz–Pauli form* $\mathrm{FP}(h, K)$ is the form displayed in the opening. A *gauge mode* at K is $K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}}$ with $\xi \in \mathbb{R}^4$.

3. For $K \neq 0$ the *transverse projector* is $\Pi_K = 1 - KK^{\mathsf{T}}/|K|^2$, and the *transverse strain* of h is $\Pi_K h \Pi_K$. The strain h is *transverse-traceless* (TT) at K when $hK = 0$ and $\operatorname{tr} h = 0$.

If A is diagonalizable with positive eigenvalues λ_i, then $J(A) = \sum_i J(\lambda_i)$: the cost of a ratio of metrics is J summed over the principal stretches. Π_K satisfies $\Pi_K^2 = \Pi_K$, $\Pi_K K = 0$ and $\operatorname{tr} \Pi_K = 3$. Reading $g_0^{-1} g$ as the comparison of two spacetime metrics, and FP as the action of linearized gravity, are identifications. The results of this section are proved in classical mathematics over the classical real numbers.

Theorem 27.2 (The exact expansion of the cost of a ratio)**.** *Let X be a real $n \times n$ matrix and ε a real number with $1 + \varepsilon X$ invertible. Then*

$$J(1 + \varepsilon X) = \frac{\varepsilon^2}{2} \operatorname{tr}\big(X^2 (1 + \varepsilon X)^{-1}\big),$$

and $J(1 + \varepsilon X)/\varepsilon^2 \to J_2(X)$ as $\varepsilon \to 0$ with $\varepsilon \neq 0$.

Proof. $(1 - \varepsilon X)(1 + \varepsilon X) = 1 - \varepsilon^2 X^2$, so $(1 + \varepsilon X)^{-1} = 1 - \varepsilon X + \varepsilon^2 X^2 (1 + \varepsilon X)^{-1}$. Take traces, add $\operatorname{tr}(1 + \varepsilon X) = n + \varepsilon \operatorname{tr} X$ and halve. The determinant of $1 + \varepsilon X$ is continuous with value 1 at $\varepsilon = 0$, so the matrix is invertible near 0, and inversion is continuous at the identity; hence $\frac{1}{2}\operatorname{tr}(X^2(1 + \varepsilon X)^{-1}) \to J_2(X)$. □

Theorem 27.3 (Fierz–Pauli is the dispersion times *J*'s long-wave form)**.** *Let h be a strain, $K \neq 0$ a wave and $A = \Pi_K h \Pi_K$ its transverse strain.*

(a) $\mathrm{FP}(h, K) = |K|^2\big(\operatorname{tr}(A^2) - (\operatorname{tr} A)^2\big) = |K|^2\big(2J_2(A) - (\operatorname{tr} A)^2\big)$.

(b) If h is TT at K, then $\mathrm{FP}(h, K) = |K|^2 \cdot 2J_2(h)$.

(c) A gauge mode $h = K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}}$ has $\mathrm{FP}(h, K) = 0$, *while $J_2(h) = (K \cdot \xi)^2 + |K|^2|\xi|^2$, which is positive when $\xi \neq 0$.*

(d) The conformal mode $h = \Pi_K$ has $\mathrm{FP}(\Pi_K, K) = -6|K|^2 < 0$, *while $J_2(\Pi_K) = \frac{3}{2} > 0$.*

Conditions and scope.

Base and premises

Real symmetric 4×4 matrices and real waves, with the Euclidean dot product, all four axes treated alike. Proved in classical mathematics.

Conclusion

On transverse-traceless strains, linearized gravity's form is exactly the dispersion times twice J's long-wave form; off them the two differ in gauge and in sign.

Not claimed

That the graviton's action is J's cost of comparing metrics: by (c) that cost is not gauge invariant. The Lorentzian form is Theorem 27.6.

Identification

h as a metric strain and FP as the action of linearized gravity.

Proof. (a) Put $s = |K|^2$ and $\Pi = \Pi_K$. Since $\Pi^2 = \Pi$, $\operatorname{tr} A = \operatorname{tr}(\Pi h) = \operatorname{tr} h - s^{-1} K^{\mathsf{T}} h K$ and $\operatorname{tr}(A^2) = \operatorname{tr}((\Pi h)^2)$. Expanding $\Pi h = h - s^{-1} K K^{\mathsf{T}} h$ and using $\operatorname{tr}(K K^{\mathsf{T}} h^2) = |hK|^2$ and $\operatorname{tr}((K K^{\mathsf{T}} h)^2) = (K^{\mathsf{T}} h K)^2$ gives $\operatorname{tr}(A^2) = \operatorname{tr}(h^2) - 2s^{-1}|hK|^2 + s^{-2}(K^{\mathsf{T}} h K)^2$. Hence

$$s\big(\operatorname{tr} A^2 - (\operatorname{tr} A)^2\big) = s \operatorname{tr} h^2 - 2|hK|^2 - s(\operatorname{tr} h)^2 + 2 \operatorname{tr} h\, K^{\mathsf{T}} h K = \mathrm{FP}(h, K),$$

the two terms $\pm s^{-1}(K^{\mathsf{T}} h K)^2$ cancelling. (b) If $hK = 0$ then $\Pi h \Pi = h$, so $A = h$ and $\operatorname{tr} A = 0$. (c) $\Pi K = 0$ gives $\Pi(K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}})\Pi = 0$, so $A = 0$ and (a) gives FP $= 0$. Also $\operatorname{tr}((K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}})^2) = 2(K \cdot \xi)^2 + 2|K|^2|\xi|^2$. (d) $A = \Pi$, $\operatorname{tr} A = \operatorname{tr} A^2 = 3$, so FP $= |K|^2(3 - 9)$, and $J_2(\Pi) = \frac{3}{2}$. □

Symmetry alone leaves the factor in front of FP free. J fixes it. To say this precisely we use the nine invariants of a strain and a wave under the symmetries of the cube (permutations and sign changes of the axes), which contain every two-derivative form a lattice of cubes can produce:

$$\begin{aligned} I_1 &= (\operatorname{tr} h)^2 |K|^2, & I_2 &= \operatorname{tr}(h^2)|K|^2, & I_3 &= \operatorname{tr} h\, K^{\mathsf{T}} h K, & I_4 &= |hK|^2, \\ A_1 &= \operatorname{tr} h \textstyle\sum_x h_{xx} K_x^2, & A_2 &= \textstyle\sum_{x,y} h_{xy}^2 K_x^2, & A_3 &= \textstyle\sum_x h_{xx} K_x (hK)_x, & A_4 &= |K|^2 \textstyle\sum_x h_{xx}^2, \end{aligned}$$

and $A_5 = \sum_x h_{xx}^2 K_x^2$, so that FP $= I_2 - 2I_4 + 2I_3 - I_1$. A form $\mathcal{Q}(h, K)$ is *gauge invariant* when $\mathcal{Q}(h + K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}}, K) = \mathcal{Q}(h, K)$ for every strain h and all K, ξ.

Theorem 27.4 (J's calibration fixes the factor)**.** *Let $\mathcal{Q} = \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_4 I_4 + \alpha_1 A_1 + \cdots + \alpha_5 A_5$ with real coefficients. If $\mathcal{Q}$ is gauge invariant, then $\mathcal{Q} = \gamma_2$ FP. Hence a gauge-invariant $\mathcal{Q}$ satisfies $\mathcal{Q}(h, K) = |K|^2 \cdot 2J_2(h)$ for every strain h that is TT at K if and only if $\mathcal{Q} =$ FP.*

Proof. Write the strain with independent entries h_{ab}, $a \le b$, and expand $\mathcal{Q}(h + K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}}, K) - \mathcal{Q}(h, K)$ as a polynomial in the h_{ab}, K_a and ξ_a. Gauge invariance makes every coefficient vanish. The coefficients of eight monomials are, by direct expansion of the nine invariants,

$$\begin{aligned} K_1^2 K_2 h_{00} \xi_2 &: \ 4\gamma_1 + 2\gamma_3, & K_0 K_1 K_2 h_{01} \xi_2 &: \ 4\gamma_3 + 4\gamma_4, \\ K_0 K_2^2 h_{01} \xi_1 &: \ 4\gamma_2 + 2\gamma_4, & K_0^4 \xi_1^2 &: \ 2\gamma_2 + \gamma_4 + \alpha_2, \\ K_1^3 h_{00} \xi_1 &: \ 4\gamma_1 + 2\gamma_3 + 2\alpha_1, & K_0^2 K_1 h_{00} \xi_1 &: \ 4\gamma_1 + 4\gamma_3 + 2\gamma_4 + 2\alpha_1 + \alpha_3, \\ K_0 K_1^2 h_{00} \xi_0 &: \ 4\gamma_1 + 4\gamma_2 + 2\gamma_3 + 2\gamma_4 + \alpha_3 + 4\alpha_4, & K_0^3 h_{00} \xi_0 &: \ 4(\gamma_1 + \cdots + \gamma_4 + \alpha_1 + \cdots + \alpha_5). \end{aligned}$$

Setting them to zero in this order gives $\gamma_3 = -2\gamma_1$, $\gamma_4 = 2\gamma_1$, $\gamma_2 = -\gamma_1$, $\alpha_2 = 0$, $\alpha_1 = 0$, $\alpha_3 = 0$, $\alpha_4 = 0$ and $\alpha_5 = 0$. So $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = \gamma_2(-1, 1, 2, -2)$ and $\mathcal{Q} = \gamma_2$FP. At the TT strain $h = E_{12} + E_{21}$ and $K = e_0$ we have $|K|^2 \cdot 2J_2(h) = 2 = \mathrm{FP}(h, K)$, so the calibration on TT strains forces $\gamma_2 = 1$. Conversely FP is gauge invariant (Theorem 27.3(a), since A does not change) and calibrated on TT strains (Theorem 27.3(b)). □

"J's calibration" is the normalization $J(x) = (x-1)^2/2 + O((x-1)^3)$ of Theorem 21.14. It fixes the factor that symmetry leaves free; it does not fix the physical strength of the action, which Proposition 27.14 takes up.

How many independent transverse-traceless strains are there? Take $D+1$ coordinates and read time along the all-ones vector $\mathbf{1} = (1, \dots, 1)$, as the lattice of Section 27.7 does. A strain H is *TT at the wave m* when it is symmetric and $H\mathbf{1} = 0$, $Hm = 0$ and $\operatorname{tr} H = 0$.

Theorem 27.5 (A graviton exactly in three or more space dimensions)**.** *Let $D \geq 0$ and let $m \in \mathbb{R}^{D+1}$ be a wave that is not a multiple of $\mathbf{1}$.*

(a) The strains TT at m form a real vector space of dimension $(D+1)(D-2)/2$, with $D-2$ read as 0 when $D < 2$.

(b) $J_2(H) > 0$ for every nonzero strain H that is TT at m.

(c) A nonzero TT strain exists at some wave that is not a multiple of $\mathbf{1}$ if and only if $D \geq 3$. For $D = 3$ there are two polarizations; for $D = 4$, five.

Proof. (a) Choose coordinates $i \neq j$ with $m_i \neq m_j$ and let R be the other $D-1$ coordinates. For $k \in R$ solve $a_k + b_k = 1$, $a_k m_i + b_k m_j = m_k$ (determinant $m_j - m_i \neq 0$) and put $r_k = e_k - a_k e_i - b_k e_j$, so that $r_k \cdot \mathbf{1} = 0 = r_k \cdot m$. For symmetric B on R, $H_B = \sum_{k,l \in R} B_{kl} r_k r_l^{\mathsf{T}}$ is symmetric with $H_B\mathbf{1} = 0 = H_B m$ and block B on R; and a symmetric H with $H\mathbf{1} = 0 = Hm$ equals H_B for its block B, since $H - H_B$ vanishes on $R \times R$ and its rows solve the same 2×2 system with zero right side. So these strains form a space of dimension $(D-1)D/2$, on which the trace is a nonzero functional when $D \geq 2$ ($\operatorname{tr} r_k r_k^{\mathsf{T}} = |r_k|^2 > 0$); for $D = 1$, R is empty. (b) $\operatorname{tr}(H^2) = \sum_{r,c} H_{rc}^2$ for symmetric H. (c) If $D \geq 3$, the wave e_0 is not a multiple of $\mathbf{1}$ and (a) gives dimension at least 2. If $D \leq 2$, (a) gives dimension 0 at every admissible wave, and when $D = 0$ there is no admissible wave. □

This is a lower bound: J gives no upper bound $D \leq 3$, and the three dimensions of Chapter 23 come from linking, not from this count.

27.2 The Lorentzian identity on $\mathbb{R}_\delta$

The identity of Theorem 27.3 is a polynomial identity once the inverse of the dispersion is given as data, and then it holds over every commutative ring. On $\mathbb{R}_\delta$ that inverse exists exactly when the dispersion is apart from 0.

For a commutative ring R, a symmetric 4×4 matrix S over R (an inverse metric), a wave $u \in R^4$ and symmetric X, Y, put $\eta = u^{\mathsf{T}} S u$ and

$$\begin{aligned} \mathrm{FP}_S(u; X, Y) = {} & \eta \operatorname{tr}(SXSY) - 2\, u^{\mathsf{T}} SXSYSu + \operatorname{tr}(SY)\, u^{\mathsf{T}} SXSu \\ & + \operatorname{tr}(SX)\, u^{\mathsf{T}} SYSu - \eta \operatorname{tr}(SX) \operatorname{tr}(SY). \end{aligned}$$

For $S = 1$ and $X = Y = h$ this is $\mathrm{FP}(h, u)$. The lattice of Recognition Science's gravity sector is taken at the cell metric $g_1 = 1 - \mathbf{1}\mathbf{1}^{\mathsf{T}}$ (the background $\kappa = 1$, a named premise; Section 27.8), in which each axis step is light-like; its inverse is $g_1^{-1} = 1 - \frac{1}{3}\mathbf{1}\mathbf{1}^{\mathsf{T}}$, and its dispersion is

$$m \cdot m = m^{\mathsf{T}} g_1^{-1} m = \sum_a m_a^2 - \tfrac{1}{3}\Big(\sum_a m_a\Big)^2.$$

We write $\mathrm{FP}_{g_1}(m; X, Y)$ for FP_S with $S = g_1^{-1}$. A strain H is TT at u when it is symmetric and $H\mathbf{1} = 0$, $Hu = 0$, $\operatorname{tr} H = 0$.

Theorem 27.6 (The Lorentzian graviton identity, over every ring)**.** *Let R be a commutative ring, S, X, Y symmetric and $u \in R^4$.*

(a) If $w\eta = 1$ for some $w \in R$, put $P = 1 - w(Su)u^{\mathsf{T}}$ and $A_X = P^{\mathsf{T}} X P$. Then

$$\mathrm{FP}_S(u; X, Y) = \eta\big(\operatorname{tr}(SA_XSA_Y) - \operatorname{tr}(SA_X)\operatorname{tr}(SA_Y)\big).$$

(b) $\mathrm{FP}_S(u; u\xi^{\mathsf{T}} + \xi u^{\mathsf{T}}, Y) = 0$ for every $\xi \in R^4$.

(c) For every $t \in R$, if X and Y are TT at u, then with $S = 1 - t\mathbf{1}\mathbf{1}^{\mathsf{T}}$, $\operatorname{tr}(SXSY) = \operatorname{tr}(XY)$ and $\operatorname{tr}(SX) = 0$.

(d) If $3t = 1$, then $(1 - \mathbf{1}\mathbf{1}^{\mathsf{T}})(1 - t\mathbf{1}\mathbf{1}^{\mathsf{T}}) = 1$, and with $S = 1 - t\mathbf{1}\mathbf{1}^{\mathsf{T}}$ the conformal mode g_1 has $\mathrm{FP}_S(u; g_1, g_1) = -6\eta$ while $\operatorname{tr}(Sg_1Sg_1) = 4$.

(e) On $\mathbb{R}_\delta$, every wave u whose dispersion $u \cdot u$ is apart from 0 has an inverse w of its dispersion.

So on $\mathbb{R}_\delta$, at every wave off the cone, $\mathrm{FP}_{g_1}(u; H, H) = (u \cdot u) \cdot 2J_2(H)$ for every TT strain H. All of this is free.

Conditions and scope.

Base and premises

Any commutative ring; on $\mathbb{R}_\delta$ the ring of Proposition 12.19. No omniscience principle and no choice principle.

Conclusion

The Lorentzian graviton form is the dispersion times J's polarized long-wave form of the transverse strain; gauge modes are null; on the lattice metric the conformal mode has the sign opposite to J's.

Not claimed

Anything on the cone $u \cdot u = 0$, where the projector P does not exist; deciding whether a wave is on the cone has a price (Theorem 27.8).

Proof. (a) Put $v = Su$ and $S_\perp = S - wvv^{\mathsf{T}}$. Using $w\eta = 1$ and $u^{\mathsf{T}}S = v^{\mathsf{T}}$: $SP^{\mathsf{T}} = S - wvv^{\mathsf{T}} = S_\perp$, and $PS_\perp = S_\perp - wvv^{\mathsf{T}} - wvv^{\mathsf{T}} + w^2\eta vv^{\mathsf{T}} = S_\perp$. Hence $\operatorname{tr}(SA_XSA_Y) = \operatorname{tr}(PS_\perp XS_\perp Y) = \operatorname{tr}(S_\perp XS_\perp Y)$ and $\operatorname{tr}(SA_X) = \operatorname{tr}(S_\perp X)$. Expanding,

$$\operatorname{tr}(S_\perp XS_\perp Y) = \operatorname{tr}(SXSY) - 2wv^{\mathsf{T}}XSYv + w^2(v^{\mathsf{T}}Xv)(v^{\mathsf{T}}Yv),$$
$$\operatorname{tr}(S_\perp X)\operatorname{tr}(S_\perp Y) = \operatorname{tr}(SX)\operatorname{tr}(SY) - w\operatorname{tr}(SX)v^{\mathsf{T}}Yv - w\operatorname{tr}(SY)v^{\mathsf{T}}Xv + w^2(v^{\mathsf{T}}Xv)(v^{\mathsf{T}}Yv).$$

The w^2 terms cancel in the difference; multiplying by η and using $w\eta = 1$ gives the five terms of FP_S, since $v^{\mathsf T}XSYv = u^{\mathsf T}SXSYSu$. (b) With $X = u\xi^{\mathsf T} + \xi u^{\mathsf T}$, $\zeta = S\xi$ and $a = \xi^{\mathsf T}Su$: $SXS = v\zeta^{\mathsf T} + \zeta v^{\mathsf T}$, so the five terms are $2\eta\,\zeta^{\mathsf T}Yv$, $-2(\eta\,\zeta^{\mathsf T}Yv + a\,v^{\mathsf T}Yv)$, $2\eta a\,\mathrm{tr}(SY)$, $2a\,v^{\mathsf T}Yv$ and $-2\eta a\,\mathrm{tr}(SY)$, which add to 0. (c) $\mathbf{1}^{\mathsf T}X = (X\mathbf{1})^{\mathsf T} = 0$, so $SX = X$ and $SY = Y$. (d) $\mathbf{1}^{\mathsf T}\mathbf{1} = 4$ gives $(1 - \mathbf{1}\mathbf{1}^{\mathsf T})(1 - t\mathbf{1}\mathbf{1}^{\mathsf T}) = 1 - (1 - 3t)\mathbf{1}\mathbf{1}^{\mathsf T}$. With $Sg_1 = 1$ the five terms are 4η, -2η, 4η, 4η and -16η. (e) is Proposition 12.19, item 3. The final sentence combines (a), (c) and (d): for TT H, $A_H = H$ because $Hu = 0$, and $\mathrm{tr}(SHSH) = \mathrm{tr}(H^2)$. □

The same identity shows that J's transverse form cannot see the background. Let $g_\kappa = 1 - \kappa\mathbf{1}\mathbf{1}^{\mathsf T}$; $\kappa = 0$ is the Euclidean metric, $\kappa = 1$ the lattice metric, and $\det g_\kappa = 1 - 4\kappa$, so $\kappa = \frac{1}{4}$ is degenerate. This κ is the background parameter of the lattice, not the Einstein coupling of Chapter 22.

Proposition 27.7 (Fierz–Pauli is affine in the background; J is blind to it)**.** *Over every commutative ring, with $S_t = 1 + t\mathbf{1}\mathbf{1}^{\mathsf T}$:*

(a) for every wave u and symmetric H, $\mathrm{FP}_{S_t}(u; H, H) = \mathrm{FP}_1(u; H, H) + t\,B(u, H)$, where, with $o = \mathbf{1}$,

$$B(u,H) = (o\cdot u)^2\big(\mathrm{tr}\,H^2 - (\mathrm{tr}\,H)^2\big) - 2|u|^2(o^{\mathsf T}Ho)\,\mathrm{tr}\,H + 2|u|^2|Ho|^2 + 4(o\cdot u)\,\mathrm{tr}\,H(o^{\mathsf T}Hu) - 4(o\cdot u)(o^{\mathsf T}H^2u) + 2(o^{\mathsf T}Ho)(u^{\mathsf T}Hu) - 2(o^{\mathsf T}Hu)^2;$$

(b) $g_\kappa S_t = 1$ whenever $t(1 - 4\kappa) = \kappa$;

(c) for symmetric X, Y with $X\mathbf{1} = 0 = Y\mathbf{1}$, $\mathrm{tr}(S_tXS_tY) = \mathrm{tr}(XY)$ and $\mathrm{tr}(S_tX) = \mathrm{tr}\,X$, whatever t.

The proofs are free on $\mathbb{R}_\delta$.

Proof. (a) Each of the five terms of FP_{S_t} is a polynomial in t of degree at most four. Expanding with $S_t = 1 + t\,oo^{\mathsf T}$ and $o^{\mathsf T}o = 4$, the coefficients of t^2, t^3 and t^4 cancel and the coefficient of t is $B(u, H)$; the expansion is a finite computation with integer coefficients. (b) $g_\kappa S_t = 1 + (t - \kappa - 4\kappa t)oo^{\mathsf T}$. (c) $o^{\mathsf T}X = 0$ removes every t. □

So the background enters Fierz–Pauli linearly and J's transverse form not at all: J does not fix κ. Section 27.8 shows what does.

On $\mathbb{R}_\delta$ the questions one asks about a wave have exact prices. The *spatial part* of a wave m is $n = m - \frac{1}{4}(\sum_a m_a)\mathbf{1}$, and when its squared length $|n|^2$ has an inverse r, the projector $\Pi = 1 - \frac{1}{4}\mathbf{1}\mathbf{1}^{\mathsf T} - r\,nn^{\mathsf T}$ and the map $\Lambda_m(H) = \Pi H\Pi - \frac{1}{2}\,\mathrm{tr}(\Pi H)\Pi$ are defined. The principles are those of Definition 7.8.

Theorem 27.8 (Cone, sign and projector on $\mathbb{R}_\delta$)**.** *On $\mathbb{R}_\delta$, with no choice principle:*

(a) $g_1 g_1^{-1} = 1 = g_1^{-1} g_1$, and the dispersion is $m\cdot m = \sum_a m_a^2 - \frac{1}{3}(\sum_a m_a)^2$;

(b) if $|n|^2$ is apart from 0, then Λ_m sends every symmetric strain to a strain TT at m, satisfies $\Lambda_m \circ \Lambda_m = \Lambda_m$ and fixes every strain TT at m; all free;

(c) deciding $m\cdot m = 0 \vee m\cdot m \neq 0$ for every wave m is equivalent to WLPO;

(d) deciding $m \cdot m > 0 \vee m \cdot m = 0 \vee m \cdot m < 0$ for every wave is equivalent to LPO;

(e) "if $|n|^2 \neq 0$, then $|n|^2$ has an inverse", for every wave, is equivalent to $\mathrm{MP}_{\mathrm{bin}}$*;*

(f) for integer waves, $m \cdot m = 0$ exactly when $3\sum_a m_a^2 = (\sum_a m_a)^2$, which is decided free.

Proof. (a) is Theorem 27.6(d) with $t = \frac{1}{3}$. (b) is a polynomial computation in r using $r|n|^2 = 1$, $\Pi\mathbf{1} = 0$, $\Pi n = 0$, $\Pi^2 = \Pi$ and $\operatorname{tr}\Pi = 2$. For (c)–(e): every point x is the dispersion of a wave, namely $m = \frac{1}{2}(1+x)\hat{n} + \frac{\sqrt{3}}{4}(1-x)\mathbf{1}$ with $\hat{n} = (e_0 - e_1)/\sqrt{2}$, since then $m \cdot m = \frac{1}{4}(1+x)^2 - \frac{1}{4}(1-x)^2 = x$; and every point $x \geq 0$ is $|n|^2$ for the wave $m = y(e_0 - e_1)$ with $y = \sqrt{x/2}$. Conversely each dispersion and each $|n|^2$ is a point. So each statement is equivalent to the corresponding statement for all points, or for all nonnegative points, and Corollary 12.18 prices those; its proofs of the WLPO and $\mathrm{MP}_{\mathrm{bin}}$ cases use only nonnegative points, and a point has an inverse exactly when it is apart from 0 (Proposition 12.19). (f) Multiply by 3. □

27.3 Completions, locality and the scale

J fixes the graviton's form on transverse-traceless strains. A full quadratic form must also say what to charge for the trace of the transverse strain, and J allows every answer. Locality decides.

Definition 27.9 (Completions and local lattice costs)**.**

1. For real β, the *completion* C_β is $C_\beta(h, K) = |K|^2\big(\operatorname{tr} A^2 + \beta(\operatorname{tr} A)^2\big)$ with $A = \Pi_K h \Pi_K$, and $C_\beta(h, 0) = 0$.
2. A *completion of J's transverse form* is a function $C(h, K)$ of strains and waves that is gauge invariant, invariant under every orthogonal R ($C(RhR^{\mathsf{T}}, RK) = C(h, K)$), homogeneous ($C(h, tK) = t^2 C(h, K)$), a quadratic form in h for each K, and equal to $|K|^2 \cdot 2J_2(h)$ whenever h is TT at K.
3. A *local lattice cost* is a finite list of terms j, each with two lattice offsets $p_j, q_j \in \mathbb{Z}^4$ and a real coefficient $c_j(H_0, H_1)$ depending on two strains. On the plane-wave pair $H_0 e^{iK\cdot x}$, $H_1 e^{-iK\cdot x}$ its cost at the site $x \in \mathbb{Z}^4$ is $\sum_j c_j e^{iK\cdot(x+p_j)} e^{-iK\cdot(x+q_j)}$. Its *symbol* is $\sum_j c_j \cos(K \cdot d_j)$ with $d_j = p_j - q_j$, and its *long-wave symbol* is $-\frac{1}{2}\sum_j c_j (K \cdot d_j)^2$.

The parameter β is DeWitt's [86]: $\beta = -1$ is Fierz–Pauli (Theorem 27.3(a)) and $\beta = 0$ is J's own long-wave form of the transverse strain. A local cost compares strains at lattice sites a fixed distance apart, with signed coefficients; the lattice actions of Section 27.4 are of this kind. The results of this section are proved in classical mathematics.

Theorem 27.10 (J allows every completion)**.**

(a) Every C_β is a completion of J's transverse form; $C_{-1} =$ FP on strains, and $C_{-1} \neq C_0$.

(b) Every completion C equals C_β on strains, with $\beta = (C(\Pi_{e_0}, e_0) - 3)/9$.

(c) Each C_β is built from J: for every strain h and wave $K \neq 0$, with $A = \Pi_K h \Pi_K$,

$$2|K|^2 \, \frac{J(1+\varepsilon A) + \beta\, J\big(\det(1+\varepsilon A)\big)}{\varepsilon^2} \longrightarrow C_\beta(h,K) \qquad (\varepsilon \to 0,\ \varepsilon \neq 0),$$

and the volume cost obeys J's composition law: for real matrices M, N with $\det M, \det N > 0$, $J(\det MN) + J(\det MN^{-1}) = 2J(\det M)J(\det N) + 2J(\det M) + 2J(\det N)$.

(d) For $\beta \geq 0$, $C_\beta(h,K) \geq 0$ for every strain, while $C_{-1}(1,K) = -6|K|^2$ for $K \neq 0$.

So no property that J imposes on the strain selects β, and J's positivity points away from Fierz–Pauli.

Proof. (a) $\Pi_K(K\xi^{\mathsf T} + \xi K^{\mathsf T})\Pi_K = 0$ gives gauge invariance; $\Pi_{RK} = R\Pi_K R^{\mathsf T}$ gives orthogonal invariance; $\Pi_{tK} = \Pi_K$ for $t \neq 0$ gives homogeneity; A is linear in h; and on TT strains $A = h$ and $\operatorname{tr} A = 0$. $C_{-1} = \mathrm{FP}$ is Theorem 27.3(a). $C_{-1} - C_0 = -|K|^2(\operatorname{tr} A)^2$ is -1 at $h = E_{00}$, $K = e_1$. (b) First take $K = e_0$. Adding the gauge mode $e_0\xi^{\mathsf T} + \xi e_0^{\mathsf T}$ with $\xi = -(\frac{1}{2}h_{00}, h_{01}, h_{02}, h_{03})$ clears the zeroth row and column, so $C(h, e_0)$ depends only on the spatial block a of h, a symmetric 3×3 matrix, through a quadratic form invariant under permutations and sign changes of the three space axes. Write $a = a_0 + \frac{1}{3}(\operatorname{tr} a)1$ with a_0 traceless. The polarization of the form between a_0 and 1 is a linear functional of a_0 invariant under the sign changes, so it ignores the off-diagonal entries, and under the permutations, so it is a multiple of $\operatorname{tr} a_0 = 0$. A traceless a, padded with zeros, is TT at e_0, so the form equals $\operatorname{tr}(a_0^2)$ there. Hence $C(h, e_0) = \operatorname{tr}(a_0^2) + \frac{1}{9}(\operatorname{tr} a)^2 C(1) = \operatorname{tr}(a^2) + \beta(\operatorname{tr} a)^2$ with $\beta = (C(1) - 3)/9$, and a is the transverse strain at e_0. For $K \neq 0$ a reflection R maps $K/|K|$ to e_0, and invariance and homogeneity give $C(h,K) = |K|^2 C(RhR^{\mathsf T}, e_0) = C_\beta(h,K)$; at $K = 0$ both vanish. (c) Theorem 27.2 gives $J(1+\varepsilon A)/\varepsilon^2 \to \frac{1}{2}\operatorname{tr} A^2$, and $\det(1+\varepsilon A) = 1 + \varepsilon \operatorname{tr} A + O(\varepsilon^2)$ with $J(1+y) = y^2/(2(1+y))$ gives $J(\det(1+\varepsilon A))/\varepsilon^2 \to \frac{1}{2}(\operatorname{tr} A)^2$. The law is J's composition law $J(xy) + J(x/y) = 2J(x)J(y) + 2J(x) + 2J(y)$ of Chapter 21 at $x = \det M$, $y = \det N$, since $\det(MN^{-1}) = \det M / \det N$. (d) For $\beta \geq 0$ both terms are nonnegative, A being symmetric. At $h = 1$, $A = \Pi_K$. □

The trace part is fixed on the other side of the comparison: on the wave covector. Chapter 21's composition law, read in a vector variable, passes to its long-wave limit as the parallelogram law.

Theorem 27.11 (The parallelogram law selects Fierz–Pauli)**.**

(a) Let F and q be real functions on $\mathbb{R}^m$ with $F(x+y) + F(x-y) = 2F(x)F(y) + 2F(x) + 2F(y)$ for all x, y and $F(\varepsilon x)/\varepsilon^2 \to q(x)$ for every x. Then $q(x+y) + q(x-y) = 2q(x) + 2q(y)$ for all x, y. No regularity of F is assumed.

(b) C_β obeys the parallelogram law in K for every strain h if and only if $\beta = -1$.

Proof. (a) Replace x, y by $\varepsilon x, \varepsilon y$ and divide by ε^2. The product term $2F(\varepsilon x)F(\varepsilon y)/\varepsilon^2 = 2\varepsilon^2 \cdot (F(\varepsilon x)/\varepsilon^2)(F(\varepsilon y)/\varepsilon^2)$ tends to 0, and the other terms tend to the parallelogram law. (b) $C_\beta = \mathrm{FP} + (1+\beta)|K|^2(\operatorname{tr} A)^2$, and FP is a quadratic polynomial in K, so it obeys the law.

The extra term is $(|K|^2 \operatorname{tr} h - K^{\mathsf{T}} h K)^2/|K|^2$. At $h = E_{00}$ it equals $(|K|^2 - K_0^2)^2/|K|^2$, with values $\frac{1}{2}$, $\frac{1}{2}$, 0, 1 at $e_0 + e_1$, $e_0 - e_1$, e_0, e_1. So the defect of the law for C_β at $K = e_0$, $K' = e_1$ is $(1+\beta)(\frac{1}{2} + \frac{1}{2} - 2 \cdot 0 - 2 \cdot 1) = -(1+\beta)$. □

Locality supplies the parallelogram law. A lattice-local cost compares strains at fixed offsets, and its symbol is built from J evaluated on unimodular record ratios. We extend J to nonzero complex z by $J(z) = \frac{1}{2}(z + z^{-1}) - 1$.

Theorem 27.12 (Locality selects the Fierz–Pauli completion)**.** *Let L be a local lattice cost, H_0, H_1 strains and K a wave.*

(a) The real part of the cost at every site x is the symbol $\sum_j c_j \cos(K \cdot d_j)$; it does not depend on x.

(b) The symbol equals $\sum_j c_j\big(1 + J(r_j)\big)$, where $r_j = e^{iK\cdot d_j}$ is the record ratio of term j, and $r_j(K+K') = r_j(K)\, r_j(K')$. Each $K \mapsto \cos(K \cdot d) - 1$ obeys the composition law in K.

(c) If $\sum_j c_j = 0$, the symbol at εK divided by ε^2 tends to the long-wave symbol, which obeys the parallelogram law in K.

(d) If for every strain h some local cost has long-wave symbol $C_\beta(h, \cdot)$ on the pair (h, h), then $\beta = -1$.

(e) Fierz–Pauli is realized: a local cost with 32 terms at the offsets $\pm(e_a \pm e_b)$ and coefficients bilinear in (H_0, H_1), with $\sum_j c_j = 0$, has long-wave symbol $\mathrm{FP}(h, K)$ on every pair (h, h).

Conditions and scope.

Base and premises
Classical real and complex numbers; Euclidean waves.

Conclusion
Every cost that compares strains at fixed lattice offsets has a long-wave symbol obeying the parallelogram law, and among the completions only Fierz–Pauli has that property; Fierz–Pauli is realized by such a cost.

Identification
That the graviton's long-wave action is the symbol of a local lattice cost (the Regge lattice of Sections 27.4 and 27.7).

Not claimed
That the realizing table of (e) is gauge null at every momentum; only its long-wave symbol is computed.

Proof. (a) $e^{iK\cdot(x+p)} e^{-iK\cdot(x+q)} = e^{iK\cdot(p-q)}$: the site drops out. (b) $\frac{1}{2}(z + z^{-1}) = \cos(K \cdot d)$ for $z = e^{iK\cdot d}$; exponents add; and $\cos(a+b) + \cos(a-b) = 2\cos a \cos b$ is the composition law for $\cos - 1$. (c) $|\cos t - 1 + t^2/2| \le t^4/24$, and $\sum c_j = 0$ removes the constant; the limit is a quadratic form in K. (d) (c) and Theorem 27.11(b). (e) Write the polarized form as $\mathrm{FP}(H_0, H_1; K) = \sum_{a,b} M_{ab}(H_0, H_1) K_a K_b$ with M bilinear. A term with coefficient c at offset d

contributes $-\frac{1}{2}c(K\cdot d)^2$ to the long-wave symbol, and $(K_a+K_b)^2-(K_a-K_b)^2=4K_aK_b$. So the coefficient $-\frac{1}{2}M_{ab}$ at offset e_a+e_b and $+\frac{1}{2}M_{ab}$ at e_a-e_b, for each ordered pair (a,b), give long-wave symbol $\sum M_{ab}K_aK_b$, and the coefficients sum to 0. □

Theorem 27.12 takes the cost to be local; a charge reads records, which reach as far as the rounds of events behind them. A *law* gives the next record (a finite bit string) of every finite list of records. The post on n records *depends on* position i when two lists of length n that differ only there get different posts; a *chain* from n is a list $i_1,\dots,i_m$ with the post on n records depending on i_1 and that on i_k records on i_{k+1}. The *flip law* posts the complement of the latest record's first bit (a missing bit read as 0), the *comparison law* whether the two latest records are equal. On $\mathbb{Z}^4$ with the sup distance, a *rule of radius r* writes each site's new record from the strains and records within r; *n rounds* from a constant record give records of depth n. A charge on strain fields has *strain range R* when its value at x depends only on strains within R of x, and *depth at most T* with radius r when it reads, within r of x, the strain and the records of T rounds of a rule of radius at most r. Charges Φ_n, one per stage n, have *stage depth*, or *uniform depth T*, when each Φ_n has depth at most n, or at most T. In the *potential world* sites keep the strain h and a potential $\psi\in\mathbb{R}$, and the cost P_β has terms at the offsets 0, e_a and e_a+e_b, coefficients bilinear in the two records and summing to zero, and long-wave symbol $\mathrm{FP}(h,K)+(1+\beta)(2s\psi-|K|^2\psi^2)$ on (r,r), $r=(h,\psi)$, $s=|K|^2\operatorname{tr}h-K^{\mathsf T}hK$; these fix its total coefficient at each offset on equal records (by the K_aK_b and K_a^2 terms). Its cost per site on plane waves of amplitude (h,ψ) is $P_\beta(h,\psi;K)=\mathscr{A}+\mathscr{B}\psi+\mathscr{C}\psi^2$, with *stationary potential* $\psi_*=-\mathscr{B}/(2\mathscr{C})$ (0 where $\mathscr{C}=0$). The *relaxation* with step ω, the rule of radius one $\psi\mapsto\psi-\omega(2\mathscr{C}\psi+\mathscr{B}h)$ with stencils $\mathscr{B}$, $\mathscr{C}$, writes on plane waves $\psi_0=0$, $\psi_{n+1}=(1-2\omega\mathscr{C})\psi_n-\omega\mathscr{B}$.

Theorem 27.13 (Locality rests on record depth and an order of limits)**.** *In classical mathematics:*

(a) every law's chains from n have length at most n; the flip law and the comparison law (which no single record decides from two records on, and which reads at most two records posted before any time t) have unbounded chains, of length n from n and from $n+1$ respectively;

(b) a charge of depth at most T with radius r has strain range $r+Tr$. So stage depth gives range $r+nr$ at stage n, and uniform depth T, or depth at most $f(n)\le T$, range $r+Tr$ at every stage;

(c) the running total $\sum_{k<n}h_{00}(x-ke_0)$ has stage depth with $r=1$, but at stage $n+1$ depends on the strain at $x-ne_0$: no strain range and no depth, for any radius, is uniform in the stage;

(d) if for every strain h some local cost L over records (h,x), $x\in\mathbb{R}^m$, with $\sum_j c_j(r,r)=0$ and each $c_j(r,r)$ continuous in r, read at $x_n(\varepsilon K)$, where $x_{k+1}=Mx_k+b$ with $M(K)$, $b(K)$ continuous in K and x_0, n fixed, has cost per site at the wave εK over ε^2 tending to $C_\beta(h,K)$ for every K, then $\beta=-1$. In the potential world, for all β, ω, n, the charge after n rounds of relaxation has strain range $1+n$, and $P_\beta(h,\psi_n(\varepsilon K);\varepsilon K)/\varepsilon^2\to\mathrm{FP}(h,K)$;

(e) in the potential world, if $\omega\ne0$ and $\mathscr{C}(K)\ne0$, then

$$P_\beta(h,\psi_n;K)=P_\beta(h,\psi_*;K)+\mathscr{C}(K)(1-2\omega\mathscr{C}(K))^{2n}\psi_*(K)^2$$

for every n, and for every β, strain h and wave K, $P_\beta(h,\psi_*(\varepsilon K);\varepsilon K)/\varepsilon^2\to C_\beta(h,K)$. If $\omega(1+\beta)<0$ and $|K|^2\ne0$, then for every $f:\mathbb{N}\to\mathbb{N}$ with $f(n)\to\infty$ and all small $\varepsilon\ne0$, $P_\beta(h,\psi_{f(n)}(\varepsilon K);\varepsilon K)\to P_\beta(h,\psi_*(\varepsilon K);\varepsilon K)$. If $\beta\ne-1$, $C_\beta(E_{00},e_1)\ne\mathrm{FP}(E_{00},e_1)$.

Proof. (a) Lists of length n that differ only at a position $i \geq n$ are equal, so the positions of a chain decrease. Changing the latest of $m+1$ records changes the flip law's post, and changing the latest of $m+2$ equal records the comparison law's, which gives the chains $n-1,\dots,0$ and $n,\dots,1$; the comparison law reads only the two latest records, each of which can change its post. (b) By induction a record of depth n at x reads only strain within nr of x. (c) The rule $\psi(x) \mapsto \psi(x-e_0)+h_{00}(x)$ writes the sum, and changing h_{00} at $x-ne_0$ alone changes it at stage $n+1$, which defeats a range $R<n$; a uniform depth would give a uniform range by (b). (d) The records $x_n(K)$ are continuous in K, so the coefficients at $(h, x_n(\varepsilon K))$ tend to those at $(h, x_n(0))$; as they sum to zero, $|\cos t - 1 + t^2/2| \leq t^4/24$ makes the limit the long-wave symbol of L at $(h, x_n(0))$, a quadratic form in K, which obeys the parallelogram law; Theorem 27.11(b) gives $\beta = -1$. In the potential world ψ_n is continuous with $\psi_n(0)=0$, as $\mathscr{B}(0)$ is a sum of coefficients, and the limit is the long-wave symbol at $\psi = 0$, $\mathrm{FP}(h,K)$. Both rules have radius one; apply (b). (e) For $\mathscr{C} \neq 0$, $P_\beta = \mathscr{A} - \mathscr{B}^2/(4\mathscr{C}) + \mathscr{C}(\psi-\psi_*)^2$ and $\psi_n - \psi_* = -(1-2\omega\mathscr{C})^n\psi_*$. As in (d), $\mathscr{A}, \mathscr{B}, \mathscr{C}$ at εK over ε^2 tend to $\mathrm{FP}(h,K)$, $2(1+\beta)s$, $-(1+\beta)|K|^2$, so for $1+\beta \neq 0 \neq |K|^2$ the stationary cost over ε^2 tends to $\mathrm{FP}+(1+\beta)s^2/|K|^2 = C_\beta(h,K)$ (proof of Theorem 27.11(b)); for $1+\beta = 0$, $\mathscr{B} = \mathscr{C} = 0$; at $K=0$ both sides vanish. If $\omega(1+\beta) < 0$, then $\omega\mathscr{C}(\varepsilon K)/\varepsilon^2 \to -\omega(1+\beta)|K|^2 > 0$ and $\mathscr{C}(\varepsilon K) \to 0$ give $0 < \omega\mathscr{C} < \frac{1}{2}$ for small $\varepsilon \neq 0$. Finally $s(E_{00}, e_1) = 1$, so there $C_\beta - \mathrm{FP} = 1+\beta$. □

Reading a round of the ledger's events as a stage is an identification. The potential world, local in its records, reaches every completion with the stage first, which for $\beta \neq -1$ no cost local in the strain does (Theorem 27.12(d)). Locality of the cost stays a named premise: the long-wave limit at a fixed stage, or a depth bound that does not grow with the stage.

Finally, the strength. J's calibration fixes the coefficient of FP in J's unit, and nothing in J converts that unit into Newton's constant. To see what would, scale the $\kappa = 1$ lattice action of Section 27.7 by s, so that its long-wave quadratic part on a wave u and amplitude h is $-(s\sqrt{3}/16)\,\mathrm{FP}_{g_1}(u;h,h)$, and read Newton's constant from Einstein's normalization, $G_s = 1/(8\pi s)$. Write $G = \varphi^5/(4\log 2)$ for the value of Newton's constant in Recognition Science's units, $\hbar = \varphi^{-5}$ for the action quantum and λ for the recognition length (Chapter 22). The *posting premise* states that one posting, a deficit of $\pi/6$, on a timelike hinge of area A carries the action $a(A) = s\,A\,\pi/6$; that a adds over area and does not decrease with it; and that a full turn of twelve postings on a face of area λ^2 counts the face's one bit, $\exp(12\,a(\lambda^2)/\hbar) = 2$.

Proposition 27.14 (J does not fix Newton's constant; the posting bit does)**.** *In classical mathematics:*

(a) if the coefficient of $-\mathrm{FP}_{g_1}$ is a number $v > 0$, then $s = 16v/\sqrt{3}$ and $G_s = \sqrt{3}/(128\pi v)$; this equals G if and only if $v = v^* = \sqrt{3}/(128\pi G) = \sqrt{3}\log 2/(32\pi\varphi^5)$;

(b) $928 < 1/v^* < 929$, so $v^* \approx 1/928.65$;

(c) the posting premise holds exactly at $s = \hbar\log 2/(2\pi\lambda^2) = 1/(8\pi G)$, that is, at $v = v^*$.

Proof. (a) $1/(8\pi\cdot 16v/\sqrt{3}) = \sqrt{3}/(128\pi v)$, and $128\pi G = 32\pi\varphi^5/\log 2$. (b) With $\varphi^5 = 11.0902$, $\pi = 3.14159$, $\sqrt{3} = 1.73205$ and $\log 2 = 0.693147$ to the digits shown, $32\pi\varphi^5/(\sqrt{3}\log 2) = 1114.89/1.20057 = 928.65$, and the errors of the rounded values move this by less than 0.1.

(c) The bit condition reads $12\,s\lambda^2\pi/6 = \hbar\log 2$, that is, $s = \hbar\log 2/(2\pi\lambda^2)$; conversely, at this s the action $a(A) = sA\pi/6$ meets every clause. Finally $\hbar\log 2/(2\pi\lambda^2) = 1/(8\pi G)$ because $G = \lambda^2/(4\hbar\log 2)$ in these units, and then $\nu = (\sqrt{3}/16)s = \nu^*$. □

No Newton constant is derived here. The number ν^* is the product of the face bit ($\log 2$ per full turn), the action quantum $\hbar = \varphi^{-5}$ and the cell factor $\sqrt{3}/16$ of the $\kappa = 1$ lattice; none of these is supplied by J. That the timelike hinge carries the posting action is a named premise.

27.4 Holonomies, deficits and the sign of the graviton term

A gauge mode is a change of coordinates, and a quantity that ignores gauge modes depends on the geometry only through what coordinates cannot remove: curvature in the continuum, holonomies around hinges on a lattice. This section proves that both descriptions see exactly the non-gauge part, that Regge's lattice action pairs a hinge's holonomy with its change of area, and that such a pairing must have both signs. All results are proved in classical mathematics.

In momentum space the linearized Riemann tensor of a strain H at the wave K is

$$R_{abcd}(H,K) = K_bK_cH_{ad} + K_aK_dH_{bc} - K_aK_cH_{bd} - K_bK_dH_{ac},$$

with Ricci tensor $R_{bd} = \sum_a R_{abad}$, scalar $R = \sum_b R_{bb}$ and Einstein tensor $G_{bd} = R_{bd} - \frac{1}{2}\delta_{bd}R$.

Theorem 27.15 (Zero curvature is gauge; Fierz–Pauli pairs a strain with its Einstein tensor)**.** *For every strain H and wave K:*

(a) $R_{abcd}(K\xi^{\mathsf T} + \xi K^{\mathsf T}, K) = 0$ *for every ξ, and* $R_{abcd}(H, tK) = t^2R_{abcd}(H,K)$*;*

(b) if $K \neq 0$*, then* $R_{abcd}(H,K) = 0$ *for all indices if and only if H is a gauge mode at K;*

(c) $\mathrm{FP}(H,K) = -\sum_{b,d} H_{bd}G_{bd}(H,K)$*;*

(d) $\sum_{b,d}(K\xi^{\mathsf T} + \xi K^{\mathsf T})_{bd}G_{bd}(H,K) = 0$ *for every ξ.*

Proof. (a) Substituting $H_{ad} = K_a\xi_d + \xi_aK_d$, the eight products cancel in pairs; homogeneity is visible. (b) Put $s = |K|^2$. Contracting $R_{abcd} = 0$ with K_c and then with K_b gives $s^2H_{ad} = sK_a(HK)_d + sK_d(HK)_a - K_aK_d(K^{\mathsf T}HK)$, so $H = K\xi^{\mathsf T} + \xi K^{\mathsf T}$ with $\xi = HK/s - (K^{\mathsf T}HK)K/(2s^2)$; the converse is (a). (c) $R_{bd} = K_b(HK)_d + K_d(HK)_b - sH_{bd} - K_bK_d\,\mathrm{tr}\,H$ and $R = 2K^{\mathsf T}HK - 2s\,\mathrm{tr}\,H$, so $\sum H_{bd}G_{bd} = 2|HK|^2 - s\,\mathrm{tr}\,H^2 - 2\,\mathrm{tr}\,HK^{\mathsf T}HK + s(\mathrm{tr}\,H)^2 = -\mathrm{FP}(H,K)$. (d) The pairing is $2\xi^{\mathsf T}GK$, and $GK = 0$ by the formulas in (c): the linearized Bianchi identity. □

On a lattice the curvature of a hinge is its holonomy: the boost or rotation acquired around it. J reads holonomies through its odd part.

Proposition 27.16 (J's odd part; every signed table is a J-built cost)**.** *For $x, y > 0$ put $J_{\mathrm{odd}}(x,y) = \frac{1}{2}\big(J(xy) - J(x/y)\big)$. Then*

(a) $J(xy) = J(x) + J(y) + J(x)J(y) + J_{\mathrm{odd}}(x,y)$ and $J(x/y) = J(x) + J(y) + J(x)J(y) - J_{\mathrm{odd}}(x,y)$;

(b) $J_{\mathrm{odd}}(x,y) = \frac{1}{4}(x - x^{-1})(y - y^{-1}) = \sinh u \sinh v$ for $x = e^u$, $y = e^v$; it is negative when $x > 1 > y$;

(c) for a real symmetric $n \times n$ table T, the cost $\sum_{i,j} T_{ij} J_{\mathrm{odd}}(r_i, r_j)$ equals $\sum_{i,j} T_{ij} \sinh u_i \sinh u_j$ at $r = e^u$; at $r_i = e^{\varepsilon u_i}$, divided by ε^2, it tends to $\sum T_{ij} u_i u_j$; and two symmetric tables give the same cost for all positive r only if they are equal.

So every quadratic table of couplings, Regge's included, is a cost built from J, and J's composition law fixes none of its weights.

Proof. (a) The law gives the even part $\frac{1}{2}(J(xy) + J(x/y)) = J(x)J(y) + J(x) + J(y)$. (b) $J(xy) - J(x/y) = \frac{1}{2}(xy + \frac{1}{xy} - \frac{x}{y} - \frac{y}{x}) = \frac{1}{2}(x - x^{-1})(y - y^{-1})$. (c) The first two claims follow from (b) and $\sinh t/t \to 1$. For the last, take u supported on $\{i, j\}$ and compare the coefficients of $\sinh u_i \sinh u_j$. □

A holonomy around a spacelike hinge is a boost, e^{ε}, and around a timelike hinge a rotation, $e^{i\varepsilon}$. J's even part treats them oppositely.

Proposition 27.17 (Rotation holonomies; on the lattice metric every hinge is timelike)**.**

(a) $J(e^{i\varepsilon}) = \cos\varepsilon - 1 \in [-2, 0]$, which equals the boost value $J(e^{\varepsilon}) = \cosh\varepsilon - 1 \geq 0$ only at $\varepsilon = 0$. $(\cos t\varepsilon - 1)/t^2 \to -\varepsilon^2/2$, and $(\cos t^2 c - 1)/t^4 \to -c^2/2$: the even part is of fourth order in a curvature of size t^2.

(b) $J_{\mathrm{odd}}(x, e^{i\varepsilon}) = \frac{i}{2}(x - x^{-1}) \sin\varepsilon$ for $x > 0$, with J_{odd} extended by (b) of Proposition 27.16, and the real part of $-i J_{\mathrm{odd}}(e^{t\alpha}, e^{it\varepsilon})$ divided by t^2 tends to $\alpha\varepsilon$, as $J_{\mathrm{odd}}(e^{t\alpha}, e^{t\varepsilon})/t^2$ does for boosts.

(c) Let $g_1(u, v) = u \cdot v - (\sum u)(\sum v)$. For nested $0/1$ vectors $d_1 \leq d_2$ with $k_1 \geq 1$ and $k_2 > k_1$ ones, $g_1(d_1, d_1) g_1(d_2, d_2) - g_1(d_1, d_2)^2 = k_1(1 - k_2)(k_2 - k_1) < 0$.

Proof. (a) $\frac{1}{2}(e^{i\varepsilon} + e^{-i\varepsilon}) = \cos\varepsilon$; $\cos\varepsilon = \cosh\varepsilon$ only at 0; the limits are Taylor's. (b) $e^{i\varepsilon} - e^{-i\varepsilon} = 2i \sin\varepsilon$, and $\sinh(t\alpha)\sin(t\varepsilon)/t^2 \to \alpha\varepsilon$. (c) $g_1(d, d) = k(1 - k)$ for a vector with k ones and $g_1(d_1, d_2) = k_1 - k_1 k_2$, so the determinant is $k_1(1 - k_1)k_2(1 - k_2) - k_1^2(1 - k_2)^2 = k_1(1 - k_2)(k_2 - k_1)$. □

By (c), every hinge of the $\kappa = 1$ lattice spanned by two of its steps is timelike, so its holonomy is a rotation of the normal plane and J's even part on it is nonpositive. A cost that is nonnegative cannot produce the graviton term. The theorem below says this for every local two-derivative form, with no symmetry assumed.

Theorem 27.18 (No nonzero local invariant form is nonnegative)**.** *Let*

$$\mathcal{Q}(h,K)=\sum_{a,b,c,d,e,f} C_{abcdef}\,h_{ab}h_{cd}K_eK_f$$

with arbitrary real coefficients. If $\mathcal{Q}$ is gauge invariant and $\mathcal{Q}(h,K)\ge 0$ for every strain h and wave K, then $\mathcal{Q}(h,K)=0$ for every strain and wave.

Proof. For fixed K, $\mathcal{Q}(\cdot,K)$ is a quadratic form: $\mathcal{Q}(h+sg,K)=\mathcal{Q}(h,K)+sX(h,g,K)+s^2\mathcal{Q}(g,K)$. An affine function $a+sb$ that is ≥ 0 for every real s has $b=0$. Put $\phi(v,K)=\mathcal{Q}(vv^{\mathsf T},K)$, of degree 4 in v and 2 in K.

1. *Gauge shear.* $(v+bK)(v+bK)^{\mathsf T}=vv^{\mathsf T}+K\xi^{\mathsf T}+\xi K^{\mathsf T}$ with $\xi=bv+\frac{b^2}{2}K$, so $\phi(v+bK,K)=\phi(v,K)$. Also $vv^{\mathsf T}$ is the gauge mode of $\xi=v/2$ at the wave v, so $\phi(v,v)=\mathcal{Q}(0,v)=0$.

2. *Wave shear.* $\phi(v,K+tv)=\phi(v,K)+tY+t^2\phi(v,v)=\phi(v,K)+tY$ is ≥ 0 for all t, so $Y=0$ and $\phi(v,K+tv)=\phi(v,K)$.

3. *Swap.* By step 1 with $b=1$, step 2 along $v+K$ with $t=-1$, step 1 again and degree 2 in the wave: $\phi(v,K)=\phi(v+K,K)=\phi(v+K,-v)=\phi(K,-v)=\phi(K,v)$.

4. *Vanishing.* $\phi(2v,K)=16\phi(v,K)$, while by step 3, $\phi(2v,K)=\phi(K,2v)=4\phi(K,v)=4\phi(v,K)$. So $\phi=0$.

5. *Every rank-one strain is invisible.* $s\mapsto\mathcal{Q}(h+svv^{\mathsf T},K)=\mathcal{Q}(h,K)+sX(h,vv^{\mathsf T},K)$ is affine and nonnegative, so $\mathcal{Q}(h+svv^{\mathsf T},K)=\mathcal{Q}(h,K)$ for every strain h. The ten strains $e_ae_a^{\mathsf T}$ and $(e_a+e_b)(e_a+e_b)^{\mathsf T}$ $(a<b)$ span the strains: $h=\sum_a(h_{aa}-\sum_{b\ne a}h_{ab})e_ae_a^{\mathsf T}+\sum_{a<b}h_{ab}(e_a+e_b)(e_a+e_b)^{\mathsf T}$. Removing these terms one at a time gives $\mathcal{Q}(h,K)=\mathcal{Q}(0,K)=0$. □

So a two-derivative graviton term must take both signs: a sum of J's with nonnegative weights has a nonnegative long-wave form, since $J(e^t)=t^2/2+O(t^4)$, and its gauge-invariant two-derivative part vanishes. The graviton term comes from J's odd part, as Regge's area times holonomy does.

We now turn to the lattice. Kuhn's triangulation [90] cuts the unit cube of $\mathbb{Z}^4$ into 24 simplices, one for each order σ of the four axes.

Definition 27.19 (The Kuhn lattice, tables and their plane-wave forms)**.**

1. The simplex σ has vertices $v_0=0$ and $v_k=e_{\sigma(1)}+\cdots+e_{\sigma(k)}$, $k=1,\dots,4$. Its ten *slots* are the edges (i,j), $i<j$, in the order 01, 02, 03, 04, 12, 13, 14, 23, 24, 34; slot f has step $s_f=v_j-v_i$, a 0/1 vector with $j-i$ ones, and midpoint $x_f=\frac12(v_i+v_j)$. Translates of the simplices tile $\mathbb{R}^4$; every lattice edge has one of the fifteen classes $d\in\{0,1\}^4\setminus\{0\}$.

2. At the background $g_\kappa=1-\kappa\mathbf{1}\mathbf{1}^{\mathsf T}$ a step d has squared length $|d|^2-\kappa(\sum d)^2$; a step with n ones has $n-\kappa n^2$.

3. A perturbation at the wave K gives each class d an amplitude $A(d)\in\mathbb{C}$, the change of squared length of the edges of class d times $e^{iK\cdot x}$ at the edge midpoint x. The vertex displacement $\eta e^{iK\cdot x}$ $(\eta\in\mathbb{C}^4)$ has the *gauge amplitude* $A(d)=4i(d\cdot\eta)\sin(K\cdot d/2)$.

Table 27.1. Regge's per-simplex tables, slots in the order $01, 02, 03, 04, 12, 13, 14, 23, 24, 34$. Left: the $\kappa = 1$ table R_1. Right: the Euclidean table E, which is 48 times the Hessian table of Theorem 27.28 at $\kappa = 0$.

$$R_1 = \begin{pmatrix} 2 & -2 & 1 & 0 & 0 & 3 & -2 & -6 & 2 & 2 \\ -2 & -4 & 4 & -1 & 6 & -6 & 2 & 3 & -2 & 2 \\ 1 & 4 & 0 & -2 & -6 & 0 & 2 & -3 & 2 & -2 \\ 0 & -1 & -2 & 2 & 2 & 2 & -2 & 2 & -1 & 0 \\ 0 & 6 & -6 & 2 & -6 & 6 & -3 & 0 & 3 & -6 \\ 3 & -6 & 0 & 2 & 6 & 0 & 0 & 6 & -6 & 3 \\ -2 & 2 & 2 & -2 & -3 & 0 & 0 & -6 & 4 & 1 \\ -6 & 3 & -3 & 2 & 0 & 6 & -6 & -6 & 6 & 0 \\ 2 & -2 & 2 & -1 & 3 & -6 & 4 & 6 & -4 & -2 \\ 2 & 2 & -2 & 0 & -6 & 3 & 1 & 0 & -2 & 2 \end{pmatrix} \qquad E = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 3 & 0 & -6 & 0 & 0 \\ 0 & -3 & 3 & 0 & 6 & -6 & 0 & 3 & 0 & 0 \\ 0 & 3 & -2 & 0 & -6 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & -6 & 0 & -6 & 6 & 0 & 0 & 3 & -6 \\ 3 & -6 & 3 & 0 & 6 & -6 & 3 & 6 & -6 & 3 \\ 0 & 0 & 0 & 0 & 0 & 3 & -2 & -6 & 3 & 0 \\ -6 & 3 & 0 & 0 & 0 & 6 & -6 & -6 & 6 & 0 \\ 0 & 0 & 0 & 0 & 3 & -6 & 3 & 6 & -3 & 0 \\ 0 & 0 & 0 & 0 & -6 & 3 & 0 & 0 & 0 & 0 \end{pmatrix}$$

4. The *plane-wave form* of a real 10×10 table T is

$$T(K; A, B) = \sum_{\sigma} \sum_{f, f'} T_{ff'}\, A(s_f)\, B(s_{f'})\, e^{iK \cdot (x_f - x_{f'})}.$$

The table is *gauge null* when $T(K; A, B) = 0$ for every gauge amplitude A, every B and every K, and of *zero symbol* when $T(K; A, B) = 0$ for all K, A, B.

Table 27.1 prints the two Regge tables: the table R_1 of the $\kappa = 1$ lattice and the Euclidean table E. Both are integer and symmetric.

Sign convention. The deficit of a hinge is $\varepsilon = 2\pi$ minus the sum of the dihedral angles of the simplices around it, and Regge's action is $\sum(\text{area}) \times (\text{deficit})$ [92]. In each simplex the deficit's share changes as $-\theta_t$ when the dihedral angle θ_t at the triangle t changes, so $\nabla\varepsilon_t = -\nabla\theta_t$. With this convention the Euclidean table is $E = +48 \sum_t \nabla\varepsilon_t \otimes \nabla A_t$; written with dihedral angles in place of deficits, every identity below changes sign. For a perturbation A and a hinge h (a lattice triangle up to translation) the *area-weighted deficit* $D_A(h)$ is the hinge's area times the linear change of its deficit under A, and $\Lambda_B(h)$ is the linear change of the logarithm of its area under B, by Heron's formula.

Theorem 27.20 (On the Euclidean Kuhn lattice the deficits see exactly the non-gauge part)**.** *At the Euclidean background:*

(a) every vertex displacement leaves every area-weighted deficit zero, at every wave;

(b) at every wave with $\sin(K_a/2) \neq 0$ *for* $a = 0, 1, 2, 3$*, a perturbation whose area-weighted deficits all vanish is a vertex displacement:* $A(d) = 4i(d \cdot \eta) \sin(K \cdot d/2)$ *for all fifteen classes and some* $\eta \in \mathbb{C}^4$*;*

(c) per simplex $E = 48 \sum_t (A_t \nabla\varepsilon_t) \otimes \nabla \log A_t$*, a symmetric table, as Schläfli's identity requires; in plane-wave form* $E(K; A, B) = 48 \sum_h D_A(h) \Lambda_B(h)$*: holonomy times area change;*

(d) a table is gauge null if and only if, at every wave with all $\sin(K_a/2) \neq 0$*, its form depends on its first argument only through that argument's area-weighted deficits;*

(e) R_1 *and* E *are both gauge null, neither has zero symbol, and no multiple of one differs from the other by a table of zero symbol; every symmetric gauge-null table is* $aR_1 + bE$ *plus a table of zero symbol.*

Conditions and scope.

Base and premises
The Euclidean Kuhn lattice, classical real and complex numbers.

Conclusion
Locality and exact gauge invariance fix the lattice graviton term up to the background (a direction in the plane of R_1 and E) and a scale. Regge's table is not singled out as a least gauge-null table.

Not claimed
A reason internal to J for invariance under vertex displacements, which is open, and (d) and (e) at $\kappa = 1$. At $\kappa = 1$, (a) and (b) are Proposition 27.37 and (c) is Theorem 27.32(c).

Proof. Every statement is finite linear algebra over the rational numbers once the phases are factored. For (a) and (c): the dihedral angles and areas of the Kuhn simplex and their derivatives in the ten squared edge lengths are rational multiples of fixed square roots (Theorem 27.26 gives the formulas), so $A_t \nabla\theta_t$ and $\nabla \log A_t$ are two rational 10×10 tables; (c) is the entrywise identity between E and 48 times their product, and (a) is the identity obtained by substituting the gauge amplitude, collected hinge by hinge over the 24 simplices. For (b), define η from the four axis classes by $\eta_a = A(e_a)/(4i \sin(K_a/2))$, which is where the hypothesis enters; subtracting the gauge amplitude of η leaves a perturbation with zero deficits that vanishes on the axis classes, and the hinge equations, a linear system in the eleven remaining amplitudes, force those to vanish. (d) follows: a gauge amplitude has the deficits of 0; and if A and A' have the same deficits, $A - A'$ is a gauge amplitude by (b). (e) reduces, after the phases are factored, to linear algebra on finitely many sums of table entries over slot pairs of equal class. The computations of (a), (b), (c) and (e) are finite, with inputs Theorem 27.26, the hinges of the 24 simplices and Table 27.1; we describe them and do not reproduce them. □

At long wavelength the Euclidean table is Fierz–Pauli on every strain, as Theorem 27.30(a) shows for R_1. Write $\mathrm{FP}_1(K; H, H')$ for the form FP_S of Section 27.2 with $S = 1$, the polarization of FP.

Proposition 27.21 (Regge's Euclidean table is Fierz–Pauli at long wavelength)**.** *In classical mathematics, for all real symmetric H, H' and every real wave K,*

$$\sum_{\sigma} \sum_{f,f'} E_{ff'} \, (s_f^{\mathsf{T}} H s_f)(s_{f'}^{\mathsf{T}} H' s_{f'}) \big(K \cdot (x_f - x_{f'})\big)^2 = 24 \, \mathrm{FP}_1(K; H, H').$$

So the order-$|K|^2$ part of the plane-wave form $E(K; A, B)$ on the amplitudes $A(d) = d^{\mathsf{T}} H d$, $B(d) = d^{\mathsf{T}} H' d$ is $-12 \, \mathrm{FP}_1(K; H, H')$.

Proof. Both sides are polynomials with rational coefficients, linear in H and in H' and quadratic in K. The steps and midpoints of the simplex σ are those of the simplex of the identity order with the coordinates permuted by σ, so the left side is one explicit polynomial, the sum over the 100 slot pairs of one simplex, summed over the 24 permutations of the axes; expanding it gives $24 \, \mathrm{FP}_1$ coefficient by coefficient. For example, at $K = e_0$ and $H = H'$ one of the three strains of the opening table, the left side is 48, 0 and -144, which are 24 times FP $= 2, 0, -6$. The last sentence is $e^{iy} = 1 + iy - \frac{1}{2}y^2 + \dots$ termwise. □

27.5 At third order the cost breaks gauge invariance

Regge's hinge term and $\mathcal{J}$'s odd part have the same second variation. A nonlinear theory needs more: the source that its cubic term makes from two first-order fields must be one that the quadratic operator can balance at second order. We compare third derivatives first and then evaluate both cubic vertices on the Kuhn lattice. Consider finitely many hinges along a deformation $t \mapsto (a_h(t), e_h(t))$ with areas $a_h > 0$ and deficits e_h, $e_h(0) = 0$, three times differentiable and obeying Schläfli's identity $\sum_h a_h e_h' = 0$ [93, 92]. Put

$$\mathcal{R}(t) = \sum_h a_h(t) e_h(t), \qquad \mathcal{J}(t) = \sum_h \frac{a_h(0)}{2}\, J_{\mathrm{odd}}\Big(\frac{a_h(t)}{a_h(0)}, e^{e_h(t)}\Big).$$

Proposition 27.22 (Regge equals $\mathcal{J}$'s odd hinge cost at second order only)**.** *In classical mathematics, $\mathcal{R}(0) = \mathcal{J}(0) = 0$, $\mathcal{R}''(0) = \mathcal{J}''(0) = \sum_h a_h' e_h'$,*

$$\mathcal{R}'''(0) = \sum_h (2a_h'' e_h' + a_h' e_h''), \qquad \mathcal{J}'''(0) = \tfrac{3}{2} \sum_h \big(a_h'' e_h' + a_h' e_h'' - a_h'^2 e_h'/a_h\big),$$

all evaluated at 0. For the two hinges $a = (e^t, 1)$, $e = (t, 1 - e^t)$, Schläfli's identity holds and $\mathcal{R}'''(0) = 2$ while $\mathcal{J}'''(0) = 0$.

Proof. $J_{\mathrm{odd}}(x, e^d) = \sinh(\log x) \sinh d$, and $\sinh(\log(a/a_0)) = \frac{1}{2}(a/a_0 - a_0/a)$ has derivatives 0, a'/a_0 and $a''/a_0 - a'^2/a_0^2$ at 0, while $\sinh e$ has 0, e', e''. The Leibniz rule gives $\mathcal{J}''$ and $\mathcal{J}'''$. For $\mathcal{R}$, $(ae)'' = 2a'e' + a_0 e''$ and $(ae)''' = 3a''e' + 3a'e'' + a_0 e'''$ at 0, and the first two derivatives of Schläfli's identity, $\sum(a'e' + ae'') = 0$ and $\sum(a''e' + 2a'e'' + ae''') = 0$, eliminate $\sum a_0 e''$ and $\sum a_0 e'''$. In the example $\sum ae' = e^t - e^t = 0$, and only the first hinge contributes: $2 \cdot 1 \cdot 1 + 0 = 2$ and $\frac{3}{2}(1 + 0 - 1) = 0$. □

$\mathcal{J}$'s odd hinge cost is even under $(a/a_0, e) \mapsto (a_0/a, -e)$, so it has no cubic term in the log-area and the deficit, whereas Regge's $a_0 e^u e$ has the cubic term $\frac{1}{2} a_0 u^2 e$. Whether this difference matters is decided on the lattice. On the Kuhn lattice of Definition 27.19 at the Euclidean background, the deficit of a hinge is 2π minus the sum of the dihedral angles of the simplices around it, and the area of a hinge depends only on its own edges, so each hinge term of Proposition 27.22 splits into shares of those simplices.

Definition 27.23 (The cubic forms and the long-wave cubic vertex)**.**

1. Number the slots $01, \dots, 34$ as $0, \dots, 9$; let L_0 be the squared edge lengths of a Kuhn simplex at the Euclidean background (slot ij has $j - i$); for $\ell \in \mathbb{R}^{10}$ let $\theta_t(s)$ and $A_t(s)$ be the dihedral angle and the area at the triangle t when the squared edge lengths are $L_0 + s\ell$, with primes for derivatives at $s = 0$. The *cubic forms* of the simplex are

$$r(\ell) = \sum_t \big(\theta_t'' A_t' + 2\theta_t' A_t''\big), \qquad d(\ell) = \sum_t \Big(\frac{A_t' \theta_t'' - A_t'' \theta_t'}{2} - \frac{3}{2}\, \frac{A_t'^2\, \theta_t'}{A_t(0)}\Big).$$

2. Three complex plane waves $H_i e^{iK_i \cdot x}$, $i = 1, 2, 3$, with H_i complex symmetric and $K_i \in \mathbb{C}^4$, give the slot f of the simplex σ of the base cell the *Taylor amplitudes* $a^i_{\sigma,n}(f) =$

$(s_f^{\mathsf T} H_i s_f)(K_i \cdot x_f)^n/n!$, where $K \cdot x = \sum_a K_a x_a$ without conjugation. For a cubic form T let $\hat T(a,b,c)$ be the coefficient of $\epsilon_1\epsilon_2\epsilon_3$ in $T(\epsilon_1 a + \epsilon_2 b + \epsilon_3 c)$. The *long-wave cubic vertex* of T is

$$V_T(H_1,H_2,H_3;K_1,K_2,K_3) = \frac{1}{6}\sum_{n_1+n_2+n_3=2}\ \sum_{\sigma} \hat T\big(a^1_{\sigma,n_1}, a^2_{\sigma,n_2}, a^3_{\sigma,n_3}\big),$$

summed over the six triples of orders with sum 2 and the 24 simplices.

Both forms are cubic with rational coefficients. They come from the closed forms of Theorem 27.26(d) at $\kappa = 0$ and their derivatives along ℓ: every term of r and d carries an even number of factors $(W_{ll}W_{mm} - W_{lm}^2)^{-1/2}$, so the square roots cancel. Table 27.2 prints them. Along a line $s \mapsto L + s\ell$ of squared edge lengths of the periodic lattice, with ℓ_σ the restriction of ℓ to the simplex σ, the proposition with $e_h' = -\sum_{\sigma\ni h}\theta'_{\sigma,h}$ and $e_h'' = -\sum_{\sigma\ni h}\theta''_{\sigma,h}$ gives

$$\mathscr R'''(0) = -\sum_\sigma r(\ell_\sigma), \qquad \mathscr J'''(0) - \mathscr R'''(0) = -\sum_\sigma d(\ell_\sigma), \tag{27.1}$$

Schläfli's identity (Theorem 27.27) holding simplex by simplex. So J's hinge cost is Regge's action plus the extra cubic term $-\frac{s^3}{6}\sum_\sigma d(\ell_\sigma)$. Since $e^{iK\cdot x_f} = \sum_n i^n (K\cdot x_f)^n/n!$, the coefficient of $\epsilon_1\epsilon_2\epsilon_3$ in the cell's share $-\frac16\sum_\sigma T(\ell_\sigma)$ of a cubic term, on the field $\sum_i \epsilon_i (s_e^{\mathsf T} H_i s_e) e^{iK_i\cdot x_e}$, is V_T at second order in the momenta; its parts of order 0 and 1 are $-\frac16$ and $-\frac{i}{6}$ times the sums with $n_1+n_2+n_3 = 0$ and 1. We write V_r and V_d for the vertices of r and d, and $\mathrm{FP}_1(K;H,H')$ for the form FP_S of Section 27.2 with $S = 1$ over $\mathbb C$.

The test data are Gaussian integers. Let $\eta = \mathrm{diag}(-1,1,1,1)$, $\xi = e_1$, the static momenta $k_1 = (0,3,4,5i)$, $k_2 = (0,5,12,13i)$, $k_3 = -k_1 - k_2$, the momenta $q_1 = (3,4,0,5i)$, $q_2 = (5,0,12,13i)$, $q_3 = -q_1 - q_2$, the gauge legs $\Xi = k_3\xi^{\mathsf T} + \xi k_3^{\mathsf T}$ and $\Xi' = q_3\xi^{\mathsf T} + \xi q_3^{\mathsf T}$, and

$$Z_1 = \begin{pmatrix} -321+240i & -228+70i & -100-125i & -200-375i \\ -228+70i & 171-240i & 75 & 150 \\ -100-125i & 75 & -75 & 75 \\ -200-375i & 150 & 75 & 225 \end{pmatrix},$$

$$Z_2 = \begin{pmatrix} 482+312i & -120+65i & -60-65i & -60+130i \\ -120+65i & -457-312i & 50 & -25 \\ -60-65i & 50 & 25 & 25 \\ -60+130i & -25 & 25 & -50 \end{pmatrix}.$$

Theorem 27.24 (J's cubic vertex breaks gauge invariance, and its second-order static field does not exist)**.** *In classical mathematics, with r and d the forms of Table 27.2:*

(a) for real symmetric H_i and real K_i, neither r nor d has a vertex term of order 0 or 1 in the momenta: the sum over σ of $\hat T(a^1_{\sigma,n_1}, a^2_{\sigma,n_2}, a^3_{\sigma,n_3})$ vanishes for $(n_1,n_2,n_3) = (0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, without momentum conservation;

(b) k_1 and k_2 are static and null ($k_{i,0} = 0$, $k_i\cdot k_i = 0$), and

$$V_r(\eta,\eta,\Xi;k_1,k_2,k_3) = 0, \qquad V_d(\eta,\eta,\Xi;k_1,k_2,k_3) = -8;$$

Table 27.2. The cubic forms of Definition 27.23 at a Kuhn simplex of the Euclidean background: $128\,r(\ell)$ (Regge's action) and $128\,d(\ell)$ ($\mathcal{J}$'s extra term). An entry ijk, c is the monomial $c\,\ell_i\ell_j\ell_k$; slots are numbered $0,\dots,9$ in the order $01, 02, 03, 04, 12, 13, 14, 23, 24, 34$.

$128\,r(\ell)$, 138 monomials

000 4	002 −12	005 24	006 4	007 −24	009 −12	011 −24	012 48	014 48	015 −48
018 −24	019 48	022 −12	024 −48	025 48	026 −24	027 −48	028 48	029 −24	035 −24
036 16	037 48	038 −24	044 −48	045 96	047 −48	055 −12	066 4	069 −24	077 −24
088 −12	089 48	099 −12	111 −4	113 12	114 12	116 −12	117 −12	118 12	119 −12
122 24	123 −48	125 −48	126 48	127 48	128 −24	133 12	134 −24	135 48	136 −24
137 −24	138 24	139 −24	144 48	145 −48	147 96	157 −48	168 −24	169 48	177 24
188 12	189 −24	222 −12	223 16	224 −24	225 24	226 −12	229 4	233 4	234 48
235 −24	238 −24	239 16	244 −24	247 −48	255 12	266 −12	268 48	269 −24	288 −12
299 4	333 −4	334 −12	336 4	337 −12	338 12	348 −24	349 48	356 −24	358 48
359 −24	366 16	367 48	368 −48	378 −24	388 12	444 −32	445 48	447 −48	448 24
449 −24	457 96	458 −48	467 −48	468 48	469 −48	477 −48	478 96	479 −48	488 −12
499 −24	555 −16	556 12	559 −12	566 24	568 −48	569 48	577 48	578 −48	579 96
589 −48	599 24	666 −12	667 −24	668 24	669 −12	677 −24	679 −48	689 48	699 −12
777 −32	778 48	779 −48	788 12	789 48	888 −4	889 −24	999 4		

$128\,d(\ell)$, 129 monomials

000 2	002 −6	005 12	007 −12	011 −12	012 24	014 24	015 −24	022 −6	024 −24
025 12	044 −24	045 48	047 −24	048 12	049 −24	055 −24	056 36	057 24	058 −24
059 24	066 −12	067 −24	068 12	078 −24	079 −24	088 6	111 1	113 −3	117 −6
118 −9	119 6	122 −6	123 12	128 12	129 12	133 −3	137 12	138 −6	144 18
145 −24	146 12	147 48	148 36	149 −24	155 12	156 −24	158 −24	159 −24	166 6
167 −24	168 12	177 −6	178 36	179 12	188 −9	222 4	223 −6	227 −18	228 6
229 −12	237 12	244 −12	245 24	246 −24	248 −24	249 −24	255 −24	256 24	257 72
258 −24	259 36	267 −24	277 −18	278 12	333 1	338 −3	344 6	345 −24	346 12
347 −24	348 12	355 12	357 −24	366 −6	368 12	377 6	388 −3	444 −10	445 24
446 −18	447 −24	448 −6	455 −36	456 72	457 24	459 24	466 −18	477 −24	478 48
479 −24	488 −6	499 −12	555 16	556 −24	557 −36	558 12	559 −24	567 24	569 12
577 24	578 −24	579 48	589 −24	599 12	666 4	668 −6	669 −6	677 −12	679 −24
689 24	699 −6	777 −10	778 18	779 −24	789 24	888 1	889 −12	999 2	

(c) q_1 *and* q_2 *are null,* Z_1 *and* Z_2 *are symmetric with* $Z_i q_i = 0$ *and* $\operatorname{tr} Z_i = 0$*, and*

$$V_r(Z_1, Z_2, \Xi'; q_1, q_2, q_3) = 0, \qquad V_d(Z_1, Z_2, \Xi'; q_1, q_2, q_3) = 13496500 + 4959400i;$$

(d) $\mathrm{FP}_1(K; H, K\zeta^{\mathsf{T}} + \zeta K^{\mathsf{T}}) = 0$ *for every complex symmetric* H *and all* $K, \zeta \in \mathbb{C}^4$*; and no complex symmetric* H_2 *satisfies*

$$-\tfrac{1}{4}\,\mathrm{FP}_1(k_1 + k_2; H_2, H') + V_d(\eta, \eta, H'; k_1, k_2, k_3) = 0 \quad \textit{for every complex symmetric } H'.$$

Conditions and scope.

Base and premises

The Kuhn lattice at the Euclidean background, complex momenta, classical real and complex numbers. Proved in classical mathematics; the values in (b) and (c) are exact arithmetic in the Gaussian integers.

Conclusion

$\mathcal{J}$'s hinge cost is Regge's action plus the extra vertex V_d. Both vertices start at second order in the momenta. On both sets of legs with one pure-gauge leg Regge's vertex vanishes and V_d does not, although the other two legs solve the linearized field equations. So $\mathcal{J}$'s second-order equation for the static field $2U\eta$ built from the two harmonic modes $U = e^{ik_1\cdot x} + e^{ik_2\cdot x}$ has no solution. The post-Newtonian parameter β_{PN} (not DeWitt's β) is read from the second-order static field, so it does not exist for $\mathcal{J}$'s cost.

Identification

The legs η at k_1, k_2 as the Euclidean form of the linearized field of a static mass; Regge's action as the gravitational action.

Not claimed

The point-mass field and real fields, and the Lorentzian $\kappa = 1$ lattice, here. That Regge's long-wave cubic vertex is one half of Einstein's is Theorem 27.61(b), on every background off the poles, and J's whole cubic vertex is written out in Theorem 27.68. For the J-built hinge costs with reference area $\rho a_h(0)$, Proposition 27.25 shows that the part of the cubic vertex linear in the deficits fails this test unless $\rho = 0$; Theorem 27.69(c) gives a hinge-local correction with which J's cost has exactly one half of Regge's cubic vertex and passes the cubic test at three witnesses, and Theorem 27.71 shows that hinge shape costs cannot carry that halved coupling to quartic order.

Proof. (a) At order 0 the amplitude $a^i_{\sigma,0}(f) = s_f^{\mathsf T} H_i s_f$ is the change of squared length of slot f under the constant strain H_i, so the sum is the polarization of $h \mapsto \sum_\sigma T\big((s_f^{\mathsf T} h s_f)_f\big)$. A constant strain h is a linear change of coordinates: along $L + s(s_e^{\mathsf T} h s_e)_e$ every deficit stays 0, so $\mathscr{R}$ and $\mathscr{J}$ vanish identically, and by (27.1), all cells being alike, $\sum_\sigma r$ and $\sum_\sigma d$ vanish on every constant strain; so does their polarization. At order 1 use the central inversion $x \mapsto \mathbf{1} - x$ of the cell. It maps the simplex of the axis order $(\sigma_1, \dots, \sigma_4)$ to that of $(\sigma_4, \dots, \sigma_1)$ and its slot ij to the slot $(4-j)(4-i)$, with the same step and the reflected midpoint $\mathbf{1} - x_f$. On slot numbers this is the permutation $(0\,9)(1\,8)(2\,6)(4\,7)$, and both forms of Table 27.2 are invariant under it, as the table shows. Reindexing the cell sum by the inversion turns the order-1 amplitude $(s_f^{\mathsf T} H_i s_f)(K_i \cdot x_f)$ into $(s_f^{\mathsf T} H_i s_f)(K_i \cdot \mathbf{1} - K_i \cdot x_f)$, so each order-1 sum S satisfies $S = (K_i \cdot \mathbf{1}) S_0 - S$ with S_0 the order-0 sum, which is 0. Hence $S = 0$.

(b) and (c) The null and static conditions are arithmetic, as are $Z_i q_i = 0$ and $\operatorname{tr} Z_i = 0$. Each vertex is a finite sum over 24 simplices, six order triples and the monomials of Table 27.2, with Gaussian-integer data, so it is exact arithmetic in $\mathbb{Z}[i]$ after multiplying by $6 \cdot 128 \cdot 8$. For (b) the six order sums $\sum_\sigma \hat T(a^1_{\sigma,n_1}, a^2_{\sigma,n_2}, a^3_{\sigma,n_3})$, in the order $(2,0,0)$, $(0,2,0)$, $(0,0,2)$, $(1,1,0)$, $(1,0,1)$, $(0,1,1)$, are 468, 1260, 0, 1536, -1248, -2016 for r, with sum 0, and -234, -630, 0, -816, 624, 1008 for d, with sum $-48 = 6 \cdot (-8)$. The static legs solve the linearized equations: by the formulas in the proof of Theorem 27.15(c), $\eta k_i = k_i$, $\operatorname{tr}\eta = 2$ and $k_i \cdot k_i = 0$ give $R_{bd} = 2k_b k_d - 2k_b k_d = 0$; transverse-traceless legs at null momenta solve them too.

(d) The first claim is Theorem 27.6(b) over $R = \mathbb{C}$ with $S = 1$, the form being symmetric in its two strains. Suppose H_2 solves the equation and take $H' = \Xi$. Since $\Xi = -\big((k_1+k_2)\xi^{\mathsf T} + \xi(k_1+k_2)^{\mathsf T}\big)$ is a gauge mode at $k_1 + k_2$, the first term vanishes, and by (b) the equation reads $-8 = 0$. The same test excludes the equation with J's whole cubic vertex $V_r + V_d$ in place of V_d, since $V_r(\eta, \eta, \Xi; k_1, k_2, k_3) = 0$ by (b). □

The test in (d) is the consistency condition of every gauge theory: the second-order field $H_2 e^{i(k_1+k_2)\cdot x}$ must make the action stationary in every direction H', and a pure-gauge direction removes the quadratic form (the linearized Bianchi identity, Theorem 27.15(d)), so a solution needs a source that vanishes on gauge directions, however the two terms are normalized. β_{PN} is read from the second-order static field, and for J's cost that equation has no solution already for this two-mode field, while Regge's vertex passes the test on both sets of legs. So J's

hinge cost agrees with Regge's action to second order only. With Regge's action read as the gravitational action (the identification named in Section 27.7), the graviton is the long-wave form of J, its self-interaction is Regge's, and nonlinear gravity on the lattice is Regge's action.

Which hinge costs avoid the defect? Take a cost linear in the deficit, $\sum_h \varphi_h(a_h)e_h$, with smooth area factors, along a deformation as in Proposition 27.22. Since $e_h(0) = 0$, Leibniz's rule gives its second and third derivatives at 0 as the left sides in (a) below, with $c_{0h} = \varphi_h(a_{0h})$, $c_{1h} = \varphi_h'(a_{0h})$, $c_{2h} = \frac{1}{2}\varphi_h''(a_{0h})$, $a_{0h} = a_h(0)$, and $a_h^{(k)}$, $e_h^{(k)}$ the kth derivatives at 0; Schläfli's identity and its first two derivatives give the three relations in (a). Regge's action has $(c_{0h}, c_{1h}, c_{2h}) = (a_{0h}, 1, 0)$, and its derivatives are the right sides with $\lambda = 1$.

Proposition 27.25 (Regge's derivatives force a linear area factor; a reference area fails the static test)**.** *In classical mathematics:*

(a) Let $n \geq 2$, $a_{0h} > 0$ and $c_{0h}, c_{1h}, c_{2h}, \lambda$ real ($h = 1, \dots, n$). Call $(a^{(1)}, a^{(2)}, e^{(1)}, e^{(2)}, e^{(3)}) \in (\mathbb{R}^n)^5$ a Schläfli jet *when $\sum_h a_{0h} e_h^{(1)} = 0$, $\sum_h (a_h^{(1)} e_h^{(1)} + a_{0h} e_h^{(2)}) = 0$ and $\sum_h (a_h^{(2)} e_h^{(1)} + 2a_h^{(1)} e_h^{(2)} + a_{0h} e_h^{(3)}) = 0$. Then*

$$\sum_h \big(c_{0h} e_h^{(2)} + 2c_{1h} a_h^{(1)} e_h^{(1)}\big) = \lambda \sum_h a_h^{(1)} e_h^{(1)},$$
$$\sum_h \big(c_{0h} e_h^{(3)} + 3c_{1h}(a_h^{(1)} e_h^{(2)} + a_h^{(2)} e_h^{(1)}) + 6c_{2h}(a_h^{(1)})^2 e_h^{(1)}\big) = \lambda \sum_h \big(2a_h^{(2)} e_h^{(1)} + a_h^{(1)} e_h^{(2)}\big)$$

hold for every Schläfli jet if and only if $c_{0h} = \lambda a_{0h}$, $c_{1h} = \lambda$ and $c_{2h} = 0$ for every h.

(b) For real r and $a_0 > 0$, the jets $c_0 = \frac{1}{4}(a_0 - r^2/a_0)$, $c_1 = \frac{1}{4}(1 + r^2/a_0^2)$, $c_2 = -r^2/(4a_0^3)$ of the area factor $\frac{1}{4}(a - r^2/a)$ of J's hinge cost with reference area r (see (c)) have the form of (a) for some λ if and only if $r = 0$.

(c) For $r, a > 0$, $\frac{r}{2} J_{\mathrm{odd}}(a/r, e^d) = \frac{1}{4}(a - r^2/a)\sinh d$. Put $\mathcal{J}_\rho = \sum_h \frac{1}{4}(a_h - \rho^2 a_h(0)^2/a_h)\sinh e_h$, so that $\mathcal{J}_1 = \mathcal{J}$. Along every deformation of Proposition 27.22, at 0, $\mathcal{J}_\rho'' = \frac{1+3\rho^2}{4}\sum_h a_h' e_h'$ and

$$\mathcal{J}_\rho''' = \sum_h \Big(\tfrac{1+5\rho^2}{4} a_h' e_h'' + \tfrac{1+2\rho^2}{2} a_h'' e_h' - \tfrac{3\rho^2}{2}\frac{a_h'^2 e_h'}{a_h} + \tfrac{1-\rho^2}{4} a_h e_h'^3\Big).$$

(d) With the static legs of Theorem 27.24(b), for every real ρ, $\frac{1+3\rho^2}{4} V_r(\eta, \eta, \Xi; k_1, k_2, k_3) + \rho^2 V_d(\eta, \eta, \Xi; k_1, k_2, k_3) = -8\rho^2$, which is 0 exactly when $\rho = 0$.

Proof. (a) If $c_{0h} = \lambda a_{0h}$, $c_{1h} = \lambda$, $c_{2h} = 0$, each left side minus its right side is λ times the second or third relation. Conversely fix h and another hinge k, and take two Schläfli jets supported on $\{h, k\}$, both with $e^{(1)} = a_{0k}$ at h and $-a_{0h}$ at k: the first with $a^{(1)} = a_{0h}$, $e^{(2)} = -a_{0k}$, $e^{(3)} = 2a_{0k}$ at h, the second with $a^{(2)} = a_{0h}$, $e^{(3)} = -a_{0k}$ at h, all other entries 0. After division by a_{0k}, the two equations for the first jet and the second for the second read $-c_0 + 2a_0 c_1 = \lambda a_0$, $2c_0 - 3a_0 c_1 + 6a_0^2 c_2 = -\lambda a_0$ and $-c_0 + 3a_0 c_1 = 2\lambda a_0$ at h, so $c_1 = \lambda$, $c_0 = \lambda a_0$ and $c_2 = 0$. (b) $c_2 = 0$ forces $r = 0$, and $r = 0$ gives $\lambda = \frac{1}{4}$. (c) $J_{\mathrm{odd}}(a/r, e^d) = \frac{1}{2}(\frac{a}{r} - \frac{r}{a})\sinh d$ by Proposition 27.16(b). Per hinge write $\mathcal{J}_\rho$ as FG with $F = \frac{1}{4}(a - \rho^2 a_0^2/a)$, $G = \sinh e$: at 0, G, G', G'', G''' are $0, e', e'', e''' + e'^3$, and F, F', F'' are $\frac{1}{4}(1 - \rho^2)a_0$, $\frac{1}{4}(1 + \rho^2)a'$,

$\frac{1}{4}((1+\rho^2)a'' - 2\rho^2 a'^2/a_0)$. Leibniz's rule and the two differentiated Schläfli identities, which remove $\sum a_0 e''$ and $\sum a_0 e'''$ as in Proposition 27.22, give the formulas. (d) $V_r = 0$ and $V_d = -8$ on these legs. □

By Proposition 27.22, (c) reads $\mathcal{J}_\rho''' = \frac{1+3\rho^2}{4}\mathcal{R}''' + \rho^2(\mathcal{J}''' - \mathcal{R}''') + \frac{1-\rho^2}{4}\sum_h a_h e_h'^3$, and by (27.1) its first two terms carry the vertex of (d): the static test fails for it at every $\rho \neq 0$, J's own cost $\rho = 1$ included. At $\rho = 0$ the cost is $\frac{1}{4}a\sinh e$; the area enters linearly and J only through the deficit. For charges built from one function of the area, the quadratic test and the cubic test at three witnesses do force Regge's derivatives (Theorem 27.66), and with one function per class of hinge they force them up to three inversion-odd terms that neither test sees (Proposition 27.67); we do not prove the long-wave order of the last term.

27.6 The Regge table for every background

The two tables of Table 27.1 are two points of one rational family. We now compute Regge's table for every background $g_\kappa = 1 - \kappa \mathbf{1}\mathbf{1}^{\mathsf{T}}$ with $\kappa \neq \frac{1}{4}$, from the Gram matrix of one Kuhn simplex. Take the simplex with vertices $v_k = e_1 + \cdots + e_k$, $k = 0, \dots, 4$ (axes numbered 1 to 4 here). Let $n^1, \dots, n^4$ be the dual basis of its edges $v_a - v_0$, and let the facet normals be $n_0 = -(n^1 + \cdots + n^4)$ and $n_i = n^i$, so that $W_{ij} = n_i \cdot n_j$ $(0 \le i, j \le 4)$ is $E^{\mathsf{T}} G^{-1} E$ for the 4×5 matrix E with columns $-\mathbf{1}, e_1, \dots, e_4$. For a triangle t, write lm for the opposite edge (the two vertices not in t) and $Q_t = 16A_t^2 = 2ab + 2bc + 2ca - a^2 - b^2 - c^2$ for Heron's expression in its squared edge lengths.

Theorem 27.26 (The Gram algebra of the Kuhn simplex)**.** *In classical mathematics, for every real κ:*

(a) the Gram matrix of the edges from v_0 is $G_{ab} = \min(a,b) - \kappa ab$ $(a, b = 1, \dots, 4)$, and $\det G = 1 - 4\kappa = \det g_\kappa$; for $\kappa \neq \frac{1}{4}$, G^{-1} is tridiagonal with off-diagonal entries -1 and diagonal $(2, 2, 2, (3\kappa - 1)/(4\kappa - 1))$;

(b) Q_t equals $4(1-2\kappa)$ on the triangles 012, 123, 234; $8(1-3\kappa)$ on 013, 023, 124, 134; $12(1-4\kappa)$ on 014, 034; and $16(1-4\kappa)$ on 024;

(c) for $\kappa \neq \frac{1}{4}$ and every triangle t, $W_{ll}W_{mm} - W_{lm}^2 = Q_t/(4(1-4\kappa))$;

(d) for $\kappa < \frac{1}{4}$, every Q_t, W_{ll} and W_{mm} is positive, the dihedral angle at t is

$$\theta_t = \arccos\!\big(-W_{lm}/\sqrt{W_{ll}W_{mm}}\big),$$

and along a direction of the squared edge lengths

$$\partial\theta_t = \frac{\mathrm{num}_t}{W_{ll}W_{mm}\sqrt{W_{ll}W_{mm} - W_{lm}^2}}, \qquad \partial A_t = \frac{\partial Q_t}{8\sqrt{Q_t}},$$

with $\mathrm{num}_t = \partial W_{lm}\, W_{ll}W_{mm} - \frac{1}{2}W_{lm}(\partial W_{ll}\, W_{mm} + W_{ll}\, \partial W_{mm})$ and $\partial W = -W(\partial G)W$ in the sense $\partial(G^{-1}) = -G^{-1}(\partial G)G^{-1}$.

Proof. (a) The edge $v_b - v_a$ has $b-a$ axis steps, so its squared length is $(b-a)-\kappa(b-a)^2$, and $G_{ab} = \frac{1}{2}(L_{0a}+L_{0b}-L_{ab}) = \min(a,b)-\kappa ab$. So $G = M - \kappa s s^{\mathsf T}$ with $M_{ab} = \min(a,b)$ and $s = (1,2,3,4) = Me_4$. $M = LL^{\mathsf T}$ with L the lower triangular matrix of ones, so $\det M = 1$, and M^{-1} is the tridiagonal matrix of the statement with last diagonal entry 1. By the matrix determinant lemma $\det G = 1 - \kappa s^{\mathsf T} M^{-1} s = 1 - 4\kappa$, and by the formula of Sherman and Morrison [94] $G^{-1} = M^{-1} + \kappa e_4 e_4^{\mathsf T}/(1-4\kappa)$, whose last entry is $(1-3\kappa)/(1-4\kappa)$. (b) A triangle ijk has edges with $j-i$, $k-j$ and $k-i$ steps; substitute the squared lengths $n-\kappa n^2$ into Heron's expression. (c) By Jacobi's theorem on complementary minors [89], a 2×2 principal minor of G^{-1} equals the complementary principal minor of G divided by $\det G$. For $l,m \geq 1$ the left side is the minor of G^{-1} on $\{l,m\}$, and the complementary minor of G is the Gram determinant of the two edges from v_0 that span t, which is $4A_t^2 = Q_t/4$. When l or m is 0, base the edges at another vertex of t; the determinant $1-4\kappa$ does not change. (d) For $\kappa < \frac{1}{4}$ the values in (b) are positive and G is positive definite, so $W_{ll} = |n_l|^2 > 0$. The angle between the outward normals of the facets l and m is $\pi - \theta_t$. Differentiating G^{-1} gives ∂W; the chain rule for $\arccos u$, $u = -W_{lm}/\sqrt{W_{ll}W_{mm}}$, uses $1-u^2 = (W_{ll}W_{mm}-W_{lm}^2)/(W_{ll}W_{mm})$; and $A_t = \sqrt{Q_t}/4$. □

Parts (c) and (d) do not need the Kuhn simplex. Let $L \in \mathbb{R}^{10}$ be any ten numbers, read as the squared lengths of the edges ij of a 4-simplex, and form from L, as above, $G_{ab} = \frac{1}{2}(L_{0a}+L_{0b}-L_{ab})$, $W = E^{\mathsf T} G^{-1} E$ when $\det G \neq 0$, Q_t, $A_t = \sqrt{Q_t}/4$, $\theta_t = \arccos(-W_{lm}/\sqrt{W_{ll}W_{mm}})$ and $D_t = W_{ll}W_{mm} - W_{lm}^2$. Write ∂_e for the derivative in the squared length L_e of the slot e.

Theorem 27.27 (Schläfli's identity for every nondegenerate 4-simplex)**.** *In classical mathematics, for every $L \in \mathbb{R}^{10}$:*

(a) if $\det G \neq 0$, then $4 \det G \cdot D_t = Q_t$ for each of the ten triangles t;

(b) if $\det G \neq 0$ and every triangle has $W_{ll} > 0$, $W_{mm} > 0$, $D_t > 0$ and $Q_t > 0$, then every θ_t has a derivative $\partial_e \theta_t$ for every slot e, and $\sum_t A_t \, \partial_e \theta_t = 0$.

A nondegenerate Euclidean simplex meets the hypotheses of (b) ($W_{ij} = n_i \cdot n_j$ for its facet normals), so (b) is Schläfli's identity [93] for a 4-simplex.

Proof. (a) For $l,m \geq 1$, D_t is the minor of G^{-1} on $\{l,m\}$, and Jacobi's theorem gives it as in Theorem 27.26(c): the complementary minor of G is $xy - \frac{1}{4}(x+y-z)^2 = \frac{1}{4}Q_t$ for the squared sides x, y, z of t. If l or m is 0, renumber the vertices so that 0 lies on t; this changes the edge basis by an integer matrix of determinant ± 1 and changes neither W nor $\det G$. (b) By (a), $\det G > 0$ and $A_t = \frac{1}{4}\sqrt{4 \det G}\sqrt{D_t}$. Near L, W is differentiable and the chain rule of Theorem 27.26(d) gives

$$A_t \, \partial_e \theta_t = \tfrac{1}{4}\sqrt{4 \det G}\Big(\partial_e W_{lm} - \tfrac{1}{2} W_{lm}(x_l + x_m)\Big), \qquad x_i = \partial_e W_{ii}/W_{ii},$$

every W_{ii} being positive since each vertex lies on an edge opposite some triangle. The ten triangles are the ten pairs $l < m$. W is symmetric with zero row sums ($E\mathbf{1} = 0$), near L, so $\partial_e W$ is too. Hence $\sum_{l<m} \partial_e W_{lm} = -\frac{1}{2}\sum_l \partial_e W_{ll}$ and $\sum_{l<m} W_{lm}(x_l + x_m) = \sum_l x_l \sum_{m \neq l} W_{lm} = -\sum_l \partial_e W_{ll}$, and the sum is $\frac{1}{4}\sqrt{4 \det G}(-\frac{1}{2}+\frac{1}{2})\sum_l \partial_e W_{ll} = 0$. □

Table 27.3. The partial-fraction tables of $R(\kappa)$, scaled to integers; slots in the order of Table 27.1. $72B_2$ has the entries 1 at $(01,03)$ and $(02,03)$, -1 at $(03,13)$, $(03,23)$, $(12,14)$ and $(13,14)$, 1 at $(14,24)$ and $(14,34)$, and their mirror images, and 0 elsewhere; $64B_3$ is diagonal with entries -1 at 02, 2 at 13, -1 at 24, and 0 elsewhere.

$$192B_0 = \begin{pmatrix} 6 & -6 & 2 & 0 & 0 & 12 & -6 & -24 & 6 & 6 \\ -6 & -12 & 14 & -3 & 24 & -24 & 6 & 12 & -6 & 6 \\ 2 & 14 & -2 & -6 & -24 & 4 & 6 & -8 & 6 & -6 \\ 0 & -3 & -6 & 6 & 6 & 6 & -6 & 6 & -3 & 0 \\ 0 & 24 & -24 & 6 & -24 & 24 & -8 & 0 & 12 & -24 \\ 12 & -24 & 4 & 6 & 24 & -12 & 4 & 24 & -24 & 12 \\ -6 & 6 & 6 & -6 & -8 & 4 & -2 & -24 & 14 & 2 \\ -24 & 12 & -8 & 6 & 0 & 24 & -24 & -24 & 24 & 0 \\ 6 & -6 & 6 & -3 & 12 & -24 & 14 & 24 & -12 & -6 \\ 6 & 6 & -6 & 0 & -24 & 12 & 2 & 0 & -6 & 6 \end{pmatrix} \qquad 256B_1 = \begin{pmatrix} 2 & -2 & -2 & 0 & 0 & 0 & -2 & 0 & 2 & 2 \\ -2 & 2 & -2 & -1 & 0 & 0 & 2 & 0 & -2 & 2 \\ -2 & -2 & 2 & -2 & 0 & 0 & 2 & 0 & 2 & -2 \\ 0 & -1 & -2 & 2 & 2 & 2 & -2 & 2 & -1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 2 & 2 & -2 & 0 & 0 & 2 & 0 & -2 & -2 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 2 & -1 & 0 & 0 & -2 & 0 & 2 & -2 \\ 2 & 2 & -2 & 0 & 0 & 0 & -2 & 0 & -2 & 2 \end{pmatrix}$$

For $\kappa \notin \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$ put

$$R(\kappa) = B_0 + \frac{B_1}{\kappa - \frac{1}{4}} + \frac{B_2}{\kappa - \frac{1}{3}} + \frac{B_3}{\kappa - \frac{1}{2}},$$

with the tables of Table 27.3. For a hinge h spanned by the steps p and q, with edges p, q, $q-p$, let $G_0(h) = |p|^2|q|^2 - (p \cdot q)^2$ be its Euclidean Gram determinant, L_κ the squared length at g_κ, and, for an edge e of h with other edges a, b, $N_\kappa(h,e) = L_\kappa(a) + L_\kappa(b) - L_\kappa(e)$, which is half the derivative of Q_h in $L_\kappa(e)$. Then $N_\kappa = N_0 - \kappa S$ with $S(h,e) = \sigma_a^2 + \sigma_b^2 - \sigma_e^2$, σ the coordinate sum of a step.

Theorem 27.28 (The Regge table for every background)**.** *In classical mathematics, for every real $\kappa \notin \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$:*

(a) $48R(0) = E$ *and* $48R(1) = R_1$*;*

(b) B_2 *and* B_3 *are nonzero tables of zero symbol, and* B_1 *is not of zero symbol: the plane-wave form of* $R(\kappa)$ *has its only pole at the degenerate background* $\kappa = \frac{1}{4}$*, and* $\kappa = 1$ *is a regular point;*

(c) for every wave K and all perturbations A, B,

$$R(\kappa)(K; A, B) = \sum_h D_A(h)\, \frac{1}{4(1-4\kappa)} \sum_{e \in h} \frac{N_\kappa(h,e)}{G_0(h)}\, B(e)\, e^{-iK \cdot x_e},$$

where D_A is the Euclidean area-weighted deficit of Theorem 27.20, and x_e is the midpoint of the edge e measured from the hinge's first vertex; at $\kappa = 0$ the inner factor is the log-area change Λ_B; by (b), the identity holds for every $\kappa \neq \frac{1}{4}$ with $B_0(K; A, B) + B_1(K; A, B)/(\kappa - \frac{1}{4})$ on the left;

(d) for $\kappa < \frac{1}{4}$, with the true derivatives of Theorem 27.26(d), $-\sum_t \partial_e \theta_t\, \partial_f A_t = \sqrt{1-4\kappa}\, R(\kappa)_{ef}$ for all slots e, f;

(e) for $\kappa \notin \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$ the algebraic expression $-\sum_t \mathrm{num}_t(e)\, \partial_f Q_t / (4 W_{ll} W_{mm} Q_t)$ equals $R(\kappa)_{ef}$.

Conditions and scope.

Base and premises

The Kuhn simplex at g_κ; classical real numbers.

Conclusion

With Schläfli's identity (Theorem 27.27), $-\sum_t \nabla\theta_t \otimes \nabla A_t$ is the Hessian of Regge's action at the flat background, with true second derivatives, so for $\kappa < \frac{1}{4}$ the Hessian is $\sqrt{\det g_\kappa}\, R(\kappa)$ (Theorem 27.29): one rational family, with the Euclidean and the lattice tables as its values at 0 and 1. The linearized deficits do not change with κ; the background enters only on the area side, through $N_\kappa/(4 \det g_\kappa)$.

Not claimed

True derivatives for $\kappa > \frac{1}{4}$, where angles are Lorentzian; there $R(\kappa)$ is the algebraic continuation. The $\kappa = 1$ derivation on $\mathbb{R}_\delta$ is Theorem 27.32.

Proof. (a) At $\kappa = 0$ the coefficients $1/(\kappa - \frac{1}{4})$, $1/(\kappa - \frac{1}{3})$, $1/(\kappa - \frac{1}{2})$ are -4, -3, -2, and at $\kappa = 1$ they are $\frac{4}{3}$, $\frac{3}{2}$, 2; the identities are entrywise arithmetic with Tables 27.1 and 27.3. (b) Group the terms of a plane-wave form by the classes of the two slots and the difference of their midpoints; a table has zero symbol exactly when every group sums to zero, which is a finite check, and it holds for B_2 and B_3 and fails for B_1. (c) The ratios N_0/G_0 and S/G_0 do not depend on the simplex in which a hinge edge is met, so the right side is $(W_0 - \kappa W_S)/(4(1 - 4\kappa))$ for two fixed rational tables W_0, W_S; and $R(\kappa)$ minus this is a combination of four fixed rational tables of zero symbol, checked by (b)'s criterion; forming them from the deficit rates, N_0, S, G_0 and Table 27.3 is a finite computation, which we describe and do not reproduce. At $\kappa = 0$, $Q_h = 4G_0$ and $\partial \log A_h/\partial L_e = N_0/(4G_0)$. (d) and (e) Assemble Theorem 27.26(d): $-\sum_t \partial\theta_t\, \partial A_t = -\sum_t \mathrm{num}_t\, \partial Q_t/(8\sqrt{Q_t}\, W_{ll} W_{mm} \sqrt{\mathrm{jac}_t})$ with $\mathrm{jac}_t = W_{ll}W_{mm} - W_{lm}^2$, and by (c) of that theorem $\sqrt{\mathrm{jac}_t} = \sqrt{Q_t}/(2\sqrt{1 - 4\kappa})$. So the left side of (d) is $\sqrt{1 - 4\kappa}$ times the expression in (e), which is a rational function of κ entry by entry and equals $R(\kappa)$ off the three poles. □

Fix $\kappa < \frac{1}{4}$, let L_κ be the squared edge lengths of the Kuhn simplex at g_κ, and put $S(L) = \sum_t A_t(L)\big(\theta_t(L_\kappa) - \theta_t(L)\big)$. The flat angles at each hinge of the lattice add to 2π, so a hinge's deficit is the sum over its simplices of $\theta_t(L_\kappa) - \theta_t(L)$, and S is one simplex's share of Regge's action.

Theorem 27.29 (Regge's Hessian at every Euclidean background, with true second derivatives)**.** *In classical mathematics, for every $\kappa < \frac{1}{4}$ and all slots e, f: for every s near 0, S has a derivative $g(s)$ in L_e at $L_\kappa + s\mathbf{1}_f$ ($\mathbf{1}_f$ the unit vector of the slot f), and $g'(0) = \sqrt{1 - 4\kappa}\, R(\kappa)_{ef}$.*

Proof. At L_κ, $\det G = 1 - 4\kappa > 0$ and Q_t, W_{ll}, W_{mm}, $D_t = Q_t/(4(1 - 4\kappa))$ are positive (Theorem 27.26), and they stay so nearby. There the product rule and Theorem 27.27(b) give $\partial_e S(L) = \sum_t \big(\theta_t(L_\kappa) - \theta_t(L)\big)\partial_e A_t(L)$. Along $L_\kappa + s\mathbf{1}_f$ the first factor vanishes at $s = 0$ with derivative $-\partial_f\theta_t$, so $g'(0) = -\sum_t \partial_f\theta_t\, \partial_e A_t = \sqrt{1 - 4\kappa}\, R(\kappa)_{fe}$ by Theorem 27.28(d), and $R(\kappa)$ is symmetric. □

27.7 The near-flat chain on $\mathbb{R}_\delta$

Everything in this section is proved on $\mathbb{R}_\delta$ with no omniscience principle and no choice principle, except the decisions whose prices are named. The cosine, π and the square root are δ's own (Definition 21.6, Theorem 21.8, Proposition 12.19). The lattice is the periodic Kuhn lattice of side N: vertices $(\mathbb{Z}/N)^4$, each unit cube cut into the 24 simplices of Definition 27.19. At the background g_1 an edge with n axis steps has squared length $n(1-n)$: the axis edges are light-like and every longer edge is timelike. For edge fields u, v the *flat Hessian* is

$$\mathscr{H}(u,v) = \frac{1}{16\sqrt{3}} \sum_{x} \sum_{\sigma} \sum_{f,f'} (R_1)_{ff'}\, u(e_{x\sigma f})\, v(e_{x\sigma f'}),$$

with $e_{x\sigma f}$ the edge of slot f of simplex σ at the vertex x. The *plane wave* of an integer wave m and a symmetric amplitude H is $u_{m,H}(e) = (s_e^{\mathsf{T}} H s_e)\cos(2\pi m\cdot x_e/N)$, with s_e the step and x_e the midpoint of the edge e. The *cell symbol* at $k \in \mathbb{R}_\delta^4$ is

$$\operatorname{sym}(k;X,Y) = \sum_{\sigma} \sum_{f,f'} (R_1)_{ff'}\,(s_f^{\mathsf{T}} X s_f)(s_{f'}^{\mathsf{T}} Y s_{f'})\cos y_{ff'}, \qquad y_{ff'} = k\cdot(x_f - x_{f'}),$$

and $\operatorname{rem}(k;X,Y)$ is the same sum with $\cos y$ replaced by $y^2 - 2 + 2\cos y$.

Theorem 27.30 (Regge's Hessian is Fierz–Pauli at long wavelength, with rate)**.** *For symmetric X, Y over $\mathbb{R}_\delta$ and $k \in \mathbb{R}_\delta^4$:*

(a) $\sum_\sigma \sum_{f,f'} (R_1)_{ff'} (s_f^{\mathsf{T}} X s_f)(s_{f'}^{\mathsf{T}} Y s_{f'}) = 0$, *and*

$$\sum_{\sigma} \sum_{f,f'} (R_1)_{ff'} (s_f^{\mathsf{T}} X s_f)(s_{f'}^{\mathsf{T}} Y s_{f'})\, y_{ff'}^2 = 24\,\mathrm{FP}_{g_1}(k;X,Y);$$

on strains TT at k the right side is $24\,(k\cdot k)\operatorname{tr}(XY)$, and when $\operatorname{tr}(XY)$ is apart from 0 it vanishes exactly when k is on the cone;

(b) $2\operatorname{sym}(k;X,Y) = -24\,\mathrm{FP}_{g_1}(k;X,Y) + \operatorname{rem}(k;X,Y)$, *and* $0 \le y^2 - 2 + 2\cos y \le y^4/12$; *so the symbol of $\mathscr{H}$ is $-\frac{\sqrt{3}}{4}\mathrm{FP}_{g_1} + O(k^4)$ with an explicit remainder;*

(c) if N does not divide $2m_c$ for some c, then $16\sqrt{3}\,\mathscr{H}(u_{m,X}, u_{m,Y}) = \frac{1}{2}N^4 \operatorname{sym}(2\pi m/N; X, Y)$;

(d) under the same condition,

$$192\,\mathscr{H}(u_{m,X},u_{m,Y}) = -96\sqrt{3}\,\pi^2 N^2\,\mathrm{FP}_{g_1}(m;X,Y) + \sqrt{3}N^4\operatorname{rem}(2\pi m/N;X,Y);$$

(e) if $m_c \neq 0$ and $2|m_c| < N$ for some c, then

$$\left|\frac{192}{N^2}\mathscr{H}(u_{m,X},u_{m,Y}) + 96\sqrt{3}\pi^2\,\mathrm{FP}_{g_1}(m;X,Y)\right| \le \frac{\sqrt{3}\pi^4}{N^2}\,B(m,X,Y),$$

with $B = \frac{1}{12}\sum_\sigma\sum_{f,f'} |(R_1)_{ff'}(s_f^{\mathsf{T}} X s_f)(s_{f'}^{\mathsf{T}} Y s_{f'})|\,(2m\cdot(x_f - x_{f'}))^4$. So

$$N^{-2}\mathscr{H}(u_{m,X},u_{m,Y}) \to -\frac{\sqrt{3}\pi^2}{2}\mathrm{FP}_{g_1}(m;X,Y)$$

at the rate $O(N^{-2})$.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ with δ's π, cosine and $\sqrt{3}$; integer tables. Free.

Conclusion

The flat Hessian of the $\kappa = 1$ Regge lattice is, at long wavelength and with an explicit error, $-\sqrt{3}\pi^2/2$ times the Fierz–Pauli form of the lattice metric; on transverse-traceless waves it is the dispersion times $\mathrm{tr}(XY)$.

Identification

The Kuhn lattice as discretized spacetime, $\kappa = 1$, the flat field $n(1-n)$ and Regge's action as the gravitational action.

Proof. (a) Both sums are polynomial identities with integer coefficients in the entries of X, Y and k, obtained by expanding over the 24 simplices and 100 slot pairs; the reader can expand them, or compare the coefficients of the 6-index integer tensor they define with those of $24\,\mathrm{FP}_{g_1}$. On TT strains $\mathrm{FP}_{g_1} = (k\cdot k)\,\mathrm{tr}(XY)$ by Theorem 27.6(c), and a product with a factor apart from 0 is 0 exactly when the other factor is. (b) Termwise $2\cos y = 2 - y^2 + (y^2 - 2 + 2\cos y)$; the constant terms cancel by the first identity of (a), and the y^2 terms give $-24\mathrm{FP}_{g_1}$ by the second. The bounds on $y^2 - 2 + 2\cos y$ are the alternating-series bounds for δ's cosine. (c) Sum over the torus: $\sum_x \cos(\theta + 2\pi m\cdot x/N)\cos(\theta' + 2\pi m\cdot x/N) = \frac{1}{2}N^4\cos(\theta-\theta')$, because the doubled-frequency terms sum to 0 over a full period when $N \nmid 2m_c$. (d) is (c) with (b) at $k = 2\pi m/N$, using $\mathrm{FP}_{g_1}(2\pi m/N) = (4\pi^2/N^2)\mathrm{FP}_{g_1}(m)$. (e) Divide (d) by N^2 and bound the remainder by (b), with $y = (\pi/N)\,2m\cdot(x_f - x_{f'})$. □

J reads the same comparison. Compare the record metric $g_1 + \varepsilon H$ with g_1: the ratio is $g_1^{-1}(g_1 + \varepsilon H) = 1 + \varepsilon X$ with the metric strain $X = g_1^{-1}H$.

Proposition 27.31 (*J's transverse form is Fierz–Pauli with coefficient one in lattice units*)**.** *On $\mathbb{R}_\delta$, free:*

(a) if $(1+\varepsilon X)B = 1$, then $J(1+\varepsilon X) = \frac{1}{2}(\mathrm{tr}(1+\varepsilon X) + \mathrm{tr}\,B) - 4 = \frac{\varepsilon^2}{2}\,\mathrm{tr}(X^2B)$;

(b) for H TT at m, $Q_J(m,H) = (m\cdot m)\cdot 2J_2(g_1^{-1}H)$ equals $\mathrm{FP}_{g_1}(m;H,H)$;

(c) for H TT at the integer wave m and $N \nmid 2m_c$ for some c, $192\,\mathscr{H}(u_{m,H},u_{m,H}) = -24\sqrt{3}N^4\,Q_J(2\pi m/N,H) + \sqrt{3}N^4\,\mathrm{rem}(2\pi m/N;H,H)$.

Proof. (a) is the identity of Theorem 27.2 with the inverse given as data. (b) $H\mathbf{1} = 0$ gives $g_1^{-1}H = H$; then Theorem 27.6. (c) is Theorem 27.30(d) with (b). □

So J's calibrated transverse form is Fierz–Pauli with coefficient one in the lattice's units; $J''(1) = 1$ fixes the shape of the action, and its strength comes from outside J (Proposition 27.14).

The flat Hessian is the second variation of Regge's action itself, computed on $\mathbb{R}_\delta$. A cell is *nondegenerate* when the denominators of its angle formulas are apart from 0.

Theorem 27.32 (Regge's action on $\mathbb{R}_\delta$ and its second variation)**.** *On $\mathbb{R}_\delta$, free:*

(a) δ's cosine has an inverse arccos *on $[-1,1]$:* $\cos(\arccos c) = c$, $0 \le \arccos c \le \pi_\delta$ *and* $\sin(\arccos c) = \sqrt{1-c^2}$*; on $[-1+\frac{1}{m+1}, 1-\frac{1}{m+1}]$ it is Lipschitz with constant $L = 64(m+1)^2$ and has a first-order rate with error at most $4L^2(c'-c)^2$;*

(b) every nondegenerate Kuhn cell has native dihedral angles with first-order rates, and Schläfli's identity $\sum_t A_t \, \mathrm{d}\theta_t = 0$ holds for it;

(c) at the flat $\kappa = 1$ cell, with ε_t the simplex's share of the deficit ($\partial\varepsilon_t = -\partial\theta_t$), $\sum_t \partial_e A_t \, \partial_f \varepsilon_t = -\sum_t \partial_e A_t \, \partial_f \theta_t = (R_1)_{fe}/(16\sqrt{3})$; the first-order rates are the derivatives of δ's own areas and angles, with second-order remainders bounded by one constant near the flat cell;

(d) there is a constant $K > 0$ such that for every side N, every edge field w with $|w_e| \le 1$ and every $|x| \le 1/40000000$ for which every cell of ℓ and of $\ell + xw$ is nondegenerate (so that S is defined at both), the Regge action S (the sum over hinges of area times deficit, with the angles of (b)) satisfies

$$\left| S(\ell + xw) - S(\ell) - \tfrac{1}{2} x^2 \, \mathscr{H}(w,w) \right| \le 24 N^4 K |x|^3,$$

where ℓ is the flat field.

Proof. The proof assembles finite identities and estimates; we give its route and the source of each estimate, and do not carry the bounds out. (a) $\arccos c = \pi_\delta/2 - \arcsin c$, with δ's arcsine, the inverse of the sine on $[-\pi_\delta/2, \pi_\delta/2]$, given on a grid of points with explicit moduli; the bounds follow from $\sin^2 + \cos^2 = 1$ and the bounds of the cosine on the closed subinterval. (b) The dihedral angle at a triangle is π_δ minus the arccos of the normalized product of the two facet normals (the angle between the normals), as in Theorem 27.26(d); its rate is the chain rule of (a). Schläfli's identity is then an identity between rational expressions in the squared edge lengths and the square roots of the cell's Gram determinants, which holds after clearing denominators. (c) At the flat cell the area rates A_t'/A_t are $1/12$ times an integer table and the weighted deficit rates $A_t\varepsilon_t'$ are $1/(8\sqrt{3})$ times another; their product over the ten triangles is $6R_1$, and $6/(12 \cdot 8\sqrt{3}) = 1/(16\sqrt{3})$. Second-order remainders come from (a) applied along a line of cells. (d) Census: the 240 slot types of the lattice fall into 50 hinge fibers, and the flat angles of each fiber add to 2π, so the flat field has zero deficit. Along $x \mapsto \ell + xw$, Schläfli's identity removes the angle-change term at each step of a telescoping sum, summation by parts handles the area term, and (c) identifies the quadratic term; the cubic bound collects the remainders of (a) and (c), cell by cell, which gives the factor $24N^4$. The constant is $K = 100K_0^2$, where K_0 bounds, on every cell whose ten squared lengths lie within $1/40000000$ of the flat cell's and along every direction with entries in $[-1,1]$, the areas, their rates, the angle rates, and the first-order remainders of areas and angles divided by x^2; K_0 exists by (a) at $m = 4999$ and by bounds on the Gram polynomials of these cells, and we do not compute it. □

So the chain from Regge's action through its Hessian, 24 times Fierz–Pauli and the long-wave limit $-(\sqrt{3}\pi^2/2)\,\mathrm{FP}_{g_1}$ runs on $\mathbb{R}_\delta$ with no choice and no omniscience.

The same assembly reaches one order higher. At the flat $\kappa = 1$ cell let A_t', A_t'' and ε_t', ε_t'' be the first and second derivatives, along a direction d of the ten squared edge lengths, of

the area of the triangle t and of the simplex's share ε_t of its deficit. The *cubic cell form* is $C(d) = \sum_t \big(2A_t''\varepsilon_t' + A_t'\varepsilon_t''\big)$, the analogue at $\kappa = 1$ of $-r$ in Definition 27.23; it is $(1152\sqrt{3})^{-1}$ times a cubic form with integer coefficients.

Theorem 27.33 (Regge's cubic vertex on $\mathbb{R}_\delta$)**.** *On $\mathbb{R}_\delta$, free: there is $C_0 > 0$ such that for every side N, every edge field w with $|w_e| \le 1$ and every $|x| \le 1/40000000$ for which every cell of ℓ and of $\ell + xw$ is nondegenerate,*

$$\Big|S(\ell + xw) - S(\ell) - \tfrac{1}{2}x^2\,\mathscr{H}(w,w) - \tfrac{1}{6}x^3 \sum_y \sum_\sigma C(w_{y\sigma})\Big| \le 24N^4 C_0\, x^4,$$

where $w_{y\sigma}$ lists the values of w on the slots of the simplex σ at the vertex y. At the flat cell the mixed differences of δ's own area and angle differ from xy times their second derivatives by at most a constant times $|xy|(|x| + |y|)$, for $|x|, |y| \le 1/80000000$ (for the angle, at nondegenerate cells).

Proof. As for Theorem 27.32(d), one order further, and again we give the route and the source of each estimate. Per simplex put $a_t(\tau) = A_t(\ell + \tau xw)$ and $b_t(\tau) = \varepsilon_t(\ell + \tau xw) - \varepsilon_t(\ell)$, so that $S(\ell + xw) - S(\ell)$ sums $\sum_t a_t(1)b_t(1)$. Summation by parts on the grid $\tau = k/m$, with Schläfli's identity at each grid point, reduces $\sum_t a_t(1)b_t(1)$ to $\sum_t \int_0^1 a_t' b_t\, d\tau$ up to $O(1/m)$. With the quadratic Taylor coefficients $a_1 = xA_t'$, $a_2 = x^2A_t''$ and b_1, b_2 likewise, this integral is $\frac{1}{2}a_1b_1 + \frac{1}{6}(2a_2b_1 + a_1b_2) + \frac{1}{8}a_2b_2$ up to the Taylor remainders; the last term and the remainders are $O(x^4)$, since δ's area and angle have cubic remainders bounded by fixed constants k_A, k_θ and a_1, b_1 are $O(|x|)$. By Theorem 27.32(c) the first term is the simplex's share of $\frac{1}{2}x^2\mathscr{H}(w,w)$, and $\sum_t(2a_2b_1 + a_1b_2) = x^3C(w_{y\sigma})$; sum over the $24N^4$ simplices. The mixed differences are bounded by Taylor's theorem in the regimes $|y| \le |x|$ and $|x| \le |y|$; the bound is a non-strict inequality, the negation of a strict one, so the case split, which holds up to double negation, needs no omniscience. The constant is $C_0 = 1 + \sum_t\big(k_A(2K_0 + \beta_t + k_\theta) + (K_0 + \alpha_t)k_\theta + \alpha_t\beta_t\big)$, with K_0 from the proof of Theorem 27.32 and α_t, β_t the sums of the absolute values of the entries of the second-order tables of A_t and ε_t at the flat cell. □

Why area times deficit? A *hinge charge* $\sigma(A)$ is the action a hinge of area A carries per unit of deficit. At background $\kappa \ge \frac{1}{2}$ the flat cell has four hinge classes with squared areas $(4\kappa - 2)/8$, $(12\kappa - 4)/8$, $3(4\kappa - 1)/4$ and $4\kappa - 1$ (Theorem 27.26(b), up to sign), all nonnegative; call their nonnegative roots $a_0, \dots, a_3$. The flat field is *stationary* at $\kappa \ge \frac{1}{2}$ for the charge σ when $\sigma(a_2)a_3 = \sigma(a_3)a_2$ and $a_0\sigma(a_0) - a_1\sigma(a_1) + \frac{1}{4}a_3\sigma(a_3) = 0$. For $\kappa > \frac{1}{2}$ these are the two conditions of the stationarity criterion [96, Prop. 4.1]: for an action that sums a continuously differentiable hinge charge $C(A, \delta)$ with $C(A, 0) = 0$ and slope $\sigma(A) = \partial_\delta C(A, 0)$ over the hinges, they hold exactly when the flat lattice is a stationary point of the action. We take them as the definition. The function σ is *pointwise continuous on* $[\frac{1}{2}, \infty)$ when for every $x \ge \frac{1}{2}$ and k there is a j with $|\sigma(y) - \sigma(x)| \le \frac{1}{k+1}$ whenever $y \ge \frac{1}{2}$ and $|y - x| \le \frac{1}{j+1}$.

Theorem 27.34 (The slope theorem is free). *On $\mathbb{R}_\delta$:*

(a) Regge's charge $\sigma(A) = \lambda A$ is stationary at every $\kappa \geq \frac{1}{2}$;

(b) if σ is pointwise continuous on $[\frac{1}{2}, \infty)$ and the flat field is stationary for σ at every $\kappa > 1$, then $\sigma(A) = \sigma(1)A$ for every $A \geq \frac{1}{2}$; free;

(c) without continuity the conclusion fails under the WLPO function: given it, the slope $\omega(A) = A$ when A^2 is rational and $\omega(A) = 0$ otherwise is a map, stationary at every $\kappa \geq \frac{1}{2}$, and not linear.

Proof. (a) Both conditions are linear identities in λ once $\sigma(a) = \lambda a$; the second is $\lambda\big(\frac{4\kappa-2}{8} - \frac{12\kappa-4}{8} + \frac{4\kappa-1}{4}\big) = 0$.

(b) Put $h(u) = \sqrt{u}\,\sigma(\sqrt{u})$ for $u \geq \frac{1}{4}$, so that $a_i\sigma(a_i) = h(a_i^2)$; h is pointwise continuous on $[\frac{1}{4}, \infty)$, as σ is on $[\frac{1}{2}, \infty)$. Write $t = 4\kappa - 1$, so that $\kappa > 1$ is $t > 3$ and the squared areas are $(t-1)/8$, $(3t-1)/8$, $3t/4$ and t. Multiplying the first condition by a_2a_3 and dividing by t, stationarity at every $\kappa > 1$ says that for every $t > 3$

$$\text{(E1)} \quad h\big(\tfrac{3}{4}t\big) = \tfrac{3}{4}h(t), \qquad\qquad \text{(E2)} \quad h\big(\tfrac{t-1}{8}\big) - h\big(\tfrac{3t-1}{8}\big) + \tfrac{1}{4}h(t) = 0.$$

The limit. Read upwards, (E1) gives $h((\frac{4}{3})^n z) = (\frac{4}{3})^n h(z)$ for $z \geq 3$. Let $x \geq 25$ and apply (E2) at $t = (\frac{4}{3})^n x$, that is, at $\kappa = ((\frac{4}{3})^n x + 1)/4$. As $(t-1)/8 = (\frac{4}{3})^n(x/8 - \varepsilon_n)$ and $(3t-1)/8 = (\frac{4}{3})^n(3x/8 - \varepsilon_n)$ with $\varepsilon_n = (\frac{3}{4})^n/8$, and $x/8 - \varepsilon_n \geq 3$, dividing by $(\frac{4}{3})^n$ gives $h(x/8 - \varepsilon_n) - h(3x/8 - \varepsilon_n) + \frac{1}{4}h(x) = 0$ for every n. As $\varepsilon_n \to 0$, continuity of h at $x/8$ and at $3x/8$ gives

$$\text{(H)} \quad h(x/8) - h(3x/8) + \tfrac{1}{4}h(x) = 0 \qquad \text{for every } x \geq 25.$$

Ratios near 1. For every $J \geq 4$ there are $n, m \geq 1$ with $\rho = 2^n(\frac{3}{4})^m$ and $|\rho - 1| \leq \frac{1}{J+1}$. Let k_j be the largest k with $2^k \leq 3^j$. Of the $J+2$ ratios $3^j 2^{-k_j} \in [1, 2)$, $j = 0, \dots, J+1$, two lie in one of the $J+1$ intervals $[1 + \frac{i}{J+1}, 1 + \frac{i+1}{J+1})$; the later over the earlier is $\rho = 3^m 2^{-k}$ with $m \geq 1$ and $|\rho - 1| \leq \frac{1}{J+1} \leq \frac{1}{5}$, so $2^k \leq \frac{5}{4}\,3^m < 4^m$ and $\rho = 2^n(\frac{3}{4})^m$ with $n = 2m - k \geq 1$.

Period two. Let $W(u) = h(4u) - 4h(2u) + 4h(u)$. (H) at $x = 8u$ with (E1) at $t = 4u$ gives $h(8u) - 3h(4u) + 4h(u) = 0$, that is, $W(2u) = -W(u)$ for $u \geq 4$; (E1) at $4t$, $2t$ and t gives $W(\frac{3}{4}t) = \frac{3}{4}W(t)$ for $t > 3$. Let $u \geq 6$ and ρ as above. The points $2^n u(\frac{3}{4})^i$, $i < m$, exceed 3, so $W(\rho u) = (\frac{3}{4})^m W(2^n u)$, while $|W(2^n u)| = |W(u)|$; as $m \geq 1$, $|W(\rho u)| \leq \frac{3}{4}|W(u)|$. (With $m = 0$ this fails: $|W(2u)| = |W(u)|$.) Continuity of h at u, $2u$ and $4u$ gives $|W(\rho u) - W(u)| \leq \frac{9}{k+1}$ once J is large, for each k; so $\frac{1}{4}|W(u)| \leq \frac{9}{k+1}$ for every k, and $W(u) = 0$. Then $V(u) = h(2u) - 2h(u)$ has $V(2u) - 2V(u) = W(u) = 0$ for $u \geq 6$, so $2h(2^n u) = 2^n(2h(u) + nV(u))$, and by (E1) $h(\rho u) = (\frac{3}{4})^m h(2^n u)$. Hence $n\rho V(u) = 2(h(\rho u) - h(u)) - 2(\rho - 1)h(u)$ with $n\rho \geq \frac{4}{5}$, and continuity of h at u gives $V(u) = 0$: $h(2u) = 2h(u)$ for every $u \geq 6$.

Linear far out. For $u \geq 6$, (E1) at $t = 4u$ gives $h(3u) = \frac{3}{4}h(4u) = 3h(u)$; so for $t \geq 49$, eight times (E2) is $h(t-1) - 3h(t - \frac{1}{3}) + 2h(t) = 0$. On the points $r_j = 49 + j/3$ the values

$R_j = h(r_j)$ satisfy $2R_{j+3} - 3R_{j+2} + R_j = 0$, whose characteristic polynomial is $(r-1)^2(2r+1)$; so $R_j = a + bj + c(-\frac{1}{2})^j$. As $2r_j = r_{147+2j}$, period two gives $R_{147+2j} = 2R_j$, which at $j = 0$ and $j = 1$ reads $147b - a + c((-\frac{1}{2})^{147} - 2) = 0$ and $147b - a + c(\frac{1}{4}(-\frac{1}{2})^{147} + 1) = 0$. The difference is $c(\frac{3}{4}(-\frac{1}{2})^{147} - 3) = 0$, so $c = 0$, $a = 147b$ and $h(r_j) = Cr_j$ with $C = 3b$. Halving, $h(r) = Cr$ at every point $r_j/2^i \geq 49$; these points lie $1/(3 \cdot 2^i)$ apart, and continuity of h at each $y \geq 49$ gives $h(y) = Cy$.

The descent. For $v > \frac{1}{4}$, (E2) at $t = 8v + 1 > 3$ reads $h(v) = h(3v + \frac{1}{4}) - \frac{1}{4}h(8v + 1)$. If $h(w) = Cw$ for every $w > \frac{1}{4}$ with $w \geq Y$, then for every $v > \frac{1}{4}$ with $v \geq Y/3$ both arguments on the right are at least Y, and $h(v) = C(3v + \frac{1}{4}) - \frac{C}{4}(8v + 1) = Cv$. From $Y = 49$, $h(v) = Cv$ for every $v > \frac{1}{4}$ after finitely many steps; so $C = h(1)$, and continuity at $\frac{1}{4}$ gives $h(\frac{1}{4}) = \frac{C}{4}$. With $u = A^2$, $\sigma(A) = \sigma(1)A$ for every $A \geq \frac{1}{2}$. Continuity is used only at the points $x/8$, $3x/8$, u, $2u$, $4u$, y and $\frac{1}{4}$, one point at a time; no modulus and no omniscience enter, as the ratios ρ come from a search over integers.

(c) The WLPO function decides, by three nested applications and the zero test it yields, whether a point is a ratio, so ω is a map. At every $\kappa \geq \frac{1}{2}$ the four squared areas are all ratios when κ is rational and none is otherwise, so each condition holds term by term, by (a) or trivially. At $A = 2^{1/4}$, $A^2 = \sqrt{2}$ is not a ratio, so $\omega(A) = 0 \neq A\,\omega(1)$. □

The theorem is the slope theorem of [96]; on $\mathbb{R}_\delta$ it costs nothing, and no route through uniform continuity is needed.

Proposition 27.35 (Regge's charge and the doublers)**.** *On $\mathbb{R}_\delta$, free:*

(a) if a hinge charge reads only the content, $C(A, \delta) = F(A\delta)$, and one posting record fixes it at the posting unit, $C(A, 2\pi_\delta/k) = sA \cdot 2\pi_\delta/k$ for every A, then $C(A, \delta) = sA\delta$ for all A and δ, with no continuity;

(b) the four corner modes of the Kuhn lattice (momentum π along one axis) are non-gauge zero modes of the flat Hessian of the background g_κ, among all $\kappa > \frac{1}{2}$ and on every torus of even side, exactly when $\kappa = 1$.

Proof. (a) $A\delta = (A\delta k/2\pi_\delta) \cdot (2\pi_\delta/k)$, so $C(A, \delta) = F(A\delta) = C(A\delta k/2\pi_\delta, 2\pi_\delta/k) = sA\delta$. (b) On a corner mode the flat Hessian reduces to a 15×15 matrix over the edge classes, with entries rational in κ and built from the flat table of the background g_κ, which is itself computed on $\mathbb{R}_\delta$; its kernel, computed as a function of κ, consists of the vertex moves unless $\kappa = 1$. For $\kappa > \frac{1}{2}$ the entries are a positive factor times integer polynomials of degree one in κ; the vertex moves are null, the doubler h_c of Proposition 27.36 is null at $\kappa = 1$ and is no vertex move, and for κ apart from 1 an integer certificate with the factor $12\kappa(\kappa-1)(2\kappa-1)(3\kappa-1)(7\kappa-2)(12\kappa^2-1)$ makes every null vector a vertex move. These polynomial identities are finite; we describe them and do not reproduce them. □

These corner modes, the lattice's doublers, are where a one-loop computation first meets trouble, and we record its exact inputs. On the torus of even side $N \geq 4$ let $\delta'_h(u)$ be the linear change of the deficit of the hinge h under the edge field u, A_h its flat area, n_h the

number of simplices around it, and $D(u,v) = \frac{1}{2}\sum_h (A_h/n_h)\,\delta_h'(u)\,\delta_h'(v)$. The doubler at the corner c is $h_c(e) = (-1)^{x_c} v_c(d)$ on the edge of class d based at x, with $v_c(d) = 2^{|d|-1}$ if $d_c = 0$ and $v_c(d) = 0$ otherwise. Let K_c be the symmetric integer matrix with diagonal entries -2 at c and 52 elsewhere and off-diagonal entries 9 in row and column c and -3 elsewhere, and $u_c(q) = q^{\mathsf{T}} K_c q$. For the residue let $T(u,v,w) = \sum_t \big(\varepsilon_t''(u,v)A_t'(w) + \varepsilon_t'(u)A_t''(v,w) + \varepsilon_t'(v)A_t''(u,w)\big)$, with the first and mixed second derivatives at the flat $\kappa = 1$ cell, so that $T(d,d,d) = C(d)$. The *probes* are the constant strains $a(d) = d^{\mathsf{T}} H_a d$, $a = E, T$: H_E is 0 on the diagonal, 2 at $(0,1)$, $(1,0)$, $(2,3)$, $(3,2)$ and -1 elsewhere; H_T is diag$(-1,-1,1,1)$ with 1 at $(0,1)$, $(1,0)$, -1 at $(2,3)$, $(3,2)$ and 0 elsewhere. For a simplex σ of the unit cell let a_σ and h_σ list the values of a and of h_c on its ten edges. The *corner pairing* is $p_a(c) = \sum_\sigma T(a_\sigma, h_\sigma, h_\sigma)$, the *residue* is $\mathrm{res}_a(c) = p_a(c)^2/\big(\sum_d v_c(d)^2\big)^2$, and the corners' *anisotropic residue* is $\rho = \sum_c \big(\mathrm{res}_E(c)/24 - \mathrm{res}_T(c)/8\big)$.

Proposition 27.36 (The doublers' weight, dispersion form and logarithm)**.** *On $\mathbb{R}_\delta$, free:*

(a) $D(h_c,h_c) = N^4 w_c$ *with* $w_c = \frac{5}{2} + \frac{3}{4}\sqrt{3} > 0$ *at every corner c, that is* $D(h_c,h_c) = \frac{10+3\sqrt{3}}{124}\sum_e h_c(e)^2$*;*

(b) $\det K_c = -1013375 = -55^2 \cdot 335$*, and each u_c has one negative and three positive directions;*

(c) the corners' anisotropic residue is $\rho = -3/7688$*, and with* $d_L = 3\sqrt{1013375}/1240^2$*,* $\rho(-\pi^2)/(2d_L(2\pi)^4) = 5/(44\sqrt{335}\,\pi^2)$*; so if* $|I(\varepsilon) + \pi^2 L(\varepsilon)| \le B$ *for* $0 < \varepsilon \le 1$*, then* $\rho I(\varepsilon)/(2d_L(2\pi)^4) - \frac{5}{44\sqrt{335}\pi^2} L(\varepsilon)$ *is bounded on* $(0,1]$*.*

Proof. (a) On an even torus each $\delta_h'(h_c)$ is $(-1)^{y_c}(m\sqrt{3}/3 + n)$, with y the first vertex of h and integers $m \in \{0, \pm1, \pm3\}$, $n \in \{0, \pm1\}$ fixed by the hinge's class through the flat deficit rates; summing $(A_h/n_h)(m\sqrt{3}/3 + n)^2$ over the fifty hinge classes at a vertex gives $2w_c$, and $\sum_d v_c(d)^2 = 3 + 12 + 16 = 31$. (b) K_c is K_0 with the coordinates 0 and c exchanged, and $6u_0(q) = -3y_0^2 + 165y_1^2 + 55y_2^2 + 335y_3^2$ with $y_0 = 2q_0 - 9(q_1+q_2+q_3)$, $y_1 = q_1 - q_2$, $y_2 = q_1 + q_2 - 2q_3$, $y_3 = q_1 + q_2 + q_3$, a substitution of determinant 12; so $\det K_0 = 12^2(-3 \cdot 165 \cdot 55 \cdot 335)/6^4$. (c) At every corner $p_E(c) = 0$ and $1152\sqrt{3}\,p_T(c) = 1728 s_c$ with $s_c = -1, -1, 1, 1$; these are finite sums over the 24 simplices with the integer coefficients of $1152\sqrt{3}\,T$, which we describe and do not reproduce. So $\mathrm{res}_E(c) = 0$, $\mathrm{res}_T(c) = \frac{3}{4}/31^2 = 3/3844$ and $\rho = -4 \cdot 3/(8 \cdot 3844)$. Then $\sqrt{1013375} = 55\sqrt{335}$ and $3 \cdot 1240^2/(7688 \cdot 2 \cdot 165 \cdot 16) = 5/44$, and the difference in the last claim is $\rho\big(I(\varepsilon) + \pi^2 L(\varepsilon)\big)/(2d_L(2\pi)^4)$. □

Read physically, on four named identifications, D is the Hessian of the weight that the ledger cost of the hinge holonomies gives a lattice history, so (a) says that this weight gaps all four doublers; u_c is the doubler's dispersion: $(\sqrt{3}/640)\,u_c$ is the second-order jet of the Schur complement onto the doubler computed from tabulated Taylor coefficients at the corner, on the named premises that these are the Taylor coefficients of the gauge-fixed $\kappa = 1$ operator (a computation not proved here) and that the doubler's eigenvalue follows the jet; and with I the imaginary part of the model integral and $L = \log(1/\varepsilon)$, (c) gives the bare action's doubler logarithm on these premises and that of the integral's asymptotics, the side going to infinity before the damping ε goes to 0. The one-loop transverse-traceless remainder itself is not computed. The quartic cell form at $\kappa = 1$ is $(576\sqrt{3})^{-1}$ times a symmetric integer table, built from third-order formulas whose identification with third derivatives is open.

Are the doublers flat? We answer this, with the $\kappa = 1$ form of Theorem 27.20(a),(b), one wave at a time over any commutative ring R. Let $z_0, \dots, z_3$ be units of R (for a wave K, $z_a = e^{iK_a}$) and $z^x = \prod_a z_a^{x_a}$. A *strain* of the wave z is a map v from the fifteen classes to R, the edge of class d from x to $x+d$ changing its squared length by $v(d)z^x$ (in Definition 27.19, $A(d) = e^{-iK\cdot d/2}v(d)$). The vertex displacement ηz^x ($\eta \in R^4$) has the strain $G_\eta(d) = 2\,(d^{\mathsf T} g_1\eta)(z^d - 1)$, as an amplitude $4i(d \cdot g_1\eta)\sin(K\cdot d/2)$, the gauge amplitude of $g_1\eta$ in Definition 27.19. The fifty hinges at the origin are the triangles $\{0, p, q\}$ with $0 < p < q$ in $\{0,1\}^4$ (coordinatewise). With $P_{tf} = 8\sqrt{3}\,A_t\,\partial_f\varepsilon_t$, the integer table of the flat $\kappa = 1$ cell (Theorem 27.32(c)), the same for every simplex, the strain vz^x has area-weighted deficit $z^y D_v(h)$ at the hinge $h + y$, where $\mathscr{D}_v(h) = 8\sqrt{3}\,D_v(h) = \sum P_{tf}\,v(s_f)\,z^{x_f}$ over the simplices having h as their triangle t and their slots f, x_f the lower end of the slot's edge.

Proposition 27.37 (Beyond gauge, the $\kappa = 1$ lattice has only the flat moduli)**.** *Over every commutative ring R, free:*

(a) at every wave, $\mathscr{D}_{G_\eta} = 0$ for every $\eta \in R^4$;

(b) if 6 is invertible, each z_a is 1 or has $z_a - 1$ invertible, and not every z_a is 1, then every strain v with $\mathscr{D}_v = 0$ is G_η for some $\eta \in R^4$;

(c) at $z = 1$, if 6 is invertible, $\mathscr{D}_v = 0$ exactly when $v(d) = d^{\mathsf T} H d$ for a symmetric 4×4 matrix H over R, and H is unique: these are the ten constant metric strains;

(d) on $\mathbb{R}_\delta$ and the torus of even side $N \geq 4$, the doubler h_c of Proposition 27.36, the strain v_c of the wave $z_c = -1$, $z_b = 1$ ($b \neq c$), is a zero mode of the flat Hessian, but $\mathscr{D}_{v_c}$ is 12, 12, 12, -8 for $c = 0, 1, 2, 3$ at the hinge $\{0, e_3, q\}$, with $q = e_2 + e_3$ for $c = 0$ and $c = 1$, $q = e_1 + e_3$ for $c = 2$ and $q = e_1 + e_2 + e_3$ for $c = 3$; so h_c is the strain of no vertex displacement.

Proof. Each step is a polynomial identity, checked coefficient by coefficient, or linear algebra over R. Write $\mathscr{D}_v(h) = \sum_d \mathscr{D}_{hd} v(d)$, $G_\eta(d) = \sum_a G_{da}\eta_a$: Laurent polynomials in z with integer coefficients, built from the rows P_t (slots in the order 01, … , 34; each column sums to 0, Schläfli's identity):

012	0	2	−1	0	0	0	0	0	1	−2
013	2	−4	2	0	0	−6	4	12	−8	4
014	0	0	0	0	0	9	−6	−18	9	0
023	−4	2	−6	4	0	12	−8	−6	4	0
024	0	−6	12	−6	12	−24	12	12	−6	0

034	0	9	−6	0	−18	9	0	0	0	0
123	0	0	3	−2	0	−6	3	0	0	0
124	0	4	−8	4	−6	12	−6	0	2	−4
134	4	−8	4	0	12	−6	2	0	−4	2
234	−2	1	0	0	0	0	−1	0	2	0

(a) is $\sum_d \mathscr{D}_{hd}G_{da} = 0$ (fifty hinges, four axes); read with z_a the translation along e_a, it covers every displacement field. (b) If every $z_a - 1$ is invertible, take η with $(g_1\eta)_a = v(e_a)/(2(z_a - 1))$, using $g_1^{-1} = 1 - \frac{1}{3}\mathbf{1}\mathbf{1}^{\mathsf T}$; $r = v - G_\eta$ vanishes on the axis classes and $\mathscr{D}_r = 0$. For each three-step class T the hinge $\{0, T, \mathbf{1}\}$ carries -36 on T and otherwise only classes below T; eliminating the $r(T)$ by these, the hinges $\{0, e_3, e_0 + e_2 + e_3\}$, $\{0, e_1, e_0 + e_1 + e_3\}$, $\{0, e_1 + e_2, e_1 + e_2 + e_3\}$, $\{0, e_1 + e_2, e_0 + e_1 + e_2\}$ and $\{0, e_a + e_3, \mathbf{1}\}$ ($a \leq 2$) give seven equations in the two-step amplitudes and $r(\mathbf{1})$ with determinant $2^{18}3^3(z_0z_2)^{-1}(z_0 - 1)^3(z_1 - 1)^3(z_2 - 1)^4(z_3 - 1)^2$, a unit; so these vanish, and then every $r(T)$. Otherwise read η off e_a ($z_a - 1$ invertible) and $e_a + e_b$ ($z_a = 1$; b least with $z_b - 1$ invertible), where $G_\eta(e_a + e_b) = 2((g_1\eta)_a + (g_1\eta)_b)(z_b - 1)$; in each of the fourteen cases a Laurent combination of hinge equations gives $\Delta r = 0$ on the other eleven

classes, Δ a unit (± 8, ± 48, ± 72 or ± 144 times a monomial and powers of the invertible $z_a - 1$); the determinant and these eliminations are finite computations with the table above, which we describe and do not reproduce. (c) At $z = 1$, $G_\eta = 0$ and $\sum_d \mathscr{D}_{hd}(1)\, d_a d_b = 0$ for all h, a, b. Conversely $H_{aa} = \nu(e_a)$, $2H_{ab} = \nu(e_a + e_b) - \nu(e_a) - \nu(e_b)$ make $r(d) = \nu(d) - d^{\mathsf{T}} H d$ vanish on the axis and two-step classes; the hinges $\{0, T, \mathbf{1}\}$ give $r(T) = 0$, and $\{0, e_3, e_2 + e_3\}$, with entry -4 at $\mathbf{1}$, gives $r(\mathbf{1}) = 0$. The same formulas recover H. (d) At the corner every entry is an integer; $\mathscr{D}_{\nu_c}$ is nonzero at 38 hinges, with the stated values, while by (a) a displacement has none. Grouped by hinges, Theorem 27.32(c) reads $\mathscr{H}(w, \nu) = \sum_h \Lambda_w(h) D_\nu(h)$, $\Lambda_w(h)$ the linear change of $\log A_h$; summed over translations, $\mathscr{H}(w, h_c) = 0$ is the rational identity $\sum_h \Lambda_{hd} \mathscr{D}_{\nu_c}(h) = 0$ for every class d, Λ_{hd} the coefficient of $\nu(d)$ in $\Lambda_\nu(h)$ at the corner. □

So the doublers are massless without being flat; and over $\mathbb{C}$, classically, (b) holds at every $K \notin (2\pi\mathbb{Z})^4$, a wider range than that of Theorem 27.20(b) at the Euclidean background.

27.8 The background $\kappa = 1$ at its exact scope

J is blind to the background (Proposition 27.7), and the native chain assumes $\kappa = 1$. What fixes it? This section proves that $\kappa = 1$ is equivalent to each of several statements about records and postings, and isolates the weakest: some record of a ledger history is definite. It is not derived. All results are proved in classical mathematics.

Records are those of Chapter 24: a 2×2 coherent record $R(p) = \frac{1}{2}(p_0 1 + \vec{p} \cdot \vec{\sigma})$ has coordinates $p = (p_0, \vec{p}) \in \mathbb{R}^4$, the record cone is $\mathscr{K} = \{p_0 \geq |\vec{p}|\}$, and $Q(p) = p_0^2 - |\vec{p}|^2 = 4 \det R(p)$, so a record is pure exactly when $Q(p) = 0$. A record with $p_0 = 1$ has *squared Bloch length* $|\vec{p}|^2$. The pairing $\langle p, q \rangle = \operatorname{tr}(R(p)R(q)) = \frac{1}{2}(p_0 q_0 + \vec{p} \cdot \vec{q})$ is the probability a test assigns to a record.

Definition 27.38 (The Kuhn form, cone identifications and definite records)**.**

1. The *Kuhn form* at κ is $q_\kappa(v) = \kappa(\sum v)^2 - |v|^2$, minus the squared length of v at g_κ; its future cone is $\{\sum v \geq 0,\ q_\kappa(v) \geq 0\}$. A single posting is a unit vector e_p; an edge class d with $|d|$ ones has $q_\kappa(d) = |d|(\kappa|d| - 1)$; the posting plane of $p \neq q$ consists of the vectors $De_p + Ce_q$.

2. A *cone identification* at $\kappa > \frac{1}{4}$ is a linear bijection Λ of $\mathbb{R}^4$ with v in the future cone of q_κ exactly when $\Lambda v \in \mathscr{K}$. It is *tick-normalized* when $(\Lambda e_i)_0 = 1$ for every i.

3. A *test* is e with $e \in \mathscr{K}$ and $(2, \vec{0}) - e \in \mathscr{K}$; it is *sharp* when moreover $e_0 = 1$ and $Q(e) = 0$. A record $p \in \mathscr{K}$ is *definite* in the sense X (extreme) if every splitting $p = \sigma + \tau$ into records has σ a multiple of p; C (certain) if some test other than $(2, \vec{0})$ has $\langle e, p \rangle = p_0$; U (undisturbed) if some sharp test's reading $R(e)R(p)R(e)$ leaves it unchanged; I (indivisible) if no two nonzero records σ, τ with $\langle \sigma, \tau \rangle = 0$ have $p - \sigma - \tau \in \mathscr{K}$; E (excluding) if $\langle p, \sigma \rangle = 0$ for some nonzero record σ.

4. An *exact readout* assigns to each posting word w a record $\rho(w)$, additive over concatenation, with each single post's record of trace 1, such that some test E satisfies $\operatorname{tr}(E\rho(w)) =$ the number of debits of w, for every w.

5. A *world of records* bundles the accepted premises (the foundation's facts, with definite comparison outcomes), the reading of a debit and a credit as two lattice steps, an exact readout, a causal bound at $\kappa = \frac{1}{4} + c_J^2$ with c_J the light speed in lattice lengths per tick, and a cone identification at that κ. It does not identify a post's own record with the image of its posting.

Let $u_1 = (1,1,1)$, $u_2 = (1,-1,-1)$, $u_3 = (-1,1,-1)$, $u_4 = (-1,-1,1)$, the vertices of a regular tetrahedron: $u_i \cdot u_i = 3$, $u_i \cdot u_j = -1$ for $i \neq j$, and $\sum u_i = 0$. For $\kappa > \frac{1}{4}$ put $T_\kappa v = \big(\sum v,\ r \sum_i v_i u_i\big)$ with $r = (4\kappa - 1)^{-1/2}$.

Theorem 27.39 (A posting's record has squared Bloch length $3/(4\kappa-1)$)**.**

(a) $q_\kappa(De_p + Ce_q) = \kappa(D+C)^2 - D^2 - C^2$ *equals* $\lambda \cdot 4DC$ *for all natural* D, C *if and only if* $\kappa = 1$ *and* $\lambda = \frac{1}{2}$*; at* $\kappa = 1$ *the lattice proper time of a word's straight displacement is its ledger proper time* $2\sqrt{DC}$ *divided by* $\sqrt{2}$*.*

(b) $Q(T_\kappa v) = \frac{4}{4\kappa-1}\, q_\kappa(v)$*, and* T_κ *is a cone identification.*

(c) $T_\kappa e_i$ *has time* 1 *and squared Bloch length* $3/(4\kappa-1)$*: a pure record exactly at* $\kappa = 1$*, a mixed record for* $\kappa > 1$*, and no record for* $\kappa < 1$*.*

(d) Every cone identification satisfies $Q(\Lambda v) = c\, q_\kappa(v)$ *for some* $c > 0$*; so under every identification a single posting's record is pure exactly when* $\kappa = 1$*.*

(e) If $\kappa \geq 1$*, a post's own record under an exact readout is* Λd *for an edge class* d *only if* $\kappa = 1$ *and* $|d| = 1$*; and at* $\kappa = 1$ *some exact readout and cone identification carry a post's own record to a single posting.*

Proof. (a) $(D,C) = (1,0)$ gives $\kappa = 1$, then $(1,1)$ gives $\lambda = \frac{1}{2}$; at $\kappa = 1$, $(D+C)^2 - D^2 - C^2 = 2DC$, and $\sqrt{2DC} = 2\sqrt{DC}/\sqrt{2}$. (b) $|\sum v_i u_i|^2 = 4|v|^2 - (\sum v)^2$ by the Gram values, so $Q(T_\kappa v) = (1+r^2)(\sum v)^2 - 4r^2|v|^2 = \frac{4}{4\kappa-1}(\kappa(\sum v)^2 - |v|^2)$; the time coordinate is $\sum v$, so the future sides correspond. (c) $|ru_i|^2 = 3r^2$. (d) ΛT_κ^{-1} keeps $\mathscr{K}$, so by the rigidity theorem of Chapter 24 it multiplies Q by a positive constant; combine with (b). (e) The test of the readout has $\mathrm{tr}(E\rho_d) = 1 = \mathrm{tr}\,\rho_d$ and $\mathrm{tr}(E\rho_c) = 0$ for the single-debit and single-credit records; if ρ_d had full rank, E would be 1 and $\mathrm{tr}(E\rho_c) = 1$. So the post's record is pure, and by (d) $0 = c\,|d|(\kappa|d| - 1)$ forces $\kappa|d| = 1$, hence $\kappa = |d| = 1$. Conversely, at $\kappa = 1$, $T_1 e_p = (1, \vec{n})$ with $|\vec{n}| = 1$ by (c); the readout with debit record $R(1,\vec{n})$, credit record $R(1,-\vec{n})$, extended additively, and the test $R(1,\vec{n})$ is exact, and with $\Lambda = T_1$ the debit's record is Λe_p. □

Theorem 27.40 (Definiteness of a posting record is $\kappa = 1$; the supplied definiteness is not)**.** *Let* $\kappa > \frac{1}{4}$ *and let* Λ *be a cone identification.*

(a) Λe_i *is a record exactly when* $\kappa \geq 1$*, and it is definite in any one of the senses X, C, U, I, E exactly when* $\kappa = 1$*.*

(b) If $\kappa \geq 1$ *and* $v \geq 0$ *is a nonzero history whose record* Λv *is definite, then* $\kappa = 1$ *and* v *lies on one account:* $v_a v_b = 0$ *for* $a \neq b$*.*

(c) If Λ is tick-normalized, every posting record has squared Bloch length $3/(4\kappa-1)$, distinct posting records have pairing $(2\kappa-1)/(4\kappa-1)$, and a sharp test that passes a posting record with probability at least $1-\varepsilon$, $\varepsilon<\frac{1}{2}$, forces $\kappa\le\frac{1}{4}+\frac{3}{4(1-2\varepsilon)^2}$.

(d) For every $\kappa>1$ there is a world of records at κ. In it the comparisons are definite and each post's own record is pure, yet the record of every nonzero posting history is mixed. If the image of each post is its own record seen through a depolarizing channel, $\Lambda(\mathit{post})=(1-r)(1,\vec{0})+r\,\rho(\mathit{post})$, then $r^2(2\kappa-1)=1$; this coupling is met at every $\kappa>1$, and it gives each posting off the posting plane a record of trace $2\kappa/(2\kappa-1)\neq 1$.

Proof. On the record cone each of X, C, U, I, E is equivalent to purity. A record with $Q>0$ is an interior point of the cone over a ball and splits into two non-proportional records; a full-rank ρ with $\operatorname{tr}(E\rho)=\operatorname{tr}\rho$ forces $E=1$; Lüders' reading by a rank-one projector P gives $\operatorname{tr}(P\rho)P$ [91], which is ρ only for pure ρ; a full-rank ρ contains ϵP and $\epsilon(1-P)$; and $\operatorname{tr}(\rho\sigma)=0$ with $\sigma\neq 0$ needs a kernel. Conversely a pure record has all five properties. (a) By Theorem 27.39(d), $Q(\Lambda e_i)=c(\kappa-1)$. (b) $Q(\Lambda v)=c\,q_\kappa(v)=c\big((\kappa-1)(\sum v)^2+2\sum_{a<b}v_av_b\big)$, a sum of nonnegative terms that must vanish. (c) With $(\Lambda v)_0=\sum v$, write $\Lambda v=(\sum v, Mv)$ with M of size 3×4; then $M^{\mathsf T}M=(1-c\kappa)\mathbf{1}\mathbf{1}^{\mathsf T}+c\,1$ has rank at most three, so its determinant $c^3(c+4(1-c\kappa))$ vanishes and $c=4/(4\kappa-1)$. Hence $|Me_i|^2=1+c(1-\kappa)=3/(4\kappa-1)$ and $Me_i\cdot Me_j=1-c\kappa=-1/(4\kappa-1)$, giving the pairing. A sharp test $(1,\vec{m})$ with $|\vec{m}|=1$ passes $(1,\vec{p})$ with probability P, $2P-1=\vec{m}\cdot\vec{p}$, so $(1-2\varepsilon)^2\le 3/(4\kappa-1)$. (d) A world at κ is built from a linear map that multiplies the forms by a positive constant and sends one future vector forward, which is a cone identification; the history records are mixed by the identity in the proof of (b), as $\kappa>1$. For the coupling: the readout makes the two posts' records pure with antipodal Bloch vectors $\pm n$, so $\Lambda e_p=(1,rn)$, $\Lambda e_q=(1,-rn)$ and $\Lambda(e_p+e_q)=(2,\vec{0})$. Rigidity at e_p+e_q gives $4=c(4\kappa-2)$, and at e_p gives $1-r^2=c(\kappa-1)=2(\kappa-1)/(2\kappa-1)$, so $r^2=1/(2\kappa-1)$. For $k\neq p,q$ the bilinear form gives $2(\Lambda e_k)_0=c\cdot 2\kappa$. □

The definiteness the foundation supplies (definite comparison outcomes, pure own records) therefore holds at every $\kappa>1$ and does not fix κ. What does is one sentence about the ledger's history.

Theorem 27.41 ($\kappa=1$ exactly when some history record is definite)**.** *In every world of records:*

(a) $\kappa=1$ if and only if some nonempty ledger history has a definite record, if and only if some positive linear reading (one that maps records to records) recovers every run's own record from its lattice record, and, under the depolarizing coupling, if and only if $r^2=1$;

(b) records are append-only, persist and can be re-examined; the event ledger is double-entry and append-only; the posts' own records keep which post occurred and a run's counts, and the lattice record keeps the counts and recovers them linearly; all of this holds at every $\kappa>1$; keeping the order of posts fails at every κ, since a debit then a credit leave the same own record as a credit then a debit;

(c) for $\kappa\ge 1$, each of the statements in Table 27.4 is equivalent to $\kappa=1$.

$\kappa=1$ is not derived: worlds of records exist at every $\kappa\ge 1$.

Conditions and scope.

Base and premises

Classical real and complex numbers; the cone identification is a named premise.

Conclusion

Every permanence statement of the foundation holds at every $\kappa > 1$; $\kappa = 1$ is exactly the named premise that the ledger's history record loses nothing, equivalently that some history record is definite.

Not claimed

A derivation of $\kappa = 1$. The doublers of Proposition 27.35(b) are massless exactly at $\kappa = 1$, the same premise, and add no further one.

Proof. (a) By Theorem 27.40(b) a definite record of a nonzero history forces $\kappa = 1$; at $\kappa = 1$ a single posting has a pure record. If a positive Φ sends an interior record x to a pure record, then for any record y and small $\epsilon > 0$, $x - \epsilon y$ is a record and $\Phi x = \epsilon \Phi y + \Phi(x - \epsilon y)$ splits a pure record, so Φy is a multiple of Φx. At $\kappa > 1$ the posting image Λe_p is interior and the debit's own record is pure, so a recovering Φ would send Λe_q to a multiple of the debit's record, while the credit's record is antipodal to it. At $\kappa = 1$ a measure-and-prepare reading recovers. The coupling: $r^2(2\kappa - 1) = 1$. (b) Each statement holds in every world of records, and worlds exist at every $\kappa > 1$ (Theorem 27.40(d)); own records add, so the two orders give $\rho_d + \rho_c$. (c) The first row holds because $q_\kappa(e_p) = \kappa - 1$ (Definition 27.38); rows two to four are Theorem 27.39(a), (c) and (e); the remaining rows are (a) and Theorem 27.40(d). □

Table 27.4. Equivalent forms of $\kappa = 1$, for $\kappa \geq 1$.

Statement	Settled by
A single lattice step is light-like: $q_\kappa(e_p) = 0$	Definition 27.38
The lattice interval on a posting plane is a fixed multiple of the ledger interval $4DC$	Theorem 27.39(a)
A posting's record under T_κ is pure: $3/(4\kappa - 1) = 1$	Theorem 27.39(c)
Some exact readout and cone identification carry a post's own record to an edge class	Theorem 27.39(e)
Some ledger history has a definite record	Theorem 27.41(a)
The history record loses nothing (a positive reading recovers every run's record)	Theorem 27.41(a)
Under the depolarizing coupling, $r^2 = 1$	Theorem 27.40(d)

The link between the posts' own records and their lattice records runs in two directions, and they differ. A real linear map Φ of record coordinates extends, through $p \mapsto R(p)$, to complex 2×2 matrices; it is *completely positive* when its Choi matrix $\sum_{i,j} E_{ij} \otimes \Phi(E_{ij})$ is positive semidefinite [23]. In a world of records let $x = \Lambda e_p$, $y = \Lambda e_q$ be the posting images and ρ_d, ρ_c the posts' own records. A *heralded discrimination* is a pair of tests e_1, e_2 with $(2, \vec{0}) - e_1 - e_2$ a record and $\langle e_1, y \rangle = \langle e_2, x \rangle = 0$; a test e *guesses* the post with success $\frac{1}{2}(\langle e, x \rangle / x_0 + \langle (2, \vec{0}) - e, y \rangle / y_0)$.

Theorem 27.42 ($\kappa = 1$ is the purity of one posting's lattice record)**.** *In classical mathematics, in every world of records (where $\kappa \geq 1$):*

(a) writing is positive: Λ sends every nonnegative account vector, so every run's lattice displacement, to a record, and the completely positive map $W(z) = \langle \rho_d, z\rangle x + \langle \rho_c, z\rangle y$ sends each run's own record to its lattice record; for every $\kappa > 1$ some world of records has all this while the reading in (b) fails;

(b) on the plane of x and y the only linear map with $x \mapsto \rho_d$, $y \mapsto \rho_c$ sends records to records exactly when $\kappa = 1$, exactly when x and y are pure, exactly when some positive reading recovers every run's own record; and a completely positive reading that recovers both posts' records, or every run's own record, exists exactly when $\kappa = 1$;

(c) if $\kappa > 1$, every heralded discrimination has $e_1 = e_2 = 0$, and every guess succeeds with probability at most $\frac{1}{2}(1 + 1/\sqrt{2\kappa - 1})$, which the depolarizing coupling of Theorem 27.40(d) attains.

Proof. (a) For $v \geq 0$, $q_\kappa(v) = (\kappa - 1)(\sum v)^2 + 2\sum_{a<b} v_a v_b \geq 0$. The own records are pure and antipodal, as in the proof of Theorem 27.39(e), so $\langle \rho_d, \rho_d\rangle = 1$, $\langle \rho_d, \rho_c\rangle = 0$, and W sends $D\rho_d + C\rho_c$ to $Dx + Cy$; its Choi matrix is a sum of Kronecker products of positive semidefinite matrices. The separating worlds are those of Theorem 27.40(d). (b) At $\kappa = 1$, $Q(sx + ty) = c\,q_1(se_p + te_q) = 2cst$, so records of the plane have $s, t \geq 0$. At $\kappa > 1$, $Q(x) = c(\kappa - 1) > 0$, so $x - \epsilon y$ is a record for small $\epsilon > 0$, while ρ_d, ρ_c are pure and antipodal and $Q(\rho_d - \epsilon\rho_c) = -4\epsilon$. Purity is $\kappa = 1$ by Theorem 27.39(d), and recovery is Theorem 27.41(a). A completely positive map is positive ($w^*\Phi(uu^*)w = \xi^* C\xi$ with $\xi = \bar{u} \otimes w$, C the Choi matrix). At $\kappa = 1$, with e_1, e_2 the sharp tests vanishing on y and on x, $z \mapsto \lambda_1\langle e_1, z\rangle\rho_d + \lambda_2\langle e_2, z\rangle\rho_c$ with $\lambda_1\langle e_1, x\rangle = \lambda_2\langle e_2, y\rangle = 1$ is completely positive and recovers. (c) A nonzero test pairs positively with the interior records x, y. A test has $|\vec{e}| \leq 1$, so a guess succeeds with probability at most $\frac{1}{2} + \frac{1}{4}\sqrt{S}/(XY)$ with $X = x_0$, $Y = y_0$, $S = |\vec{x}Y - \vec{y}X|^2$. Rigidity gives $X^2 - |\vec{x}|^2 = Y^2 - |\vec{y}|^2 = c(\kappa - 1)$ and $XY - \vec{x}\cdot\vec{y} = c\kappa$, so $S = 2c\kappa XY - c(\kappa - 1)(X^2 + Y^2)$ and $4X^2Y^2 - (2\kappa - 1)S = |\vec{x}Y + \vec{y}X|^2 + 2c\kappa(\kappa - 1)(X - Y)^2 \geq 0$. Under the coupling $x = (1, \vec{m})$, $y = (1, -\vec{m})$, $r^2(2\kappa - 1) = 1$, the test $(1, \vec{n})$ succeeds with probability $\frac{1}{2}(1 + r)$. □

So writing the ledger into the lattice picture is positive at every $\kappa \geq 1$, and reading the posts back is positive exactly at $\kappa = 1$: the premise of Theorem 27.41, unchanged by asking for complete positivity.

27.9 The locality clause is a named premise

The lattice of Section 27.7 posts at sites, and its gravity sector assumes a locality clause: each posting is decided by one comparison of the site's record with one partner's. This section proves what that clause is, and that nothing proved so far supplies it. The results are proved in classical mathematics; the definitions and the implication from one comparison are choice-free.

Definition 27.43 (Posting laws and the locality clause)**.** A *posting law* on a set V of sites, with records in R, assigns to each site x and each assignment $r : V \to R$ the post $F(x, r)$. The law *reads only* $A \subseteq V$ at x when $F(x, r) = F(x, r')$ whenever r and r' agree on A, and the site *recognizes* when the law does not read only $\{x\}$ at x. The *locality clause* (pair-reading) holds when every site x has a partner y such that $F(x, r)$ is determined by $r(x)$ and $r(y)$; *least read* holds when every x has a set $A \ni x$ of at most two sites that the law reads only. The law is decided by *one comparison* when $F(x, r) = g(x, \mathrm{cmp}(r(x), r(\pi x)))$ for a partner map π, a Boolean test cmp and some g.

Theorem 27.44 (The locality clause is least read)**.** *For every posting law F:*

(a) one comparison implies the locality clause;

(b) every set that contains a recognizing site and that the law reads only at it has at least two sites; so with records of $L + 1$ bits, a recognizing posting that reads its own record needs a window of at least $2(L + 1)$ bits, which pair-reading laws attain;

(c) the locality clause is equivalent to least read;

(d) the parity law on three sites (post a debit exactly when $r_0 + r_1 + r_2$ is even) is not pair-reading, yet each of its posts costs exactly what any single unit post costs on the ledger.

Proof. (a) Take $y = \pi x$. (b) If the set were $\{x\}$, the site would not recognize; x belongs to the set by hypothesis. (c) Given a partner y, take $A = \{x, y\}$; given A, take its other element, or $y = x$. (d) For any site x and partner y, changing the third record flips the parity. Each post is one unit post on one account, whose cost does not depend on what decided it. □

Proposition 27.45 (Nothing proved so far selects the clause)**.** *In classical mathematics:*

(a) in the foundation's semantics, the three facts (distinguishable occurrences exist; comparing needs information about both; a later examination is an occurrence) are equivalent in every world to the first; the world in which every post, once three records exist, records whether records 0, 1 and 2 all agree meets them; and for any two record positions it has two permitted histories that agree at both positions and whose forced next posts differ, so no pair of records decides a post;

(b) $J(x) = 0$ exactly when $x = 1$, so any number of agreeing comparisons costs 0, and $J(1) + J(1) < J(2) = \frac{1}{4}$; every unit post costs the same whatever decided it; so J does not count comparisons, and a uniform positive price per comparison only renames the clause.

Proof. (a) In the foundation's interpreter comparisons read two present records and re-examination is an event, so the second and third facts hold in every world. For positions i, j choose two seeds that agree at i and j and differ at the third of the positions $0, 1, 2$ so that the all-equal test changes. (b) $J(x) = (x - 1)^2/(2x)$, and $J(2) = \frac{1}{4}$. □

The clause can still be written into a world's rule. We describe such worlds and show that none of the standard worlds of the foundation's interpreter is one.

Definition 27.46 (Worlds, clause worlds and the standard worlds)**.**

1. Records and asking histories (lists of the acts post r, retain and compare (i, j)) are those of Chapter 30. A *world* is a set of histories, its permitted ones, containing the empty history and closed under prefixes; it is *rich* when for every n it permits a history with at least n records and n outcomes.

2. Let V be a set of sites with decidable equality, $\sigma : \mathbb{N} \to V$ a placement (post k is made at $\sigma(k)$) and $\pi : V \to V$ a partner map. With k records, $\ell_k(y)$ is the last position $i < k$ with $\sigma(i) = y$, if any; its record is the *current record* of y. The *clause world* $W_{\sigma,\pi}$ permits a history when each post made with $k \geq 3$ records comes right after compare $(\ell_k(\sigma k), \ell_k(\pi\sigma k))$, both positions existing, and is the one-letter record of its outcome. A world is *inside a clause world* when some clause world permits every history it permits.

3. The eighteen *standard worlds*, for all values of their parameters, are: all histories, and the two worlds of a located probe (a probe read as the post of its one-bit result, a move as a retain), which permit every history; the prefixes of one history, of the empty history, of the canonical query (post the codes of a sentence and of its negation, retain, compare; two constructions), and of the histories of a finite record of events; the clock of a sequence s of records (the prefixes of post $s(0)$, ..., post $s(n-1)$ followed by one comparison), its truncation to $n \leq N$, the tally clock (s constantly 1) and the uncompared tally (posts of 1 only); the capacity world of B (after each nonempty prefix the ledger's standard code has at most B bits) and the patch world of N (posts of 1 only, or the capacity world of N); the retain world (retains only); the free tally (posts of 1 and comparisons); and the window world of capacity N of Chapter 30 and its loose form (each address of a comparison that holds a record lies in the window), whose conditions bind only comparisons.

Theorem 27.47 (The clause can be a world's rule, and no standard world has that rule)**.** *In classical mathematics, for every set of sites, placement σ and partner map π:*

(a) $W_{\sigma,\pi}$ meets the three facts of Proposition 27.45(a), and in each history it permits, a post made with at least three records is the record of $[\, c(\sigma k) = c(\pi\sigma k) \,]$, c the current records: a post of a law that satisfies the locality clause with partner π;

(b) $W_{\sigma,\pi}$ permits every one-post history and no history consisting of four posts; on two sites posted in turn, each the other's partner, it is rich, while with a fresh site for every post it permits no post once three records exist;

(c) no clause world equals a world that forbids some one-post history, permits only finitely many, or permits some history consisting of four posts; no rich world inside a clause world equals a world that permits such a history or bounds the numbers of records or of outcomes of its histories; and some rich world inside a clause world, whose first three acts are fixed, permits exactly one one-post history;

(d) every standard world, for every value of its parameters, meets a condition of each list in (c); so none equals a clause world or a rich world inside one.

Proof. (a) Post 1, post 0, compare $(0, 1)$ is permitted and compares two present records that differ, which gives the facts by Proposition 27.45(a). A permitted post made with $k \geq 3$ records follows the comparison of positions $\ell_k(\sigma k)$ and $\ell_k(\pi\sigma k)$ and is its outcome; the law reads only x and πx. (b) The rule binds only posts made with three or more records, and in four posts the fourth follows a post. On two sites the comparison is of positions $k-2$ and $k-1$, so three posts followed by n rounds of comparing the last two records and posting the outcome give $n+3$ records and n outcomes. A fresh site has no current record. (c) A clause world permits the one-post history of every record, so infinitely many, and no history consisting of four posts; nor does a world inside it, and a rich world bounds neither count. The world of the histories of (b) on two sites that begin post 0, post 0, post 1 (where those acts are present) is inside the two-site clause world, rich, and its only one-post history is post 0. (d) All histories, the probe worlds, the uncompared and free tallies, the patch worlds and both window worlds permit the history of four posts of 1, which meets a condition of both lists; so do the clock and the tally clock, which permit the history of their first four posts. The retain world forbids every post and has no records; the capacity world of B forbids posting 1^{B+1} and has at most B records; the other prefix worlds and the truncated clock permit finitely many one-post histories and bound the length of their histories. □

The separation uses the first four acts of a history; it does not cover encodings of records and histories, intersections or unions of standard worlds, or classes of worlds. The clause is not derived.

In a synchronous tick every site updates at once from the previous tick's records. Take sites $0, \dots, N-1$, seeds c_0 and posting laws F_t (Definition 27.43), and put $c_t(x) = F_t(x, c_{t-1})$. In the *posting run*, records $0, \dots, N-1$ are the seeds and record $Nt + x$ is F_t at x applied to $y \mapsto$ record $(t-1)N + y$. A post *depends on* a site or position when changing that record alone, from some assignment, changes the post, and *depends on it at r* when changing it alone from r does.

Proposition 27.48 (Synchronous ticks run as posts, and who reads whom)**.** *In classical mathematics:*

(a) record $Nt + x$ of the posting run is $c_t(x)$; if F_t reads only $\{x, p_t(x)\}$ at x, post $Nt + x$ reads only the positions $(t-1)N + x$ and $(t-1)N + p_t(x)$;

(b) if a law reads only $\{x, p(x)\}$ at every x with p injective (a pairing, say), at most two posts depend on each site; yet for every site y_0 a law decided by one comparison reads only $\{x, y_0\}$ at every x and has all N posts depending on y_0;

(c) let each site pick its contact from its own record, $F(x, r) = G(x, r(x), r(c(x, r(x))))$. At every r the post of x is fixed by $r(x)$ and $r(c(x, r(x)))$, and in the posting run post $Nt + x$ is fixed by two positions of tick $t-1$; but if no $G(x, u, \cdot)$ is constant and $c(x, \cdot)$ takes two values other than x, no fixed y makes F read only $\{x, y\}$ at x. Where the picks $x \mapsto c(x, r(x))$ are injective, at most two posts depend on each site at r; if the picks pair the sites off at every r, each $c(x, \cdot)$ is constant;

(d) let site x decide from $r(x)$ and $r(q(x))$, by a test a, whether to contact one candidate $q(x)$: $F(x, r) = G(x, r(x), r(qx))$ when the test holds and $H(x, r(x))$ otherwise. Then F satisfies the locality clause with partner q; and if $q(q(x)) = x$ and $a(qx, v, u) = a(x, u, v)$, the contacts made pair the sites off at every r, and for the agreement test they change with the records.

The statements other than the counts of dependent posts hold with no omniscience principle and no choice principle.

Proof. (a) Induction on t: post $Nt + x$ applies $F_t(x, \cdot)$ to the records $(t-1)N + y$, which are $c_{t-1}(y)$. (b) A post of x that depends on y has $y \in \{x, p(x)\}$, and $p(x) = y$ for at most one x. For the hub, post the outcome of the test that $r(x)r(y_0)$ is empty. (c) The first claims are the form of F. Given y, pick u with $c(x,u) \notin \{x, y\}$, give every site u, and change only the record at $c(x,u)$. If $c(x,u) = z \neq x$, the assignments giving x the record u force $c(z, \cdot) = x$, and then for each record b the assignment giving every site b forces $c(x,b) = z$. (d) The test and both branches use only $r(x)$ and $r(qx)$; if the test holds at x it holds at $q(x)$, whose contact is $q(q(x)) = x$. Under the agreement test on two sites, equal records give the contacts $1, 0$ and different ones $0, 1$. □

The run itself has no comparisons, so no clause world permits it: here the clause is a property of the tick law. Injective partners, not the clause alone, keep each record to two readers.

A site may also read its candidates' records before picking. For a posting law F on sites $0, \dots, N-1$ with records finite binary words (Definitions 27.43 and 27.46), the post of x is *on two records at* an assignment r if some site y has $F(x, r') = F(x, r)$ for all r' equal to r at x and y (the locality clause asks for one y for all r).

Proposition 27.49 (Picks that keep each post on two records)**.** *In classical mathematics:*

(a) for three different sites x, c_1x, c_2x, let a test $b(x, r(x), r(c_1x), r(c_2x)) \in \{1, 2\}$ pick c_bx, and let $F(x,r) = G(x, r(x), r(c_bx))$, G any law. The post of x is on two records at every r exactly when, for some map χ from records to $\{1, 2\}$, the tie $G(x, a, d_1) = G(x, a, d_2)$ holds whenever $b(x, a, d_1, d_2) \neq \chi(a)$. If every $G(x, a, \cdot)$ is injective, a tie is $d_1 = d_2$; if $G(x, a, d)$ is the one-letter record of $[\, a = d \,]$ (the comparison), it is $a = d_1 \Leftrightarrow a = d_2$. On sites $0, 1, 2$, with $x = 0$, $c_1x = 1$, $c_2x = 2$, the comparison and the pick of site 1 exactly when $r(1) = r(0)$ leave the post on no two records at some r;

(b) for different sites $c_1, \dots, c_k \neq x$, $k \geq 1$, let $F(x,r) = P(r(x), v)$ with $v_i = r(c_i)$. The post is on two records at every r exactly when for each a some i has $P(a, v) = P(a, w)$ whenever $v_i = w_i$;

(c) let every site x pick $c(x, r(x))$ from its own record and post $F(x,r) = L(x, r(x), r(c(x, r(x))))$, where $L(x, a, d)$ is $G(x, a, d)$ if a test $\mathrm{acc}(x, a, d)$ holds (the site accepts) and $H(x, a)$ if not. At every r each post is on two records; for $y \neq x$ the post of x depends on y at r exactly when $y = c(x, r(x))$ and $L(x, r(x), d) \neq F(x, r)$ for some record d; where $x \mapsto c(x, r(x))$ is injective, at most two posts depend on each site at r, whatever is accepted; and for each site y_0 some such law has, at some r, at most one site accepting and all N posts depending on y_0.

The first two statements of (c) hold with no omniscience principle and no choice principle.

Proof. (a) Given χ, the post is $G(x, r(x), r(c_{\chi(r(x))}x))$, as the values tie where the pick differs from χ. Conversely, fix a, place a, d_1, d_2 at x, c_1x, c_2x, and let $T(d_1, d_2)$ be the post (row d_1, column d_2); so $T(d, d) = G(x, a, d)$. The site kept with x is not both candidates, and changing the other keeps the post, so the row of d_1 or the column of d_2 is constant. If some row is not constant, each column meets it at a point whose row is not, so every column is constant.

Constant rows give $G(x,a,d_1) = T(d_1,d_1) = T(d_1,d_2) = G(x,a,d_2)$ wherever $b = 2$, so $\chi(a) = 1$ serves; constant columns give $\chi(a) = 2$. In the example, the records empty, 1, 0 at sites 0, 1, 2 give the post 0; emptying $r(1)$ makes it 1, as does emptying $r(2)$, and one of the two keeps the record of the site kept with site 0. (b) Given i for $a = r(x)$, the records at x and c_i fix the post. Conversely, fix a; each v has an i that *serves* it: $P(a,w) = P(a,v)$ whenever $w_i = v_i$ (the candidate kept with x, or any i if none is). If $i \neq j$ serve v and w, then $P(a,v) = P(a,u) = P(a,w)$ for u equal to v except $u_j = w_j$. So if $P(a,v) \neq P(a,w)$, one i alone serves both, and also every u, as $P(a,u)$ differs from $P(a,v)$ or $P(a,w)$; if $P(a,\cdot)$ is constant, any i will do. (c) The own record fixes the pick, so it and the picked record fix the post. Changing $r(y)$ alone, $y \neq x$, keeps the pick, so the post stays unless y is the pick, and then becomes $L(x, r(x), d)$ for the new record d. With injective picks only the post of y and the one post that picks y can depend on y. For the hub, every site picks y_0; site z holds 1^z, accepts if its record is the hub's, and posts the hub's record if so and 1 followed by its own if not. Only y_0 accepts; the post of y_0 moves when its record becomes 0, and that of $z \neq y_0$ when the hub's record becomes 1^z, as z then accepts. □

The clause can also carry the golden ratio. On two sites posted in turn, let each post after the third be the partner's current record followed by the site's own, a law that reads only the site and its partner, seeded with 0, 0, 1. Then $R_0 = R_1 = 0$, $R_2 = 1$ and $R_k = R_{k-1}R_{k-2}$ for $k \geq 3$: 10, 101, 10110, ... Write $\#_1 R$ for the number of letters 1, $|R|$ for the length, and $\nu(a + b\varphi) = a^2 + ab - b^2$ for the norm of $\mathbb{Z}[\varphi]$.

Proposition 27.50 (The Fibonacci clause law posts the powers of φ)**.** *In classical mathematics:*

(a) $\#_1 R_n + |R_n|\varphi = \varphi^n$ *in* $\mathbb{Z}[\varphi]$ *and* $\nu(\#_1 R_n + |R_n|\varphi) = (-1)^n$*, for every* $n \geq 1$*;*

(b) $|R_{k+2}|$ *is the number of postings of the* $(k+1)$*st recognition of Definition 21.24, for every* k*;*

(c) $\varphi^n + \varphi^{-n} - 2$*, which is* $2J(\varphi^n)$*, is the integer* $2\#_1 R_n + |R_n| - 2$ *for even* $n \geq 1$ *and* $-|R_n| - 2 + 2|R_n|\varphi = |R_n|\sqrt{5} - 2$ *for odd* n*.*

Proof. (a) $\#_1$ and $|\cdot|$ add over $R_k = R_{k-1}R_{k-2}$, as $\varphi^k = \varphi^{k-1} + \varphi^{k-2}$; for $n = 1, 2$ the values are φ and $1 + \varphi = \varphi^2$. The norm is multiplicative and $\nu(\varphi) = -1$. So $\#_1 R_n = F(n-1)$ and $|R_n| = F(n)$, as $\varphi^n = F(n)\varphi + F(n-1)$ (Section 21.5). (b) $F(k+2)$ is that posting count in Section 21.5. (c) $\varphi^{-1} = -\psi$, so $\varphi^{-n} = (-1)^n\psi^n = (-1)^n(F(n+1) - F(n)\varphi)$; add, and use $2\varphi - 1 = \sqrt{5}$ and $2J(x) = x + x^{-1} - 2$. □

So the locality clause is consistent with everything proved, follows from none of it, is exactly least read, and can be written into a world's rule that no standard world has. It is a named premise of the lattice's gravity sector, in the same position as $\kappa = 1$.

27.10 What reading the cost as the graviton's action needs

The comparison of a geometry with a reference that this chapter began with is *reading one*: for positive definite g_0 and g, $C(g_0, g) = J(g_0^{-1}g)$, J summed over the principal stretches. Reading one is an identification, and we close by stating exactly what it needs. A cost of comparing geometries is *frame-free* when $C(P^{\mathsf{T}}g_0P, P^{\mathsf{T}}gP) = C(g_0, g)$ for every invertible P and all positive definite g_0, g, and *axis-additive* when $C(1, \operatorname{diag}\lambda) = \sum_i J(\lambda_i)$ for all positive λ_i.

On four coordinates with **1** as time, the posting frame is $p_1 = (1,1,-1,-1)$, $p_2 = (1,-1,1,-1)$, $p_3 = (1,-1,-1,1)$, and the two polarizations at the face wave p_1 are $h_+ = p_2 p_2^{\mathsf{T}} - p_3 p_3^{\mathsf{T}}$ and $h_\times = p_2 p_3^{\mathsf{T}} + p_3 p_2^{\mathsf{T}}$. For weights w_k, the *step-class form* is $\sum_S w_{|S|} (\mathbf{1}_S^{\mathsf{T}} H \mathbf{1}_S)^2$, summed over the nonempty sets S of axes.

Proposition 27.51 (The reading is two numbers at one wave)**.** *In classical mathematics:*

(a) a cost is reading one if and only if it is frame-free and axis-additive, and neither property alone suffices;

(b) a quadratic cost $\mathcal{Q}$ of strains with the lattice's symmetry equals J_2 on all TT strains if and only if $\mathcal{Q}(h_+) = 16$ (calibration at the face) and $\mathcal{Q}(h_\times) = \mathcal{Q}(h_+)$ (no birefringence); for the step-class form, this holds exactly when $w_2 = \frac{1}{4}$ and $w_1 + w_3 = 1$, so equal weights fail it, while the per-posting weights $\frac{1}{2}, \frac{1}{4}, \frac{1}{2}$ meet it.

Proof. (a) A positive definite pair is simultaneously diagonalizable: $P^{\mathsf{T}} g_0 P = 1$ and $P^{\mathsf{T}} g P = \operatorname{diag}\lambda$, so frame freedom reduces C to $C(1, \operatorname{diag}\lambda)$, and axis additivity gives $\sum J(\lambda_i) = J(g_0^{-1} g)$. The axis reading $\sum_i J(g_{ii}/(g_0)_{ii})$ is axis-additive and not frame-free; the volume reading $n J(\det(g_0^{-1} g)^{1/n})$ is frame-free and not axis-additive. (b) With $f_a = \frac{1}{2} p_a$, a strain with $H\mathbf{1} = 0$ is $\sum_{a,b} M_{ab} f_a f_b^{\mathsf{T}}$, M symmetric, with $\operatorname{tr} H = \operatorname{tr} M$ and $J_2(H) = \frac{1}{2} \operatorname{tr} M^2$; every TT strain is one with $\operatorname{tr} M = 0$. Exchanging axes 1 and 2, or 1 and 3, exchanges f_1 with f_2, or with f_3, and exchanging 0 with 1 and 2 with 3 negates f_2 and f_3; so relabellings permute the f_a at will and change any two of their signs. Each product of an off-diagonal M_{ab} with another entry is reversed by one of these sign changes, so $\mathcal{Q}$, a quadratic form in the entries M_{ab} ($a \le b$), has none; by the permutations, and as $\sum_{a \ne b} M_{aa} M_{bb} = -\sum_a M_{aa}^2$ when $\operatorname{tr} M = 0$, $\mathcal{Q} = c \sum_a M_{aa}^2 + c' \sum_{a \ne b} M_{ab}^2$ on these strains. Here h_+ has $M = \operatorname{diag}(0,4,-4)$ and $h_\times$ has $M_{23} = M_{32} = 4$, so $\mathcal{Q}(h_+) = 32c$, $\mathcal{Q}(h_\times) = 32c'$ and $J_2(h_+) = J_2(h_\times) = 16$; as both are TT at p_1, $\mathcal{Q} = J_2$ on all TT strains exactly when $\mathcal{Q}(h_+) = \mathcal{Q}(h_\times) = 16$. For the step-class form, $\mathbf{1}_S^{\mathsf{T}} H \mathbf{1}_S$ on the fifteen steps gives $\mathcal{Q}(h_+) = 64 w_2$ and $\mathcal{Q}(h_\times) = 16(w_1 + w_3)$. □

The per-posting weights were selected with the target in view. Each number rests on a premise, through the changes $U \in \mathrm{SU}(2)$ of the coherent record carrier of Section 27.8 or through the per-posting count. The *face frame* $f_0 = \frac{1}{2}\mathbf{1}$, $f_a = \frac{1}{2} p_a$ is orthonormal; for a real 4×4 matrix X put $v_a = f_0^{\mathsf{T}} X f_a$ and $M_{ab} = f_a^{\mathsf{T}} X f_b$ ($a, b = 1, 2, 3$). Let S_U be the rotation with $U \sigma_b U^\dagger = \sum_a (S_U)_{ab} \sigma_a$, $R_U = f_0 f_0^{\mathsf{T}} + \sum_{a,b} (S_U)_{ab} f_a f_b^{\mathsf{T}}$ and $\iota = \frac{1}{2}\mathbf{1}\mathbf{1}^{\mathsf{T}} - 1$. A cost Q of real 4×4 matrices has the *coherent reading* when $Q(R_U^{\mathsf{T}} X R_U) = Q(X)$ for all U and X, and is *quadratic on strains* when $Q(H) = B(H,H)$ for a symmetric bilinear B and all strains H.

Theorem 27.52 (The carrier's changes are exactly the rotations of space that fix time)**.** *In classical mathematics:*

*(a) the R_U are exactly the orthogonal R with $R\mathbf{1} = \mathbf{1}$ and $\det R = 1$; ι is orthogonal, fixes **1** and has determinant -1, and every orthogonal R fixing **1** has $\det R = 1$ or $\det(R\iota) = 1$;*

(b) Q has the coherent reading if and only if $Q(R^{\mathsf{T}} X R) = Q(X)$ for all X and all R as in (a); then $Q(h_\times) = Q(h_+)$;

(c) if Q is quadratic on strains and has the coherent reading, then $Q(R^{\mathsf T}HR) = Q(H)$ for every strain H and every orthogonal R with $R\mathbf{1} = \mathbf{1}$, improper ones included;

(d) the cost $\pi(X) = \det[\,v,\ Mv,\ M^2v\,]$, of degree six, has the coherent reading and $\pi(\iota X\iota) = -\pi(X)$ for every X, and $\pi \neq 0$ on strains.

Proof. With $F = [f_0\ f_1\ f_2\ f_3]$, $R_U = F\,\mathrm{diag}(1, S_U)F^{\mathsf T}$ and $\iota = F\,\mathrm{diag}(1,-1,-1,-1)F^{\mathsf T}$, and every orthogonal R fixing $\mathbf{1}$ is $F\,\mathrm{diag}(1, S)F^{\mathsf T}$ with S orthogonal, $\det S = \det R = \pm 1$. (a) is then the standard fact that $U \mapsto S_U$ maps $\mathrm{SU}(2)$ onto the rotations of $\mathbb{R}^3$ [88]. (b) follows, and the rotation by $\pi/4$ about f_1, $f_2 \mapsto (f_2 + f_3)/\sqrt{2}$, $f_3 \mapsto (f_3 - f_2)/\sqrt{2}$, carries $h_+ = 4(f_2 f_2^{\mathsf T} - f_3 f_3^{\mathsf T})$ to $h_\times = 4(f_2 f_3^{\mathsf T} + f_3 f_2^{\mathsf T})$. (c) It remains to treat ι. Split a strain as $H = E + O$ with $O = \sum_a v_a(f_0 f_a^{\mathsf T} + f_a f_0^{\mathsf T})$; then $\iota H\iota = E - O$ and $Q(H) - Q(\iota H\iota) = 4B(E, O)$. The coherent turns by π and $\pi/2$ about the face axes leave B invariant and reverse each pairing of $f_0 f_a^{\mathsf T} + f_a f_0^{\mathsf T}$ with $f_0 f_0^{\mathsf T}$ or with $f_b f_d^{\mathsf T} + f_d f_b^{\mathsf T}$, so $B(E, O) = 0$. (d) $X \mapsto R^{\mathsf T}XR$ sends $v \mapsto S^{\mathsf T}v$ and $M \mapsto S^{\mathsf T}MS$, so $\pi \mapsto (\det S)\pi$; at a strain with $v = (1,1,1)$ and $M = \mathrm{diag}(1,2,3)$, π is a Vandermonde determinant, 2. □

So for quadratic costs the coherent reading frees the whole posting frame on strains, and π shows that it does not beyond them; we do not prove that six is the least such degree.

The coherent reading is a named premise about frames of the carrier. Let $L_U = \mathrm{diag}(1, S_U)$, so that $UR(p)U^\dagger = R(L_U p)$. As $\sum_i v_i = 2f_0^{\mathsf T}v$ and $(\sum_i v_i u_i)_a = 2f_a^{\mathsf T}v$, $T_1 = DF^{\mathsf T}$ with $D = \mathrm{diag}(2, \frac{2}{\sqrt3}, \frac{2}{\sqrt3}, \frac{2}{\sqrt3})$, so $L_U T_1 = T_1 R_U$; as L_U keeps $\mathscr{K}$, it is a cone identification at $\kappa = 1$ putting posting i at the pure record $UR(T_1 e_i)U^\dagger$ (Theorem 27.39). As $(L_U T_1)^{\mathsf T} X L_U T_1 = R_U^{\mathsf T}(T_1^{\mathsf T} X T_1)R_U$, Q has the coherent reading exactly when the charge $Q(\Lambda^{\mathsf T}X\Lambda)$ of every real 4×4 matrix X is the same in all frames $\Lambda = L_U T_1$, $U \in \mathrm{SU}(2)$. Call Q *quadratic* when $Q(X) = B(X, X)$ for a symmetric bilinear B and every real 4×4 matrix X, *lattice-symmetric* when $Q(X)$ is unchanged by every relabelling $X_{ij} \mapsto X_{\sigma(i)\sigma(j)}$, and say it has *parity* when $Q(\iota X\iota) = Q(X)$ for all X. Under a common relabelling and exchange of the pairs ij, kl, the quadruples $ijkl$ form eleven classes $[ijkl]$, numbered $0, \dots, 10$ by 0000, 0001, 0010, 0011, 0012, 0101, 0102, 0110, 0112, 0121, 0123; put $Q_a(X) = \sum_{ijkl} a_{[ijkl]} X_{ij} X_{kl}$ for $a \in \mathbb{R}^{11}$, $E_2 = a_0 - a_3 - a_5 - a_7 + 2a_{10}$, $E_3 = a_1 - a_2 - a_6 + a_9$ and

$$E_1 = a_0 - 2a_1 - 2a_2 - a_3 + 4a_4 - a_5 + 2a_6 - a_7 + 4a_8 + 2a_9 - 6a_{10}.$$

Let $s_\pm = \mathbf{1}p_1^{\mathsf T} \pm p_1\mathbf{1}^{\mathsf T}$, and let R_* fix f_0, f_1 and send $f_2 \mapsto (f_2 + f_3)/\sqrt{2}$, $f_3 \mapsto (f_3 - f_2)/\sqrt{2}$.

Proposition 27.53 (What the coherent reading adds: three linear numbers)**.** *In classical mathematics:*

(a) Every quadratic lattice-symmetric Q is a Q_a, and for it these are equivalent: the coherent reading; $Q(R_*^{\mathsf T}XR_*) = Q(X)$ for all X; $Q(R^{\mathsf T}XR) = Q(X)$ for all X and all orthogonal R with $R\mathbf{1} = \mathbf{1}$; the three numbers $Q(h_\times) = Q(h_+)$, $Q(s_+ + h_\times) = Q(s_+ - h_\times)$, $Q(s_- + h_\times) = Q(s_- - h_\times)$; $E_1 = E_2 = E_3 = 0$; parity with $Q(h_\times) = Q(h_+)$. Parity is equivalent to the last two numbers, and if $Q(X) = Q(\Pi_{\mathbf{1}} X \Pi_{\mathbf{1}})$ for all X, the coherent reading is equivalent to $Q(h_\times) = Q(h_+)$.

(b) Q_a with $a_4 = 1$, the step-class form with $w_1, \dots, w_4 = \frac12, \frac14, \frac12, 0$, and Q_a with $a_2 = 1$, $a_4 = \frac12$ (other $a_m = 0$) fail exactly the first, the second and the third number.

(c) A step-class form has the coherent reading exactly when $w_1 = 3w_3$ and $w_1 + w_3 = 4w_2$. The weights $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ and $\frac{1}{2}, \frac{1}{4}, \frac{1}{2}, 0$ give nonnegative forms with $Q(h_+) = 16$, the second equal to J_2 on all TT strains, and neither has the coherent reading, so each charges some matrix differently in two frames $L_U T_1$; $\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, 0$ gives both.

Proof. (a) $B(X,Y) = \frac{1}{4}(Q(X+Y) - Q(X-Y))$ inherits both symmetries, so its coefficient of $X_{ij}Y_{kl}$ depends only on $[ijkl]$. Summing 256 terms, $Q_a(h_\times) - Q_a(h_+) = 16E_1$, $Q_a(s_+ + h_\times) - Q_a(s_+ - h_\times) = 64E_2$, $Q_a(s_- + h_\times) - Q_a(s_- - h_\times) = -128E_3$. R_* is an R_U (Theorem 27.52(a)), $R_* h_+ R_*^{\mathsf{T}} = h_\times$, R_*^2 fixes $s_\pm$ and turns $h_\times$ into $-h_\times$, and ι fixes $h_\times$ and turns $s_\pm$ into $-s_\pm$; so the coherent reading gives invariance under R_*, which gives the numbers, and parity gives the last two. The zeros of E_2, E_3 form a space of dimension 9 spanned by $\frac{1}{16}\sum_a (p_a^{\mathsf{T}} X p_a)^2$ ($a_m = \frac{3}{16}$ for $m = 0, 3, 5, 7, 10$, otherwise $-\frac{1}{16}$; $E_1 = -2$) and eight forms with $E_1 = 0$: $(\mathbf{1}^{\mathsf{T}} X \mathbf{1})^2$ (all $a_m = 1$), $(\mathbf{1}^{\mathsf{T}} X \mathbf{1}) \operatorname{tr} X$ ($a_0 = a_3 = 1$, $a_1 = a_2 = a_4 = \frac{1}{2}$), $(\operatorname{tr} X)^2$, $\operatorname{tr} X^2$, $\operatorname{tr} X^{\mathsf{T}} X$ ($a_0 = 1$ and $a_3, a_7, a_5 = 1$ respectively), $|X\mathbf{1}|^2$, $|X^{\mathsf{T}}\mathbf{1}|^2$ ($a_0 = a_5 = 1$ and $a_1 = a_6 = 1$, $a_2 = a_9 = 1$ respectively), $(X\mathbf{1}) \cdot (X^{\mathsf{T}}\mathbf{1})$ ($a_0 = a_7 = 1$, $a_1 = a_2 = a_8 = \frac{1}{2}$); other coefficients are 0. They are independent (in a vanishing combination $a_{10}, a_4, a_6, a_9, a_8$ remove the first, second, sixth, seventh, eighth in turn, and a_3, a_7, a_5 the rest), and are unchanged by every orthogonal R with $R\mathbf{1} = \mathbf{1}$: so the numbers give frame freedom, hence the coherent reading. All nine have parity, and $\Pi_{\mathbf{1}} \iota = -\Pi_{\mathbf{1}}$. (b) (E_1, E_2, E_3) is $(4, 0, 0)$ and $(0, 0, -1)$ for the two Q_a. (c) A step-class form depends only on $X + X^{\mathsf{T}}$, which gives the third number; $Q(h_+) = 64w_2$, $Q(h_\times) = 16(w_1 + w_3)$ (Proposition 27.54). With $(p_a)_S = p_a^{\mathsf{T}} \mathbf{1}_S$, $\mathbf{1}_S^{\mathsf{T}} s_+ \mathbf{1}_S = 2|S|(p_1)_S$ and $\mathbf{1}_S^{\mathsf{T}} h_\times \mathbf{1}_S = 2(p_2)_S(p_3)_S$; as $(p_1)_i(p_2)_i(p_3)_i = 1$, a double step has one nonzero $(p_a)_S$ and $(p_a)_S = 0$ at $|S| = 4$, $Q(s_+ + h_\times) - Q(s_+ - h_\times) = 64(w_1 - 3w_3)$. Apply (a) and Proposition 27.51(b). □

On the per-posting route a step S, $0 < |S| < 4$, is charged $\frac{1}{2}u^2$ on each of $c(|S|)$ postings, with shares u summing to $\mathbf{1}_S^{\mathsf{T}} H \mathbf{1}_S$; even shares give $w_k = 1/(2c(k))$. The four *count rules* are $(c(1), c(2), c(3))$ with $c(2) = 2$ and $c(1), c(3) \in \{1, 3\}$: the *fewest* $(1, 2, 1)$, forward $(1, 2, 3)$, reversed $(3, 2, 1)$ and longest $(3, 2, 3)$. Read each counter as an account's signed balance (an identification), a posting being a debit $+1$ or a credit -1 on one account. A word of postings with $t \in \mathbb{Z}$ ticks *realizes* S when its balance plus $t\mathbf{1}$ is $\mathbf{1}_S$; the *tick-free count* of S is the least length of a realizing word, and the *tick-posted count* the least of the length plus $4|t|$. The named premise *tick-free least postings* for c says that $c(|S|)$ is the tick-free count of every S with $0 < |S| < 4$. With $E_a = f_a f_a^{\mathsf{T}}$, a cost Q is *axis-calibrated* when $Q(\mu E_a) = \frac{1}{2}\mu^2$, has *independent axes* when $Q(E_a + E_b) = Q(E_a) + Q(E_b)$ for $a \neq b$, and is *face-axis additive* when $Q(\sum_a \mu_a E_a) = \frac{1}{2}\sum_a \mu_a^2$.

Proposition 27.54 (What the count and the face calibration rest on)**.** *In classical mathematics:*

(a) a share rule unchanged by relabelling the counters is even inside each step and even outside it, and m shares with sum s cost at least $s^2/(2m)$, with equality exactly at equal shares;

(b) every count rule has the face calibration, and no birefringence holds exactly for the fewest count; least charge picks the longest count, which is birefringent, as are equal weights;

(c) for $0 < |S| = k < 4$ the tick-free count of S is $\min(k, 4-k)$, as for its complement, and the tick-posted count is k; tick-free least postings holds for the fewest count, forces it, and then gives the per-posting form and both numbers, while the forward count fails it;

(d) for Q quadratic on strains: Q is face-axis additive exactly when it is axis-calibrated with independent axes; if it is axis-calibrated, $Q(h_+) = 16$ exactly when $Q(E_2 + E_3) = Q(E_2) + Q(E_3)$, and under the coherent reading exactly when it is face-axis additive; every cost $(\frac{1}{2} - \alpha)\operatorname{tr} M^2 + \alpha(\operatorname{tr} M)^2$ has the coherent reading, is quadratic on strains and axis-calibrated, and has independent axes only at $\alpha = 0$.

Proof. (a) A transposition of two counters both in S, or both outside it, fixes S; and $\sum_i \frac{1}{2}u_i^2 - s^2/(2m) = \frac{1}{2}\sum_i (u_i - s/m)^2$. (b) Number the counters 1 to 4. Only the double steps $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$ strain h_+, by ± 4, and only the single and triple steps strain $h_\times$, by ± 2; so $Q(h_+) = 64w_2 = 16$, while $Q(h_\times) = 16(w_1 + w_3)$ is 16, $\frac{32}{3}$, $\frac{32}{3}$, $\frac{16}{3}$ for the fewest, forward, reversed and longest counts. Least charge minimizes $s^2/(2c)$, so takes $c = 3$. (c) A word is at least as long as $\sum_i |b_i|$ for its balance b, which $|b_i|$ postings on each account attain; with t ticks the balance is $\mathbf{1}_S - t\mathbf{1}$, which needs $k|1-t| + (4-k)|t|$ postings, least at $t = 0$ or $t = 1$, and adding $4|t|$ makes $t = 0$ best. Exchanging debits with credits and t with $1-t$ maps the realizations of S to those of its complement. Then apply (b). (d) $Q(\sum_a \mu_a E_a) = \sum_a \mu_a^2 Q(E_a) + 2\sum_{a<b} \mu_a \mu_b B(E_a, E_b)$, and $Q(h_+) = Q(4E_2 - 4E_3) = 16 - 32B(E_2, E_3)$ for an axis-calibrated Q; the coherent turns by $\pi/2$ about the face axes carry $\{E_2, E_3\}$ to every other pair. The family depends only on $\operatorname{tr} M$ and $\operatorname{tr} M^2$ and charges $1 + 2\alpha$ for $E_a + E_b$. □

A count rule comes from how the ledger writes a step (writing each step by a least writing is an identification, as are the credit conditions of (d)). A *writing* of S is a list of postings (a debit $+1$ or a credit -1 on one counter) and *ticks* (one event adding 1 to every counter) with net change $\mathbf{1}_S$. A *charge* prices events, a writing costs the sum, and a *least writing* of S costs no more than any writing of S. A charge *reads off* c when for every S with $0 < |S| < 4$ least writings exist and all have $c(|S|)$ postings, and has *prices* $p_\mathrm{d}, p_\mathrm{c}, q$ when debits cost p_d, credits p_c and the tick q. The *comparison charge* of a change $\Delta \in \mathbb{R}^4$ is $\sum_{i,j} J(e^{\Delta_i - \Delta_j})$ over ordered pairs; the *retained-count charge* of an event is $\sum_i (J(e^{\delta_i}) + J(e^{\gamma_i}))$, δ_i, γ_i the changes of account i's debit and credit counts, a tick being one debit on each account.

Proposition 27.55 (The price of the tick decides the count)**.** *In classical mathematics:*

(a) for prices $p_\mathrm{d}, p_\mathrm{c} > 0$, $q \geq 0$ and $0 < |S| = k < 4$, least writings of S exist; all have k postings if $p_\mathrm{d} k < p_\mathrm{c}(4-k) + q$, all $4-k$ if $p_\mathrm{c}(4-k) + q < p_\mathrm{d} k$, and some of each at equality. A rule read off it satisfies tick-free least postings exactly when $p_\mathrm{c} + q < 3p_\mathrm{d}$ and $p_\mathrm{d} < 3p_\mathrm{c} + q$;

(b) if $p_\mathrm{d} = p_\mathrm{c} = p > 0$ and $q \geq 0$, the charge reads off the fewest count when $q < 2p$, the forward count when $q > 2p$, and no rule when $q = 2p$;

(c) the comparison charge has prices $6J(e)$, $6J(e)$, 0 and reads off the fewest count, and each such rule satisfies tick-free least postings, has the per-posting form and meets both numbers of Proposition 27.51(b); the retained-count charge has prices $J(e)$, $J(e)$, $4J(e)$ and reads off the forward count, and each such rule fails tick-free least postings and is birefringent;

(d) for a counter a and $\nu \in \mathbb{R}^4$ with $\sum_j \nu_j = 1$ and $(\mathbf{1}_{\{a\}})_i - (\mathbf{1}_{\{a\}})_j + \nu_i - \nu_j = 0$ for all i, j, $\nu = \frac{1}{2}\mathbf{1} - \mathbf{1}_{\{a\}}$; this antipode meets both conditions, $\mathbf{1}_{\{a\}} + \nu = \frac{1}{2}\mathbf{1}$, ν has the differences of the credit $-\mathbf{1}_{\{a\}}$, and the comparison charge is $6J(e)$ at ν and at $\mathbf{1}_{\{a\}}$, 0 at constants.

Proof. (a) A writing with t ticks needs $|(\mathbf{1}_S)_i - t|$ debits or credits on counter i, and more postings cost more, so its least charge is $p_d k$ for $t = 0$, $p_c(4-k)+q$ for $t = 1$, and $p_c(4t-k)+qt$, larger by $(t-1)(4p_c+q)$, for $t \geq 2$: the least writings are the k debits on S or the tick with the $4-k$ credits off S, whichever costs less. Then $c(2) = 2$, $c(1) = 1$ exactly when $p_d < 3p_c + q$, $c(3) = 1$ exactly when $p_c + q < 3p_d$, and tick-free least postings is $c = (1, 2, 1)$ (Proposition 27.54(c)). (b) At $k = 1$, $p < 3p + q$; at $k = 3$ compare $3p$ with $p + q$. (c) A posting on a changes six comparisons by $e^{\pm 1}$, and $J(x) = J(1/x)$; a tick changes none, and $J(1) = 0$. A posting changes one count by one, a tick four. $J(e) = \cosh 1 - 1 > 0$. Apply (b) and Proposition 27.54(b),(c). (d) The condition at (a, j) gives $v_j = v_a + 1$ for $j \neq a$, and the sum $4v_a + 3 = 1$; the rest is direct. □

Independent axes rest on a named statistical premise. A *law* (of probabilities) is $\varpi \in \mathbb{R}^4$ with all $\varpi_i > 0$ and $\sum_i \varpi_i = 1$; the *face bit* a of posting i is $(p_a)_i = \pm 1$, the ath coordinate of u_i. The state $R(1, \vec{r})$ gives the law $\varpi_i(\vec{r}) = \frac{1}{2}\langle T_1 e_i, (1, \vec{r})\rangle = \frac{1}{4}(1 + \vec{r}\cdot u_i/\sqrt{3})$; $\varpi(\vec{0})$ is the *uniform law*. For $\mu \in \mathbb{R}^3$ and a gain γ, $\delta = \frac{\gamma}{4}\sum_a \mu_a p_a$ moves the mean of face bit b by $\gamma\mu_b$; as $J(1+x) = x^2/(2(1+x))$, its charge $\frac{1}{2}\sum_i \delta_i^2/\varpi_i$ is the limit of $\varepsilon^{-2}\sum_i \varpi_i J(1 + \varepsilon\delta_i/\varpi_i)$. A cost Q has the *statistical face reading* at (ϖ, γ) when $Q(\sum_a \mu_a E_a) = \frac{1}{2}\sum_i \delta_i^2/\varpi_i$ for every μ.

Proposition 27.56 (Independent axes are the uniform law of the face bits)**.** *In classical mathematics:*

(a) Under $\varpi(\vec{r})$ face bit a has mean $r_a/\sqrt{3}$, for every $\vec{r}$. The uniform law is the only law unchanged by relabelling the postings, and under it any two face bits are independent.

(b) If Q has the statistical face reading at (ϖ, γ), then $Q(E_a+E_b) - Q(E_a) - Q(E_b) = \frac{\gamma^2}{16}\sum_i (p_c)_i/\varpi_i$ for $\{a, b, c\} = \{1, 2, 3\}$. So if $\gamma \neq 0$, Q has independent axes exactly when ϖ is uniform; if $\varpi = \varpi(u_1/\sqrt{3}) = (\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, the law of the pure record $R(T_1 e_1)$, every such cross term is $-\gamma^2/4$.

Proof. (a) $\sum_i (p_a)_i = 0$ and $p_a \cdot p_b = 4\delta_{ab}$. A law unchanged by relabelling is constant. For $a \neq b$ the pairs $((p_a)_i, (p_b)_i)$ are the four sign pairs once each. (b) As $(p_a)_i^2 = 1$ and $(p_1)_i(p_2)_i(p_3)_i = 1$,

$$Q\Big(\sum_a \mu_a E_a\Big) = \frac{\gamma^2}{32}\sum_i \frac{(\sum_a \mu_a (p_a)_i)^2}{\varpi_i} = \frac{\gamma^2}{32}\Big(\sum_a \mu_a^2 \sum_i \frac{1}{\varpi_i} + 2\sum_{a<b} \mu_a\mu_b \sum_i \frac{(p_c)_i}{\varpi_i}\Big);$$

the cross terms vanish exactly when $(1/\varpi_i)_i$ is orthogonal to p_1, p_2, p_3, a multiple of $\mathbf{1}$, so ϖ is uniform. At $(\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, $\sum_i (p_c)_i/\varpi_i = 2 - 6$. □

So the face bits read the complementary components $\sigma_1, \sigma_2, \sigma_3$, as $\mathrm{tr}(R(1, \vec{r})\sigma_a) = r_a$, but complementarity alone does not give independent axes.

Two more facts locate the remaining premise. A local lattice cost (Definition 27.9) is *one-contact* when each offset $p_j - q_j$ is a Kuhn step $\mathbf{1}_S$, as for an additive law of posting events that each read one record and one a step away; its coefficients are *bilinear tables* when $c_j(H_0, H_1) = \sum T_{abcd}(H_0)_{ab}(H_1)_{cd}$.

Proposition 27.57 (Locality is the premise; reading one is J_2 at quadratic order only)**.** *In classical mathematics:*

(a) every local lattice cost with bilinear-table coefficients has, on all pairs of matrices and at all waves, the long-wave symbol of a one-contact local cost with bilinear-table coefficients of either sign;

(b) if $1+\varepsilon X$ is invertible, then $J(1+\varepsilon X) = \varepsilon^2 J_2(X) - \frac{1}{2}\varepsilon^3 \operatorname{tr} X^3 + \frac{1}{2}\varepsilon^4 \operatorname{tr}(X^4(1+\varepsilon X)^{-1})$, and $\operatorname{tr} H^3 = 0$ for every strain H that is TT at a wave not a multiple of $\mathbf{1}$.

Proof. (a) The long-wave symbol is $\sum_{e,f} g_{ef} K_e K_f$ with bilinear tables g_{ef}; $K_e^2 = (K\cdot \mathbf{1}_{\{e\}})^2$ and $2K_eK_f = (K\cdot\mathbf{1}_{\{e,f\}})^2 - K_e^2 - K_f^2$, and a table g times $(K\cdot\mathbf{1}_S)^2$ is one term at the offset $\mathbf{1}_S$ with coefficient $-2g$. (b) Expand $(1+\varepsilon X)^{-1}$ one order further than in Theorem 27.2. A TT strain is $\sum_{a,b} M_{ab} f_a f_b^{\mathsf T}$ with M traceless and singular, and $\operatorname{tr} M^3 = 3\det M$ for traceless 3×3 matrices. □

So the premise that selects the Fierz–Pauli trace coefficient (Theorem 27.12(d)) is locality, and that the ledger's charge is local is not derived. At cubic order reading one parts from the static form $V(h_1,h_2,h_3) = -\frac{1}{4}\sum k_b^{\mathsf T} h_a h_b h_c k_b$, summed like reading one's cubic term $-\frac{1}{2}\sum \operatorname{tr}(h_a h_b h_c)$ over the six orderings (a,b,c) of three legs. Take the space waves $k_1=(1,0,0)$, $k_2=(0,1,0)$, $k_3=(-1,-1,0)$ and the transverse-traceless legs $h_1^+ = \operatorname{diag}(0,1,-1)$, $h_2^+ = \operatorname{diag}(1,0,-1)$, $h_3^+ = \left(\begin{smallmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & -2 \end{smallmatrix}\right)$, $h_2^\times = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right)$ and $h_3^\times = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \end{smallmatrix}\right)$. On (h_1^+,h_2^+,h_3^+), V is $\frac{1}{2}$ and reading one is 6; on $(h_1^+,h_2^\times,h_3^\times)$ they are 1 and 3. The ratios $\frac{1}{12}$ and $\frac{1}{3}$ differ, so no constant turns one form into the other. That V is the static cubic vertex of Einstein's action, the coefficient of $\mu_1\mu_2\mu_3$ in $\sqrt{\gamma}\,R(\gamma)$ for the spatial metric $\gamma = 1 + \sum_a \mu_a h_a e^{ik_a\cdot x}$, rests on an expansion we do not reproduce, and we do not claim it; read with it, reading one matches Einstein's graviton at quadratic order only. The coherent reading (three linear numbers, Proposition 27.53), independent axes (the statistical face reading at the uniform law, Proposition 27.56), the per-posting charge, tick-free least postings (a tick cheaper than two postings, Proposition 27.55) and locality (a depth bound or an order of limits, Theorem 27.13) are named premises, each sharpened here and none derived; the preprint [95] develops linear gravity on them, and Section 27.16 states where each premise stands.

27.11 Regge's action is one half of Einstein's through quartic order

The earlier sections compared Regge's action with linearized gravity: its Hessian is the Fierz–Pauli form at long wavelength (Proposition 27.21, Theorem 27.30). Gravity is not linear. A graviton scatters off a graviton through the cubic and quartic terms of the action, and those terms decide whether a lattice action is Einstein's theory or only shares its propagator. So the question here is the next one: are the cubic and quartic long-wave vertices of Regge's action on the Kuhn lattice those of $\sqrt{|g|}\,R$?

They are, each with the factor $\frac{1}{2}$, the same at every order. On every Euclidean background g_κ, $\kappa < \frac{1}{4}$, this holds with Regge's own derivatives at quadratic, cubic and quartic order; on the

Lorentzian backgrounds it holds for the continued tables; and the identities of the tables hold on every background g_κ off three degenerate values. The full treatment is the preprint [196] (Zenodo, 2026), and the tables, with programs that check the identities exactly, are the data set [210]; this section states the results, proves the steps that fit on a page, and describes the finite computations that close the rest. Everything here is proved in classical mathematics: the identities are identities of polynomials and rational functions over a field, and the derivatives are derivatives of real functions.

Definition 27.58 (Tables, long-wave vertices and Einstein's vertices)**.**

1. A *table* of degree n is a homogeneous polynomial $T(\lambda)$ of degree n in the ten slot variables $\lambda = (\lambda_f)$ of one Kuhn simplex (slots as in Definition 27.19). For n plane waves $H_i e^{iK_i\cdot x}$ with H_i symmetric, the Taylor amplitudes $a^i_{\sigma,m}(f) = (s_f^{\mathsf T} H_i s_f)(K_i\cdot x_f)^m/m!$ are those of Definition 27.23, and $\hat T(b_1,\dots,b_n)$ is the coefficient of $\epsilon_1\cdots\epsilon_n$ in $T(\epsilon_1 b_1+\cdots+\epsilon_n b_n)$. The *long-wave vertex* of T is

$$\mathscr{V}[T](H;K) = \frac{1}{n!}\sum_{m_1+\cdots+m_n=2}\ \sum_{\sigma}\hat T\big(a^1_{\sigma,m_1},\dots,a^n_{\sigma,m_n}\big),$$

   summed over the 24 simplices and the ways of putting two momentum factors on the legs. For $n=3$ it is the vertex V_T of Definition 27.23.

2. For a density $\mathscr{L}(g,\partial g,\partial\partial g)$ and a constant symmetric g_0, the *vertex* $V_n[\mathscr{L}](H;K)$ is the coefficient of $\epsilon_1\cdots\epsilon_n$ in $\mathscr{L}$ at the metric $g = g_0+\sum_i \epsilon_i H_i e^{iK_i\cdot x}$; when $K_1+\cdots+K_n=0$ it does not depend on x.

3. With $\Gamma_{e,ab} = \tfrac12(\partial_a g_{eb}+\partial_b g_{ea}-\partial_e g_{ab})$, *Einstein's first-order Lagrangian* is

$$\mathscr{L}_{\Gamma\Gamma} = \sqrt{|\det g|}\; g^{ab}g^{ce}g^{df}\big(\Gamma_{e,ad}\Gamma_{f,bc}-\Gamma_{e,ab}\Gamma_{f,cd}\big),$$

   and *Einstein's vertices* on the background with inverse metric $\gamma = g_0^{-1}$ are $W_n(\gamma;H;K) = V_n[\mathscr{L}_{\Gamma\Gamma}](H;K)/\sqrt{|\det g_0|}$.

4. Let Ψ be the *simplex function*, the share $\sum_t A_t(\theta_t-\theta_t^0)$ of one Kuhn simplex, so that Regge's action near the flat lattice is $-\sum_\sigma \Psi_\sigma$, and let ℓ_κ be the squared edge lengths of the Kuhn simplex at g_κ. For $\kappa<\frac14$, $n=2,3,4$ and $\lambda\in\mathbb{R}^{10}$ the *Regge tables* of the background g_κ are

$$T_n(\kappa;\lambda) = (1-4\kappa)^{-1/2}\,\frac{d^n}{dr^n}\Psi(\ell_\kappa+r\lambda)\Big|_{r=0}.$$

   By Theorem 27.59 each $T_n(\kappa)$ is a table of degree n whose coefficients are rational functions of κ: $T_2(\kappa;\lambda) = -\lambda^{\mathsf T}R(\kappa)\lambda$, with $R(\kappa)$ the rational family of Theorem 27.28; the coefficients of $T_3(\kappa)$ have poles only at $\frac14,\frac13,\frac12$, and $T_3(0)=r$, the form of Table 27.2; and the coefficients of $T_4(\kappa)$ are $N_\mu(\kappa)/D(\kappa)$ with integer polynomials N_μ and

$$D(\kappa) = 1536\,(2\kappa-1)^3(3\kappa-1)^3(4\kappa-1)^3.$$

   At every other real or complex κ off these poles, $T_n(\kappa)$ is the table of the continued rational functions, which is unique. The coefficients of T_2, T_3 and T_4 are listed in the data set [210].

The vertex of a density is an actual coefficient of the action: one cell of the lattice and one unit of coordinate volume are the same region, and a vertex with two derivatives is of second order in the momenta. For a lattice action whose simplex shares are $-\Psi_\sigma$, the part of second order in the momenta of the $\epsilon_1 \cdots \epsilon_n$ coefficient of one cell is $\mathscr{V}[Q_n]$ when the share's part of order n is $-Q_n/n!$. So the vertices of Regge's action are those of the tables once the tables, which are defined from its derivatives, are known to be the rational tables described above. That is the first theorem.

Theorem 27.59 (Regge's derivatives on every Euclidean background)**.** *In classical mathematics, for every $\kappa < \frac{1}{4}$, every direction $\lambda \in \mathbb{R}^{10}$ and $n = 2, 3, 4$,*

$$\frac{d^n}{dr^n}\Psi(\ell_\kappa + r\lambda)\Big|_{r=0} = \sqrt{1-4\kappa}\; T_n(\kappa;\lambda),$$

where $T_n(\kappa;\lambda)$ is a homogeneous polynomial of degree n in λ whose coefficients are rational functions of κ: $T_2(\kappa;\lambda) = -\lambda^{\mathsf{T}} R(\kappa)\lambda$; the coefficients of T_3 have poles only at $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}$, and $T_3(0) = r$; and the coefficients of T_4 are $N_\mu(\kappa)/D(\kappa)$ with integer polynomials N_μ.

Conditions and scope.

Base and premises

The Kuhn simplex at the Euclidean backgrounds g_κ, $\kappa < \frac{1}{4}$; classical real numbers.

Conclusion

The three tables are Regge's own second, third and fourth derivatives, with the volume factor $\sqrt{\det g_\kappa}$ in front, on every Euclidean member of the family. With Definition 27.58 the long-wave vertices of Regge's action at g_κ are $V_n[S_{\mathrm{R}}] = \sqrt{\det g_\kappa}\,\mathscr{V}[T_n(\kappa)]$.

Exact computation

The third- and fourth-order identities are long polynomial identities, checked by exact computation as the proof describes and not reproduced here; the coefficients of $T_3(\kappa)$ and $T_4(\kappa)$ as rational functions of κ, and programs that check the identities exactly, are in the data set [210].

Not claimed

True derivatives for $\kappa > \frac{1}{4}$, where the angles are Lorentzian; there the tables are continued as rational functions.

Proof. For $n = 2$ this is Theorem 27.29, since $S = -\Psi$ there and $\Psi'' = \sum_t A_t'\theta_t'$. For $n = 3$ and $n = 4$ the route is the same, carried further. Schläfli's identity (Theorem 27.27) gives $\Psi' = \sum_t A_t'(\theta_t - \theta_t^0)$ along every line, and differentiating it again,

$$\Psi'' = \sum_t A_t'\theta_t', \qquad \Psi''' = \sum_t \big(2A_t''\theta_t' + A_t'\theta_t''\big), \qquad \Psi^{(4)} = \sum_t \big(3A_t'''\theta_t' + 3A_t''\theta_t'' + A_t'\theta_t'''\big)$$

at $r = 0$, where the angles' flat values drop out. Each derivative of an area or an angle comes from the Gram algebra of Theorem 27.26: rational functions of κ and λ times the square roots $\sqrt{Q_t}$ and $\sqrt{W_{ll}W_{mm} - W_{lm}^2}$, and by part (c) of that theorem every product above carries the square roots in pairs that reduce to one factor $\sqrt{1-4\kappa}$. So each identity, cleared

of denominators, is a polynomial identity in κ and the ten λ_f, one per triangle. At third order the ten triangles give an exact computation of moderate size. At fourth order the reversal of the simplex's vertex order maps four of the ten triangles onto four others, so six identities remain; each expands, after clearing denominators, into a polynomial in eleven variables whose verification by exact computation takes between one and one and a half million monomial merges, and all six vanish identically. □

The second step moves from Einstein's first-order Lagrangian to $\sqrt{|g|}\,R$ itself.

Theorem 27.60 (The vertices of $\sqrt{|g|}\,R$ are Einstein's first-order vertices)**.** *In classical mathematics, let $\sigma \in \{1,-1\}$ and let g_0 be a symmetric 4×4 real matrix with $\sigma \det g_0 > 0$, $\gamma = g_0^{-1}$. For symmetric H_i and momenta with $\sum_i K_i = 0$, and $n = 2,3,4$,*

$$V_n\big[\sqrt{\sigma \det g}\,R\big](H;K) = \sqrt{\sigma \det g_0}\; W_n(\gamma;H;K),$$

and $W_2(\gamma;X,X';k,-k) = -\frac{1}{2}\mathrm{FP}_\gamma(k;X,X')$, with FP_S the form of Section 27.2. In particular, for every $\kappa < \frac{1}{4}$ the quartic vertex of $\sqrt{\det g}\,R$ around g_κ is $\sqrt{1-4\kappa}\,W_4(g_\kappa^{-1})$, and around g_1 the vertices of $\sqrt{-\det g}\,R$ are $\sqrt{3}\,W_n(g_1^{-1})$.

Proof. Three links. First, the classical identity $\sqrt{|g|}\,R = \mathscr{L}_{\Gamma\Gamma} + \partial_c\big(\sqrt{|g|}\,Y^c\big)$ with $Y^c = g^{bd}\Gamma^c_{bd} - g^{bc}\Gamma^a_{ba}$ [200, §93], which holds over every commutative ring with derivations once $\partial g^{-1} = -g^{-1}(\partial g)g^{-1}$ and Jacobi's formula $\partial\sqrt{|g|} = \sqrt{|g|}\,\Gamma^e_{ec}$ are supplied. Second, on plane waves the derivative ∂_c acts on the coefficient of $\epsilon_1\cdots\epsilon_n$ as multiplication by $i\sum_i K_{i,c}$, so a total derivative has no vertex when the momenta sum to zero. Third, the vertex is the top coefficient of the density in the ring $\mathbb{R}[\epsilon_1,\dots,\epsilon_n]/(\epsilon_1^2,\dots,\epsilon_n^2)$ with $g = g_0 + \sum \epsilon_i H_i$, $\partial_c g = \sum \epsilon_i\, iK_{i,c}H_i$: the map that sends a smooth function to this jet is a ring homomorphism, it sends $\sqrt{\sigma\det g}$ to a square root of $\sigma\det g$ and g^{-1} to the inverse, and so the mixed derivative is computed inside that ring. The quadratic case is a direct expansion of $\mathscr{L}_{\Gamma\Gamma}$ to second order; for the two backgrounds use $\det g_\kappa = 1-4\kappa$ (Theorem 27.26(a)). □

Theorem 27.61 (Regge's tables are one half of Einstein's through quartic order)**.** *In classical mathematics:*

(a) Quartic, every background. *Let F be a field of characteristic zero and $\kappa \in F$ with $(2\kappa-1)(3\kappa-1)(4\kappa-1)\neq 0$. For all symmetric $H_1,\dots,H_4$ over F and all $K_1,\dots,K_4 \in F^4$ with $K_1+\cdots+K_4 = 0$,*

$$\mathscr{V}[T_4(\kappa)](H;K) = \tfrac{1}{2}\,W_4\big(g_\kappa^{-1};H;K\big), \qquad g_\kappa^{-1} = 1 + \tfrac{\kappa}{1-4\kappa}\mathbf{1}\mathbf{1}^{\mathsf T}.$$

(b) Cubic, every background. *For every real κ with $(2\kappa-1)(3\kappa-1)(4\kappa-1)\neq 0$, all complex symmetric H_1,H_2,H_3 and all $K_1,K_2,K_3 \in \mathbb{C}^4$ with $K_1+K_2+K_3 = 0$, $\mathscr{V}[T_3(\kappa)](H;K) = \frac{1}{2}W_3(g_\kappa^{-1};H;K)$; in particular $\mathscr{V}[T_3(0)] = \frac{1}{2}W_3(1)$ and $\mathscr{V}[T_3(1)] = \frac{1}{2}W_3(g_1^{-1})$, with $g_1^{-1} = 1 - \frac{1}{3}\mathbf{1}\mathbf{1}^{\mathsf T}$.*

(c) Quadratic, every background. *For every real κ with $(2\kappa-1)(3\kappa-1)(4\kappa-1)\neq 0$ and all complex symmetric X, X' and $P \in \mathbb{C}^4$, $\mathscr{V}[T_2(\kappa)](X,X';P,-P) = -\frac{1}{4}\mathrm{FP}_{g_\kappa^{-1}}(P;X,X')$, which is $\frac{1}{2}W_2(g_\kappa^{-1})$.*

Conditions and scope.

Base and premises
The Kuhn simplex and its tables; classical algebra over a field. Part (a) holds over $\mathbb{Q}(\kappa)$, $\mathbb{R}$ and $\mathbb{C}$, so for complex momenta too; parts (b) and (c) hold at every real κ off the poles, for complex legs.

Conclusion
At every order through four, on every background g_κ off the poles, the long-wave vertex of the lattice tables is one half of Einstein's. With Theorems 27.59 and 27.60, Regge's action on the Kuhn lattice is $\frac{1}{2}\int \sqrt{g}\,R$ through quartic order at long wavelength around every Euclidean background g_κ, $\kappa < \frac{1}{4}$ (Corollary 27.62).

Exact computation
Each part ends in a finite computation that the proof describes and does not reproduce: for (a), fifty-three comparisons of integers of about 200,000 digits; for (b), polynomial identities in κ and the eight free momentum components on the 20 representative triples of basis matrices; for (c), the same on the 100 ordered pairs. The tables are part S1 of the data set [210]; its parts S4 and S3 hold programs that check the identities of (a) on the fifty-three representatives, and those of (b) and (c) with κ kept as a variable, exactly.

Identification
The midpoint rule that turns a metric wave into edge lengths, $d^{\mathsf{T}} H d\, e^{iK\cdot(y+d/2)}$, and Regge's action as the gravitational action.

Not claimed
The metric's five missing edge modes, which Section 27.14 treats.

Proof. (a) Both sides are linear in each leg, invariant under permutations of the legs and under the 24 relabellings of the axes, which permute the simplices of a cell and fix g_κ; and W_4 is homogeneous of degree five in γ and bilinear in the two Christoffel factors. A functional with these properties that vanishes on 53 representative 4-tuples of basis polarizations $E_{ab}+E_{ba}$, with the fourth momentum eliminated, vanishes on all symmetric legs and conserved momenta; that the 53 representatives reach all 10^4 ordered tuples is a finite check. On a representative, the lattice side times $192\,D(\kappa)$ is a polynomial of degree below ten in κ and quadratic in the free momentum components, and Einstein's side times $16(1-4\kappa)^5$ is a polynomial of degree at most five in κ; the identity $(1-4\kappa)^5 \cdot 192 D(\kappa) \cdot \big(\mathscr{V}[T_4] - \frac{1}{2}W_4\big) = 0$ is checked by exact computation for each of the 53 representatives. Each check packs all coefficients of both polynomials into one integer, with distinct monomials at distinct powers of a large base, and compares two integers of about 200,000 decimal digits; bounds on the coefficients make the comparison equivalent to the polynomial identity. Since $(1-4\kappa)D(\kappa) \neq 0$, the identity follows. (b) Both sides are trilinear and symmetric in the legs and quadratic in the momenta. Take real data first. Both sides are sums, over the entries of the three legs, of quadratic forms in the momenta whose coefficients are rational in κ; with the third momentum eliminated only the projected coefficients count, and clearing the denominators $(4\kappa-1)^4$ on Einstein's side and those of $T_3(\kappa)$ on the lattice side turns the identity into finitely many polynomial identities in κ with integer coefficients, checked by exact computation; they hold at every real κ, so the identity holds wherever both sides are defined. Both sides are then polynomials

in the entries of the legs and two free momenta, with real coefficients, that agree at every real point; so they agree at every complex point. At $\kappa = 0$ the table is that of Table 27.2. (c) At each κ the quadratic vertex of $T_2(\kappa)$ and $-\frac{1}{4}\mathrm{FP}$ are polynomials in the entries of X, X', P whose coefficients are rational in κ; clearing the poles of $R(\kappa)$ turns the identity into finitely many polynomial identities in κ, checked by exact computation. The last equality is Theorem 27.60. At $\kappa = 0$ the quadratic identity can be read off Proposition 27.21: since $\hat{T}_2(b, b') = -2\, b^{\mathsf{T}} R(0) b'$, $48R(0) = E$ and the two momentum factors give $\frac{1}{2}\big(K \cdot (x_f - x_{f'})\big)^2$, the vertex is $-\frac{1}{96}\sum_\sigma \sum_{f,f'} E_{ff'} (s_f^{\mathsf{T}} X s_f)(s_{f'}^{\mathsf{T}} X' s_{f'}) \big(K \cdot (x_f - x_{f'})\big)^2 = -\frac{24}{96}\mathrm{FP}_1(K; X, X')$. □

Corollary 27.62 (One half of Einstein–Hilbert, through quartic order)**.** *In classical mathematics, on the Kuhn lattice, for $n = 2, 3, 4$, all real symmetric polarizations and real momenta summing to zero,*

$$V_n[S_{\mathrm{R}}](H; K) = \tfrac{1}{2}\, V_n\big[\sqrt{\det g}\, R\big](H; K) \quad \text{around every Euclidean member } g_\kappa,\ \kappa < \tfrac{1}{4},$$

in particular around $g_0 = 1$. On the Lorentzian member $\kappa = 1$ the same holds for the tables continued to $\kappa = 1$: $\sqrt{3}\, \mathscr{V}[T_n(1)] = \frac{1}{2} V_n[\sqrt{-\det g}\, R]$ around g_1, $n = 2, 3, 4$.

Proof. Theorem 27.59 gives $V_n[S_{\mathrm{R}}] = \sqrt{\det g_\kappa}\, \mathscr{V}[T_n(\kappa)]$, Theorem 27.61 gives $\mathscr{V}[T_n] = \frac{1}{2} W_n$, and Theorem 27.60 gives $\sqrt{|\det g_\kappa|}\, W_n = V_n[\sqrt{|g|}\, R]$; at $\kappa = 1$, $\sqrt{|\det g_1|} = \sqrt{3}$. □

The tables at $\kappa = 1$ are the Lorentzian lattice's own. The cell forms of Section 27.7, built on the $\kappa = 1$ cell from the rates of its areas and angles, are at orders two, three and four the forms of $T_2(1)$, $T_3(1)$ and $T_4(1)$ times one and the same factor, of absolute value $\sqrt{3} = \sqrt{|\det g_1|}$; and for every $\kappa > \frac{1}{2}$ the cell forms of the background g_κ are the continued tables times one factor of absolute value $\sqrt{4\kappa - 1}$ (Proposition 27.63). That the quartic cell form at $\kappa = 1$ is the fourth derivative of δ's own Regge action on $\mathbb{R}_\delta$ rests, as before, on its third-order rates, whose identification with third derivatives is open.

Proposition 27.63 (One normalization at every order)**.** *In classical mathematics: on the $\kappa = 1$ cell, the quadratic, cubic and quartic cell forms equal $\sqrt{3}$ times the forms of the tables $-T_2(1)$, $-T_3(1)$ and $-T_4(1)$, for every argument; and for every $\kappa > \frac{1}{2}$ the corresponding cell forms of the background g_κ equal $\sqrt{4\kappa - 1} = \sqrt{|\det g_\kappa|}$ times the forms of $-T_n(\kappa)$.*

Proof. Each cell form is assembled from closed forms for the first, second and third rates of the Lorentzian areas $\frac{1}{4}\sqrt{-Q_t}$ and angles of the cell, as the Euclidean derivatives are in the proof of Theorem 27.59. Jacobi's identity gives $A_t/\sqrt{|D_t|} = \frac{1}{2}\sqrt{|\det G|}$ on both sides of the light cone, so the square roots again collect into one factor, $\sqrt{4\kappa - 1}$, and the remaining identities are polynomial identities in κ and the slot variables, checked by exact computation; at $\kappa = 1$ they are integer identities. □

Remark 27.64 (Why the half is a normalization and not a change of coupling)**.** The factor $\frac{1}{2}$ is the classical normalization of Regge calculus, in which $\sum_h A_h \delta_h$ corresponds to $\frac{1}{2}\int \sqrt{g}\, R$ [199]. It is the same at orders two, three and four. A common factor in front of an action fixes its units; a factor that differed between orders would change the ratio of the cubic to

the quadratic vertex, which is the gravitational coupling. With an action scale s, $s\sum_h A_h\delta_h = \frac{s}{2}\int\sqrt{|g|}\,R = (16\pi G)^{-1}\int\sqrt{|g|}\,R$ at $G = 1/(8\pi s)$. On the Euclidean members the identities also follow from the convergence theorem of Cheeger, Müller and Schrader [199], by an argument that uses fatness bounds and the analyticity of the action in the wave amplitudes; the proofs above use no convergence theorem, and that theorem does not reach the Lorentzian member.

The answer to the opening question is therefore yes, at the order Einstein's theory is written in. Reading the lattice action as gravity's action remains an identification, and the factor $\kappa = 1$ remains the named premise of Section 27.8; given both, the lattice's gravitons interact at long wavelength through cubic and quartic order as general relativity's do.

27.12 Among area charges, gauge consistency selects Regge's action through cubic order

Section 27.5 showed that the hinge cost built from J fails at third order: its cubic vertex breaks gauge invariance, and Proposition 27.25 showed that a hinge cost linear in the deficit matches Regge's derivatives exactly when its area factor is linear to second order. This section turns the question around. Which hinge costs pass the gauge tests at all, and what exactly is wrong with J's? The answers are: among hinge charges built from one function of the area, through cubic order only Regge's charge passes, up to its scale, and for shape costs the tests fix Regge's action up to its coupling; J's cubic vertex is one quarter of Einstein's plus a gauge-breaking part Δ that we write out; and one hinge-local correction removes Δ and leaves exactly half of Regge's cubic vertex. All results are proved in classical mathematics, on the Euclidean Kuhn lattice.

Definition 27.65 (Hinge charges, jets and the gauge tests)**.**

1. A *hinge charge* built from one function of the area is the cost $\sum_h \varphi(A_h)\,\delta_h$, δ_h the deficit. At the Euclidean flat lattice the hinges have the four areas $a \in \{\frac{1}{2}, \frac{\sqrt{2}}{2}, \frac{\sqrt{3}}{2}, 1\}$, and the *jets* of φ are $\big(\varphi(a)/a, \varphi'(a), \varphi''(a)\big)$ at these four areas. Regge's charge $\lambda A\delta$ has the jets $(\lambda, \lambda, 0)$ at every area. A *ratio charge* is $a_0\psi(A/a_0)\,\delta$, a_0 the flat area of the hinge, with jets $(c_0, c_1, c_2) = (\psi(1), \psi'(1), \frac{1}{2}\psi''(1))$; Regge's are $(1, 1, 0)$.

2. The *quadratic test*: the charge's per-simplex quadratic table is gauge null (Definition 27.19) at every momentum. The *cubic test at a witness*: on a configuration of three legs, two of them solutions of the linearized field equations and one a pure-gauge leg, the long-wave cubic vertex of the charge's cubic table vanishes. The three *witnesses*, with Gaussian-integer null momenta for the Euclidean form, are: W_1, the legs $(\eta', \eta', K_3 \otimes e_1 + e_1 \otimes K_3)$ with $\eta' = \mathrm{diag}(-1, 1, 1, 1)$ at the static momenta $K_1 = (0, 3, 4, 5i)$, $K_2 = (0, 5, 12, 13i)$, $K_3 = -K_1 - K_2$; W_2, the same with e_2 in place of e_1; and W_3, the transverse-traceless legs

$$h_1 = \begin{pmatrix} -321 + 240i & -228 + 70i & -100 - 125i & -200 - 375i \\ -228 + 70i & 171 - 240i & 75 & 150 \\ -100 - 125i & 75 & -75 & 75 \\ -200 - 375i & 150 & 75 & 225 \end{pmatrix},$$

$$h_2 = \begin{pmatrix} 482+312i & -120+65i & -60-65i & -60+130i \\ -120+65i & -457-312i & 50 & -25 \\ -60-65i & 50 & 25 & 25 \\ -60+130i & -25 & 25 & -50 \end{pmatrix}$$

at $K_1 = (3,4,0,5i)$, $K_2 = (5,0,12,13i)$, with the gauge leg $K_3 \otimes e_1 + e_1 \otimes K_3$, $K_3 = -K_1 - K_2$. Each h_j is traceless and transverse to its momentum, $\sum_a K_{j,a} h_{j,ab} = 0$ for every b, and both momenta are null, $\sum_a K_{j,a}^2 = 0$.

The tests read a charge only through its jets: along a line of squared lengths the quadratic and cubic parts of $\varphi(A)\delta$ involve φ, φ' and φ'' at the flat areas and nothing else, since δ vanishes on the flat lattice.

Theorem 27.66 (Regge is the only consistent area charge through cubic order)**.** *In classical mathematics, for hinge charges built from one function of the area on the Euclidean Kuhn lattice:*

(a) the quadratic table is gauge null at every momentum exactly when the value-and-slope jets $(\varphi(a)/a, \varphi'(a))$, listed over the four areas, lie in the span of $v_1 = (2,1\,|\,2,1\,|\,2,1\,|\,2,1)$, $v_2 = (1,0\,|\,1,0\,|\,1,0\,|\,1,0)$ and $v_3 = (2,1\,|\,\frac{5}{2},\frac{1}{2}\,|\,3,0\,|\,3,0)$; the table of v_1 has zero symbol, that of v_2 is $-\frac{1}{48}E$ and that of v_3 is $-\frac{1}{16}R_1$, each up to a table of zero symbol, and the table of the constant jets $(1,1)$ is $\frac{1}{48}E$;

(b) the charge passes the quadratic test and the cubic test at W_1, W_2 and W_3 if and only if its jets are $(\lambda,\lambda,0)$ at all four areas for one real λ: Regge's jets at scale λ.

Conditions and scope.

Base and premises

The Euclidean Kuhn lattice; hinge charges built from one function of the area; classical real numbers.

Conclusion

Among these charges, the quadratic test and the cubic test at three witnesses leave exactly Regge's charge, up to its scale.

Exact computation

Part (a) is exact linear algebra on eight rational 10×10 tables; part (b) evaluates the per-area cubic tables exactly at the three witnesses of Definition 27.65 and checks that the six real linear equations of Table 27.5 have full rank: their determinant, computed exactly from the printed coefficients, is $-27\,866\,927\,400\,397\,976\,309\,514\,240$. The witnesses and the coefficients are printed here; the evaluation of the tables at the witnesses that produces the coefficients is the exact computation of [197, Table 1].

Not claimed

Charges with one function for each class of hinge, and shape costs (Proposition 27.67); costs nonlinear in the deficit.

Table 27.5. The cubic test of Theorem 27.66(b) as six real equations. Each row is the real or imaginary part of the test at one witness of Definition 27.65; the columns multiply $u_1 + u_2$, u_3 and the four curvature jets z_a. The static witnesses do not see the curvature jet at area 1; the transverse-traceless witness does. The determinant is $-27\,866\,927\,400\,397\,976\,309\,514\,240 \neq 0$.

	$u_1 + u_2$	u_3	$z_{1/2}$	$z_{\sqrt{2}/2}$	$z_{\sqrt{3}/2}$	z_1
W_1, real	-56	-360	-8	0	48	0
W_1, imag.	$-\frac{80}{3}$	-2356	-864	80	$\frac{2432}{3}$	0
W_2, real	-280	228	992	-504	-240	0
W_2, imag.	$-\frac{352}{3}$	-4139	-1344	352	$\frac{3328}{3}$	0
W_3, real	$-\frac{946463965}{8}$	$-\frac{1504452249}{4}$	14665886	$\frac{971809893}{8}$	$-\frac{37294969}{2}$	$\frac{55612715}{2}$
W_3, imag.	$-\frac{998974207}{8}$	$-\frac{41233057}{4}$	390527308	$-\frac{1786358293}{8}$	1780182	$-\frac{68444255}{2}$

Proof. (a) The quadratic table is linear in the eight value-and-slope jets, through eight rational 10×10 tables obtained from the first derivatives of the areas and angles of Theorem 27.26. Gauge nullity at every momentum is, after the phases are factored, a finite set of linear conditions (Theorem 27.20(d)), and the span is exact linear algebra on them. (b) Write the value-and-slope jets as $u_1 v_1 + u_2 v_2 + u_3 v_3$ and the curvature jets as z_a. The cubic table is linear in the twelve jets, and three exact identities among the per-area cubic tables reduce the cubic test at a witness to $\frac{3}{2}(u_1 + u_2)\mathscr{V}(X_1) + u_3\mathscr{V}(V_3) + \sum_a z_a \mathscr{V}(S_{2,a})$, with fixed tables X_1, V_3, $S_{2,a}$ evaluated exactly at the witness. The real and imaginary parts at the three witnesses are six real linear equations in $u_1 + u_2$, u_3 and the four z_a, of full rank (Table 27.5). So $u_1 + u_2 = u_3 = 0$ and every $z_a = 0$, which is $(\lambda, \lambda, 0)$ with $\lambda = u_1$. Conversely Regge's charge passes both tests (Theorem 27.20(e) and Theorem 27.24(b),(c)). □

The tests see the jets and nothing more: $[A + (A - a_0)^3/a_0^2]\,\delta$ has Regge's jets at the flat area a_0 without being Regge's charge. Rotation invariance does not do the selecting either: J's own cubic vertex below is rotation invariant and still fails. The gravity preprint [197] adds a second route, the record clause, which fixes the whole function and its scale [197, Theorem 5.5]. Wider classes behave as one expects once the inversion symmetry of the cell is taken into account.

Proposition 27.67 (Wider classes of hinge costs)**.** *In classical mathematics, on the Euclidean Kuhn lattice:*

(a) the central inversion of the cell leaves the long-wave cubic vertex of every cubic table unchanged under the reversal of its slots, at every momentum-conserving triple of legs;

(b) with one function of the area for each class of hinge, the quadratic test and the cubic test at W_1, W_2, W_3 leave exactly Regge's jets on the inversion-even part and three inversion-odd terms that neither test can see;

(c) for shape costs, $\sum_h f_c(L_h)\delta_h$ with one smooth function of the hinge's three squared lengths for each of its six classes, the quadratic test and the cubic test at eleven on-shell witnesses leave, on the inversion-even part, exactly λ times Regge's action plus μ times the coupling direction Y of Proposition 27.70, and inversion-odd terms of zero symbol.

Proof. (a) The inversion $x \mapsto \mathbf{1} - x$ maps the simplices of a cell to simplices, reverses the slots by the permutation $(0\,9)(1\,8)(2\,6)(4\,7)$ and reflects the midpoints; reindexing the cell sum as in Theorem 27.24(a) changes the phases $K_i \cdot x_f$ into $K_i \cdot \mathbf{1} - K_i \cdot x_f$, and the vertex, quadratic in the momenta, is unchanged when $\sum K_i = 0$. (b) and (c) The tests are linear in the jets; an inversion-odd table has, by (a), no long-wave vertex. The even part reduces to exact linear algebra on rational tables at the witnesses, as in Theorem 27.66. □

Now $\mathcal{J}$'s charge. By (27.1), $\mathcal{J}$'s odd hinge cost has the cubic form $-(r + d)$ per simplex, with r Regge's and d the extra term of Table 27.2. For a cubic monomial m in two momenta and three polarizations, write

$$\langle m \rangle = - \sum_{(i,j,k)} m(K_i, H_i, K_j, H_j, H_k),$$

summed over the six orderings of the legs; the sign is the i^2 of two derivatives.

Theorem 27.68 ($\mathcal{J}$'s cubic vertex is one quarter of Einstein's plus Δ)**.** *In classical mathematics, for all real symmetric H_1, H_2, H_3 and real momenta with $K_1+K_2+K_3 = 0$, with $W_3 = W_3(1; H; K)$:*

(a) $V_r = \frac{1}{2}W_3$, $V_d = -\frac{1}{4}W_3 + \Delta$ *and so* $V_r + V_d = \frac{1}{4}W_3 + \Delta$, *where*

$$\Delta = \tfrac{1}{8}\big\langle (K_i^{\mathsf{T}} H_i K_j)\,\mathrm{tr}\,H_j\,\mathrm{tr}\,H_k \big\rangle - \tfrac{1}{8}\big\langle (K_i^{\mathsf{T}} H_i H_j K_j)\,\mathrm{tr}\,H_k \big\rangle - \tfrac{1}{8}\big\langle (H_i K_i)^{\mathsf{T}} H_k (H_j K_j) \big\rangle$$
$$- \tfrac{1}{16}\big\langle (K_i \cdot K_j)\,\mathrm{tr}\,H_i\,\mathrm{tr}\,H_j\,\mathrm{tr}\,H_k \big\rangle + \tfrac{1}{16}\big\langle (K_i \cdot K_j)\,\mathrm{tr}(H_i H_j)\,\mathrm{tr}\,H_k \big\rangle + \tfrac{1}{8}\big\langle K_i^{\mathsf{T}} H_j H_k H_i K_j \big\rangle;$$

(b) $\Delta = -\frac{1}{8}\mathscr{E}_1 - \frac{1}{32}\mathscr{E}_3 + \mathscr{R}$, *where* $\mathscr{E}_1 = \sum \mathrm{tr}\,H_i\,(H_j : G^1(H_k))$ *and* $\mathscr{E}_3 = \sum \mathrm{tr}\,H_i\,\mathrm{tr}\,H_j\,S^1(H_k)$ *are field-equation vertices (G^1, S^1 the linearized Einstein tensor and scalar of a leg, summed over the leg assignments) and $\mathscr{R}$ is a sum of five invariant monomials; and Δ is not a combination of W_3 and the four field-equation vertices.*

Conditions and scope.

Base and premises

The Euclidean Kuhn lattice; classical real numbers.

Conclusion

Regge's cubic vertex is one half of Einstein's. $\mathcal{J}$'s odd hinge cost has the cubic form $-(r+d)$ and so the vertex $-(\frac{1}{4}W_3 + \Delta)$, against Regge's $-\frac{1}{2}W_3$: half of Regge's coupling plus a part Δ that no redefinition of the metric (which adds field-equation vertices) and no change of the coupling (which adds multiples of W_3) removes. This is the defect that Theorem 27.24 detected on two configurations.

Proof. (a) $V_r = \frac{1}{2}W_3$ is Theorem 27.61(b) at $\kappa = 0$. For d, both sides are trilinear in the legs, symmetric under their permutations and quadratic in the momenta; as in Theorem 27.61(a) they agree once they agree on representative triples of basis polarizations with two free momenta, and there both are explicit polynomials compared by exact computation. (b) The field-equation vertices and the five monomials of $\mathscr{R}$ are explicit; the identity is again a polynomial identity. For the last claim, one configuration makes W_3 and all four field-equation vertices vanish while $\Delta = -\frac{1}{4}$. □

A ratio correction can repair $\mathcal{J}$'s jets, and a hinge-local correction can do better.

Theorem 27.69 (The counterterm, and J without Δ)**.** *In classical mathematics, on the Euclidean Kuhn lattice:*

(a) a ratio charge passes the cubic test at W_1 if and only if $c_0 = c_1$ and $c_2 = 0$;

(b) J's charge has the ratio jets $(0, \frac{1}{2}, -\frac{1}{4})$; a ratio correction with jets (c_0^*, c_1^*, c_2^*) has a quadratic table of zero symbol exactly when $c_0^* = 2c_1^*$, and the only such correction with which J's charge passes the cubic test at W_1 is $(1, \frac{1}{2}, \frac{1}{4})$, for example $\big[a_0 + \frac{1}{2}(A-a_0) + \frac{1}{4}(A-a_0)^2/a_0\big]\delta$; J plus it has Regge's jets $(1, 1, 0)$, so the halved coupling is lost;

(c) the hinge-local correction $C^* + \frac{1}{2}Y$, with C^* the ratio correction of (b) and Y the coupling direction of Proposition 27.70, has a quadratic table of zero symbol and the cubic form $d + \frac{1}{2}r$; J's charge plus it has the cubic form $-\frac{1}{2}r$, exactly one half of Regge's form $-r$, so its cubic vertex is one half of Regge's with Δ removed, and it passes the cubic test at W_1, W_2 and W_3.

Proof. (a) A ratio charge has the per-simplex cubic table $c_0(Y_1 + 2X_1) - 3c_1(X_1 + Y_1) - 6c_2 Z_1$ with $X_1 = \sum_t A_t' \theta_t''$, $Y_1 = \sum_t A_t'' \theta_t'$ and $Z_1 = \sum_t A_t'^2 \theta_t' / A_t$, by Leibniz's rule and the twice differentiated Schläfli identity. At W_1 the three vertices are $\mathscr{V}(X_1) = -\frac{112}{3} - \frac{160}{9}i$, $\mathscr{V}(Y_1) = \frac{56}{3} + \frac{80}{9}i$ and $\mathscr{V}(Z_1) = -\frac{40}{3} - \frac{80}{9}i$, exact sums over the 24 simplices; so the witness value is $(c_0 - c_1)(-56 - \frac{80}{3}i) + c_2(80 + \frac{160}{3}i)$, whose imaginary part vanishes exactly when $c_0 - c_1 = 2c_2$, and then its real part is $-32c_2$. (b) $J_{\mathrm{odd}}(x, e^{\delta}) = \frac{1}{2}(x - x^{-1})\sinh\delta$ (Proposition 27.16(b)), whose ratio part $\psi_J(x) = \frac{1}{4}(x - x^{-1})$ has $\psi_J(1) = 0$, $\psi_J'(1) = \frac{1}{2}$, $\psi_J''(1) = -\frac{1}{2}$; the rest, $\psi_J(\sinh\delta - \delta)$, enters neither the quadratic nor the cubic action. The zero-symbol condition is the direction v_1 of Theorem 27.66(a), and (a) applied to $(c_0^*, c_1^* + \frac{1}{2}, c_2^* - \frac{1}{4})$ with $c_0^* = 2c_1^*$ gives $(1, \frac{1}{2}, \frac{1}{4})$. (c) In the sign of the cost forms Regge's charge has the cubic form $-r$ and J's the form $-(r + d)$ (by (27.1)); C^* has the form d, since J's charge plus C^* has Regge's jets, and Y has no quadratic part and the form r, minus Regge's (Proposition 27.70). So $C^* + \frac{1}{2}Y$ has the form $d + \frac{1}{2}r$, and J's form plus it is $-\frac{1}{2}r$, whose vertex is $-\frac{1}{2}V_r = -\frac{1}{4}W_3$ by Theorem 27.68, one half of Regge's $-\frac{1}{2}W_3$. The witness values are exact evaluations. □

The coupling direction behind part (c) is a change of Regge's coupling, and it is a hinge cost to first order.

Proposition 27.70 (The coupling direction)**.** *In classical mathematics, let L_0 be a reference simplex, L a nondegenerate simplex near it, $\Psi_{L_0}(L) = \sum_t A_t(L)(\theta_t(L) - \theta_t(L_0))$ and $\Psi_Y(L) = \sum_t \big(A_t(L) + L_0 \cdot \nabla A_t(L)\big)\big(\theta_t(L) - \theta_t(L_0)\big)$.*

(a) $\dfrac{d}{dg}\Big[g^{-2}\Psi_{L_0}\big(L_0 + g(L - L_0)\big)\Big]_{g=1} = -\Psi_Y(L)$. Hence the hinge cost $Y = \sum_h (A_h + L_0 \cdot \nabla A_h)\delta_h$, with L_0 the flat squared lengths of each hinge, has an identically zero quadratic table and minus Regge's cubic table: $S_{\mathrm{R}} + (1 - \nu)Y$ has Regge's quadratic action and ν times Regge's coupling through cubic order.

(b) The second derivative at $g = 1$ of $g^{-2}\Psi_{L_0}(L_0 + g(L - L_0))$ is a hinge cost plus the simplex-local product $\sum_t (D_{L_0} A_t)(D_{L_0}\theta_t)$, with D_{L_0} the derivative in the direction L_0 at L.

Proof. (a) Put $F(g) = \Psi_{L_0}(L_0 + g\ell)$, $\ell = L - L_0$. Then $\frac{d}{dg}[g^{-2}F]_{g=1} = F'(1) - 2F(1)$ and $F'(1) = \sum_t [A_t'(\theta_t - \theta_t(L_0)) + A_t\theta_t']$; Schläfli's identity removes $\sum_t A_t\theta_t'$. Heron's area is homogeneous of degree one in the squared lengths, so $L \cdot \nabla A_t = A_t$ by Euler's identity and $A_t' = A_t - L_0 \cdot \nabla A_t$; hence $F'(1) - 2F(1) = -\Psi_Y(L)$. Summed over the simplices at $L_0 = \ell_\kappa$, the function $\nu \mapsto \nu^{-2} S_{\mathrm{R}}(\ell_\kappa + \nu\lambda) = \sum_n \nu^{n-2} S_n$ has derivative $-Y$ at $\nu = 1$, so $Y_n = -(n-2)S_n$: zero at $n = 2$ and $-S_3$ at $n = 3$. (b) Differentiating once more, Euler's identity for the homogeneous functions A_t and ∇A_t turns every term but one into a function of the triangle's squared lengths times $\theta_t(L) - \theta_t(L_0)$, a hinge cost; the remaining term is $\sum_t (D_{L_0} A_t)(D_{L_0}\theta_t)$. □

So Y is exactly as consistent as Regge's action to first order in the change of coupling, and part (b) shows why it stops there: the second-order term contains a product of two rates on the same simplex, which no hinge cost has. The next section proves that at quartic order no hinge shape cost makes up the difference.

27.13 Hinge shape costs cannot complete a changed coupling

Rescaling every departure from flat space by ν and the action by ν^{-2} gives $\nu^{-2} S_{\mathrm{R}}(\ell_\kappa + \nu\lambda) = S_2 + \nu S_3 + \nu^2 S_4 + \dots$, Regge's action with ν times its coupling. The hinge cost Y gives $S_{\mathrm{R}} + (1-\nu)Y = S_2 + \nu S_3 + (2\nu - 1)S_4 + \dots$, since $Y_n = -(n-2)S_n$, which falls short at quartic order by $(1-\nu)^2 S_4$; at $\nu = \frac{1}{2}$, the coupling that J's charge carries, the shortfall is one quarter of Regge's quartic part. Can hinge costs supply it? A *shape cost* gives each hinge h of class c the cost $f_c(L_h)\delta_h$, with L_h its three squared lengths and one smooth function f_c for each of the six classes $(|a|, |b|) \in \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$ of hinges with vertices y, $y + a$, $y + a + b$. The values, gradients and Hessians of the f_c fix the quadratic and cubic parts; their third derivatives first enter at quartic order, through sixty *third-derivative directions* (c, j), one for each class and each of the ten cubic monomials in the three changes of squared length. Let $F_{c,j}(\kappa)$ be their per-simplex quartic tables divided by $\frac{1}{2}\sqrt{|1 - 4\kappa|}$, the normalization in which Regge's quartic part per simplex is $-\frac{1}{12} T_4(\kappa)$; they are rational in κ. A *physical configuration* on $\gamma = g_\kappa^{-1}$ has three gravitons, K_i null for γ and H_i symmetric, transverse and traceless for γ $(i = 1, 2, 3)$, and one pure-gauge leg $H_4 = K_4 \xi^{\mathsf{T}} + \xi K_4^{\mathsf{T}}$ with $K_4 = -K_1 - K_2 - K_3$ null.

Theorem 27.71 (The quartic obstruction). *In classical mathematics, let $\kappa \in \{0, \frac{1}{10}, 1\}$.*

(a) No complex numbers $z_{c,j}$ satisfy $\sum_{c,j} z_{c,j} \mathscr{V}[F_{c,j}(\kappa)](H; K) = \mathscr{V}[-\frac{1}{48} T_4(\kappa)](H; K)$ on every physical configuration.

(b) At $\kappa = 1$ and $\kappa = \frac{1}{10}$ the same holds with the lattice's five non-metric edge modes included, in the corrected gauge variation of each table (the variation with the gravitons' non-metric admixtures and the gauge leg's non-metric part, which solve the lattice's field equation to the needed order); the corrected variation does not depend on how the admixtures and the gauge part are chosen, and at $\kappa = 1$ it is the leading Taylor coefficient, of third order in the momenta, of the exact lattice vertex.

Conditions and scope.

Base and premises
The Kuhn lattice at $\kappa = 0, \frac{1}{10}, 1$; classical algebra over the Gaussian rationals and the integers.

Conclusion
The space of the sixty shape variations is linear, so no nonzero multiple of Regge's quartic variation lies in it: no change of the coupling, $\nu \neq 1$, is completed by shape costs at quartic order.

Inference
Reading (a) as the impossibility of a changed coupling uses one stated inference, which this book does not prove: that completing the coupling requires exactly the quartic shortfall in the gauge variation on physical configurations. The preprint argues it [196, Remark 10.4].

Exact computation
Each claim is an exact rank certificate on 17 or 33 explicit configurations, described in the proof and not printed. The quartic tables and the sixty shape-cost tables are parts S1 and S2 of the data set [210]; its part S8 holds the configurations and certificates of (a) at $\kappa = 0, \frac{1}{10}, 1$ and a program that recomputes every vertex and checks $y^{\mathsf{T}} M = 0$ and $y^{\mathsf{T}} t \neq 0$ exactly.

Not claimed
Costs nonlinear in the deficit, or depending on more than a hinge's three squared lengths.

Proof. Each claim is an exact certificate. Choose finitely many physical configurations, evaluate the sixty-one vertices exactly, and form the matrix M with one column per direction and the target column t. An integer vector y with $y^{\mathsf{T}} M = 0$ and $y^{\mathsf{T}} t \neq 0$ excludes every combination with real coefficients, since applying y to the supposed equalities gives $0 = y^{\mathsf{T}} t$. At $\kappa = 0$ and $\kappa = \frac{1}{10}$ the metric is Euclidean, null momenta are complex, and 17 configurations with Gaussian-integer entries suffice; each contributes a real and an imaginary row, and the sixty columns have rank 32, raised to 33 by the target. Complex coefficients are excluded through real ones: the complex conjugate of a physical configuration is physical and the tables are rational, so the real part of a complex solution would be a real solution. At $\kappa = 1$ real null momenta exist and 33 real integer configurations are used. For (b) the admixtures are computed exactly from the 15×15 form of the lattice at zero momentum, which is nondegenerate on the five non-metric classes (Theorem 27.74(e)), and the same kind of certificate excludes the target. □

At $\kappa = 0$ the lattice has one further mode with no quadratic term, the body diagonal of the cube.

Proposition 27.72 (The body diagonal at the Euclidean background)**.** *In classical mathematics, at* $\kappa = 0$*: every triangle of a Kuhn simplex that contains the body diagonal* 04 *is right-angled opposite it, so the first derivatives of the areas along the diagonal, the row and column* 04 *of the Euclidean table E and the true second derivatives of the simplex function along the diagonal vanish; the edge field with amplitude one on the class* 1111 *is a zero mode of the plane-wave form*

of E at every momentum, and it is no gauge amplitude at any wave with every $\sin(K_a/2) \neq 0$. *Two constant metric strains do not couple to it, while at an explicit on-shell configuration two graviton waves give it a nonzero source at long wavelength. At $\kappa = 1$ the row 04 of R_1 is not zero.*

Proof. The triangle $0j4$ has squared sides j, $4-j$ and 4, and $j+(4-j) = 4$: by Thales it is right-angled at v_j, since the steps $v_j - v_0$ and $v_4 - v_j$ have disjoint supports. As $\partial A/\partial z = (x+y-z)/(16A)$ for squared sides x, y, z, the areas are stationary in ℓ_{04}, and by Schläfli's identity the Hessian $\sum_t \partial A_t \, \partial\theta_t$ has a zero row and column 04. The body diagonal is slot 04 of every simplex and of no other slot, so its column of the plane-wave form vanishes. A gauge amplitude vanishing on the four axis classes has $\eta_a \sin(K_a/2) = 0$, hence is zero. The source values are exact sums over the 24 simplices with the cubic table r. Finally $(R_1)_{04,04} = 2$ (Table 27.1). □

So the flat $\kappa = 0$ lattice is a degenerate critical point on the fifteen-class space, and statements there about four-point amplitudes are statements about the metric truncation. At $\kappa = 1$ and, as the next section shows, at every other admissible background the five non-metric modes are massive.

27.14 Four gravitons at tree level

Once the quadratic form, the cubic vertex and the quartic vertex are fixed, the tree-level scattering of four gravitons is fixed. Since the lattice's vertices are one half of Einstein's, its amplitude should be general relativity's at one half of Einstein's normalization, gauge invariant and with general relativity's helicity structure. This section proves that, on every flat background g_κ of the lattice family off its poles, at long wavelength (second order in the momenta), for the amplitude built from the tables and for the one built from the Lorentzian lattice's own cell forms. The backgrounds g_κ are related by linear changes of coordinates in the continuum theory, and the proofs use that. All statements are proved in classical mathematics over the complex numbers.

Definition 27.73 (The tree four-graviton amplitude)**.** Fix a real κ with $(2\kappa-1)(3\kappa-1)(4\kappa-1) \neq 0$, $\gamma = g_\kappa^{-1}$, $g = g_\kappa$ and $P^2 = P^{\mathsf{T}} \gamma P$. For four legs (H_i, K_i), H_i complex symmetric and $\sum K_i = 0$: the *contact term* is $\mathscr{V}[T_4(\kappa)](H;K)$; the *current* J^{ij} of legs i, j is the matrix with $\mathscr{V}[T_3(\kappa)](H_i, H_j, Z; K_i, K_j, -K_i - K_j) = \sum_{ab} J^{ij}_{ab} Z_{ab}$ for every symmetric Z; the *propagator* of a current is $\Pi(J;P) = \frac{4}{P^2}\big(gJg - \frac{1}{2}(g:J)\,g\big)$, $g:J = \sum g_{ab}J_{ab}$ (the gauge-fixed inverse of the quadratic vertex in de Donder gauge); and the *amplitude* is

$$\mathscr{A}(H;K) = \mathscr{V}[T_4(\kappa)](H;K) + E(J^{12}, J^{34}; K_1+K_2) + E(J^{14}, J^{23}; K_1+K_4) + E(J^{13}, J^{24}; K_1+K_3),$$

$E(J, J'; P) = \sum_{ab} J_{ab} \Pi(J'; P)_{ab}$, defined when $s = (K_1+K_2)^2$, $t = (K_1+K_4)^2$ and $u = (K_1+K_3)^2$ are nonzero.

The exchange is the stationary value of the channel's gauge-fixed quadratic-plus-cubic action, so contact and exchanges enter with weight one, and the amplitude is the coefficient of $\epsilon_1 \cdots \epsilon_4$ in the tree-level effective action of one cell of the metric sector at second order in the momenta. On the Euclidean lattice the null momenta are complex and spinors describe them. For $a, b \in \mathbb{C}^2$, $v(a,b) \in \mathbb{C}^4$ is the vector y whose matrix $\left(\begin{smallmatrix} y_4+iy_3 & y_2+iy_1 \\ iy_1-y_2 & y_4-iy_3 \end{smallmatrix}\right)$ is ab^{T}; the

brackets are $\langle ac\rangle = a_1c_2 - a_2c_1$ and $[bd] = b_1d_2 - b_2d_1$, so that $v(a,b)\cdot v(c,d) = \frac{1}{2}\langle ac\rangle[bd]$. Leg i has $K_i = v(\lambda_i, \tilde{\lambda}_i)$; with reference spinors μ_i, $\tilde{\mu}_i$ its polarizations are $\epsilon_i^+ = v(\mu_i, \tilde{\lambda}_i)/\langle \mu_i\lambda_i\rangle$ and $\epsilon_i^- = v(\lambda_i, \tilde{\mu}_i)/[\tilde{\lambda}_i\tilde{\mu}_i]$, the graviton of helicity $\pm$ is $H_i = \epsilon_i^{\pm} \otimes \epsilon_i^{\pm}$, and $s = \langle 12\rangle[12]$, $t = \langle 14\rangle[14]$, $u = \langle 13\rangle[13]$. These are the conventions of [196, Section 9.3] and of part S7 of the data set [210]. The data are *generic* when momentum is conserved, $s, t, u \neq 0$, and every reference bracket is nonzero.

Theorem 27.74 (Gauge invariance, helicity selection and the MHV amplitude)**.** *In classical mathematics:*

(a) Noether identities. *For every invertible complex symmetric γ, Einstein's vertices obey the order-by-order identities of invariance under a change of coordinates: W_2 vanishes on a pure-gauge leg, and W_3, W_4 with one pure-gauge leg equal minus the lower vertex applied to the linear gauge variations of the other legs.*

(b) Gauge invariance. *For every real κ off the poles, the amplitude vanishes on every physical configuration with $s, t, u \neq 0$ (pure-gauge gravitons decouple); at $\kappa = 0$ and $\kappa = 1$ it does not depend on the gauge parameter of the propagator.*

(c) Helicity selection and MHV. *At $\kappa = 0$, for generic spinor data, the amplitude vanishes unless exactly two of the four helicities are positive, and if legs m, m' are negative and p, p' positive,*

$$\mathcal{A} = -\frac{1}{256}\,\frac{\langle mm'\rangle^4\,[pp']^4}{stu}.$$

The same selection rule and the same MHV value hold at every real κ off the poles, for spinor data carried to the background g_κ by a complex frame of g_κ.

(d) The lattice's own amplitude. *For every real κ with $(2\kappa-1)(3\kappa-1)(4\kappa-1) \neq 0$ the lattice's own long-wave quadratic form is one half of Einstein's, so the amplitude built from the lattice's own propagator is the amplitude above: general relativity's tree amplitude at one half of Einstein's normalization. For every $\kappa > \frac{1}{2}$ the amplitude built from the Lorentzian lattice's own Feynman rules, its cell forms of orders two, three and four, contact term included, is $\sqrt{4\kappa-1} = \sqrt{|\det g_\kappa|}$ times it, for all real legs whose channel momenta are off the cone; it is gauge invariant.*

(e) The other five edge modes. *For every $\kappa \notin \{0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$ the five non-metric edge modes (the four classes with three steps and the class 1111) are nondegenerate at zero momentum; eliminating them at tree level is solvable in every channel with nonzero invariant, and the amplitude with all fifteen edge classes on the internal lines and in the legs equals the amplitude of the metric sector above, so it is Einstein's and gauge invariant. At $\kappa = 0$ the decoupling fails (Proposition 27.72).*

Conditions and scope.

Base and premises

The Kuhn lattice on every flat background g_κ off the poles; complex legs; classical algebra.

Conclusion

The lattice's gravitons scatter at tree level and long wavelength (second order in the momenta) exactly as general relativity's do, on every flat background g_κ of the family off the poles, the Lorentzian $\kappa = 1$ lattice included; and for $\kappa < \frac{1}{4}$ (Theorem 27.59) and $\kappa > \frac{1}{2}$ (part (d)), with the lattice's own Feynman rules the only change is the volume factor $\sqrt{|\det g_\kappa|}$, the same at every order.

Exact computation

The proofs end in finite computations that they describe and do not reproduce: the 40, 400 and 4000 polynomial identities of (a), the canonical MHV value $-\frac{1}{256}$ of (c), the contact-term identity of (d) and the determinant and Schur complement of (e). The tables are part S1 of the data set [210]. Its part S7 holds programs that check (a) on every basis case at three backgrounds, (b) at $\kappa = 0$ and $\kappa = 1$, and (c) at $\kappa = 0$ and $\kappa = 1$ exactly, with the spinor conventions of Definition 27.73; its part S6 holds the fifteen-class quadratic form and its determinant used in (e). The contact-term identity of (d) is not in the data set.

Not claimed

Loops; amplitudes at higher order in the momenta (Section 27.15).

Proof. (a) At $\gamma = 1$ each identity is linear in every polarization and in the gauge vector and polynomial in the momenta, so it holds once it holds for basis polarizations and unit gauge vectors with the last momentum eliminated: 40, 400 and 4000 polynomial identities of degree three, checked by exact expansion. Every invertible complex symmetric γ is $E^{-\mathsf{T}}E^{-1}$ for an invertible complex E, and the vertices transform covariantly under the change of coordinates E, which carries the identities from $\gamma = 1$ to γ. (b) The tables' vertices are one half of Einstein's at every κ off the poles, at all three orders (Theorem 27.61(a)–(c)); the Ward reduction expresses the amplitude with a gauge leg through the identities of (a), and every term cancels once the gravitons are on shell, the gauge-fixed form annihilates the external legs (K_4 null) and the propagator solves the gauge-fixed equation. Independence of the gauge parameter is the same cancellation with the gauge term varied. (c) The amplitude does not depend on the reference spinors, since changing a reference spinor adds a pure-gauge part to that leg and (b) applies. Choosing the references adapted to the helicities makes every non-MHV configuration vanish term by term after the spinor identities, and reduces the MHV case to one canonical configuration, where the value is computed exactly; covariance under the action of $\mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C})$ on the spinors carries it to all generic data, and the complex change of coordinates of (a) carries it to every background. (d) The lattice's quadratic vertex is the identity of Theorem 27.61(c). For the Lorentzian cell forms, Proposition 27.63 gives each of the three vertices as $\sqrt{4\kappa - 1}$ times the tables' vertex; in the amplitude the propagator carries $(\sqrt{4\kappa - 1})^{-1}$ and each exchange two cubic vertices, so each exchange and the contact term scale by $\sqrt{4\kappa - 1}$. The contact term needs one algebraic identity in κ and the slot variables, proved by exact computation. (e) On the five classes the zero-momentum form has determinant $3\kappa(2\kappa - 1)^3/(16(4\kappa - 1)^4)$, defined for $\kappa \neq \frac{1}{4}$ and nonzero exactly off $\{0, \frac{1}{2}\}$; at $\kappa = \frac{1}{3}$ it is $-\frac{3}{16}$, and $\kappa = \frac{1}{3}$ is excluded as a pole of the tables, not as a zero of the determinant. A Schur complement onto the metric sector then changes the quadratic, cubic and quartic terms in the metric fields only at fourth order in the momenta, because the zero-momentum form annihilates the metric fields and a cell vertex with all legs metric but one is of second order

in the momenta; so the fifteen-class amplitude at second order in the momenta is the metric one. □

27.15 The graviton's dispersion beyond long wavelength

At order K^2 every lattice of the family is one half of Einstein's theory on its background g_κ, so its graviton travels on the light cone $P^{\mathsf T} g_\kappa^{-1} P = 0$. The next order decides whether the lattice keeps Lorentz invariance at all and whether the two polarizations travel alike. Real light-cone directions exist only on the Lorentzian backgrounds $\kappa > \frac{1}{4}$, where g_κ has the one negative eigenvalue $1 - 4\kappa$. Take a real null covector $P \neq 0$ of g_κ, a direction Q with $g_\kappa^{-1}(P,Q) \neq 0$ and the momentum $K(s) = sP + s^3 tQ$. A *non-gauge wave through order* s^4 at (P,Q,t) is a symmetric polarization H that is not a pure-gauge polarization at P, with edge amplitude $\Phi(H) + s^2 a_2 + s^4 a_4$ on all fifteen classes (a_2, a_4 arbitrary, the non-metric admixtures included), whose plane-wave form of the background's table $R(\kappa)$ against every test amplitude vanishes at orders s^2 and s^4. Along $K(s)$ the dispersion is $K(s)^{\mathsf T} g_\kappa^{-1} K(s) = s^4 \mu + O(s^6)$ with $\mu = 2t\, g_\kappa^{-1}(P,Q)$, the *branch value*: Lorentz invariance at this order means $\mu = 0$, and no birefringence means one value for both polarizations. With $p_k = \sum_a P_a^k$ put

$$
\begin{aligned}
A_\kappa(P) &= f_1(\kappa)\, p_4 + f_2(\kappa)\, p_1 p_3 + f_3(\kappa)\, p_1^4, \qquad \operatorname{disc}_\kappa(P) = f_2(\kappa)^2\, B_\kappa(P),\\
f_1(\kappa) &= \frac{60\kappa^3 - 40\kappa^2 + 10\kappa - 1}{12(2\kappa - 1)}, \qquad f_2(\kappa) = -\frac{\kappa(3\kappa - 1)(5\kappa - 1)}{3(2\kappa - 1)},\\
f_3(\kappa) &= \frac{\kappa^3(45\kappa^2 - 22\kappa + 2)}{12(2\kappa - 1)(4\kappa - 1)^2},\\
B_\kappa(P) &= p_4^2 - 2 p_1 p_3 p_4 + \frac{4(3\kappa - 1)}{3(4\kappa - 1)} p_1^2 p_3^2 + \frac{6\kappa^2 - 1}{4(4\kappa - 1)^2} p_1^4 p_4\\
&\quad - \frac{9\kappa^2 - \kappa - 2}{6(4\kappa - 1)^2} p_1^5 p_3 + \frac{27\kappa^4 - 4\kappa^3 - 12\kappa^2 + 1}{48(4\kappa - 1)^4} p_1^8 .
\end{aligned}
$$

At $\kappa = 1$ these are

$$
a(P) = \tfrac{29}{12} p_4 - \tfrac{8}{3} p_1 p_3 + \tfrac{25}{108} p_1^4, \quad \operatorname{disc}(P) = \tfrac{16}{729} p_1^8 - \tfrac{64}{81} p_1^5 p_3 + \tfrac{80}{81} p_1^4 p_4 + \tfrac{512}{81} p_1^2 p_3^2 - \tfrac{128}{9} p_1 p_3 p_4 + \tfrac{64}{9} p_4^2 .
$$

Theorem 27.75 (The graviton's order-K^4 dispersion on the lattice family)**.** *In classical mathematics:*

(a) for every real $\kappa \notin \{0, \frac{1}{4}, \frac{1}{2}\}$, every real null $P \neq 0$ of g_κ, every real Q with $g_\kappa^{-1}(P,Q) \neq 0$ and every complex t, a non-gauge wave through order s^4 exists at (P,Q,t) exactly when $\mu = 2t\, g_\kappa^{-1}(P,Q)$ is a root of $\mu^2 - 2A_\kappa(P)\mu + A_\kappa(P)^2 - \operatorname{disc}_\kappa(P)$, that is, $\mu = A_\kappa(P) \pm \sqrt{\operatorname{disc}_\kappa(P)}$. The two branch values differ by $2 f_2(\kappa)\sqrt{B_\kappa(P)}$, up to sign, so the polarization splitting is proportional to $\kappa(3\kappa - 1)(5\kappa - 1)/(2\kappa - 1)$ times a function of the direction; at $\kappa = \frac{1}{3}$ the two values coincide in every direction;

(b) at $\kappa = 1$, where $A_1 = a$ and $\operatorname{disc}_1 = \operatorname{disc}$: along $P = (1,4,1,3)$ with $Q = e_1$ ($g_1^{-1}(P,Q) = 1$) such a wave exists exactly for $t = 109$ and $t = -3$, so the branch values are 218 and -6, the two polarizations disperse differently and neither follows the light cone at this order; along the three-fold axis $(1,1,1,0)$ the two share one value; and the odd orders of the expansion vanish;

(c) every background with $\kappa > \frac{1}{2}$ is birefringent at order K^4: for $0 < x < \sqrt{2}$ the covector $P(x) = (1 + x, 1 - x, 1, 1)$ is null for g_κ at $\kappa = \frac{1}{4} + 1/(2x^2)$, the direction $Q = (1, 1, 1, 1)$ has $g_\kappa^{-1}(P(x), Q) = -2x^2$, and non-gauge waves through order s^4 exist at the two different shifts

$$t_1 = -\frac{x^6 + 3x^4 - 34x^2 - 48}{24x^2(x^2 - 2)}, \qquad t_2 = -\frac{5x^6 + 6x^4 - 116x^2 - 24}{192x^2(x^2 - 2)};$$

as x runs over $(0, \sqrt{2})$, κ runs over every value above $\frac{1}{2}$.

At the background $\kappa = \frac{3}{4}$, the point $x = 1$ of (c), the branch values along $(2, 0, 1, 1)$ are 13 and $\frac{43}{16}$.

Conditions and scope.

Base and premises

The linearized lattice action of the background g_κ (the continued Regge table $R(\kappa)$) with all fifteen edge classes and its gauge modes; classical algebra over $\mathbb{C}$.

Conclusion

The long-wave results of the earlier sections are unchanged: they concern order K^2. At order K^4, which on a lattice of spacing a is order k^2a^2, the two branch values of every Lorentzian background off $\kappa = \frac{1}{2}$ are $A_\kappa \pm \sqrt{\mathrm{disc}_\kappa}$; every background with $\kappa > \frac{1}{2}$, the $\kappa = 1$ lattice included, is birefringent and not Lorentz invariant; the $\kappa = 1$ lattice is anisotropic and birefringent wherever disc $\neq 0$; and at $\kappa = \frac{1}{3}$ the two polarizations travel alike at this order.

Exact computation

The identities behind (a) and (c) are polynomial identities, in the direction after the normalization of the proof and in x, checked by exact computation as the proof describes; they are not reproduced here. The data: the table $R(\kappa)$ is this book's (Theorem 27.28); the fifteen-class quadratic form at zero momentum is part S6 of the data set [210]; the metric polarizations of the two waves of (c) are printed in the proof. Their admixtures, and the computation of (a) and (c), are not printed and are not part of that data set. Part (b) is checked by hand in the proof from the closed forms.

Proof. (a) The branch does not change under $(P, Q) \mapsto (cP, c^3Q)$, $c \neq 0$. On the cone $p_2 = \kappa p_1^2/(4\kappa - 1)$, so $p_1 \neq 0$ when $P \neq 0$; take $p_1 = 4$ and write $P = \mathbf{1} + w$ with $\sum_a w_a = 0$, so that the cone is $|w|^2 = 4/(4\kappa - 1)$. Every Taylor part of the plane-wave form of $R(\kappa)$ is affine in $1/(\kappa - \frac{1}{4})$ (Theorem 27.28(b)), so with $|w|^2$ in its place every identity below is a polynomial identity in w. The order-s^2 equation is the admixture equation: the zero-momentum form on the five non-metric classes is invertible for $\kappa \notin \{0, \frac{1}{4}, \frac{1}{2}\}$, its inverse being polynomial after one scalar factor, and the metric leg is on shell, since the order-K^2 form is $-\frac{1}{4}$ times the Fierz–Pauli form of g_κ. Projecting the order-s^4 equation onto the metric sector by the Schur complement leaves a linear condition on the polarization modulo gauge, the gauge modes decoupling; a non-gauge solution exists exactly when the determinant of that condition on the two physical polarizations vanishes, and that determinant is a nonzero multiple of $\mu^2 - 2A_\kappa\mu + A_\kappa^2 - \mathrm{disc}_\kappa$ at $\mu = 2t\, g_\kappa^{-1}(P, Q)$. These are polynomial identities, proved by exact computation. At $\kappa = \frac{1}{3}$, $f_2 = 0$ and $\mathrm{disc}_\kappa = 0$. (b) At $\kappa = 1$ the odd Taylor orders of the plane-wave form vanish by the

central inversion. At $P = (1, 4, 1, 3)$, $p_1 = 9$ and $p_2 = 27$, so P is null; $p_3 = 93$ and $p_4 = 339$ give $a(P) = 106$ and $\mathrm{disc}(P) = 12544 = 112^2$, so the roots are 218 and -6; and the background identity $K(s)^{\mathsf{T}} g_1^{-1} K(s) = 2ts^4 + \frac{2}{3}t^2 s^6$ is direct. (c) Here $p_1 = 4$ and $p_2 = 4 + 2x^2$, which is $\kappa p_1^2/(4\kappa - 1)$ at $\kappa = \frac{1}{4} + 1/(2x^2)$, so $P(x)$ is null; and $g_\kappa^{-1}(P, Q) = p_1/(1 - 4\kappa) = -2x^2$. The waves at t_1 and t_2 are explicit. Their metric polarizations are

$$H_1(x) = \begin{pmatrix} -(x^2+3x+2) & x^2-2 & -(x+2) & 0 \\ x^2-2 & -(x^2-3x+2) & x-2 & 0 \\ -(x+2) & x-2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$H_2(x) = \begin{pmatrix} x^2-2 & -(x^2+2) & 4(x+1) & 0 \\ -(x^2+2) & x^2-2 & 4(1-x) & 0 \\ 4(x+1) & 4(1-x) & 8 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and neither is pure gauge at $P(x)$: a polarization $P \otimes \xi + \xi \otimes P$ whose fourth row vanishes has $2\xi_4 = 0$ and then $\xi_a = 0$ for every a, since $P_4 = 1$, so it is zero, and $H_1(x)$, $H_2(x)$ are not ($8 \neq 0$ and $x^2 - 2 \neq 0$). Their admixtures on the fifteen edge classes, at orders s^2 and s^4, are polynomials in x over the common denominator $192x^2(x-2)(x^2-2)^2$; checking the waves is a polynomial identity in x, proved by exact computation. Finally $t_1 - t_2 = -(x^6+6x^4-52x^2-120)/(64x^2(x^2-2))$, and for $0 < x^2 < 2$ the numerator is below $8+24-120 < 0$, so $t_1 \neq t_2$. At $x = 1$, $\mu = -4t$ gives 13 and $\frac{43}{16}$. □

In plain terms: on every lattice of the family with $\kappa > \frac{1}{2}$ the two graviton polarizations travel at slightly different speeds at wavelengths approaching the lattice spacing, and on the $\kappa = 1$ lattice the speed depends on the direction. Along $(1, 4, 1, 3)$ one polarization is subluminal and the other superluminal at this order. At $\kappa = \frac{1}{3}$ the two travel alike at this order. The guess that the exact branch is the zero set of one edge-class Laplacian is excluded, because such a branch has one value for both polarizations.

27.16 The premises of reading the cost as the graviton's action

Section 27.10 named the premises that turn J's comparison into the graviton's charge: tick-free least postings, the coherent reading, independent axes (the statistical face reading), locality and $\kappa = 1$. None of them follows from the accepted foundation: each is either separated from it by an explicit model of the foundation's facts in which the premise fails, or reduced to one sharper named reading.

Proposition 27.76 (Where the named premises stand)**.** *In classical mathematics:*

(a) Tick-free writing. *No accepted fact makes the ledger write a step by its cheapest record: counting the extra comparisons a longer writing makes orders writings by their number of postings and so restates least writing; every writing of a step leaves the same record for a recognition chain to read, which chooses nothing; and a world that writes the three-counter step with three postings meets the foundation's facts. One structural reading forces the count* $(1, 2, 1)$ *without minimizing: if each step is one comparison event that writes at most one unit on each of its two parties' records, charged per unit, parity and the tick-free lower bound leave exactly one, two and one postings.*

(b) The coherent reading. *With a credit read as lowering its counter (a signed balance, an identification), double entry together with the reading that the charge reads a strain only through what comparison events post forces the charge to read only spatial strains, and so gives it the face-frame parity; neither premise alone does. For such charges the coherent reading is exactly no birefringence at the face wave.*

(c) Locality. *If the derived scalar record is a ledger balance and its updates are double entry, it cannot carry the gap that would force Einstein's trace coefficient; every value of the coefficient, Einstein's included, is reached by a double-entry record, so nothing is selected. Double entry itself is a separate premise.*

(d) $\kappa = 1$. *The readings of the ledger's records into the Kuhn lattice (the sixteen-record cell, whose histories are the Kuhn simplices, and the posting-lattice reading) are positive with both posts in their domain at every* $\kappa \geq 1$*, and a completely positive writing map onto the posts' Kuhn records exists in every record world; so these readings do not give* $\kappa = 1$*. In the reading direction, on the plane of the two posting images, the only linear recovery is positive exactly when* $\kappa = 1$*, which is the purity of each post's lattice record (Theorem 27.42).*

(e) Independent axes. *In the* $\kappa = 1$ *record geometry the Born rule is the coherent carrier's posting law, and reading a strain through the same rule moves each face direction's record only with its own stretch, the sign set by the probing record. Which records probe the strain is not fixed by the accepted results: the unbiased reference state sees no face stretch, single postings couple the face directions, and only the records of the double steps charge the three face directions independently. The statistical face reading is therefore separated, and reading it through the double steps is a named statement that restates face-axis additivity.*

Conditions and scope.

Base and premises
The foundation's accepted facts and the named readings; finite linear algebra; classical mathematics.

Route given
Each separation rests on an explicit model that the proof names and describes but does not print in full: the three-posting world of (a), the double-entry records of (c) that reach every value of the coefficient, the writing maps of (d) with their Choi matrices, and the probes of (e). The reductions are finite linear algebra on the four counters, the six transfers and the fifteen steps, described in the proof.

Proof. Each separation exhibits an explicit model of the accepted facts (for (a), a world of the foundation's interpreter that meets the three facts of Proposition 27.45(a)) in which the premise fails; each reduction is an equivalence between the premise and the named reading, proved by finite linear algebra on the four counters, the six transfers between accounts, or the fifteen steps. (a) A unit posting changes exactly three comparisons by one unit, so a writing with c postings makes $3c$ unit comparison changes, of which a step with k counters keeps $k(4-k)$; the three-posting world writes the triple step forward. (b) Double entry keeps the total, which makes a charge read an event and its reversal alike; with the posting reading this kills the time-space part. (c) A translation-invariant update keeps every total exactly when its zero-wavelength multiplier is one, and a gradient update conserves its total exactly when that multiplier's coefficient vanishes. (d) The writing maps are explicit, with Choi matrices checked positive semidefinite; the recovery on the plane of the two images is a 2×2 problem whose positivity is $\kappa = 1$. (e) The posting law of the strained geometry shifts by $\frac{1}{3} v^{\mathsf{T}} H e_i$ for a probe v, and the cross terms vanish exactly for the double-step probes. □

Two of these readings rest on double entry, the ledger's reversal symmetry, which is itself a premise about writing and not one of the foundation's three facts.

Remark 27.77 (The one-loop graviton)**.** The near-flat one-loop graviton of the $\kappa = 1$ lattice has been computed numerically, and those numbers are not theorems of this book. One tool behind them is: an adaptive integrator over the Brillouin zone with the long-wave corner cut out, built on $\mathbb{R}_\delta$ from interval arithmetic with δ's own π, cosine and sine, proved sound with no omniscience principle and no choice, together with the rule that decides the sign of a zone integral when the enclosure's error plus a bound on the corner is smaller than the enclosure's distance from zero. The enclosure the one-loop question needs has not been certified with it: near the corner the propagator's soft modes turn with the direction of the wave, which defeats every cell-by-cell interval rule of feasible size. The gravity preprint states what is established and what is open at one loop [197].

Questions to work through

1. (Theorem 27.3.) At $K = e_0$ take the strain $h = E_{11} - E_{22}$. Compute $J(1 + \varepsilon h)$ exactly, its long-wave limit and $\mathrm{FP}(h, e_0)$. Then add the gauge mode $E_{03} + E_{30}$ and compute both again.

2. (Theorem 27.5.) For $D = 3$ and $m = (1, -1, 0, 0)$, find two strains that span the TT strains at m.

3. (Theorem 27.24.) Check that k_1 and k_2 are static and null and that Ξ is a gauge mode at $k_1 + k_2$. Using the slot permutation $(0\,9)(1\,8)(2\,6)(4\,7)$ of the central inversion, check on the monomials 000, 444 and 044 of Table 27.2 that both cubic forms are invariant. Finally, explain why the test of part (d) cannot exclude a second-order static field when V_d is replaced by Regge's vertex V_r alone.

4. (Theorem 27.26.) At $\kappa = 1$ compute $\det G$, the four values of Q_t, and check the identity of part (c) for the triangle 012.

5. (Theorem 27.40.) Take $\kappa = \frac{5}{4}$ and a tick-normalized identification. Compute the squared Bloch length of a posting record, the largest probability with which a sharp test can pass it, and the depolarizing strength r.

6. (Theorem 27.44.) Show that the parity law of part (d) is not pair-reading at the site 0, by exhibiting, for each partner, two assignments that agree on the site and the partner and give different posts.

Part VIII

The library, learning and open problems

Part VIII prices the conventions of classical mathematics as it is used in practice, prices the act of learning a law from records, and states what remains open. Chapter 28 shows what the conventions of a classical library cost on $\mathbb{R}_\delta$: a total inverse with $0^{-1} = 0$ costs the LPO function, choosing a representative for every point costs the WLPO function, and a nonnegative derivative gives monotonicity at no more than that price. Chapter 29 treats time, chance and learning: the logic of one recorded history, exchangeable records and de Finetti's theorem, and the exact prices of learning a law in the limit, with finitely many errors and by committing once. Chapter 28 ends with the price table of the mathematics that physics uses, 1467 statements, of which 1166 are free, 279 have an exact price in the table's sense and 22 are open, listed in full in Appendix Q. Chapter 30 collects proved and refuted statements at the frontier and lists the open problems of the whole book exactly.

Chapter 28

The classical library on the line

A standard classical library of algebra, order and analysis divides by zero. Its inverse is a total map with $0^{-1} = 0$, so that $1/0 = 0$, $x/0 = 0$ and $\log 0 = 0$ are values like any other; its linear orders carry stored decisions of $\leq$, $=$ and $<$; its derivative is "the derivative where one exists and 0 elsewhere". These conventions make classical algebra smooth, and classically they cost nothing. What does one of them cost on $\mathbb{R}_\delta$?

Here is the worked case. Suppose i is a map on $\mathbb{R}_\delta$ with $i(0) = 0$ and $x \cdot i(x) = 1$ for every $x \# 0$. Take a binary sequence α and its jump point x_α (Lemma 12.4): $x_\alpha = 2^{-N}$ when N is the first index with $\alpha(N) = 1$, and $x_\alpha = 0$ when there is none. For $\alpha = 0, 0, 0, 1, 0, \ldots$ we get $x_\alpha = \frac{1}{8}$ and $x_\alpha \cdot i(x_\alpha) = \frac{1}{8} \cdot 8 = 1$; for the zero sequence we get $x_\alpha = 0$ and $x_\alpha \cdot i(x_\alpha) = 0$. In general we do not know which case we are in, yet the point $e = x_\alpha \cdot i(x_\alpha)$ is 0 or 1 all the same. The order is cotransitive: since $0 < 1$, either $0 < e$ or $e < 1$. If $0 < e$, then x_α has the inverse $i(x_\alpha)/e$, so $x_\alpha \# 0$ and $e = 1$. If $e < 1$, then $x_\alpha \# 0$ is impossible, because it would give $e = 1$; so $x_\alpha = 0$ and $e = 0$. Stage 7 of any regular sequence naming e lies within $\frac{1}{8}$ of e, so comparing it with $\frac{1}{2}$ reads off the bit, and the bit does not depend on the sequence used. The rule

$$d(\alpha) = \begin{cases} 1, & \text{stage 7 of } x_\alpha \cdot i(x_\alpha) \text{ exceeds } \frac{1}{2}, \\ 0, & \text{otherwise,} \end{cases}$$

decides for every binary sequence whether it ever takes the value 1. It is the LPO function of Definition 21.1: one map that answers the question LPO poses, for all sequences at once. Conversely, the LPO function decides apartness from 0 and so builds the inverse. The convention $0^{-1} = 0$ costs exactly the LPO function.

Not every convention is this expensive. Setting $\sqrt{x} = 0$ for $x \leq 0$ costs nothing, because the extended square root is continuous. The sign function, which one would predict to cost the LPO function as well, costs only the WLPO function: its jump at 0 does not reveal on which side of 0 a point lies. One rule gives a lower bound for every such price (Theorem 28.8), and Table 28.1 lists eighteen conventions with the exact price of each.

The chapter opens Part VIII by reading the rest of the classical library on $\mathbb{R}_\delta$ in the same way. A map that picks a regular sequence for every point, the classical library's choice of representatives, costs exactly the WLPO function (Section 28.3). The classical linear order costs exactly the WLPO function (Section 28.4); the ordered field costs the LPO function and the complete ordered field Markov's principle with a double-negation decision (Section 28.5). The derivative convention costs exactly Markov's principle with a decision of differentiability (Section 28.6), and monotonicity from a nonnegative pointwise derivative costs at most the WLPO function (Section 28.7). Two models show where these principles stop: over intuitionistic second-order arithmetic WLPO does not give Π^0_2 weak excluded middle (Section 28.8), and, in a calculus of streams, WLPO does not collect the signs of the tails of a signed-digit stream (Section 28.9). Section 28.10 proves two theorems that move whole areas of the library at no cost, by density and by algebraic certificates, and reports an audit of the library read on

$\mathbb{R}_\delta$; every count of library theorems or of their uses in this chapter comes from that audit and is not a theorem of the book. Finally, Section 28.11 adds what is new about programs from classical proofs beyond Section 20.4.

Throughout, maps are read as in Section 21.1: a map is a rule given as data, and a proof that every input has exactly one output does not by itself produce a map. "Free", "costs exactly", "at most" and "at least" have the meanings fixed there. Every theorem of this chapter is proved with no choice principle unless its statement names one; the two model constructions of Sections 28.8 and 28.9 reason in classical mathematics with the axiom of choice, and we say so where they begin. A ratio compared with a point stands for its point, as in Chapter 13, and we write 0 and 1 for $\eta(0)$ and $\eta(1)$.

28.1 Tests and decisions as data

Chapter 12 priced the decisions on the line as propositions: deciding $x = 0$ for every point is WLPO, deciding $x > 0$ or $x = 0$ or $x < 0$ is LPO, and a sign for every nonzero point is MP (Corollary 12.18). A classical library stores decisions as data. We record the data forms once and use them in every later section.

Definition 28.1 (Decisions stored as data)**.** Let R be a relation between points of $\mathbb{R}_\delta$. A *decision of R* is a map d from pairs of points to $\{0,1\}$ with $d(x,y) = 1$ exactly when $R(x,y)$. We speak of *decided equality*, a *decided order* $\le$ and a *decided strict order* $<$. A *floor map* is a map $\lfloor\cdot\rfloor$ from $\mathbb{R}_\delta$ to the integers with $\lfloor x\rfloor \le x < \lfloor x\rfloor + 1$ for every x. A map a from $\mathbb{R}_\delta$ to $\{0,1\}$ is a *positivity test* if $a(x) = 1$ exactly when $x > 0$.

A decision procedure for a family of propositions, stored as data, is the same thing as the data form of its principle. For the statements $\exists n\, \alpha(n) = 1$, one for each binary sequence α, it is the LPO function; for $\forall n\, \alpha(n) = 0$ it is the WLPO function; for the propositions $\neg\neg P$ it is the WEM function. A decision used at one fixed proposition inside a proof is different: it costs only excluded middle for that proposition.

Two facts about the jump point carry most reductions in this chapter. For a binary sequence α let x_α be the point of its positive jump presentation (Lemma 12.4, with every sign $+1$). Then

$$x_\alpha = 0 \iff \forall n\, \alpha(n) = 0, \qquad x_\alpha > 0 \iff \exists n\, \alpha(n) = 1, \qquad \neg(x_\alpha < 0). \tag{28.1}$$

The first is Lemma 12.5. For the second, an event at N makes every stage from N on equal to 2^{-N}, which exceeds $2/(n+1)$ at $n = 2^{N+2}$; conversely a positive stage is nonzero, so the bounded search behind it has found an event. The third holds because no stage is negative. The map $\alpha \mapsto x_\alpha$ is a rule: it computes the stages from α.

Lemma 28.2 (Tests descend to points)**.** *For a regular sequence s let $o_s(n) = 1$ when $|s_n| > 2/(n+1)$ and $p_s(n) = 1$ when $s_n > 2/(n+1)$, and 0 otherwise. If s names x, then $x = 0$ exactly when $o_s(n) = 0$ for every n, and $x > 0$ exactly when $p_s(n) = 1$ for some n. Consequently:*

1. *if w is the WLPO function, then $z(x) = w(o_s)$ is a zero test, and $w(p_s)$ is a map whose value is 1 exactly when $x \le 0$;*
2. *if l is the LPO function, then $l(p_s)$ is a positivity test;*

3. *if $x = 0$ or $x = 1$, then the bit $[s_7 > \frac{1}{2}]$ is 1 exactly when $x = 1$.*

In each case the value depends only on x, not on the sequence s that names it.

Proof. The first equivalence is Bishop's zero condition, Lemma 12.15; the second is the definition of $x > 0$, which does not depend on the sequence. In item 1 the bit $w(o_s)$ is 1 exactly when $x = 0$. If s' also names x, the bits $w(o_s)$ and $w(o_{s'})$ are each 1 exactly when $x = 0$, so they are equal: two bits with the same condition for being 1 coincide. So the rule is a map on points in the sense of Section 21.1. The bit $w(p_s)$ is 1 exactly when no stage of s is a positive witness, that is, when $\neg(x > 0)$, which is $x \leq 0$. Item 2 is the same argument with l. For item 3, regularity gives $|s_7 - s_m| \leq \frac{1}{8} + \frac{1}{m+1}$ for every m, so $|x - s_7| \leq \frac{1}{8}$. If $x = 0$ then $s_7 \leq \frac{1}{8} < \frac{1}{2}$, and if $x = 1$ then $s_7 \geq \frac{7}{8} > \frac{1}{2}$; the bit is therefore fixed by which of the two holds. □

Theorem 28.3 (Decisions on the line, as data)**.**

1. *The LPO function exists exactly when the WLPO function exists and MP holds.*
2. *A zero test, decided equality and a decided order $\leq$ each exist exactly when the WLPO function exists.*
3. *An apartness test, a positivity test, a decided strict order $<$ and a floor map each exist exactly when the LPO function exists.*

No choice principle is used in either direction.

Proof. 1. If l is the LPO function, $w(\alpha) = 1 - l(\alpha)$ is the WLPO function, and MP holds: if α is not all zero, $l(\alpha) = 0$ would make every $\alpha(n)$ zero, so $l(\alpha) = 1$. Conversely put $l(\alpha) = 1 - w(\alpha)$. If $l(\alpha) = 1$, then α is not all zero and MP gives an index with $\alpha(n) = 1$; if some $\alpha(n) = 1$, then $w(\alpha) = 0$.

2. From the WLPO function, Lemma 28.2 gives a zero test, and a decided order by $d(x, y) = w(p_s)$ for s naming $x - y$, since $x \leq y$ is $\neg(x - y > 0)$. Decided equality is $d(x, y)\, d(y, x)$, because $x \leq y \leq x$ gives $x = y$ (Proposition 12.19, item 2). Conversely each of the three gives a zero test, $z(x) = d(x, 0)$ or $z(x) = d(x, 0)\, d(0, x)$, and a zero test gives the WLPO function $\alpha \mapsto z(x_\alpha)$ by (28.1).

3. From the LPO function, Lemma 28.2 gives a positivity test a; then $a(x) + a(-x) - a(x)a(-x)$ is an apartness test and $a(y - x)$ decides $x < y$. For a floor map, let x be a point. Search the integers in the order $0, -1, 1, -2, 2, \ldots$ for the first m with $\neg(x < m)$ and $x < m + 1$, each decided by the decided order. The search ends: a stage of a regular sequence naming x gives an integer B with $-B < x < B$, and the largest m between $-B$ and B with $\neg(x < m)$, found by a bounded search, satisfies $x < m + 1$. Any two integers m, m' with this property satisfy $m < m' + 1$ and $m' < m + 1$, so they are equal; the result depends only on x. Conversely, an apartness test a gives $\alpha \mapsto a(x_\alpha)$, which is 1 exactly when $x_\alpha > 0$ by (28.1); a positivity test or a decided $<$ gives the same at once. A floor map gives $\lfloor -x_\alpha \rfloor < 0$ exactly when $x_\alpha > 0$: if $\lfloor -x_\alpha \rfloor = m < 0$ then $-x_\alpha < m + 1 \leq 0$, and if $x_\alpha > 0$ then $m \geq 0$ would give $0 \leq m \leq -x_\alpha < 0$. □

The second item contrasts with Corollary 12.18, which gives the same price for the proposition that every point is 0 or not 0. The difference is real under the reading of this part: by Section 21.1 the data form and the proposition coincide once unique choice is assumed, and the next proposition shows what the data form collects that the proposition does not.

Proposition 28.4 (What a data principle collects)**.**

1. *The LPO function gives* $\mathrm{LPO}_{\mathrm{seq}}$*: for every double sequence of bits A a binary sequence f with $f(i) = 1$ exactly when $A(i, n) = 1$ for some n.*
2. *The WLPO function gives* $\mathrm{WLPO}_{\mathrm{seq}}$ *(Definition 13.8), and countable unique choice into* $\{0, 1\}$ *for* Π^0_1 *relations: if for each i exactly one bit b has $r(i, b, n) = 1$ for every n, then some f has $r(i, f(i), n) = 1$ for all i and n.*
3. *The WEM function gives the WLPO function, and the WEM function with MP gives the LPO function.*

No choice principle is used.

Proof. 1 and 2. Apply the map to each row: $f(i) = l(A(i, \cdot))$, or $f(i) = w(A(i, \cdot))$. For the choice statement put $f(i) = w(\beta_i)$ with $\beta_i(n) = 1 - r(i, 1, n)$, so $f(i) = 1$ exactly when $r(i, 1, n) = 1$ for every n. If $f(i) = 0$, the unique bit is not 1, so it is 0. 3. Put $w(\alpha) = d(\forall n\, \alpha(n) = 0)$; a universal statement about decidable equations is equivalent to its double negation. With MP, Theorem 28.3, item 1, applies. □

In the model of named objects of the computable calculus (Definition 16.20) this is a separation. There LPO holds, as in the proof of Theorem 16.27, but the WLPO function fails: a WLPO function run on the computed codes of the rows of a computable double sequence would compute the sequence f of item 2, and the proof of Theorem 16.27 exhibits a computable double sequence, built from bounded halting, with no computable such f. The same argument refutes the LPO function. So in that model, which has no unique choice, LPO holds and its data form fails; the argument reasons in a classical metatheory.

Remark 28.5 (Under the default reading)**.** Section 21.1 shows that unique choice turns LPO into the LPO function, WLPO into the WLPO function and WEM into the WEM function. Under the default reading of this book, which contains unique choice, every price in this chapter that names a data form therefore becomes the price of the principle: a zero test costs exactly WLPO, a floor map exactly LPO, and so on. We state the theorems in their data form because that is what is proved, and it holds a fortiori under the default reading. Where unique choice changes a statement beyond this translation, the text says so.

28.2 Conventions at excluded points

A convention gives an operation a value where mathematics leaves it undefined: the inverse at 0, the logarithm at 0, the square root of a negative number. On $\mathbb{R}_\delta$ a total map has to produce its value from a regular sequence, without being told whether the point is the excluded one. The conventions we meet fall into three classes, according to how the map behaves near the excluded point.

Definition 28.6 (Jumps)**.** A map $f : \mathbb{R}_\delta \to \mathbb{R}_\delta$ *jumps at* 0 if there are $\varepsilon > 0$ and $r > 0$ with $|f(y) - f(0)| \geq \varepsilon$ for every $y \# 0$ with $|y| < r$. It *jumps on the right* at 0 if this holds for every y with $0 < y < r$. A jump *carries a test* if, in addition, $|f(y) - f(0)| > 0$ implies $y \# 0$ for every y.

The inverse i with $i(0) = 0$ jumps and carries a test. For $y \# 0$ with $|y| < 1$ we have $|i(y)| = 1/|y| > 1$. And $y \cdot i(y)$ is 0 or 1 and is 0 only at $y = 0$, by the worked case at the start of the chapter; so $|i(y)| > 0$ forces $y \cdot i(y) = 1$, hence $y \# 0$. The map with value 1 at 0 and 0 at every $y \neq 0$ jumps but carries no test, since a value can be known to differ from 1 without its argument being known apart from 0. A continuous map does not jump at all. The rule below gives a lower bound for each class with a jump and shows that a continuous convention needs no principle.

Lemma 28.7 (The jump reading is two-valued)**.** *Let $f : \mathbb{R}_\delta \to \mathbb{R}_\delta$, $\varepsilon > 0$ and $r > 0$, and suppose $|f(y) - f(0)| \geq \varepsilon$ whenever $0 < y < r$. For a point y put $y' = \min(|y|, r/2)$ and*

$$e(y) = \min\Big(1,\ \frac{|f(y') - f(0)|}{\varepsilon}\Big).$$

Then e is a map, and for every y: $e(y) = 0$ or $e(y) = 1$, and $e(y) = 0$ exactly when $y = 0$. If moreover $|f(u) - f(0)| > 0$ implies $u \# 0$ for every u, then $e(y) = 1$ exactly when $y \# 0$.

Proof. The point y' satisfies $0 \leq y' \leq r/2 < r$, and $y' > 0$ exactly when $|y| > 0$, that is, when $y \# 0$, since $r/2 > 0$. So $y \# 0$ gives $0 < y' < r$, hence $|f(y') - f(0)| \geq \varepsilon$ and $e(y) = 1$; and $y = 0$ gives $y' = 0$ and $e(y) = 0$.

Now let y be arbitrary. Since $0 < 1$, cotransitivity gives $0 < e(y)$ or $e(y) < 1$. If $e(y) < 1$, then $y \# 0$ is impossible, by the first paragraph; a point that is not apart from 0 equals 0, because every stage of a regular sequence naming it then satisfies $|s_n| \leq 2/(n+1)$ (Lemma 12.15); so $y = 0$ and $e(y) = 0$. If $0 < e(y)$, then $|f(y') - f(0)| > 0$, so $y' \neq 0$ and $y \neq 0$. We show $\varepsilon \leq |f(y') - f(0)|$, which is the negation of $|f(y') - f(0)| < \varepsilon$: under that strict inequality $y \# 0$ is impossible, so $y = 0$, against $y \neq 0$. Hence $e(y) = 1$. No step uses Markov's principle: the conclusion $\varepsilon \leq |f(y') - f(0)|$ is itself a negation.

For the last assertion, $e(y) = 1$ gives $|f(y') - f(0)| > 0$, hence $y' \# 0$ by the hypothesis, hence $y' > 0$ and $y \# 0$. □

Theorem 28.8 (The convention rule)**.** *On $\mathbb{R}_\delta$:*

1. *If some map jumps at* 0*, or jumps on the right at* 0*, the WLPO function exists; conversely, if the WLPO function exists, some map jumps at* 0*.*
2. *If some map jumps at* 0 *and its jump carries a test, the LPO function exists; conversely, if the LPO function exists, some map has such a jump.*
3. *A map continuous at* 0 *(for every k some j has $|f(y) - f(0)| \leq 1/(k+1)$ whenever $|y| \leq 1/(j+1)$) does not jump on the right at* 0*. Every map g of the ratios into the ratios that is uniformly continuous with a modulus extends to a uniformly continuous map of $\mathbb{R}_\delta$, with a modulus, agreeing with g at every ratio.*

No choice principle is used.

Proof. 1. A jump at 0 is in particular a jump on the right. By Lemma 28.7 and Lemma 28.2, item 3, the map $y \mapsto 1 - [\text{ stage 7 of } e(y) > \frac{1}{2}]$ is a zero test, and Theorem 28.3 gives the WLPO function. Conversely, a zero test z gives the map $f(y) = z(y)$, read as a point, which is 1 at 0 and 0 at every $y \# 0$: a jump of size 1.

2. Now $e(y) = 1$ exactly when $y \# 0$, so the bit of $e(y)$ is an apartness test, and Theorem 28.3 gives the LPO function. Conversely the apartness test a, read as a point, is 0 at 0 and 1 at every $y \# 0$, and $|a(y) - a(0)| > 0$ means $a(y) = 1$, that is, $y \# 0$.

3. Suppose f is continuous at 0 and $|f(y) - f(0)| \geq \varepsilon$ for $0 < y < r$. Take k with $1/(k+1) < \varepsilon$, its j, and $y = \min(1/(j+1), r/2)$. Then $0 < y < r$, and $\varepsilon \leq |f(y) - f(0)| \leq 1/(k+1) < \varepsilon$.

For the extension, let ω be a modulus of g: $|q - q'| \leq 1/(\omega(k)+1)$ gives $|g(q) - g(q')| \leq 1/(k+1)$. Put $M(k) = \max_{i \leq 2k+1} \omega(i)$, and for a regular sequence s let $t_k = g(s_{4M(k)+3})$. For $k \leq k'$, $|s_{4M(k)+3} - s_{4M(k')+3}| \leq 1/(2M(k)+2) \leq 1/(\omega(2k+1)+1)$, so $|t_k - t_{k'}| \leq 1/(2k+2)$ and t is regular; the point it names lies within $1/(2k+2)$ of t_k for every k. Two sequences naming the same point have $|s_n - s'_n| \leq 2/(n+1)$ at every stage (Theorem 12.17 and its proof), so their sequences t and t' satisfy $|t_k - t'_k| \leq 1/(2k+2)$ and name the same point. The map $F(x) =$ the point of t is therefore a map on points, and at a ratio q it is $g(q)$. For the modulus, let $|x - y| \leq 1/(2M(2k+1)+2)$ and put $j = 2k+1$. Stage $n = 4M(j)+3$ of sequences s, s' naming x, y satisfies $|s_n - s'_n| \leq 2/(n+1) + |x - y| \leq 1/(M(j)+1) \leq 1/(\omega(2k+1)+1)$, so $|t_j - t'_j| \leq 1/(2k+2)$ and $|F(x) - F(y)| \leq 1/(4k+4) + 1/(2k+2) + 1/(4k+4) = 1/(k+1)$. The modulus is $k \mapsto 2M(2k+1)+1$. □

The second assertion of part 3 is why continuous conventions are free: a convention given by a uniformly continuous formula on the ratios is already a map on the line. Part 3 leaves a gap: a map can be discontinuous at 0 without jumping, and then the rule does not apply.

Theorem 28.9 (The prices of the conventions)**.** *On $\mathbb{R}_\delta$, with no choice principle in either direction:*

1. *A map i with $i(0) = 0$ and $x \cdot i(x) = 1$ for every $x \# 0$ exists exactly when the LPO function exists. The same holds for a map with $i(0) = 0$ and $x \cdot i(x) = 1$ for every $x \neq 0$; for a division map v with $v(a, 0) = 0$ and $v(a, b) \cdot b = a$ for every $b \# 0$; and, for each n, for a map q with $q(0) = 0$ and $x^{n+1} q(x) = 1$ for every $x \# 0$.*

2. *A map L with $L(0) = 0$ and $\exp(L(x)) = |x|$ for every $x \# 0$ exists exactly when the LPO function exists.*

3. *A map P with $P(0) = 1$ and $P(y) = 0$ for every $y \neq 0$, the convention $0^0 = 1$, $0^y = 0$, exists exactly when the WLPO function exists.*

4. *A sign map S with $S(0) = 0$, $S(x) = 1$ for $x > 0$ and $S(x) = -1$ for $x < 0$ exists exactly when the WLPO function exists.*

5. *The following exist with no principle: a map R with $R(x) \geq 0$ for every x, $R(x)^2 = x$ for $x \geq 0$ and $R(x) = 0$ for $x \leq 0$; the natural powers x^n; truncated subtraction on the natural numbers; the truncation of the integers to the natural numbers.*

6. *Every $x \neq 0$ has an inverse exactly when MP holds.*

Proof. 1. Let i be given and put $e(x) = x \cdot i(x)$. Cotransitivity at $0 < 1$ gives $0 < e(x)$ or $e(x) < 1$. In the first case x has the inverse $i(x)/e(x)$, so $x \# 0$ by Proposition 12.19, item 3, and $e(x) = 1$. In the second, $x \# 0$ is impossible, so $x = 0$ and $e(x) = 0$. So e is two-valued and equals 1 exactly at the points apart from 0; its bit (Lemma 28.2, item 3) is an apartness test, and Theorem 28.3 gives the LPO function. Conversely, the LPO function gives an apartness test a; put $i(x) = x^{-1}$ if $a(x) = 1$, which is defined because $a(x) = 1$ supplies $x \# 0$, and $i(x) = 0$ otherwise. A map specified at the points $x \neq 0$ is specified in particular at the points $x \# 0$; conversely the LPO function gives MP (Theorem 28.3, item 1), which turns $x \neq 0$ into $x \# 0$ (Corollary 12.18), so the map just built serves. For division, $x \mapsto v(1, x)$ is an inverse of the first kind and $v(a, b) = a \cdot i(b)$ is a division map. For negative powers, $x \mapsto x^n q(x)$ is an inverse of the first kind, and $q(x) = i(x)^{n+1}$ is a map of the required kind.

2. Let L be given and put $g(y) = |y| \exp(-L(y))$. At $y \# 0$, $g(y) = \exp(L(y)) \exp(-L(y)) = 1$; at $y = 0$, $g(y) = 0$. As in part 1, $0 < g(y)$ gives $|y| > 0$, because $\exp(-L(y)) > 0$, hence $y \# 0$ and $g(y) = 1$, while $g(y) < 1$ excludes $y \# 0$, so $y = 0$ and $g(y) = 0$. So the bit of g is an apartness test. Conversely, with an apartness test a put $L(x) = \log |x|$ when $a(x) = 1$, using the logarithm of a positive point (Definition 21.4), and $L(x) = 0$ otherwise; Section 21.2 proves $\exp(\log |x|) = |x|$.

3. Every $y \# 0$ is $\neq 0$, so $|P(y) - P(0)| = 1$ there: P jumps at 0, and Theorem 28.8, part 1, gives the WLPO function. Conversely, a zero test, read as a point, is such a map.

4. For $y \# 0$, $|S(y) - S(0)| = 1$, so S jumps at 0 and part 1 of Theorem 28.8 applies. Conversely, let w be the WLPO function and $u(x) = w(p_s)$ for s naming x, so that $u(x) = 1$ exactly when $\neg(x > 0)$ (Lemma 28.2). Put $S(x) = 1$ if $u(x) = 0$; otherwise $S(x) = -1$ if $u(-x) = 0$; otherwise $S(x) = 0$. If $x > 0$ then $u(x) = 0$ and $S(x) = 1$. If $x < 0$ then $\neg(x > 0)$ and $-x > 0$, so $S(x) = -1$. If $x = 0$ both tests return 1 and $S(x) = 0$.

5. Put $R(x) = \sqrt{\max(x, 0)}$. The square root of a nonnegative point (Proposition 12.19, item 5) is computed stage by stage from a regular sequence naming the point, by bisection on ratios, and the bound $|\sqrt{a} - \sqrt{b}| \leq \sqrt{|a - b|}$ for $a, b \geq 0$ makes the result independent of the sequence; so R is a map, uniformly continuous, with no jump. Then $R(x)^2 = \max(x, 0)$, which is x for $x \geq 0$ and 0 for $x \leq 0$. Natural powers are defined by recursion, and the two truncations are decided by comparisons of integers.

6. By Proposition 12.19, item 3, a point has an inverse exactly when it is apart from 0, and Corollary 12.18 prices $x \neq 0 \to x \# 0$ at MP. □

The sign is the instructive case. Its jump at 0 looks like the jump of the inverse, and a first guess prices it at the LPO function. But the specification of the sign says nothing at a point that is merely not 0: the map built in part 4 returns 1 at every point that is not nonpositive, a condition weaker than positivity unless MP holds. Its jump carries no test, and part 1 of the rule gives the WLPO function, which part 4 shows is the exact price.

Table 28.1 collects eighteen conventions of a classical library. Rows 13 to 16 are decisions stored as data, priced by Theorem 28.3; a ceiling map is $x \mapsto -\lfloor -x \rfloor$ and costs what a floor map costs. Row 10, the derivative convention, is an operator on all maps, and its exact price

Table 28.1. Conventions of a classical library at excluded points, and their exact prices on $\mathbb{R}_\delta$. "Jump, test" is part 2 of Theorem 28.8, "jump" part 1, "continuous" part 3; a decision is a map to $\{0, 1\}$. Every price is an equivalence proved with no choice principle.

	Convention	Kind	Price	Result
1	inverse, $0^{-1} = 0$, specified at $x \# 0$	jump, test	LPO function	28.9.1
2	inverse, $0^{-1} = 0$, specified at $x \neq 0$	jump, test	LPO function	28.9.1
3	division, $a/0 = 0$	jump, test	LPO function	28.9.1
4	$x^{-(n+1)}$ with value 0 at 0, each n	jump, test	LPO function	28.9.1
5	$\log 0 = 0$ and $\log x = \log \lvert x \rvert$	jump, test	LPO function	28.9.2
6	$0^0 = 1$ and $0^y = 0$ for $y \neq 0$	jump	WLPO function	28.9.3
7	natural powers x^n	none	free	28.9.5
8	$\sqrt{x} = 0$ for $x \leq 0$	continuous	free	28.9.5
9	sign, with sign of 0 equal to 0	jump	WLPO function	28.9.4
10	derivative where one exists, 0 elsewhere	operator	MP and the differentiability decision	28.24
11	truncated subtraction on $\mathbb{N}$	none	free	28.9.5
12	truncation of $\mathbb{Z}$ to $\mathbb{N}$	none	free	28.9.5
13	floor and ceiling	decision	LPO function	28.3.3
14	decided equality	decision	WLPO function	28.3.2
15	decided $\leq$	decision	WLPO function	28.3.2
16	decided $<$	decision	LPO function	28.3.3
17	a hypothesis $x \neq 0$ read as $x \# 0$	reading	MP	28.9.6
18	$\lvert x \rvert$, max, min	none	free	12.19

is the subject of Section 28.6: the rule gives the WLPO function as a lower bound, and the operator costs more. Row 17 is not a convention at all; it is the reading of a hypothesis $x \neq 0$ as apartness, and its price is Markov's principle.

The rule gives the lower bound of every row it covers, and Theorem 28.9 supplies the matching construction. Rows 1 to 5 are jumps whose size is positive only away from the excluded point, so their price is at least the LPO function, and exactly that by Theorem 28.9; rows 6 and 9 are jumps without that test, at least and exactly at the WLPO function; rows 7, 8, 11, 12 and 18 do not jump. The decisions of rows 13 to 16 are indicators of jumps: equality and $\leq$ give zero tests, and $<$ and the floor give apartness tests. Real powers of negative numbers with non-integer exponents are left unpriced; no row of the audit reported in Section 28.10 uses them. Among the uses of the classical library that the audit counted for Recognition Science's forcing chain and its gravity sector, 584 uses of 72 library theorems concern inverse or division, where the value at 0 can enter; 27 uses of 4 theorems concern integer powers, 9 uses of 6 theorems the logarithm, 12 uses of 7 theorems real powers, 21 uses of 9 theorems the square root, 4 uses of 3 theorems the derivative, 2 uses of one theorem truncated subtraction, and 27 uses of 6 theorems stored decisions. No use involves the sign, the integer truncation or the floor.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ with its free ordered-ring structure; maps read as in Section 21.1.

Conclusion

A convention that jumps at the excluded point costs at least the WLPO function, and at least the LPO function when its jump reveals apartness; a map that jumps exists exactly when the WLPO function does, and one whose jump carries a test exactly when the LPO function does; a continuous convention is free. For the eighteen conventions of Table 28.1 the price is exact (Theorems 28.9, 28.3 and 28.24, and Proposition 12.19 for row 18); under unique choice each price is the corresponding principle (Remark 28.5).

Not claimed

That the rule decides every discontinuous convention: a map can be discontinuous at a point without jumping there. Nor that a jump fixes the price: row 10 costs more than the lower bound its jump gives.

28.3 Choosing representatives

A point of $\mathbb{R}_\delta$ is a class of regular sequences. Every point has a regular sequence naming it, and for one given point we may use one, because an assertion about a single point is a proposition. A classical library goes further: with the axiom of choice it fixes one representative of every class at once, and many of its constructions read a point through that representative. On $\mathbb{R}_\delta$ such a choice would be a map s from points to regular sequences with $s(x)$ naming x, a *section* of the passage from sequences to points. Chapter 12 showed that countable choice supplies representatives for one sequence of points at a time and decides nothing (Theorem 12.22). A section is a different object, and it has a price.

The price comes from one observation. A section lets us look at a stage of the chosen sequence: the bit "stage 2 of $s(x)$ is at most $\frac{1}{2}$" is a map from points to $\{0,1\}$, and it is 1 at 0 and 0 at 1. A two-valued map on the line is already strong.

For a binary sequence β write $\nu_n(\beta) = \sum_{k<n} \beta(k)2^{n-1-k}$ for the numeral of its first n digits and $[\beta]$ for the point $\sum_k \beta(k)2^{-(k+1)}$, named by the regular sequence of the ratios $\nu_n(\beta)/2^n$; so β is a binary expansion of $[\beta]$ in the sense of Section 12.9. If $\beta(k) = 1$ for every $k \geq N$, then $[\beta] = (\nu_N(\beta) + 1)/2^N$, and if $\beta(k) = 0$ for every $k \geq N$, then $[\beta] = \nu_N(\beta)/2^N$.

Theorem 28.10 (Two values give the WLPO function)**.** *Let g be a map from $\mathbb{R}_\delta$ to $\{0,1\}$, and let a and b be points with $g(a) \neq g(b)$. Then the WLPO function exists. No choice principle is used.*

Proof. Put $G(t) = g(a + t(b-a))$ if $g(a) = 0$ and $G(t) = 1 - g(a + t(b-a))$ if $g(a) = 1$. Then G is a map to $\{0,1\}$ with $G(0) = 0$ and $G(1) = 1$.

Bisection. Define digits $\beta(n)$ by recursion. Let $\ell_n = \nu_n(\beta)$, so $\ell_0 = 0$, and look at the midpoint $m_n = (2\ell_n+1)/2^{n+1}$ of $[\ell_n/2^n, (\ell_n+1)/2^n]$: put $\beta(n) = 1$ if $G(m_n) = 0$ and $\beta(n) = 0$ if $G(m_n) = 1$. By induction on n,

$$G\big(\ell_n/2^n\big) = 0 \qquad \text{and} \qquad G\big((\ell_n+1)/2^n\big) = 1, \tag{28.2}$$

since a midpoint with value 0 becomes the new left end and one with value 1 the new right end. Each step is a case distinction on a bit. Let $c = [\beta]$ and $v = G(c)$.

Modified names. For a binary sequence α let $\sigma_\alpha(k) = 1$ when $\alpha(i) = 1$ for some $i \le k$, a bounded search. If $v = 0$, let $\mu(\alpha)$ be the sequence $\max(\beta(k), \sigma_\alpha(k))$; if $v = 1$, the sequence $\min(\beta(k), 1 - \sigma_\alpha(k))$. If α is all zero, then $\mu(\alpha) = \beta$. If α has a first 1 at N, then $\mu(\alpha)$ agrees with β below N and is constant from N on: all ones when $v = 0$, all zeros when $v = 1$. So $[\mu(\alpha)]$ is the right end $(\ell_N + 1)/2^N$ in the first case and the left end $\ell_N/2^N$ in the second, and by (28.2) $G([\mu(\alpha)]) = 1 - v$ in both.

The function. Put $d(\alpha) = 1$ if $G([\mu(\alpha)]) = v$ and $d(\alpha) = 0$ otherwise; d is a composite of maps. If α is all zero, then $[\mu(\alpha)] = c$ and $d(\alpha) = 1$. If $d(\alpha) = 1$ and $\alpha(n) = 1$ for some n, the first 1 of α lies at some $N \le n$, and then $G([\mu(\alpha)]) = 1 - v \neq v$; so every $\alpha(n)$ is 0. Hence $d(\alpha) = 1$ exactly when α is all zero. □

The bisection is the familiar proof that a two-valued function on an interval has no continuous version; here it returns a map rather than a proposition. Three consequences follow at once. A map from $\mathbb{R}_\delta$ into any set with decidable equality, such as the integers, that takes two different values gives the WLPO function: compose with the decided equality to one of the values. A map assigning to every point of $[0, 1]$ a binary expansion gives it: the first digit of an expansion of 0 is 0, since $\eta(e_0/2) \le 0$, and of 1 is 1, since $1 \le \eta(e_0/2 + \frac{1}{2})$; compose with the clamp $x \mapsto \min(\max(x, 0), 1)$. A map assigning to every point of $[-1, 1]$ a signed-digit expansion gives it in the same way, since the first digit of an expansion of 1 is 1 and of -1 is -1. Every point of $[-1, 1]$ has a signed-digit expansion with no principle (Proposition 12.62); what costs the WLPO function is a map that writes one out for every point.

Theorem 28.11 (Choosing representatives costs exactly the WLPO function)**.** *Each of the following exists exactly when the WLPO function exists, with no choice principle in either direction:*

1. *a map from $\mathbb{R}_\delta$ to $\{0, 1\}$ taking two different values;*
2. *a map from $\mathbb{R}_\delta$ to the integers taking two different values;*
3. *a map assigning to every point of $[0, 1]$ a binary expansion of it;*
4. *a map assigning to every point of $[-1, 1]$ a signed-digit expansion of it;*
5. *a section: a map s from points to regular sequences with $s(x)$ naming x for every point x;*
6. *a section over $[0, 1]$;*
7. *a map s from points to regular sequences with $s(0)$ naming 0 and $s(1)$ naming 1;*
8. *a map q from points to ratios with $|x - q(x)| \le 1$ for every x;*
9. *a map assigning to every point x a sequence of ratios that converges to x: for every k some M has $|x - q_n| \le 1/(k + 1)$ for all $n \ge M$.*

Proof. Each object gives two values. Items 1 to 4 were treated above. For item 7, the bit $g(x) = [\, s(x)_2 \le \frac{1}{2} \,]$ compares two ratios, so it is a map; stage 2 of a sequence naming x lies

within $\frac{1}{3}$ of x, so $g(0) = 1$ and $g(1) = 0$. Items 5 and 6 give item 7, composing with the clamp for item 6. For item 8, $|q(0)| \le 1$ and $q(3) \ge 2$, so $[\, q(x) \le 1 \,]$ separates 0 from 3. For item 9, the sequences for 0 and for 3 come within 1 of their points from some indices on; at one index n beyond both, the bit $[\, q_n \le \frac{3}{2} \,]$ separates 0 from 3. Theorem 28.10 gives the WLPO function in every case.

Conversely, let w be the WLPO function. Theorem 28.3 gives a zero test, which is 1 at 0 and 0 at 1: items 1 and 2. It also gives a decided order $\le$. For item 3 run the bisection of the proof of Theorem 12.61 with the decided comparisons $B(n, j) = [\, j/2^n \le x \,]$ in place of the table chosen there: if $B(n, j) = 0$ then $\neg(j/2^n \le x)$, which gives $x \le j/2^n$ because $j/2^n \le x$ is $\neg(x < j/2^n)$; nothing is chosen. For item 4, a binary expansion (e_i) of $(x + 1)/2$ gives the signed digits $2e_i - 1$ of x: their partial sums are $2c_n - 1 + 2^{-n}$, and $2c_n - 1 \le x \le 2c_n - 1 + 2^{1-n}$.

For item 8, search the integers in a fixed order for the first m with $m - 1 \le x$ and $x \le m + 1$, both decided. The search ends, because stage 1 of a sequence naming x lies within $\frac{1}{2}$ of x and the integer nearest to it qualifies; the result depends only on x. For item 5 we turn q into a section. Put $s(x)_n = q\big(2(n + 1)x\big)/\big(2(n + 1)\big)$. Then $|s(x)_n - x| \le 1/(2n + 2)$, so $|s(x)_m - s(x)_n| \le 1/(2m + 2) + 1/(2n + 2)$ and $s(x)$ is regular; and a regular sequence t naming x has $|t_n - x| \le 1/(n + 1)$, so $|s(x)_n - t_n| \le 2/(n + 1)$ for every n and $s(x)$ names x, as in the proof of Lemma 12.15. Items 6, 7 and 9 follow from item 5. □

A section gives representatives to every family of points, indexed by any set, by composition. With representatives the limit of a Cauchy sequence can be written down, and so the WLPO function buys completeness without countable choice.

Corollary 28.12 (Completeness at the WLPO function)**.** *Assume the WLPO function. Then every family of points $(F_i)_{i \in I}$ has a family of regular sequences naming it, and every sequence of points with $|F_m - F_n| \le 1/(k + 1)$ for all $m, n \ge k$ has a limit L with $|F_m - L| \le 1/(k + 1)$ for all $m \ge k$. No choice principle is used.*

Proof. Let s be a section and $f(i) = s(F_i)$. For the limit, put $t_n = s(F_{4n+3})_{4n+3}$. Then $|t_n - F_{4n+3}| \le 1/(4n + 4)$, and for $m \ge n$ the modulus gives $|F_{4n+3} - F_{4m+3}| \le 1/(4n + 4)$, so $|t_n - t_m| \le 3/(4n + 4) < 1/(n + 1)$ and t is regular. Let L be its point. For $m \ge k$ and every $n \ge k$, $|F_m - L| \le |F_m - F_{4n+3}| + |F_{4n+3} - t_n| + |t_n - L| \le 1/(k + 1) + 2/(n + 1)$, and since n is arbitrary, $|F_m - L| \le 1/(k + 1)$. □

Chapter 12 reached the same limit with countable choice, which selects representatives sequence by sequence and decides nothing. A section selects them point by point, uniformly, and that costs exactly the WLPO function. In classical mathematics the choice operator that picks a member of each equivalence class is a section of $\mathbb{R}_\delta$; so the object behind every representative choice of a classical library, read on $\mathbb{R}_\delta$, costs exactly the WLPO function. Under unique choice (Remark 28.5) Theorem 28.11 says that each of the nine objects exists exactly when WLPO holds.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ as a set of classes of regular sequences; maps read as in Section 21.1.

Conclusion

A two-valued map on the line, a map writing out a name or a representative for every point, or a rational approximation within 1 of every point, each costs exactly the WLPO function; with it the line is complete for Cauchy sequences with a modulus, with no countable choice.

Not claimed

That countable choice of representatives, one sequence of points at a time, gives the WLPO function; it decides none of LPO, WLPO, LLPO and MP (Theorem 12.22).

28.4 The linear order and the ordered field

On $\mathbb{R}_\delta$ the strict order comes first: $x < y$ when $y - x$ has a positive stage, and $x \le y$ is defined as $\neg(y < x)$. A classical library goes the other way. Its basic notion is a total order $\le$, and it defines the strict order from it. We read its structures by their content.

Definition 28.13 (The classical structures, by content)**.** A *linear-order structure* on a set consists of a relation $\le$ that is reflexive, transitive, antisymmetric and total; a relation $<$ required to satisfy $x < y$ exactly when $x \le y$ and not $y \le x$; maps max and min with $\max(a,b) = b$ if $a \le b$ and $\max(a,b) = a$ otherwise, and dually; and stored decisions of $\le$, $=$ and $<$. A *field structure* on the ring $\mathbb{R}_\delta$ is a map $x \mapsto x^{-1}$ with $0^{-1} = 0$ and $x \cdot x^{-1} = 1$ for every $x \neq 0$. An *ordered field* is a field structure with a linear-order structure satisfying the strict-ordered-ring laws: $a < b$ gives $a + c < b + c$, $0 < c$ and $a < b$ give $ac < bc$, and $0 < 1$, together with the corresponding laws for $\le$ (addition preserves and cancels $\le$, and $0 \le 1$), which hold on $\mathbb{R}_\delta$ with no principle. A *floor structure* is a linear-order structure together with maps $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ to the integers with $z \le \lfloor a \rfloor$ exactly when $z \le a$, and $\lceil a \rceil \le z$ exactly when $a \le z$, for all integers z. The *Archimedean property* states that for every x and every $y > 0$ some natural number n has $x \le ny$.

Read on $\mathbb{R}_\delta$ with δ's own $\le$, the definition forces the strict order: $x \le y$ and $\neg(y \le x)$ is $\neg(y < x) \wedge \neg\neg(x < y)$, which is $\neg\neg(x < y)$, since $x < y$ already excludes $y < x$. So there are two ways to put a classical linear order on the line: with δ's $\le$ and the strict order $\neg\neg <$ that the definition forces, or with δ's own $<$, which agrees with $\neg\neg <$ exactly when MP holds. They have different prices.

Theorem 28.14 (The classical linear order costs exactly the WLPO function)**.**

1. *If a linear-order structure on $\mathbb{R}_\delta$ has δ's $\le$ as its $\le$, the WLPO function exists, and the structure's $<$ is $\neg\neg(x < y)$.*
2. *Conversely, the WLPO function gives a linear-order structure whose $\le$ is δ's $\le$ and whose $<$ is $\neg\neg(x < y)$; with δ's sum and product it satisfies the strict-ordered-ring laws.*

No choice principle is used.

Proof. 1. The structure stores a decision of δ's $\le$; Theorem 28.3, item 2, gives the WLPO function. The form of $<$ was computed above.

2. Theorem 28.3 gives a decided order d. Totality: if $d(x,y) = 1$ then $x \le y$; if $d(x,y) = 0$ then $\neg(x \le y)$, that is, $\neg\neg(y < x)$, and then $x < y$ is impossible, so $y \le x$. Decided equality is $d(x,y)d(y,x)$, by antisymmetry (Proposition 12.19, item 2), and the decided strict order is $d(x,y)(1 - d(y,x))$. For max take δ's own max (Proposition 12.19, item 5): if $a \le b$ it equals b, and if not, then $b \le a$ by totality and it equals a; so it is the case split the definition asks for, and min likewise. Reflexivity, transitivity and antisymmetry of $\le$ are Proposition 12.19, item 2. The structure's $<$ is $\neg\neg(x < y)$, and the strict-ordered-ring laws for it follow from the laws for δ's $<$ in Proposition 12.19, item 1, applied under the double negation: from $\neg\neg(a < b)$ we get $\neg\neg(a + c < b + c)$, and from $\neg\neg(0 < c)$ and $\neg\neg(a < b)$ we get $\neg\neg(ac < bc)$. □

Theorem 28.15 (The classical ordered field costs exactly the LPO function)**.** *Each of the following exists on $\mathbb{R}_\delta$, with δ's sum and product and, where an order is part of the structure, with δ's $\le$ and δ's $<$, exactly when the LPO function exists: a field structure; a linear-order structure; an ordered field; a floor structure. The strict-ordered-ring laws for δ's $\le$ and the strict order it forces are free. The Archimedean property for δ's $\le$ and the strict order it forces holds exactly when MP holds. No choice principle is used.*

Proof. A field structure is row 2 of Table 28.1 (Theorem 28.9, item 1). A linear-order structure with δ's $<$ stores a decided $<$, hence gives the LPO function (Theorem 28.3, item 3); an ordered field and a floor structure contain one. Conversely, the LPO function gives decided $\le$ and $<$, and MP (Theorem 28.3). Build the structure of Theorem 28.14, item 2, but with δ's $<$: the required identity of $x < y$ with $x \le y \wedge \neg(y \le x)$, which is $\neg\neg(x < y)$, is MP for the points $y - x$ (Corollary 12.18). The field structure is Theorem 28.9, item 1, and the strict-ordered-ring laws for δ's $<$ are Proposition 12.19, item 1. For the floor structure take the floor map of Theorem 28.3 and $\lceil a \rceil = -\lfloor -a \rfloor$. If $z \le \lfloor a \rfloor$ then $z \le a$; if $z \le a$ and $\lfloor a \rfloor < z$, then $\lfloor a \rfloor + 1 \le z \le a$, against $a < \lfloor a \rfloor + 1$.

The strict-ordered-ring laws for $\le$ and $\neg\neg <$ are the last step of the proof of Theorem 28.14 and use no decision. For the Archimedean property, MP turns $\neg\neg(0 < y)$ into $0 < y$, and then Proposition 12.19, item 4, gives n with $x/y < n$, so $x \le ny$. Conversely, let α be a binary sequence that is not all zero. Then $\neg\neg(0 < x_\alpha)$ by (28.1), and the property gives n with $1 \le nx_\alpha$. Cotransitivity at $0 < 1$ gives $0 < nx_\alpha$ or $nx_\alpha < 1$; the second is excluded, so $nx_\alpha > 0$, hence $x_\alpha > 0$ and $\alpha(k) = 1$ for some k. □

The two theorems separate what a lemma about a linear order asks of the line. Its hypotheses may be stated with δ's $<$ at no cost, and its conclusions transfer at no cost when they are stable under double negation. Only a conclusion that has to produce δ's $<$ from the library's $<$ costs more, and the extra price is exactly Markov's principle.

Proposition 28.16 (Where the strict order costs Markov's principle)**.**

1. *For all points, $x < y$ gives $x \le y$ and $\neg(y \le x)$. No principle is used.*
2. *The converse, for all points x and y, holds exactly when MP holds.*
3. *The statement "for every positive a, $1 < \max(a, a^{-1})$ exactly when $a \neq 1$" holds exactly when MP holds.*

Proof. 1. $x < y$ excludes $y < x$, so $x \leq y$, and $y \leq x$, which is $\neg(x < y)$, is refuted. 2. The hypothesis $x \leq y \wedge \neg(y \leq x)$ is $\neg\neg(x < y)$, and $\neg\neg(z > 0) \to z > 0$ for all points is MP (Corollary 12.18). 3. Under MP, $a \neq 1$ gives $a \# 1$, so $a > 1$ or $a^{-1} > 1$; and $1 < \max(a, a^{-1})$ excludes $a = 1$. Conversely, let $z \neq 0$ and $a = 1 + z^2$. Then $a \neq 1$, since $z^2 = 0$ would give $z = 0$, so $1 < \max(a, a^{-1}) = a$, hence $z^2 > 0$ and $z \# 0$. That every $z \neq 0$ is apart from 0 is MP (Corollary 12.18). □

With the linear order of Theorem 28.14 in place, a classical lemma about ordered groups, rings and fields can be read on $\mathbb{R}_\delta$ in one of three ways: at the price of the order alone, the WLPO function, when its statement only assumes strict inequalities and concludes something stable under double negation; at Markov's principle on top of the WLPO function, which together make the LPO function, when it produces a strict inequality; or at the price of a convention it states, such as $0^{-1} = 0$. The audit reported in Section 28.10 applied this criterion to 102 lemmas of the classical theory of ordered groups, rings and fields: those that its later routes through the ordered field of Theorem 28.15 price at the LPO function, among the rows its first transfer does not move. For each row it recorded a proof of an implication from the price to the δ statement, or of the δ statement outright, with no choice principle except in the two rows noted below. We report its counts; the row-by-row arguments are not reproduced. Fifty-six rows become free: 14 are class statements whose structure has a free instance, 3 are free classes on $\mathbb{R}_\delta$, 5 hold by definition, 6 are strict conclusions about positive points proved directly, 16 are proved free on the line although the transfer through the linear order prices them at the WLPO function, and 12 more, eight about fields and four about parity, are proved free on the line directly. A typical row of the sixteen is $f(\max(x, y)) = \max(f(x), f(y))$ for monotone f: an equation, so the case split on $x \leq y$, which holds up to double negation, costs nothing. Twenty-seven rows are proved at the WLPO function, as an upper bound: two state a decision of $\leq$ themselves, two others are proved only with a choice principle, which enters through the integer absolute value in their statement, and for the remaining 23 no lower bound is proved. One row costs exactly MP, item 3 of Proposition 28.16. Two rows cost exactly LLPO: both say that $a^n = b^n$, for $n \neq 0$, holds exactly when $a = b$ or, for even n, $a = -b$; at $n = 2$ this is the sign of a square root, which is LLPO. Sixteen rows remain at the LPO function, every one of them through the total inverse $0^{-1} = 0$ in its own statement; for these the LPO function is the price of the stated convention, not a lower bound for the mathematics, because the points inverted in them are apart from 0. Across these lemmas the LPO function never enters through a decided strict order.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ with δ's ring, $\leq$ and $<$; classical structures read by the content of Definition 28.13.

Conclusion

The classical linear order with δ's $\leq$ costs exactly the WLPO function; with δ's $<$, and the field and ordered field, exactly the LPO function. Producing δ's $<$ from the library's $<$ is exactly MP.

Not claimed

That the sixteen rows left at the LPO function need it: their price is that of the inverse

convention they state. The row counts report the audit of Section 28.10; they are not theorems here.

28.5 Suprema, separation and compactness

A classical library completes its ordered field by a supremum map defined on every set of points. On $\mathbb{R}_\delta$ every inhabited, order-located set of points that is bounded above has a supremum with countable choice (Theorem 12.36 and Section 12.9), but a map that produces a supremum for every set, located or not, is a different object. Its first price is Markov's principle, and the argument needs only one law of the supremum.

Theorem 28.17 (A total supremum gives Markov's principle)**.** *Let* sup *be a map from sets of points to points such that $a \le \sup S$ whenever a is a member of a set S that has an upper bound. Then MP holds. No choice principle is used.*

Proof. Let α be a binary sequence that is not all zero, and let S be the set of points $\eta(n)$ with $\alpha(n) = 1$ and $\alpha(m) = 0$ for every $m < n$. It has at most one member, so whenever a is a member, a is an upper bound and $a \le \sup S$. The point $u = \sup S$ lies below $\eta(N)$ for some natural number N (Proposition 12.19, item 4). If $\alpha(k) = 1$ for some k, a bounded search finds the first index $n_0 \le k$ with $\alpha(n_0) = 1$; then $\eta(n_0) \in S$, so $\eta(n_0) \le u < \eta(N)$ and $n_0 < N$. Hence $\exists k\, \alpha(k) = 1$ implies that $\alpha(n) = 1$ for some $n < N$. Since α is not all zero, the latter is not false, and a bounded search over $n < N$ decides it; so it is true. □

The supremum map was used at one set whose inhabitation nobody can show; the classical law "a member lies below the supremum" does the rest. With a decision of double negations as well, the supremum can be built.

Lemma 28.18 (A monotone test has a point)**.** *Let u be a map from ratios to $\{0, 1\}$ such that $u(q) = 1$ and $q \le q'$ give $u(q') = 1$, and suppose $u(-m) = 0$ and $u(m') = 1$ for some natural numbers m, m'. Then a point c is computed from u with no choice, and for every set S of points such that $u(q) = 1$ exactly when $\eta(q)$ is an upper bound of S, the point c is an upper bound of S below every upper bound of S.*

Proof. Searches find the least such m and m'; put $A = -m$. For each n search the grid $A + i/(4n + 4)$, $i = 1, 2, \ldots$, for the first point where u is 1; it exists because $u(m') = 1$, and say it is $A + (i_n + 1)/(4n + 4)$. Then $u(A + i_n/(4n + 4)) = 0$. Call a ratio *low* if u is 0 at some larger ratio and *high* if u is 1 at some smaller one. By monotonicity every low ratio lies below every high ratio. The ratio $\ell_n = A + (i_n - 1)/(4n + 4)$ is low and $\ell_n + 3/(4n + 4)$ is high, so $|\ell_n - \ell_m| \le 3/(4n + 4) + 3/(4m + 4)$ and (ℓ_n) is a regular sequence; let c be its point. A ratio q with $\eta(q) > c$ is high, and one with $\eta(q) < c$ is low.

Now let S be as stated. If a member a had $c < a$, take a ratio q with $c < \eta(q) < a$; it is high, so $u(q) = 1$, so $\eta(q)$ is an upper bound and $a \le \eta(q)$, a contradiction. So c is an upper bound. If an upper bound y had $y < c$, take q with $y < \eta(q) < c$; it is low, so $u(q') = 0$ for some $q' > q$; but $\eta(q') > y$ is an upper bound, so $u(q') = 1$, a contradiction. So $c \le y$. □

Theorem 28.19 (The complete ordered field)**.** *Call a* complete ordered field *on* $\mathbb{R}_\delta$ *an ordered field in the sense of Definition 28.13, with δ's sum, product,* $\leq$ *and* $<$*, together with maps* sup *and* inf *from all sets of points to points such that, for every inhabited set S with an upper bound,* sup S *is an upper bound of S below every upper bound, and* $\sup S = \sup \varnothing$ *for every set S with no upper bound; dually for* inf*. Then:*

1. *A complete ordered field exists exactly when MP holds and the WEM function exists, and exactly when an ordered field exists and the WEM function exists.*
2. *The same holds with δ's* $\leq$ *and the strict order it forces,* $\neg\neg <$*, in place of δ's* $<$*, for the linear order with* sup *and* inf *alone.*
3. *The WEM function gives* WEM*;* WEM *together with unique choice from propositions into* $\{0, 1\}$ *gives the WEM function.*

No choice principle is used beyond the one named in item 3.

Proof. The structure gives the WEM function. For a proposition P let S_P consist of 0, and of 1 if P. It is inhabited and bounded by 1. If P then $1 \leq \sup S_P$; if $\neg P$ then 0 is an upper bound and $\sup S_P = 0$. Put $d(P) = 1$ when the stored decision says $\sup S_P \neq 0$. If $\sup S_P \neq 0$ then $\neg P$ is impossible, so $\neg\neg P$; if $\neg\neg P$ and $\sup S_P = 0$, then P would give $1 \leq 0$, so $\neg P$, a contradiction. So $d(P) = 1$ exactly when $\neg\neg P$. *The structure gives MP:* Theorem 28.17. Together with Theorem 28.15 this proves the forward directions of items 1 and 2.

The construction. Let d be a WEM function and assume MP. Every statement stable under double negation is decided by d: $x \leq y$, $x = y$, and, with MP, $x < y$ and $x \# 0$. This gives the stored decisions, and the inverse x^{-1} on the points where d reports $x \neq 0$, and 0 elsewhere. For the supremum of a set S let $u_S(q) = d(\eta(q)$ is an upper bound of $S)$; "is an upper bound" is a universal statement about $\leq$, stable under double negation, so $u_S(q) = 1$ exactly when $\eta(q)$ is an upper bound. If d reports that S is not not inhabited and bounded, then it is not not true that $u_S(-m) = 0$ for some m and $u_S(m') = 1$ for some m', and MP, applied to the binary sequences $m \mapsto 1 - u_S(-m)$ and $m \mapsto u_S(m)$, makes both true; put $\sup S = c$, the point of Lemma 28.18. Otherwise put $\sup S = 0$. On an inhabited set with an upper bound d reports the guard, and the lemma gives the laws; a set with no upper bound and the empty set both fail the guard, so both have supremum 0. The infimum is $\inf S = -\sup(-S)$. For item 2 the same construction with the strict order $\neg\neg <$ needs no MP for the order itself, but the supremum still uses MP for its integer brackets.

Item 3. $d(P) = 1$ gives $\neg\neg P$ and $d(P) = 0$ gives $\neg P$. Conversely WEM says that each P has exactly one bit b with $b = 1$ exactly when $\neg\neg P$, and unique choice collects these bits into one map. □

By Proposition 28.4, item 3, the WEM function with MP gives the LPO function, so the complete ordered field sits above the ordered field by exactly a double-negation decision. Under unique choice (Remark 28.5) item 3 identifies the WEM function with WEM, and the complete ordered field costs exactly MP together with WEM.

Two further laws of a classical library have prices of their own. Interpolation between two sets asks for a point between them; it holds in every complete ordered field (take $c = \sup s$)

and already gives WEM. The witness property, that a point below the supremum lies below some member, goes higher: it gives excluded middle.

Proposition 28.20 (Interpolation and witnesses)**.**

1. *Suppose that for any two inhabited sets s, t of points with $a \le b$ for all $a \in s$ and $b \in t$ some point c has $a \le c \le b$ for all such a, b. Then* WEM *holds.*
2. *Let* sup *be as in Theorem 28.17, and suppose that for every inhabited set s with an upper bound and every point b with $\neg\neg(b < \sup s)$ some member a has $\neg\neg(b < a)$. Then excluded middle holds.*
3. *In a complete ordered field, if every inhabited set s and every $b < \sup s$ have a member a with $b < a$, excluded middle holds.*

No choice principle is used.

Proof. 1. Let s consist of 0, and of 1 if P; let t consist of 1, and of 0 if $\neg P$. Every member of s lies below every member of t, since P and $\neg P$ do not both hold. Let c lie between them. Cotransitivity at $0 < 1$ gives $0 < c$ or $c < 1$. If $0 < c$, then $\neg P$ would put 0 in t and give $c \le 0$; so $\neg\neg P$. If $c < 1$, then P would put 1 in s and give $1 \le c$; so $\neg P$.

2. Let Q be a proposition with $\neg\neg Q$ and let s consist of 0, and of 1 if Q. Then Q gives $1 \le \sup s$ and hence $0 < \sup s$, so $\neg\neg(0 < \sup s)$. The hypothesis gives a member a with $\neg\neg(0 < a)$. The member 0 is excluded, so $a = 1$ and Q holds. So $\neg\neg Q \to Q$ for every Q, which gives excluded middle at $Q = P \vee \neg P$.

3. The same set, with MP from Theorem 28.17 turning $\neg\neg(0 < \sup s)$ into $0 < \sup s$. □

Separation and compactness

A classical library builds its topology on the line from the open intervals and expresses compactness through filters. Read on $\mathbb{R}_\delta$: the *interval topology* is the family of sets generated by the intervals $\{x : a < x < b\}$ under finite intersections and arbitrary unions. It is *Hausdorff* when any two points $x \neq y$ lie in disjoint open sets. A *filter* is a family of sets closed under finite intersections and supersets and containing the whole line; it is *proper* when the empty set is not a member. A point x is a *cluster point* of a filter F when every open set containing x meets every member of F, in the sense that their intersection is not empty. A set K is *filter compact* when every proper filter containing K has a cluster point in K.

Every open set of the interval topology contains, with each of its points x, all points within $1/(K+1)$ of x for some K; this holds for the generating intervals and is preserved by finite intersections and unions. Conversely, the points within $1/(K+1)$ of x contain an open interval around x. So x is a cluster point of F exactly when for every K and every $A \in F$ it is not the case that no point within $1/(K+1)$ of x lies in A.

Theorem 28.21 (Hausdorff separation and filter compactness)**.**

1. *The interval topology is Hausdorff exactly when MP holds.*
2. *If a filter-compact set contains two points $p < q$, then* WEM *holds.*
3. $[0,1]$ *is filter compact exactly when* WEM *holds and countable unique choice into* $\{0,1\}$ *holds: whenever every n has exactly one bit b with $R(n,b)$, some binary sequence f has $R(n, f(n))$ for every n.*

No choice principle is used beyond the one named.

Proof. 1. Under MP, $x \neq y$ gives $x < y$ or $y < x$ (Corollary 12.18); a ratio q between them separates them by the intervals below and above q. Conversely, let $x \neq y$ lie in disjoint open sets U and V. Some interval $a < z < b$ around y lies in V. Cotransitivity at $a < y$ and at $y < b$ gives $a < x$ or $x < y$, and $x < b$ or $y < x$. If $a < x < b$ then $x \in U \cap V$, which is empty; so $x < y$ or $y < x$, and every point $x \neq y$ is apart from y. This is MP by Corollary 12.18.

2. Let K be the set. For a proposition P let F_P consist of the sets A with $P \to p \in A$ and $\neg P \to q \in A$. It is a filter, it contains K, and it is proper, since $\emptyset \in F_P$ would give $\neg P$ and $\neg\neg P$. Let $x \in K$ be a cluster point. Under P the set $\{p\}$ is a member, and $p < x$ would give the open set $\{z : p < z < x + 1\}$ around x, which misses it; so $P \to \neg(p < x)$. Likewise $\neg P \to \neg(x < q)$. Cotransitivity at $p < q$ gives $p < x$, hence $\neg P$, or $x < q$, hence $\neg\neg P$.

3. Filter compactness of $[0,1]$ gives WEM by item 2. For unique choice, let R be given. For a bit string c of length M let $I(c)$ be the interval of length 3^{-M} with left end $\sum_{i<M} 2c_i 3^{-(i+1)}$, the interval of the Cantor set coded by c, and call c *valid* if $R(i, c_i)$ for every $i < M$. Valid strings of every length exist, and by uniqueness there is only one of each length. Let C_N be the set of points of $[0,1]$ that lie, for every $M \leq N$, in $I(c)$ for some valid c of length M, and let F consist of the sets containing some C_N. The left end of the valid string of length N lies in C_N, so F is proper; let x be a cluster point in $[0,1]$. Then $x \in I(c)$ for every valid c: if x lay outside, an open set around x would miss $I(c)$, which contains C_M. Fix a regular sequence t naming x; the goal is a proposition. Define f by recursion: $f(n) = 1$ when stage $12 \cdot 3^n - 1$ of t exceeds the midpoint of $I(f(0) \dots f(n-1))$, and $f(n) = 0$ otherwise. If the first n values are valid and $R(n,b)$, then x lies in the third of that interval coded by b; the two thirds lie $1/(6 \cdot 3^n)$ from the midpoint, and the stage lies within $1/(12 \cdot 3^n)$ of x, so $f(n) = b$. By induction $R(n, f(n))$ for every n.

Conversely, assume WEM and countable unique choice into $\{0,1\}$, and let F be a proper filter containing $[0,1]$. Call a closed dyadic interval I *good* if no member A of F has every point of I outside A. Goodness is stable under double negation, so WEM decides it. The interval $[0,1]$ is good, since F is proper. If I is good and its left half is not, its right half is good: otherwise members A_L and A_R miss the two halves, and $A_L \cap A_R$ misses I, because every point of I lies, up to double negation, in one half and the conclusion is a negation. Let the halving path go left when the left half is good and right otherwise. For each n exactly one bit is the nth step of this path, so countable unique choice gives the path as a binary sequence f, and every interval on it is good. Its point $x = [f]$ lies in $[0,1]$; the points within $1/(K+1)$ of x contain the Kth interval of the path, which is good. So x is a cluster point. □

Filter compactness is therefore a logical principle, weak excluded middle for all propositions together with a little choice, unlike Heine–Borel for $[0,1]$ with presented covers, which is the decidable fan theorem (Theorem 13.6). Under unique choice (Remark 28.5) filter compactness of $[0,1]$ is exactly WEM.

Conditions and scope.

Base and premises
$\mathbb{R}_\delta$; the classical supremum, topology and filters read by their content.

Conclusion
A total supremum gives MP; the complete ordered field costs exactly MP and the WEM function; interpolation gives WEM, and the witness property gives excluded middle; Hausdorff separation is exactly MP; filter compactness of $[0,1]$ is exactly WEM with countable unique choice into $\{0,1\}$.

Not claimed
That the witness property is needed by the mathematics that uses it: it is the price of one way of stating suprema.

28.6 The derivative convention

A classical library has a derivative operator defined on every function: it returns the derivative where one exists and 0 where none does. The convention is a jump of an operator, and the rule of Section 28.2 gives a first price. Applied to the square root, the operator also predicts when an event will occur, and that costs Markov's principle.

Definition 28.22 (Pointwise derivative and the convention)**.** A point d is a *derivative* of $f : \mathbb{R}_\delta \to \mathbb{R}_\delta$ at x if for every k some j has

$$|f(y) - f(x) - d(y-x)| \le \frac{|y-x|}{k+1} \qquad \text{whenever } |y-x| \le \frac{1}{j+1}.$$

A *derivative convention* is an operator $\mathscr{D}$ assigning to every map f a map $\mathscr{D}(f)$ such that $\mathscr{D}(f)(x) = d$ whenever d is a derivative of f at x, and $\mathscr{D}(f)(x) = 0$ whenever f has no derivative at x. A *differentiability decision* is a map c from pairs (f,x) to $\{0,1\}$ with $c(f,x) = 1$ exactly when it is not false that f has a derivative at x.

Derivatives are unique: if d and d' are derivatives at x, then $|d-d'|\,|y-x| \le 2|y-x|/(k+1)$ for y close to x, and $y = x + 1/(j+1)$ gives $|d-d'| \le 2/(k+1)$ for every k. Under the classical display of $\mathbb{R}_\delta$ on the classical real numbers this derivative is the classical one, and so is pointwise continuity; that is a statement of classical mathematics, Theorem 28.56, item 4.

Theorem 28.23 (The derivative convention gives Markov's principle)**.** *If a derivative convention exists, MP holds. No choice principle is used.*

Proof. Let $R(y) = \sqrt{\max(y,0)}$ be the free map of Theorem 28.9, item 5. First, for $c > 0$ the point $e = 1/(2c)$ is a derivative of R at c^2. If $|y - c^2| \le c^2$ then $y \ge 0$, and with $s = R(y) = \sqrt{y}$,

$$R(y) - c - e(y - c^2) = (s-c) - e(s-c)(s+c) = -e(s-c)^2,$$

while $|s-c|(s+c) = |y-c^2|$ and $s+c \geq c$ give $|s-c| \leq |y-c^2|/c$. So the error is at most $|y-c^2|^2/(2c^3)$, which is at most $|y-c^2|/(k+1)$ once $|y-c^2| \leq 2c^3/(k+1)$. At $c = 2^{-(n+1)}$ the derivative is 2^n.

Now let α be a binary sequence that is not all zero. Let $\beta(n) = 1$ when $\alpha(n) = 1$ and $\alpha(m) = 0$ for every $m < n$, so that $c = [\beta]$ is $2^{-(n_0+1)}$ if the first 1 of α is at n_0, and 0 if α has no 1. The value $u = \mathscr{D}(R)(c^2)$ is one point, so $|u| < B$ for some natural number B. Search $\alpha(0), \ldots, \alpha(B)$. If one of them is 1, it is the witness. If none is, and $\alpha(n) = 1$ for some n, the first 1 lies at some $n_0 > B$; then $c = 2^{-(n_0+1)} > 0$, the convention forces $u = 2^{n_0} > n_0 > B$, and this contradicts $|u| < B$. So α would be all zero, which it is not; the search therefore finds the witness. □

The operator, evaluated at one point that depends on α, returns a number that bounds in advance where the first event of α can be.

Theorem 28.24 (The exact price of the derivative convention)**.**

1. *A derivative convention exists exactly when MP holds and a differentiability decision exists.*
2. *A differentiability decision gives the WLPO function, and the WEM function gives a differentiability decision.*
3. *A derivative convention gives the LPO function, hence* $\mathrm{LPO}_{\mathrm{seq}}$ *and* Π^0_2 *weak excluded middle (Definition 28.25).*
4. *MP and the WEM function together give a derivative convention.*

No choice principle is used.

Proof. The convention gives the WLPO function. For $t \geq 0$ let $g_t(x) = x + t|x|$. At $t = 0$ it is the identity, with derivative 1 at 0. For $t > 0$ it has no derivative at 0: take n with $1/(n+1) < t$; a derivative d at tolerance $1/(2n+2)$, tested at $y = \pm 1/(j+1)$, gives $|1 + t - d| \leq 1/(2n+2)$ and $|1 - t - d| \leq 1/(2n+2)$, hence $2t \leq 1/(n+1) < t$, which is impossible for $t > 0$. So $h(y) = \mathscr{D}(g_{|y|})(0)$ is 1 at $y = 0$ and 0 at every $y \# 0$: a jump of size 1, and Theorem 28.8 gives the WLPO function and with it a zero test z.

Item 1, forward. MP is Theorem 28.23. A map f has a derivative at x exactly when $f + \mathrm{id}$ has one, the two differing by 1. Put $g(f,x) = \mathscr{D}(f + \mathrm{id})(x) - \mathscr{D}(f)(x)$: it is 1 when f has a derivative at x and 0 when it has none. Let $c(f,x) = 1 - z(g(f,x))$. If $c(f,x) = 1$ then $g(f,x) \neq 0$, and since "no derivative" would give $g = 0$, it is not false that a derivative exists. If it is not false that a derivative exists and $g(f,x) = 0$, then a derivative would give $g = 1$, so there is none, a contradiction; so $g(f,x) \neq 0$ and $c(f,x) = 1$.

Item 1, backward. Fix f and x, and for a ratio q put

$$R_q(y) = \max\bigl(\min\bigl(f(y) - f(x) - q(y-x),\ y - x\bigr),\ 0\bigr).$$

Let d be a derivative of f at x. (i) If $d \leq q$, then 0 is a derivative of R_q at x: for $y \leq x$ the inner minimum is at most $y - x \leq 0$, so $R_q(y) = 0$; for $y \geq x$, $f(y) - f(x) - q(y-x) \leq |y-x|/(k+1) + (d-q)(y-x) \leq |y-x|/(k+1)$, so $0 \leq R_q(y) \leq |y-x|/(k+1)$; the bound is a non-strict inequality, so the case split $y \leq x$ or $x \leq y$, which holds up to double negation,

is allowed. (ii) If R_q has a derivative e at x, then $d \le q$: points $y < x$ give $|e| \le 1/(k+1)$ for every k, so $e = 0$; if $q < d$, points $y > x$ close to x give $R_q(y) \ge \min\big((d-q)/2, 1\big)(y-x)$, which exceeds $|y-x|/(k+1)$ for large k.

Now let c be a differentiability decision and assume MP. If $c(f,x) = 0$, put $\mathscr{D}(f)(x) = 0$; then f has no derivative at x. If $c(f,x) = 1$, let $u(q) = c(R_q, x)$ and let S be the set of derivatives of f at x, which is not empty up to double negation and has at most one member. By (i) and (ii), $u(q) = 1$ exactly when $\eta(q)$ is an upper bound of S: if $\eta(q)$ bounds S, a member d makes 0 a derivative of R_q, so it is not false that R_q has one; and if $u(q) = 1$, every member d has $\neg\neg(d \le q)$ by (ii), hence $d \le q$. Up to double negation a member d exists, and then $u(-m) = 0$ for $-m < d$ and $u(m') = 1$ for $m' > d$; MP makes these integers exist. Let $\mathscr{D}(f)(x)$ be the point that Lemma 28.18 computes from the test u; it is an upper bound of S below every upper bound. When f has a derivative d, $S = \{d\}$, and that point is d.

2. For a WEM function d put $c(f,x) = d(f \text{ has a derivative at } x)$. A differentiability decision c gives the zero test $y \mapsto c(g_{|y|}, 0)$: by the first paragraph $g_{|y|}$ has a derivative at 0 exactly when $y = 0$, and $y = 0$ is stable under double negation. Theorem 28.3 gives the WLPO function.

3. The first paragraph and Theorem 28.23 give the WLPO function and MP, which make the LPO function (Theorem 28.3, item 1). Proposition 28.4 gives $\mathrm{LPO}_{\mathrm{seq}}$: a sequence h with $h(N) = 1$ exactly when $M(N,m) = 1$ for some m; the LPO function applied to $1 - h$ decides whether $h(N) = 1$ for every N, that is, it decides Π_M outright.

4. Item 2 and item 1. □

Applied row by row and then to the resulting sequence, as in item 3, the LPO function decides every arithmetic statement whose matrix is decidable, by induction on the number of quantifiers. So the derivative convention carries every arithmetic principle of the price lattice, far above the Π^0_2 weak excluded middle that its jump alone suggests. Its exact price places it between the LPO function, which it implies, and MP with the WEM function, which the complete ordered field of Theorem 28.19 pays: the library's own complete field pays for its own derivative. Under unique choice (Remark 28.5) the upper bound is MP with WEM. Whether a differentiability decision alone gives Π^0_2 weak excluded middle, and whether it lies strictly above the WLPO function, is open.

Definition 28.25 (Π^0_2 weak excluded middle and the shift)**.** For a double sequence of bits M write Π_M for the statement that every row has a 1: $\forall N \exists m\, M(N,m) = 1$. Π^0_2 *weak excluded middle* states $\neg\Pi_M \vee \neg\neg\Pi_M$ for every M. The *shift* states the same for every M whose rows are each not not inhabited: $\forall N \neg\neg\exists m\, M(N,m) = 1$. The *double-negation shift* DNS states that $\forall n\, \neg\neg Q(n)$ gives $\neg\neg\forall n\, Q(n)$ for every family of propositions Q.

Theorem 28.26 (The Π^0_2 gap above the WLPO function)**.** *Assume the WLPO function. Then* Π^0_2 *weak excluded middle holds exactly when the shift holds. The shift follows from MP and from* DNS. *No choice principle is used.*

Proof. Let w be the WLPO function and M a double sequence of bits. The bit $t = w\big(N \mapsto w(M(N,\cdot))\big)$ is 1 exactly when no row is all zero, that is, when every row is not not inhabited. If $t = 0$, then Π_M is refuted, since it would make every row inhabited. If $t = 1$, the shift applies.

The converse is the special case. Under MP a row that is not not inhabited is inhabited, so Π_M holds; under DNS the rows give $\neg\neg\Pi_M$. □

So the question whether the WLPO function gives Π^0_2 weak excluded middle is exactly the question whether it gives the shift. For WLPO as a sentence of intuitionistic second-order arithmetic the answer is no (Section 28.8): there the shift follows from MP but not from WLPO. For the WLPO function the question stays open.

Conditions and scope.

Base and premises
$\mathbb{R}_\delta$; the pointwise derivative of Definition 28.22; an operator defined on all maps.

Conclusion
The derivative convention costs exactly MP together with a decision of differentiability up to double negation. It gives the LPO function, and the complete ordered field pays for it.

Not claimed
That the decision of differentiability is strictly above the WLPO function, or that it gives Π^0_2 weak excluded middle; both are open. Reading derivatives through supplied derivatives, as in Section 28.7, avoids the convention and its price; what monotonicity then costs is open.

28.7 Monotonicity from a pointwise derivative

"A function with a nonnegative derivative is increasing" is a classical theorem proved through the mean value theorem. On $\mathbb{R}_\delta$ we read it with the derivative supplied: f and a map f' are given, and $f'(z)$ is a derivative of f at z for every z in $[0,1]$. No convention is involved. Call this statement MON:

> for all maps f, f' with $f'(z)$ a derivative of f at z and $f'(z) \geq 0$ for every $z \in [0,1]$, $f(x) \leq f(y)$ whenever $0 \leq x \leq y \leq 1$.

Its strict form asks $f(x) < f(y)$ for $x < y$ when $f'(z) > 0$ on $[0,1]$, and its segment form asks the conclusion on every segment $[x,y]$ of the line with the hypotheses on that segment only. The difficulty is familiar from compactness: the derivative gives, at each point, a radius on which f is nearly linear, and nothing says how the radii vary. We give two routes around it.

Lemma 28.27 (The affine chain rule)**.** *If d is a derivative of f at $x + tc$, then cd is a derivative of $s \mapsto f(x + sc)$ at t. No principle is used.*

Proof. Given k, take a natural number $B \geq |c|$ and the j that the derivative of f at $u = x + tc$ gives for the tolerance $1/((B+1)(k+1))$. If $|s - t| \leq 1/((B+1)(j+1))$ and $u' = x + sc$, then $|u' - u| = |s-t|\,|c| \leq 1/(j+1)$, so $|f(u') - f(u) - d(u'-u)| \leq |u'-u|/((B+1)(k+1)) \leq |s-t|/(k+1)$, and $d(u' - u) = cd(s - t)$. □

Lemma 28.28 (Monotonicity from one sequence of representatives)**.** *Let $x \le y$, let $f'(z)$ be a derivative of f at z with $f'(z) \ge 0$ for every $z \in [x, y]$, and suppose regular sequences $G(n, m)$ name the values $f\big(x + (m/2^n)(y - x)\big)$ for all natural numbers n and m. Then $f(x) \le f(y)$. No principle is used.*

Proof. The conclusion is a negation, so we may assume $f(y) - f(x) \le -c$ with $c = 1/(M + 1)$ and derive a contradiction. Write $a_n = x + (m_n/2^n)(y - x)$ and $b_n = a_n + 2^{-n}(y - x)$ for the current interval, and $\beta_n = c\,2^{-(n+1)}(1 + 2^{-n})$ for its budget. We choose m_n so that $f(b_n) - f(a_n) \le -\beta_n$; at $n = 0$ this is the assumption. Let p be the midpoint. Since $f(p) - f(a_n)$ and $f(b_n) - f(p)$ add up to at most $-\beta_n$, one of them is at most $-\beta_n/2$, and $\beta_n/2 - \beta_{n+1} = c\,2^{-(2n+3)}$. Read a ratio r within $c\,2^{-(2n+5)}$ of $f(p) - f(a_n)$ off late stages of the sequences G. If $r \le -\beta_n/2 + c\,2^{-(2n+5)}$, take the left half: then $f(p) - f(a_n) \le -\beta_n/2 + c\,2^{-(2n+4)} \le -\beta_{n+1}$. Otherwise $f(p) - f(a_n) > -\beta_n/2$, so $f(b_n) - f(p) < -\beta_n/2 \le -\beta_{n+1}$, and we take the right half. Each step compares two ratios, so the bisection is a rule.

The intervals shrink to a point $z \in [x, y]$. Choose k with $(y - x)/(k + 1) < c/4$ (take $B \ge y - x$ and $k + 1 > 4B/c$) and the j of the derivative at z for the tolerance $1/(k + 1)$, and n with $2^{-n}(y - x) \le 1/(j + 1)$. Then a_n and b_n lie within $1/(j + 1)$ of z, on either side of it, and

$$f(b_n) - f(a_n) \ge f'(z)(b_n - a_n) - \frac{b_n - a_n}{k + 1} \ge -\frac{2^{-n}(y - x)}{k + 1} > -c\,2^{-(n+2)},$$

while the budget gives $f(b_n) - f(a_n) \le -\beta_n \le -c\,2^{-(n+1)}$. This is a contradiction. □

The only object the proof uses beyond f and f' is the double sequence of representatives G. A section supplies it, and since the conclusion is a negation, it is enough that such representatives are not impossible. Say that *representation choice up to double negation* holds when for every sequence of points it is not false that some sequence of regular sequences names it. A map f is *presented* when it is computed by an operation F on regular sequences: $F(s)$ names $f(x)$ whenever s names x.

Theorem 28.29 (Monotonicity at the WLPO function)**.**

1. MON *holds under each of: the WLPO function; representation choice up to double negation, and in particular countable choice; the full fan theorem* FAN *of Definition 13.2.*
2. MON *holds with no principle for every presented map f, and in particular for every map written with constants, $+$, $\cdot$, $-$, $|\cdot|$,* max, min, exp *and π.*
3. *With no principle,* MON*, its strict form and its segment form are equivalent.*
4. *Under the WLPO function, on every segment $[x, y]$: if f' is a derivative of f and f'' of f' with $f'' > 0$ on $[x, y]$ and $x < y$, then $f((1 - t)x + ty) < (1 - t)f(x) + tf(y)$ for $0 < t < 1$; and if $f' = 0$ on $[x, y]$ then $f(x) = f(y)$.*

No choice principle is used beyond the one named.

Proof. 1. The WLPO function gives a section (Theorem 28.11), hence G, and Lemma 28.28 applies. Under representation choice up to double negation, it is not false that G exists,

and the conclusion is a negation. For FAN, assume again a gap $f(y) - f(x) \leq -c$, choose k with $(y - x)/(k + 1) < c$, and call the dyadic subinterval I of $[x, y]$ coded by a binary string *fine* if some $z \in I$ and some j satisfy the derivative estimate at z for the tolerance $1/(k + 1)$ within $1/(j + 1)$, and I has length at most $1/(j + 1)$. Fineness is a property of strings, not a decidable one. Every binary sequence names a point z of $[x, y]$, and the strings along it become fine once their intervals are shorter than the j of z; so the fine strings form a bar, and FAN makes it uniform at some depth N. Since the goal is a negation, we may decide fineness for the finitely many strings of length at most N; the shortest fine prefixes of the strings of length N tile $[x, y]$ by finitely many fine intervals $[a, b]$ with tags z, and on each, $f(b) - f(a) \geq f'(z)(b - a) - (b - a)/(k + 1) \geq -(b - a)/(k + 1)$. Summing gives $f(y) - f(x) \geq -(y - x)/(k + 1) > -c$.

2. For a presented f, fix regular sequences naming x and y; the dyadic points $x + (m/2^n)(y - x)$ have sequences computed from them, and F turns these into G. A map written with the listed operations is presented, since each operation acts on regular sequences.

3. Strict from MON: let $x < y$ in $[0, 1]$. Since $f'(x) > 0$, there is m with $f'(x) \geq 1/(m + 1)$; the derivative at x for the tolerance $1/(2m+2)$ gives j, and $h = (y-x)/(j+1)$ satisfies $0 < h \leq y-x$ and $f(x + h) - f(x) \geq h/(m + 1) - h/(2m + 2) > 0$. MON on $[x + h, y]$ gives $f(x + h) \leq f(y)$, so $f(x) < f(y)$. MON from the strict form: assume a gap $f(y) - f(x) \leq -1/(M + 1)$ and put $\varepsilon = 1/(2M + 2)$; $f + \varepsilon\mathrm{id}$ has the positive derivative $f' + \varepsilon$, and $x < y$ would give $f(x) + \varepsilon x < f(y) + \varepsilon y \leq f(x) - 1/(M + 1) + \varepsilon$, which is false; and $x = y$ contradicts the gap. Since $x \leq y$ gives $x < y$ or $x = y$ up to double negation, the gap is impossible. The segment form from MON: apply MON to $g(t) = f(x + t(y - x))$, which has the derivative $(y - x)f'(x + t(y - x)) \geq 0$ by Lemma 28.27; $g(0) \leq g(1)$ is $f(x) \leq f(y)$, and no case split on $x = y$ is needed.

4. Constancy: apply the segment form to f and to $-f$. Strict convexity: the strict segment form, applied to f', makes f' strictly increasing on $[x, y]$. Let $z = (1 - t)x + ty$ and $\varphi(s) = f(s) - f(z) - f'(z)(s - z)$, with derivative $\varphi'(s) = f'(s) - f'(z)$. On $[z, y]$ this is nonnegative, so $\varphi \geq 0$ there; beyond the midpoint z' of $[z, y]$ it is at least $f'(z') - f'(z) > 0$, so the strict form gives $\varphi(y) > \varphi(z') \geq 0$. The same argument on $[x, z]$, with $s \mapsto \varphi(x + z - s)$, gives $\varphi(x) > 0$. Since $(1 - t)(x - z) + t(y - z) = 0$, the difference $(1 - t)f(x) + tf(y) - f(z)$ equals $(1 - t)\varphi(x) + t\varphi(y) > 0$. □

The fan route needs a bar given by an arbitrary property. It follows from a logical principle with a little choice.

Theorem 28.30 (Excluded middle with unique choice gives the full fan theorem)**.** *Assume excluded middle and countable unique choice into* $\{0, 1\}$*. Then every bar, given by an arbitrary property of binary strings, is uniform: the full fan theorem* FAN *holds.*

Proof. Let B be a bar and call a string s *bounded* if some N has every binary sequence extending s meet B at a length at most N. If both one-digit extensions of s are bounded, so is s. Excluded middle decides at each string whether its left extension is bounded, and so defines a path from the root: go left when the left extension is not bounded, and right otherwise. For each n exactly one bit is the nth step, so countable unique choice collects the steps into one binary sequence p. Suppose the root is not bounded. Then by induction no string along p is bounded:

if s is not bounded and its left extension is bounded, the right one is not. But p meets the bar at some string, and a string in the bar is bounded with N its length. So the root is bounded, which is the conclusion; excluded middle decides whether it is. □

This agrees with Theorem 13.11: without unique choice, excluded middle does not give even the decidable fan theorem. Under the default reading of this book (Remark 28.5) unique choice is present, and excluded middle gives FAN. So MON costs at most the WLPO function, at most countable choice of representatives, and at most excluded middle with countable unique choice into $\{0, 1\}$; every principle that it implies is implied by the WLPO function.

Whether MON holds with no principle at all is open. One model shows that monotonicity can hold where these routes fail. In the model of named objects of the computable calculus (Definition 16.20), sequences range over computable sequences and points over points with computable names. There every computable sequence of points has a computed sequence of representatives, so representation choice holds; the bisection of Lemma 28.28 uses the derivative only at the point it constructs and the values of f only at dyadic points, so the statement of monotonicity holds; Kleene's tree (Lemma 16.29) refutes every fan theorem; and the WLPO and LPO functions fail, as after Proposition 28.4. These are facts about that model, proved in a classical metatheory: the bisection runs in a world without fan theorems and without omniscience functions. They are not a measurement of MON inside the computable calculus, whose language states order relations through atoms that none of its rules inspects; there the calculus derives from monotonicity no statement whose order atoms range over sets no larger than the continuum, such as the fan theorems, that it does not already derive outright.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; a map with a supplied pointwise derivative on $[0, 1]$ or on a segment.

Conclusion

Monotonicity, its strict form, monotonicity on every segment, strict convexity and constancy cost at most the WLPO function; monotonicity is free for presented maps; it also follows from the full fan theorem, which excluded middle with countable unique choice into $\{0, 1\}$ gives.

Not claimed

That monotonicity costs anything: whether it is free is open.

28.8 WLPO and Π^0_2 weak excluded middle

Theorem 28.26 reduced one question to another: given the WLPO function, Π^0_2 weak excluded middle holds exactly when its shift does, and the shift follows from MP and from the double-negation shift. Does WLPO alone give it? We answer for WLPO as a sentence of intuitionistic second-order arithmetic HAS (Definition 7.14), with the systems of truth values of Definition 7.19: it does not, and the separating model needs no hypothesis beyond the usual axioms of set theory. The theorem concerns HAS, whose comprehension axiom makes every

definable set available; it is not transferred here to the WLPO function in the reading of maps of this part, for which the question stays open in the form Theorem 28.26 gives it.

Definition 28.31 (Π^0_2 sentences of HAS)**.** With the pairing $\langle N, m\rangle = (N+m)\cdot(N+m)+m$ and the formulas of Definition 7.18, put $\mathrm{Row}(X,N) :\equiv \exists m\, \langle N,m\rangle \in X$ and $\mathrm{All}(X) :\equiv \forall N \mathrm{Row}(X,N)$, and

$$\Pi^0_2\text{-WEM} :\equiv \forall X\big(\mathrm{Dec}(X) \to \neg\mathrm{All}(X) \vee \neg\neg\mathrm{All}(X)\big),$$
$$\text{Shift} :\equiv \forall X\big(\mathrm{Dec}(X) \to \forall N \neg\neg\mathrm{Row}(X,N) \to \neg\mathrm{All}(X) \vee \neg\neg\mathrm{All}(X)\big).$$

The model is built from a quotient of a power set. A *σ-ideal* on a set Ω is a family J of subsets that contains $\emptyset$, contains every subset of each of its members, and contains every countable union of its members; we assume $\Omega \notin J$. Its *Stone space* X_J is the set of ultrafilters on Ω that contain no member of J, with the topology whose basic open sets are $[S] = \{u \in X_J : S \in u\}$. Four standard facts of Stone duality [110, 106] hold. (a) $[S]\cap[T] = [S\cap T]$, $[S]\cup[T] = [S\cup T]$ and the complement of $[S]$ is $[\Omega \setminus S]$, so every basic set is closed as well as open. (b) $[S] = \emptyset$ exactly when $S \in J$, so $[S] \subseteq [T]$ exactly when $S \setminus T \in J$; a set $S \notin J$ lies in some point of X_J, because S together with the complements of the members of J has the finite intersection property. (c) X_J is compact, as a closed subspace of the space of all ultrafilters on Ω. (d) An open subset of a basic set $[c]$ whose complement in $[c]$ is also open is basic: it is closed in the compact set $[c]$, hence a finite union of basic sets.

For maps $g, h : \mathbb{N} \to \mathbb{N}$ write $g \le^* h$ when $g(n) \le h(n)$ for all large n. A family of maps is *bounded* when one h satisfies $g \le^* h$ for all its members. The *bounding number* $\mathfrak{b}$ is the least cardinality of an unbounded family [111]. Let $J_{\mathfrak{b}}$ be the σ-ideal on $\Omega = \mathbb{N}^{\mathbb{N}}$ generated by the sets of cardinality less than $\mathfrak{b}$: its members are the sets covered by countably many such sets. Every member of $J_{\mathfrak{b}}$ is bounded, because a set of size less than $\mathfrak{b}$ is bounded and countably many bounding maps h_k are bounded by $n \mapsto \max_{k\le n} h_k(n)$. Since $\mathbb{N}^{\mathbb{N}}$ is unbounded, $\Omega \notin J_{\mathfrak{b}}$.

Theorem 28.32 (The bounding frame)**.** *In the system of truth values formed by the open sets of the Stone space of $J_{\mathfrak{b}}$, WLPO has the value $\top$ and* Shift *does not. Consequently HAS derives neither* WLPO $\to$ Shift *nor* WLPO $\to \Pi^0_2$-WEM*. The proof reasons in classical set theory with the axiom of choice and uses no further hypothesis.*

Proof. WLPO is valid in the Stone space of every σ-ideal. Let X_0 be valued by open sets A_n, and let a basic set $[c]$ lie in the value of $\mathrm{Dec}(X_0)$. For each n, $[c]$ is covered by the disjoint open sets A_n and $\neg A_n$, so by (d) $A_n \cap [c] = [b_n]$ with $b_n \subseteq c$. Let $T = \bigcup_n b_n$. The set $[c \setminus T]$ meets no A_n, so it lies in the value Z' of $\mathrm{Z}(X_0)$. The set $[c \cap T]$ is disjoint from Z': if a basic $[R] \subseteq Z'$ met it, then $R \cap c \cap T = \bigcup_n (R \cap b_n)$ would be outside J, so some $R \cap b_n$ would be outside J, and $[R \cap b_n]$ would be a nonempty set inside both $Z' \subseteq \neg A_n$ and A_n. So $[c \cap T] \subseteq \neg Z'$, and $[c] = [c \setminus T] \cup [c \cap T]$ lies in the value of $\mathrm{Z}(X_0) \vee \neg\mathrm{Z}(X_0)$. The basic sets form a base, so the value of the implication is $\top$ for every X_0.

A criterion for the shift. Let $a(N,m) \subseteq \Omega$ satisfy (i) $\Omega \setminus \bigcup_m a(N,m) \in J$ for every N, and (ii) with $K_f = \bigcap_N \bigcup_{m \le f(N)} a(N,m)$ for $f : \mathbb{N} \to \mathbb{N}$: every $T \subseteq \Omega$ with $K_f \setminus T \in J$ for all f contains some $Y \notin J$ with $Y \cap K_f \in J$ for all f. We show that the value of Shift is not $\top$. Value $\langle N,m\rangle \in X_0$ by $[a(N,m)]$ and every other number by $\emptyset$. Every value is clopen, so $\mathrm{Dec}(X_0)$ has the value $\top$. Each union $\bigcup_m [a(N,m)]$ is dense: a basic $[R] \neq \emptyset$ has $R \notin J$, so by (i) and countable additivity

some $R \cap a(N,m) \notin J$. So $\forall N \neg\neg \mathrm{Row}(X_0, N)$ has the value $\top$. Let V be the value of $\mathrm{All}(X_0)$, the interior of $\bigcap_N \bigcup_m [a(N,m)]$. Every $[K_f]$ lies in V, since $K_f \subseteq \bigcup_{m \le f(N)} a(N,m)$ for each N. Conversely, a basic $[R] \subseteq V$ is covered, for each N, by the sets $[a(N,m)]$, hence by compactness by finitely many, $m \le f(N)$; so $R \setminus K_f \in J$ for that f, by countable additivity. If Shift had the value $\top$, its instance at X_0 would make $\neg V \cup \neg\neg V$ the whole space. Then $\neg V$ would be clopen, $\neg V = [b]$ and $\neg\neg V = [\Omega \setminus b]$. Since $[K_f] \subseteq V \subseteq [\Omega \setminus b]$, the set $T = \Omega \setminus b$ satisfies $K_f \setminus T \in J$ for all f; let $Y \subseteq T$ be given by (ii). A basic set inside $[Y] \cap V$ has the form $[R]$ with $R \setminus Y \in J$ and $R \setminus K_f \in J$ for some f, so $R \subseteq (R \setminus Y) \cup (Y \cap K_f) \cup (R \setminus K_f) \in J$ and $[R] = \emptyset$. So $[Y] \cap V = \emptyset$, $[Y] \subseteq \neg V = [b]$; but $Y \subseteq \Omega \setminus b$, so $[Y] = \emptyset$ and $Y \in J$, a contradiction.

The instance. Let $J = J_{\mathfrak{b}}$ and $a(N,m) = \{g : g(N) = m\}$. The rows cover Ω, so (i) holds, and K_f is the set of g with $g \le f$ everywhere. For (ii), let T satisfy $K_f \setminus T \in J$ for all f. First, for every h some $y \in T$ has $y \ge h$ everywhere: otherwise the box of maps g with $h \le g \le h+1$ lies in $K_{h+1} \setminus T$, hence in J; but the map $t \mapsto h + \chi_t$, where χ_t is the indicator of the set of codes $\langle n, t(n)\rangle$, sends $\mathbb{N}^{\mathbb{N}}$ injectively into the box, and the preimages of countably many sets of size less than $\mathfrak{b}$ would cover $\mathbb{N}^{\mathbb{N}}$ by bounded sets, making it bounded. Next, index an unbounded family as $(u_\gamma)_{\gamma<\mathfrak{b}}$. For each γ the earlier members are fewer than $\mathfrak{b}$, hence bounded by some e_γ; choose $y_\gamma \in T$ with $y_\gamma \ge \max(e_\gamma, u_\gamma)$ everywhere. The set $Y = \{y_\gamma\}$ lies in T and is unbounded, since a bound would bound every u_γ; so $Y \notin J$. For each f take γ_0 with $u_{\gamma_0} \not\le^* f$. For $\gamma \ge \gamma_0$, $u_{\gamma_0} \le^* y_\gamma$, so $y_\gamma \not\le f$ and $y_\gamma \notin K_f$. So $Y \cap K_f$ has fewer than $\mathfrak{b}$ members and lies in J.

Consequences. By Lemma 7.21 a derivable sentence has the value $\top$ in every system of truth values. The value of WLPO $\to$ Shift is the value of Shift, which is not $\top$. HAS derives Π^0_2-WEM $\to$ Shift, since the shift only adds a hypothesis, so Π^0_2-WEM has a value below that of Shift, and WLPO $\to$ Π^0_2-WEM is not derivable either. □

The ideal $J_{\mathfrak{b}}$ is generated by the small sets rather than given as a family of small sets. That is why the proof needs no assumption on the cofinality of $\mathfrak{b}$ and no scale: condition (ii) is checked one f at a time, and each check meets fewer than $\mathfrak{b}$ points.

Proposition 28.33 (Placement over HAS)**.** *HAS derives* MP $\to$ Shift, WLPO $\to$ (Shift $\to$ Π^0_2-WEM), Π^0_2-WEM $\to$ Shift *and* LPO $\to$ WLPO $\wedge$ Shift. *It derives none of* WLPO $\wedge$ Shift $\to$ LPO, Shift $\to$ WLPO, Shift *and* $\neg$Shift. *So, over HAS, WLPO is strictly weaker than* WLPO $\wedge$ Shift, *which is equivalent to* WLPO $\wedge$ Π^0_2-WEM, *that is, to* Π^0_2-WEM *itself (Theorem 7.62), and strictly weaker than LPO;* Shift *is comparable with WLPO in neither direction; and MP does not hold in the bounding frame (its value there is not* $\top$*).*

Proof. Let X be decidable with every row not not inhabited. Under MP, comprehension gives for each N the decidable set Y_N of the m with $\langle N, m\rangle \in X$; it is not empty up to double negation, so MP gives a member, and $\mathrm{All}(X)$ follows, hence $\neg\neg\mathrm{All}(X)$. Under WLPO, let X be decidable and let comprehension give B, the set of N with $\mathrm{Z}(Y_N)$; WLPO at each Y_N shows that B is decidable, and WLPO at B gives $\mathrm{Z}(B)$, in which case every row is not not inhabited and the shift applies, or $\neg\mathrm{Z}(B)$, in which case $\mathrm{All}(X)$ would empty B, so $\neg\mathrm{All}(X)$. The third derivation drops a hypothesis, and the fourth combines LPO $\to$ WLPO and LPO $\to$ MP (Proposition 7.35) with the first.

For the non-derivations use Lemma 7.21. In $\beta\mathbb{N}$ the closure of an open set is open, so every instance of $\neg U \vee \neg\neg U$ has the value $\top$ (Section 7.13); hence Shift has the value $\top$, as does

WLPO, while LPO does not (Table 7.2). In the one-point compactification of an uncountable set MP has the value $\top$ and WLPO does not (Table 7.2); by the first derivation the value of Shift is at least that of MP, hence $\top$. The bounding frame gives Shift a value other than $\top$, and $\beta\mathbb{N}$ gives $\neg$Shift the value $\bot$. In the bounding frame the value of MP lies below that of Shift, so it is not $\top$ either. □

Conditions and scope.

Base and premises
Intuitionistic second-order arithmetic HAS with full comprehension; models are systems of truth values built in classical set theory with choice.

Conclusion
Over HAS, WLPO does not imply Π^0_2 weak excluded middle, not even for decidable matrices whose rows are each not not inhabited; MP does imply that restricted form.

Not claimed
The same separation for the WLPO function in the reading of maps of this part, without unique choice; there the question is the one Theorem 28.26 reduces to the shift, and it is open.

28.9 Names, a space-filling curve and tail signs

Chapter 21 priced the decision whether a given cost obeys Recognition Science's composition law. For a cost given as a bare map, parts 3 and 4 of its price ladder (Theorem 21.29) place the price between two zero tests: the *function zero test* FZ, which decides for every map H on $\mathbb{R}_\delta$ whether $H(x) = 0$ for every x, and the *pair zero test* PZ, the same for maps of two points. A map of two points could hide a nonzero value where no map of one point can see it. We show that it cannot: the two tests are equivalent, with no choice principle, and so deciding the law costs exactly one zero test, as Theorem 21.29 records. The proof splits on whether a map can vanish on a dense grid without vanishing, and in the case where it cannot, a space-filling curve reaches every pair of points.

Lemma 28.34 (A hidden value gives the WLPO function)**.** *If a map f on $\mathbb{R}_\delta$ vanishes at every triadic point $z/3^n$, with z an integer, and $f(x_0) > 0$ for some point x_0, then the WLPO function exists. No choice principle is used.*

Proof. Fix a regular sequence s naming x_0; the goal is a proposition. For a binary sequence α with first 1 at b, let g_b be a triadic ratio within $1/(4b+4)$ of s_{4b+3}. The *switch point* of α is named by the sequence whose stage k is s_{4k+3} if $\alpha(0) = \cdots = \alpha(k) = 0$, and g_b once the first 1 has appeared at some $b \le k$. It is regular: two stages before the switch differ by at most $1/(4k+4) + 1/(4m+4)$, a stage k before it and a later one differ by at most $1/(4k+4) + 2/(4b+4) < 1/(k+1)$ since $b > k$, and stages after it are equal. If α is all zero the switch point is x_0; otherwise it is the triadic point g_b. So $v = f(\text{switch}(\alpha))$ is $c = f(x_0)$ or 0, up to double negation, and it is c exactly when α is all zero. Choose ratios $0 < q_1 < q_2 < c$. Cotransitivity at $q_1 < q_2$ gives $q_1 < v$ or $v < q_2$, and a search through the stages of a sequence naming v finds one of the two. If $q_1 < v$, then $v \neq 0$ and so $v = c$; if $v < q_2$, then $v \neq c$ and so

$v = 0$. The search therefore decides the proposition $v = c$, and its answer depends only on v. Put $d(\alpha) = 1$ exactly when $v = c$; then $d(\alpha) = 1$ exactly when α is all zero. □

Lemma 28.35 (Schoenberg's curve on the line)**.** *Let p be the map of $\mathbb{R}_\delta$ with $p(t) = 0$ for $t \le \frac{1}{3}$, $p(t) = 3t - 1$ on $[\frac{1}{3}, \frac{2}{3}]$, $p = 1$ on $[\frac{2}{3}, \frac{4}{3}]$, $p(t) = 5 - 3t$ on $[\frac{4}{3}, \frac{5}{3}]$ and $p = 0$ on $[\frac{5}{3}, 2]$, continued with period 2 for $t \ge 0$, and put*

$$\sigma_1(t) = \sum_{k \ge 0} \frac{p(3^{2k} t)}{2^{k+1}}, \qquad \sigma_2(t) = \sum_{k \ge 0} \frac{p(3^{2k+1} t)}{2^{k+1}}.$$

Then $\sigma = (\sigma_1, \sigma_2)$ is a map from $\mathbb{R}_\delta$ to pairs of points, and whenever $x, y \in [0, 1]$ have binary expansions (b_n) and (c_n), the point $t = \sum_k 2a_k 3^{-(k+1)}$ with $a_{2n} = b_n$ and $a_{2n+1} = c_n$ lies in $[0, 1]$ and has $\sigma(t) = (x, y)$. No principle is used.

Proof. p is the piecewise-linear interpolation of its values at the thirds $j/3$, $j \ge 0$, clamped to $[0, 1]$, with slope at most 3, so each term is a map, its value lies in $[0, 2^{-(k+1)}]$, and the series converge with the explicit bound 2^{-K} on the tail after K terms; σ is the limit. For $t = \sum_k 2a_k 3^{-(k+1)}$ write $3^k t = E + 2a_k/3 + r$, where $E = \sum_{i<k} 2a_i 3^{k-1-i}$ is an even integer and $0 \le r \le \sum_{i>k} 2 \cdot 3^{k-1-i} = \frac{1}{3}$. By periodicity $p(3^k t) = p(2a_k/3 + r)$, which is 0 if $a_k = 0$ and 1 if $a_k = 1$. Hence $\sigma_1(t) = \sum_n b_n 2^{-(n+1)} = x$ and $\sigma_2(t) = y$. □

This is Schoenberg's form of Lebesgue's curve [108]. On $\mathbb{R}_\delta$ the curve itself is free; what it needs from a point is a binary expansion, and Theorem 12.61 prices those at LLPO with countable choice. A map of two points can be read along the curve wherever names exist.

Theorem 28.36 (One zero test decides every law)**.** *With no choice principle, the function zero test* FZ *is equivalent to the pair zero test* PZ*, and both are equivalent to deciding, for every map C on $\mathbb{R}_\delta$, whether C obeys the composition law $C(s + t) + C(s - t) = 2C(s)C(t) + 2C(s) + 2C(t)$ for all s, t.*

Proof. Chapter 21 shows, with no choice principle, that deciding the law for every map gives FZ, that PZ gives the decision of the law, and that FZ decides countably many function zero tests at once. It remains to derive PZ from FZ.

A dichotomy up to double negation. Say that grids are dense if every map vanishing at every triadic point vanishes everywhere. If there is no WLPO function, then by Lemma 28.34, applied to f and to $-f$, a map vanishing on the triadic points has no value apart from 0, so every value is 0: grids are dense. Hence it is not false that the WLPO function exists or grids are dense.

A countable test. Let H be a map of two points. We claim that $H(x, y) = 0$ for all x, y exactly when (a) H vanishes at every pair of triadic points and (b) for every pair of integers (i, j) and every t, $H(\sigma_1(t) + i, \sigma_2(t) + j) = 0$. One direction is clear. For the other, $H(x, y) = 0$ is stable under double negation, so we may use the dichotomy. With the WLPO function, a decided order places x in $[i, i + 1]$ and y in $[j, j + 1]$ for integers i, j, Theorem 28.11 gives binary expansions of $x - i$ and $y - j$, and Lemma 28.35 gives t with $\sigma(t) = (x - i, y - j)$; (b) gives $H(x, y) = 0$. If grids are dense, then for a triadic y_0 the map $x \mapsto H(x, y_0)$ vanishes on the

grid by (a), hence everywhere, and then for each x the map $y \mapsto H(x, y)$ vanishes on the grid, hence everywhere.

The decision. Condition (a) is a sequence zero test, which is the function zero test of the interpolation of the sequence, and (b) is a countable family of function zero tests, indexed by the pairs (i, j). With FZ all of them are decided at once, so FZ decides whether H vanishes. □

So the price of deciding the law for a bare cost, calibrated or not, is exactly the function zero test: the two bounds that Chapter 21 proves for it, FZ below and PZ above, meet. The proof used names of points only up to double negation, and the WLPO function only to obtain them. Do names up to double negation need more than WLPO? The answer turns on one sequence of signs.

Definition 28.37 (Tail signs)**.** A *signed-digit stream* is a sequence d with digits in $\{-1, 0, 1\}$. For $n \le m$ put $P_{n,m} = \sum_{n \le k < m} d_k 2^{m-1-k}$, an integer, so that $2^{-m} P_{n,m}$ is the part of the tail $\sum_{k \ge n} d_k 2^{-(k+1)}$ before position m. A binary sequence f *reads the tail signs* of d if for all $n \le m$: $f(n) = 1$ gives $P_{n,m} \ge -1$, and $f(n) = 0$ gives $P_{n,m} \le 1$. The *tail-sign principle* states that for every signed-digit stream it is not false that some binary sequence reads its tail signs. *Names up to double negation* states that for every point x of $[0, 1]$ it is not false that x has a binary expansion.

So $f(n) = 1$ asserts that the tail from n is nonnegative and $f(n) = 0$ that it is nonpositive, each checked at every finite stage with a slack of one unit of the last place. One sequence f reads the signs of all tails of one stream.

Theorem 28.38 (Names and tail signs)**.**

1. *Names up to double negation hold exactly when the tail-sign principle holds; so WLPO gives names up to double negation exactly when it gives the tail-sign principle.*
2. *Under names up to double negation, every pair of points of $[0, 1]^2$ is not not a value of Schoenberg's curve on $[0, 1]$.*
3. *Names up to double negation follow from MP, from the double negation of MP, and from the double-negation shift for Σ^0_1 rows: if every row of a double sequence of bits is not not inhabited, then it is not false that every row is inhabited. WLPO with countable unique choice into $\{0, 1\}$ gives every point of $[0, 1]$ a binary expansion.*
4. *WLPO decides each tail sign: for every stream d and every n some bit c satisfies the condition on $f(n)$ for all $m \ge n$.*
5. *No continuous rule reads tail signs on any cylinder: if Φ maps streams to binary sequences and each value $\Phi(d)(n)$ depends only on finitely many digits of d, then for every stream d_0 and every L some stream that agrees with d_0 below L has tail signs that Φ misreads.*

No choice principle is used beyond the one named.

Proof. 1. Assume the tail-sign principle and let $x \in [0, 1]$. It has a signed-digit expansion d with partial sums S_n (Proposition 12.62), so $S_n - 2^{-n} \le x \le S_n + 2^{-n}$. Since the goal is

a negation, we may take f reading the tail signs of d. Decide each comparison $j/2^n \le x$ by the table $B(n,j) = 1$ if $j < 2^n S_n$, $B(n,j) = 0$ if $j > 2^n S_n$, and $B(n,j) = f(n)$ if $j = 2^n S_n$. The first two entries are correct by the bounds on x; in the third, $f(n) = 1$ says that every later window lies at or above S_n, so $S_n \le x$, and $f(n) = 0$ says $x \le S_n$. The bisection of Theorem 12.61, run on this table, gives a binary expansion. Conversely, assume names up to double negation and let d be a stream. The point $x = \frac{1}{2} + \sum_k d_k 2^{-(k+2)}$ lies in $[0,1]$ and in the window $[(2^m + 2^m S_m - 1)/2^{m+1}, (2^m + 2^m S_m + 1)/2^{m+1}]$ for every m. Take a binary expansion with numerators a_n, and put $f(n) = 1$ exactly when $2^n + 2^n S_n \le a_{n+1}$. If $f(n) = 1$, then x lies at or above $(2^n + 2^n S_n)/2^{n+1}$, and comparing with the window at $m \ge n$ gives $P_{n,m} \ge -1$; if $f(n) = 0$, then x lies at or below it, and the window gives $P_{n,m} \le 1$.

2. Under the goal's double negation, take expansions of x and y and apply Lemma 28.35.

3. Let $x \in [0,1]$ and suppose x has no binary expansion. Then x differs from every dyadic ratio q in $(0,1)$, since a dyadic ratio has an expansion; so it is not false that some stage of a sequence naming x witnesses $x < q$ or $q < x$. These are the rows of a double sequence of bits, indexed by q. If every row were inhabited, a search would find the first witness in each row and give a table of correct comparisons, and the bisection would give an expansion, a contradiction. So the shift for Σ^0_1 rows gives the contradiction. MP gives that shift, since it makes every row inhabited, and the double negation of MP suffices because the conclusion is a negation. With WLPO each comparison $q \le x$ is decided by a unique bit, and countable unique choice collects the bits over the countably many dyadic q into a table.

4. Fix n. WLPO applied to the decidable sequence "$P_{n,m} < -1$", $m \ge n$, says that all $P_{n,m} \ge -1$, and then $c = 1$ serves, or that not all are. In the second case $c = 0$ serves: $P_{n,m} \le 1$ is decidable, and $P_{n,m} \ge 2$ for one m would make all pieces at least -1, because $P_{n,m+e} = 2^e P_{n,m} + \sum_{i<e} d_{m+i} 2^{e-1-i} \ge 2^{e+1} - (2^e - 1)$ and, for earlier m', $2^{m-m'} P_{n,m'} \ge P_{n,m} - (2^{m-m'} - 1) > -2^{m-m'}$.

5. Let d_1 agree with d_0 below L and be 0 from L on. The value $\Phi(d_1)(L)$ depends on the first K digits; let $K' = \max(K, L)$ and change d_1 only at position K', to -1 if $\Phi(d_1)(L) = 1$ and to $+1$ otherwise. The new stream d has $\Phi(d)(L) = \Phi(d_1)(L)$, while $P_{L,K'+2} = \mp 2$ violates the condition that $\Phi(d)(L)$ imposes. □

Item 4 shows that WLPO decides each sign separately. Item 5 shows why collecting the signs is hard: a rule that reads them cannot be continuous in the stream. A model in which every function is computed continuously from a generic stream should therefore refute the tail-sign principle while keeping WLPO. We build one for a calculus of numbers and functions.

The *stream calculus* has two sorts, numbers and functions from numbers to numbers. Number terms are variables, 0, successors, and a function variable applied to a number term. The atomic formulas are equations and a tail atom $\mathrm{Tail}(d, f, n, m)$, which states the condition of Definition 28.37 on $f(n)$ at m for the stream d, with each value v of d read as the digit $v - 1$ and each value of $f(n)$ other than 1 read as 0. Formulas are built with $\bot$, $\wedge$, $\vee$, $\to$ and quantifiers over both sorts. The rules are those of intuitionistic natural deduction for both sorts, reflexivity and substitution of equals, decidability of every atom, the successor axioms, induction for every formula, explicit definition of a function by a number term, and choice for atomic matrices: from $\forall n\, \exists m\, A(n,m)$ with A atomic infer $\exists F\, \forall n\, A(n, F(n))$, which is what a least-number search gives. Read with numbers in $\mathbb{N}$ and functions ranging over all maps,

every rule is sound, so the calculus is a fragment of ordinary mathematics; its sentence WLPO, $\forall\alpha(\forall n\,\alpha(n) = 0 \vee \neg\forall n\,\alpha(n) = 0)$, reads as WLPO and its tail-sign sentence as the tail-sign principle.

Theorem 28.39 (WLPO without tail signs)**.** *In the stream calculus, the tail-sign sentence is not derivable from WLPO, and neither are MP and its double negation. The proof reasons in classical mathematics, with the axiom of choice.*

Proof. Let $Y = \{0,1,2\}^{\mathbb{N}}$ with its product topology, the space of streams. A *condition* is an open set $U \subseteq Y$. A number datum on U is a map $a : Y \to \mathbb{N}$ that is locally constant on the closure of U: near each stream of the closure it depends only on finitely many digits. A function datum is a map $Y \to \mathbb{N}^{\mathbb{N}}$ each of whose values at a fixed argument is a number datum. Forcing is defined by recursion. U forces an atom when the atom holds at every stream of U, and forces $\bot$ when U is empty; $\wedge$ is pointwise; U forces $A \vee B$ when open $V, W \subseteq U$ force A and B and $V \cup W$ is dense in U; U forces $A \to B$, or a universal statement, when every open $V \subseteq U$ that forces A forces B, or every open $V \subseteq U$ forces every instance by a datum on V; and U forces an existential statement when finitely many open pieces of U, with data, force instances and their union is dense in U. The finiteness in the last clause is what makes Markov's principle fail below; the failure of the tail-sign principle does not use it.

Soundness. By induction on derivations, every derivable sequent is forced by every condition with data that forces its hypotheses. The induction uses that forcing passes to smaller open sets, that the empty set forces everything, that a statement forced on finitely many open pieces whose union is dense in U is forced by U, that terms built from locally constant data are locally constant, and that a decidable atom splits U into an open part where it holds and an open part where it fails, with dense union. The compactness of Y reduces each locally constant datum to finitely many constant pieces, which is what the finite existential clause needs. The choice rule holds because the least witness of an atomic matrix, computed stream by stream, is again locally constant.

WLPO is forced everywhere. Let α be a function datum on an open V. Let P be the interior of the set of streams in V at which α is all zero, and $Q = V \setminus \overline{P}$. Then P forces $\forall n\,\alpha(n) = 0$. An open $W \subseteq Q$ forcing it lies in that set, hence in P, hence is empty; so Q forces $\neg\forall n\,\alpha(n) = 0$. $P \cup Q$ is dense in V.

Tail signs are not forced. Interpret d by the generic stream, the datum $y \mapsto y$; its digits are forced to lie in $\{0,1,2\}$. Suppose a nonempty condition W forces that some f reads the tail signs. Then a nonempty open piece of W carries a function datum f with $f(y)$ reading the tail signs of y at every stream y of the piece. On a cylinder inside the piece, each value $f(y)(n)$ depends on finitely many digits of y, so f is a continuous rule reading tail signs on that cylinder, which Theorem 28.38, item 5, excludes. So every condition forcing the existence of f is empty, Y forces its negation, and Y does not force its double negation. Since Y forces WLPO, soundness shows that the tail-sign sentence is not derivable from WLPO.

The negation of MP is forced everywhere. Let a nonempty open V force MP and take a cylinder $C \subseteq V$ that fixes the first L digits. Let α be the function datum with $\alpha(y)(n) = 1$ when $n \geq L$ and digit n of y is 2, and 0 otherwise; each value reads one digit. Every nonempty open $W \subseteq C$ contains a cylinder that also fixes some digit $n \geq L$ to 2, and that cylinder forces $\exists n\,\alpha(n) = 1$ with a constant witness; so only the empty set forces $\neg\exists n\,\alpha(n) = 1$ below C, and C forces

$\neg\neg\exists n\,\alpha(n) = 1$. Since C forces MP, it forces $\exists n\,\alpha(n) = 1$: finitely many open pieces of C, with number data b_i, force $\alpha(b_i) = 1$, and their union is dense in C. The stream $z \in C$ that is 0 from place L on lies in the closure of that finite union, hence in the closure of one piece. Near z the datum b_i is constant on that closure, say equal to N, and the streams of the piece near z agree with z at place N; there $\alpha(b_i) = 1$ asks for $N \geq L$ and digit 2 at place N, while z has a digit other than 2 there. A nonempty open part of the piece fails the atom, a contradiction. So Y forces the negation of MP, and since Y forces WLPO, neither MP nor its double negation follows from WLPO. □

The theorem is exact for the stream calculus, a fragment of ordinary mathematics with induction and choice for atomic (decidable) matrices. It does not cover quotients, propositions as objects or universes, so it is not a theorem about the full base of this part; there, whether WLPO gives names up to double negation remains open, and by Theorem 28.38 it is exactly the question whether WLPO gives the tail-sign principle.

Conditions and scope.

Base and premises
$\mathbb{R}_\delta$ for Lemmas 28.34 and 28.35 and Theorems 28.36 and 28.38; the stream calculus and a forcing model built in classical mathematics with the axiom of choice for Theorem 28.39.

Conclusion
Deciding the composition law for every bare cost costs exactly one function zero test. Names up to double negation are exactly the tail-sign principle; WLPO decides each tail sign but, in the stream calculus, does not give the principle.

Not claimed
The separation for the full base of this part.

28.10 Moving the library: density, cones and counts

The sections so far priced the structures of a classical library. Its lemmas are another matter: there are thousands of them, and reading each on $\mathbb{R}_\delta$ by hand is not a method. Two theorems move whole classes of lemmas at no cost, and their failures mark where the remaining prices sit.

Take the triangle inequality $|a + b| \leq |a| + |b|$. At two ratios it is a statement about integers, true by a case distinction on signs, which is decidable for ratios. On $\mathbb{R}_\delta$ the case distinction is not available. But every point is the limit of its own rational stages, both sides are continuous, and an inequality $\leq$ is a negation, which survives a limit. So the inequality holds on the line, with no principle. The same argument fails for $(x \cdot x - 2)^2 > 0$: it holds at every ratio, because $a^2 \neq 2b^2$ for integers with $b \neq 0$, and it fails at $\sqrt{2}$. A strict inequality does not survive a limit.

Definition 28.40 (Terms of ordered algebra)**.** For a natural number n, the *terms in n variables* are built from the variables $x_1, \dots, x_n$ and ratios by $+$, $\cdot$, $-$, $|\cdot|$, max and min. A term t has a value $t(x)$ at every point $x \in \mathbb{R}_\delta^n$, and a ratio value at every $q \in \mathbb{Q}_\delta^n$, with $t(\eta(q)) = \eta(t(q))$. A *strict atom* is $s < t$; a *closed atom* is $s \le t$ or $s = t$. *Terms with inverse* may also contain u^{-1}; such a term has a value at x when every subterm u that is inverted has a value apart from 0 there, and an atom over terms with inverse is *defined* at x when both its sides have values.

Ratios are where a classical lemma can be checked as it stands. Their order is decided, and on them the classical structures cost nothing.

Proposition 28.41 (The ratios carry the classical order)**.** *The ratios $\mathbb{Q}_\delta$, with the order that η induces from $\mathbb{R}_\delta$, form a linear-order structure in the sense of Definition 28.13, with decided $\le$, $=$ and $<$, and satisfy the strict-ordered-ring laws; η preserves sums, products, $<$ and $\le$. No principle is used.*

Proof. For ratios with positive denominators, $\eta(u/d) < \eta(v/e)$ exactly when $ue < vd$ (Proposition 12.19), a comparison of integers, so $<$ is decided; $\le$ is its negated converse and $=$ the conjunction of two $\le$, both decided. Totality is trichotomy of integers. The strict-ordered-ring laws are the corresponding facts about integers after clearing denominators, and η preserves sums and products by Proposition 12.19. □

Theorem 28.42 (Density transfer)**.** *Let $s_1 < t_1, \dots, s_k < t_k$ be strict atoms and $a_1 \trianglelefteq_1 b_1, \dots, a_l \trianglelefteq_l b_l$ closed atoms, all over terms in n variables.*

1. *If the closed atoms hold at every $q \in \mathbb{Q}_\delta^n$ at which the strict atoms hold, they hold at every $x \in \mathbb{R}_\delta^n$ at which the strict atoms hold.*
2. *The same is true with one further hypothesis $a \ne b$, or $\neg(a \le b)$, for terms a, b, added on both sides.*
3. *For terms with inverse: if at every $q \in \mathbb{Q}_\delta^n$ at which the strict atoms hold, with both sides defined, and the closed atoms are defined, the closed atoms hold, then the same is true at every $x \in \mathbb{R}_\delta^n$.*

No principle is used.

Proof. 1. Three facts carry the proof. (i) A point x named by a regular sequence s is the limit of the ratios $\eta(s_m)$: $|x - \eta(s_m)| \le 1/(m+1)$. For a tuple, use one sequence per coordinate; the conclusion is a proposition, so the sequences may be fixed for the proof. (ii) Every term is continuous with an explicit modulus, by recursion on the term: sums and negatives are 1-Lipschitz in each argument, a product is Lipschitz on a bounded region, and $|\max(a,b) - \max(c,d)| \le \max(|a-c|, |b-d|)$, likewise for min and $|\cdot|$. So $t(\eta(s_m)) \to t(x)$ for every term. (iii) If $u < v$, then $u_m < v_m$ for all late m whenever $u_m \to u$ and $v_m \to v$: a positive rational lower bound ε of $v - u$ is beaten once both approximations lie within $\varepsilon/3$.

Now let the strict atoms hold at x, and let q_m be the tuple of mth stages. By (ii) and (iii) the strict atoms hold at q_m for all late m, so the closed atoms hold there. A closed atom $a \le b$ passes

to the limit: if $b(x) < a(x)$, then (iii) gives $b(q_m) < a(q_m)$ for late m, against $a(q_m) \le b(q_m)$; so $\neg(b(x) < a(x))$, which is $a(x) \le b(x)$. An equation is two such inequalities.

2. The conclusions are stable under double negation, so we may argue under one. The hypothesis $a \neq b$ is $\neg(a = b)$, which is $\neg\neg(a \# b)$ because $a = b$ is $\neg(a \# b)$; under the double negation $a < b$ or $b < a$. In each case add it as a strict atom and apply part 1: at a ratio point where the enlarged list holds, $a \neq b$ holds, so the conclusions hold there by the hypothesis. The hypothesis $\neg(a \le b)$ is $\neg\neg(b < a)$ and is treated the same way.

3. Apartness from 0 is an open condition in the sense of (iii), so at q_m for late m every inverted subterm of the strict atoms and of the closed atoms has a value apart from 0, bounded away from 0 by a fixed ratio. On such a region the inverse is Lipschitz: $w_m - w = w_m w(u - u_m)$ when $u_m w_m = uw = 1$, and $|u| \ge 1/(M+1)$ gives $|w| \le M + 1$. So (ii) holds for terms with inverse at points where they are defined, and the argument of part 1 applies. □

The theorem is free because the conclusion is closed. Its two limits are exact.

Proposition 28.43 (What density cannot move)**.**

1. Strict conclusions. *The term $b(x) = (x \cdot x - 2)^2$ has $0 < b(q)$ at every ratio q, and $b(x) = 0$ for the point $x = \sqrt{2}$.*
2. Closed hypotheses. *The implication $x \cdot x = 2 \to 0 = 1$ holds at every ratio and fails at $\sqrt{2}$.*

No principle is used.

Proof. For a ratio u/d with $d > 0$, $b(u/d) = (u^2 - 2d^2)^2/d^4$, and $u^2 \neq 2d^2$ because the exponent of 2 in u^2 is even and in $2d^2$ odd; so $b(u/d)$ is a positive ratio. The square root of 2 (Proposition 12.19, item 5) satisfies $\sqrt{2} \cdot \sqrt{2} = 2$, so $b(\sqrt{2}) = 0$, and the closed hypothesis of item 2 holds there while $0 = 1$ does not. □

Closed hypotheses in solved form are the exception to item 2. A hypothesis $x_i \le t$, $t \le x_i$ or $x_i = t$, where t does not contain x_i, is removed by substituting $x_i := t - |y|$, $t + |y|$ or t with a new variable y; the result is a statement of Theorem 28.42.

Strict conclusions need a different argument, and an old one supplies it. Krivine and Stengle showed that a polynomial inequality holds on a semialgebraic set exactly when an algebraic identity certifies it [107, 109]. We need only the easy half, the identity itself, and on $\mathbb{R}_\delta$ it costs nothing.

Definition 28.44 (Cone expressions)**.** Fix strict hypotheses $l_i < r_i$ ($i \le m$) and closed hypotheses $l'_j \le r'_j$ ($j \le m'$) over terms in n variables. The *nonnegative expressions* are built from the slacks $r_i - l_i$ and $r'_j - l'_j$, from 1 and from squares $t \cdot t$ of terms, by sums and products. The *cone expressions* are built from the strict slacks $r_i - l_i$ and 1 by sums and products, and by adding a nonnegative expression to a cone expression. Each expression has a value at every $x \in \mathbb{R}_\delta^n$.

Theorem 28.45 (Cone transfer and cone refutation). *Let l, r be terms in n variables, and let $x \in \mathbb{R}_\delta^n$ satisfy the strict and the closed hypotheses.*

1. *If cone expressions k_1, k_2 satisfy $k_1 \cdot (r - l) = k_2$ at every point of $\mathbb{R}_\delta^n$, then $l(x) < r(x)$. The same holds when k_1 is only a nonnegative expression.*
2. *If a cone expression k, formed over the strict hypotheses enlarged by $r < l$, vanishes at every point of $\mathbb{R}_\delta^n$, then $l(x) \le r(x)$.*

No principle is used, and no density argument.

Proof. At x every strict slack is positive and every closed slack and every square is nonnegative. By recursion on the expression, every nonnegative expression is ≥ 0 at x and every cone expression is > 0 at x: sums and products of positive points are positive, and a positive point plus a nonnegative one is positive.

1. At x, $k_1 \cdot (r - l) = k_2 > 0$ with $k_1 \ge 0$. A positive product has both factors apart from 0, since $|k_1|\,|r - l| > 0$ gives each factor a positive absolute value; with $k_1 \ge 0$ this makes $k_1 > 0$, and then $r - l = k_2 \cdot k_1^{-1} > 0$.

2. The conclusion $l \le r$ is $\neg(r < l)$. If $r(x) < l(x)$, the enlarged hypotheses hold at x, so $k(x) > 0$; but $k(x) = 0$. □

For $0 < a < b$ the identity $1 \cdot (b \cdot b - a \cdot a) = (b - a) \cdot \big((b - a) + a + a\big)$ is a certificate for $a \cdot a < b \cdot b$: its right side is a product of cone expressions over the slacks $b - a$ and $a - 0$. Once such an identity is written down, checking it is algebra; finding it is a search, and nothing in the theorem depends on how it was found.

Proposition 28.46 (Positive cones). *The positive points of $\mathbb{R}_\delta$ form a commutative group under δ's product, with no principle. A linear-order structure on the positive points whose strict order is δ's $<$ exists exactly when the LPO function exists, with no choice principle in either direction.*

Proof. A positive point is apart from 0 and has an inverse (Proposition 12.19, item 3), which is positive, and products of positive points are positive. A linear-order structure stores a decided $<$; the bit at the pair $(1, 1 + x_\alpha)$ is 1 exactly when $x_\alpha > 0$, that is, by (28.1), exactly when α takes the value 1, so it is the LPO function. Conversely, restrict the structure of Theorem 28.15 to the positive points. □

What moves, and what it costs: a reported audit

The counts in this subsection and in Table 28.2, like those in Sections 28.2 and 28.4, report an audit. They are not theorems of this book, and we do not reproduce the audit's row-by-row arguments. The audit read on $\mathbb{R}_\delta$ the theorems of five areas of algebra and order in the classical library that opens this chapter, its theorems on complete orders and fields, and every library theorem used in a set of classical proofs, written with that library, of the steps of the forcing chain restated in Table 22.1 and of the near-flat sector of gravity; those proofs are not part of this book either. A *row* is one library theorem, or one use of it. For each row it priced, the audit recorded a proof of an implication from a price to the statement read on $\mathbb{R}_\delta$, or a proof

Table 28.2. The classical library moved to $\mathbb{R}_\delta$: the counts of the audit described in the text, reported and not proved here. "Route" is the price an upper bound has proved, not a lower bound; "convention" marks rows that hold only through $0^{-1} = 0$ or $c \neq 0$ read as apartness.

Body of results	Rows	Free	Exact price	Other
Ordered fields, groups, rings and lattices	804 theorems	239		389 at the LPO function (route); 176 not moved by that route
The 389 rows above, by Theorems 28.42 and 28.45	389	311	9 LLPO, 5 MP	2 free in one direction; 62 convention
Lemmas of the linear order (Section 28.4)	102	56	2 LLPO, 1 MP	27 at the WLPO function (route); 16 at the LPO function, each through $0^{-1} = 0$
Complete ordered fields	324 theorems	3	8 EM with the WEM function	157 at MP and the WEM function, one of them at least WEM; 3 at the LPO function and 3 at the LPO function with EM (route); 67 no instance; 83 not read
Library theorems used by the forcing chain and gravity	560 theorems, 4,530 uses	483	3 LLPO, 2 MP	21 convention; 51 of analysis, topology and homology: 39 free, 3 through the derivative convention, 1 at most countable choice of representatives, 8 open
Uses in the forcing chain and the near-flat sector of gravity	851 uses of 134 theorems	822	1 LLPO	27 convention, at the steps' own $c \neq 0$; 1 not read

of that statement outright, and for an exact price also the converse; rows marked not read, without an instance or open carry none. Below we name the kind of argument behind each kind of row and give an example. A row *read through the convention* holds at an excluded point only through $0^{-1} = 0$, or reads a hypothesis $c \neq 0$ as apartness; its form guarded by apartness is free, and its unguarded form carries the price of the convention (Table 28.1). A route's price is an upper bound and nothing more. Table 28.2 collects the counts.

In the audit, the first two rows tell the main story. A transfer through the ordered field built from a positivity test (Theorem 28.15) moves 628 of the 804 generic theorems of five areas: ordered fields, ordered groups with zero, ordered rings and groups with absolute value, and lattices. It moves 239 free and 389 at the LPO function, the price of that ordered field. That price belongs to the route. Density moves 21 of the 389 free, and 2 more in one direction of their equivalence. Of the remaining 366, cone certificates, cone refutations, inverses guarded by apartness and induction prove 290 free; 62 hold only through the convention; and 14 carry exact prices. Nine cost exactly LLPO: six state a closed disjunction, and three split an arbitrary property into the cases of max, min or $|\cdot|$. A typical one is $a \leq \max(b, c)$ exactly when $a \leq b$ or $a \leq c$; at $a = 0$, $b = x$, $c = -x$ the left side always holds, and the right side is the dichotomy

$0 \le x \vee x \le 0$, which is LLPO (Corollary 12.18). Five cost exactly MP, such as $0 < |a|$ exactly when $a \neq 0$: since $0 < |a|$ is $a \# 0$, this is the reading of $\neq$ as apartness (Corollary 12.18). No row is left open. The 176 rows the first transfer does not move are moved by other routes, 71 of them free, except three that concern only the coherence of the classical structures; along those routes no row met a lower bound above its route's price, and two cost exactly LLPO.

So, as far as the audit reaches, the classical algebra of ordered fields is free on $\mathbb{R}_\delta$, except where it decides a closed disjunction (LLPO), reads $\neq$ as apartness (MP), or relies on $0^{-1} = 0$ at an excluded point. Completeness is different. Of the 324 theorems on complete orders and fields, 160 are proved at the price of the complete ordered field of Theorem 28.19, MP with the WEM function, or below it: three, which do not mention the supremum, are free, and one, interpolation, costs at least WEM (Proposition 28.20). Eight more, the witness rows "$b < \sup s$ gives a member $a \in s$ with $b < a$" in their several forms, cost exactly excluded middle together with the WEM function; under unique choice, excluded middle alone (Remark 28.5). Six theorems about ordered fields go through the LPO function, three of them with excluded middle as well, as an upper bound. Sixty-seven have no instance on the line: they assume a least or greatest point, a well-founded order, or a monotone product. The remaining 83 are not read.

For Recognition Science the last two rows matter. Of the 560 library theorems used in the steps of the forcing chain and in the near-flat sector of gravity, 509 are read on $\mathbb{R}_\delta$, and 483 of them are free; 21 hold only through the convention. Five carry exact prices. Three are forms of the domain law, $xy = 0$ exactly when $x = 0$ or $y = 0$, which costs LLPO (Table 12.2) and enters the equal-cost law of the second rung. One is the positivity of J away from 1, which costs exactly MP in the form "$J(x) > 0$ for $x \neq 1$" and is free for x apart from 1 (Chapter 22). One is the law "$\neg(b \le a)$ gives $a < b$", which costs exactly MP, as in Proposition 28.16, item 2; every step that uses it also has a free proof. The remaining 51, from analysis, topology and homology, are read separately. Thirty-nine are free. Three read the derivative convention through a supplied derivative (Section 28.6). One, strict convexity from a positive second derivative, costs at most countable choice of representatives, and the step that uses it, the strict convexity of J, has a free proof (Chapter 22). Eight, facts of singular homology and of embeddings, are open: none is moved to the line, while the steps that use them are free in Chapter 23: that detecting linking forces three dimensions, that a loop linking the equatorial circle of the 3-sphere bounds nothing in its complement, and that the first homology of the circle is $\mathbb{Z}$ as a relation (steps T4.4, T4.5 and R2-35 of the table in Chapter 22); the data form of the last, one winding function on all chains, costs exactly the WLPO function (Theorem 22.65). Counted by use, the 134 theorems outside analysis and topology that the transfer priced only at the LPO function or did not move have 851 uses in the forcing chain and the near-flat gravity sector; 822 of them are free; 27 read the steps' own hypotheses $c \neq 0$, and are free when an inverse is supplied with c; one is the domain law, at LLPO; and one, in the gravity sector, is not read.

28.11 Classical proofs as programs of Gödel's T

Section 20.4 showed that a classical proof of $\forall x \exists y A$, with A quantifier-free, yields a derivation without excluded middle and a program that computes y from x. That program came from realizability, argued in a classical metatheory, and it was a code of a partial recursive

function whose totality the argument supplied. For three quantifiers the strategy against counterexamples was found by unbounded search, and its form as a term of Gödel's system T was cited, not proved. This section builds the programs as terms of T, by maps on derivations, with no classical metatheory. Then it asks the question that the existence of a program leaves open: how large is the program, measured against the proof it came from?

Here is the case to keep in mind. Every number is even or odd, and here is a classical proof that differs from the one in Section 20.4. Induct on x; $0 = 0 + 0$. For the step from n to $S(n)$, split on whether n is even, which excluded middle permits although nothing decides it. If $n = y + y$, then $S(n) = S(y + y)$. If n is not even, the induction hypothesis gives $n = y + y$ or $n = S(y + y)$; the first is excluded, so $S(n) = S(S(y + y)) = S(y) + S(y)$. Written out in Peano arithmetic, the proof is a finite tree of rule applications. The compiler of Theorem 28.51 turns it into a closed term of T, of size bounded by one polynomial of degree five in the size of the proof. Its value at 7 is 3, and Theorem 28.53 says why: the compiled term meets the formula at every input, and 3 is the only y with $7 = y + y$ or $7 = S(y + y)$. The value is not the point; anyone can halve 7. The point is that a program, with a proved bound on its size, comes out of a proof whose one classical step is a case split that nothing decided.

Throughout this section A is a quantifier-free formula of arithmetic, Peano arithmetic is the calculus of Definition 8.10, and $\vdash_{\mathrm{PA}}$, $\vdash_{\mathrm{HA}}$ and the negative translation φ^N are as in Section 20.4. Every result of the section is proved with no omniscience principle and no choice principle, except Theorem 28.56 and Proposition 28.57, which reason in classical mathematics, and the last assertion of Theorem 28.55, as their statements say.

Programs and their size

Definition 28.47 (Programs of Gödel's T)**.** The finite types are those of Subsection 20.3.2, generated from O by $A \to B$. The *programs* of T are built by application, of a program of type $A \to B$ to a program of type A, from the constants

$$\mathsf{K}_{A,B} : A \to (B \to A), \qquad \mathsf{S}_{A,B,C} : (A \to (B \to C)) \to ((A \to B) \to (A \to C)),$$
$$0 : O, \qquad \mathsf{Succ} : O \to O, \qquad \mathsf{R}_A : A \to \big((O \to (A \to A)) \to (O \to A)\big),$$

one for each choice of the types in the subscripts. The value $\mathrm{val}(t)$ of a program t is defined by recursion on t: a number for type O, a function from values of A to values of B for type $A \to B$, with

$$\mathsf{K}\,a\,b = a, \qquad \mathsf{S}\,f\,g\,a = f\,a\,(g\,a), \qquad \mathsf{Succ}\,n = n + 1,$$
$$\mathsf{R}\,b\,g\,0 = b, \qquad \mathsf{R}\,b\,g\,(n+1) = g\,n\,(\mathsf{R}\,b\,g\,n),$$

and the value of an application is the value of the function at the value of the argument. A program of type $O \to O$ therefore computes a function from numbers to numbers, defined at every input.

The system is Gödel's [70], in the combinator form that needs no variables. Its recursor is primitive recursion at every finite type, and nothing in it searches without a bound. A program with variables of types $A_1, \dots, A_L$ is written as a closed program of type $A_1 \to \cdots \to A_L \to B$: bracket abstraction removes a variable, $[x]x = \mathsf{S}\,\mathsf{K}\,\mathsf{K}$, $[x]c = \mathsf{K}\,c$ for c without x, and $[x](u\,v) = \mathsf{S}\,([x]u)\,([x]v)$, and keeps the value.

Definition 28.48 (Size and shape)**.** Derivations, formulas, terms, finite types, programs, lists and natural numbers are generated by constructors. The *size* $\|o\|$ of such an object is 1 plus the sizes of the arguments of its outermost constructor; a list is built from the empty list by adding members, and a natural number counts as its value. The arguments of a node of a derivation are its list of hypotheses, the formulas and terms its rule mentions, and its subderivations; the arguments of a constant or an application of T include the types that fix its typing. The *shape* $\mathrm{sh}(A)$ of a formula is its number of connectives and quantifiers plus its number of equations and occurrences of $\bot$, an equation counting once whatever its terms. For a derivation d, $\mathrm{sh}(d)$ is the largest shape of a formula concluded at a node of d, and $\mathrm{cx}(d)$ is the largest length of a list of hypotheses at a node of d.

Because every node carries its hypotheses, a derivation that keeps K hypotheses open along K steps has size of order K^2. This matters for the exponents of Theorem 28.52, which are stated for exactly this measure.

The compilers

Theorem 28.49 (Classical proofs compile to programs of T)**.** *Let A be quantifier-free.*

1. *A map on derivations sends every derivation of φ from Γ in Peano arithmetic to a derivation of φ^N from Γ^N in Heyting arithmetic.*
2. *A map assigns to every derivation d of $\forall x\,\exists y\,A$ in Peano arithmetic from no hypotheses, and every assignment ρ of numbers to the other free variables of A, a closed program t of type $O \to O$ such that, for every number x, A holds at x, at $y = \mathrm{val}(t)(x)$ and at ρ.*
3. *A map assigns to every such derivation of $\forall x\,\exists y\,\forall z\,A$ and every ρ a closed program Φ of type $O \to ((O \to O) \to O)$ such that, for every number x and every function f from numbers to numbers, A holds at x, at $y = \mathrm{val}(\Phi)(x)(f)$, at $z = f(y)$ and at ρ: Kreisel's no-counterexample functional [55].*
4. *Read with O as the orbit $\mathbb{N}_\delta$ of Chapter 1 and the constants by the same equations, every closed program has a value, and the reading of the orbit as numbers carries it to the value above.*

No omniscience principle and no choice principle is used.

Proof. 1. The proof of Theorem 20.57 is a recursion on derivations that builds, at each rule, a derivation of the translated conclusion from derivations of the translated premises, for every rule including excluded middle. Read as a definition, it is the map. Its one arithmetic input is stability of translated equations, which uses a derivation of $s = t \vee \neg(s = t)$ in Heyting arithmetic for each pair of terms; the proof of Lemma E.1, item 2, builds one by recursion on the terms.

2 and 3. The programs come from realizability relative to a challenge. Fix a property P of numbers. To each formula assign a list of finite types: one O for an equation and for $\bot$; the two lists side by side for a conjunction; for $\varphi \to \psi$, the types of ψ, each taking the types of φ as arguments; for $\forall x\,\varphi$, those of φ, each taking an argument of type O; for $\exists x\,\varphi$ and for a

disjunction, one O in front. A list w of values of these types *realizes* a formula under P, at an assignment of its variables, by recursion on the formula: an equation when it holds or P holds of the number w; $\bot$ when P holds of w; a conjunction when the two parts of w realize the two conjuncts; $\varphi \to \psi$ when w, applied to every list realizing φ, realizes ψ; $\forall x\,\varphi$ when w applied to every number n realizes φ at $x = n$; $\exists x\,\varphi$ when the rest of w realizes φ at x equal to the first entry; a disjunction when the first entry selects a side that the rest realizes. This is Kreisel's modified realizability [7] with $\bot$ read as "P holds of a number", Friedman's translation [68] made computational, as in [105]. By recursion on a derivation of φ from Γ in Heyting arithmetic we build programs, taking as arguments the realizers of Γ and the values of the free variables, that realize φ under every P whenever their arguments realize Γ. A hypothesis is a projection. Implication introduction keeps the programs of its premise and reads the arguments of the discharged hypothesis as arguments of the type; implication elimination applies each program of the major premise, argument by argument, to the programs of the minor premise. Universal introduction lifts the programs of its premise into a context with one more number and applies them to the arguments of that context; the other quantifier rules take or supply a number. Equality substitution is a conditional on a program that decides the equation: where the equation holds it returns the programs of the substituted premise, and where it fails it copies the number that realizes the equation into every component. Induction is the recursor R; the impossible case copies the number that realizes $\bot$ into every component, which realizes every formula, since an equation is realized by any number with P. Bracket abstraction turns the arguments into a closed program.

Apply this to the translation, by item 1, of a derivation of $\forall x\,\exists y\,\forall z\,A$; it derives $\forall x\,\neg\forall y\,\neg\forall z\,A^N$. Let r be its realizer. Fix x and f, and let $P(n)$ say that A holds at x, n and $f(n)$. We need a realizer W of $\forall y(\forall z\,A^N \to \bot)$: given y and a realizer w of $\forall z\,A^N$, it must return a number with P. A program χ_A, built from A by recursion, has value 0 exactly when A holds. If A holds at x, y, $f(y)$, then W returns y. Otherwise w applied to $f(y)$ realizes A^N at $z = f(y)$, the translation of a false quantifier-free formula. At a false equation the identity realizes its negation, because a realizer of a false equation is a number with P; from these, by recursion on A, a realizer of $\neg A^N$ is assembled, and applying the realizer of A^N to it returns a number with P, which W returns. Then r applied to x and W realizes $\bot$, that is, returns a number y with $P(y)$. The program Φ of item 3 is $x, f \mapsto r\,x\,W_{x,f}$, closed by bracket abstraction. A program cannot read the assignment ρ; it receives the finitely many values of ρ that A mentions, as numerals. For item 2, apply to d, by implication elimination, a fixed derivation of $\forall x\,\exists y\,A \to \forall x\,\exists y\,\forall z\,A$, with z a variable that A does not mention; apply item 3 to the result, and apply Φ to the constant function 0: the program is $\mathsf{S}\,\Phi\,(\mathsf{K}\,(\mathsf{K}\,0))$.

4. The orbit has a start, a step and recursion along the step (Chapter 1). Reading 0, Succ and R by them and K, S by their equations, induction on the program shows that the value at the orbit element of n is the orbit element of the value at n, at every finite type. $\square$

The programs are terms of T, so they are total by their types and contain no search without a bound. Item 3 is the form that Section 20.4 cited from Gödel's functional interpretation [70, 56]; here it is proved, by a different route, and every step is a map on derivations or on programs.

Conditions and scope.

Base and premises
Peano and Heyting arithmetic as in Section 20.4; a quantifier-free matrix A; the programs of Definition 28.47.

Conclusion
Maps on derivations: classical derivations to intuitionistic ones, and classical proofs of $\forall x\,\exists y\,A$ and $\forall x\,\exists y\,\forall z\,A$ to closed programs of T computing a witness and a strategy against every counterexample function. The programs run on the orbit $\mathbb{N}_\delta$ with the same values. Free.

Not claimed
Any bound on the size of the programs; the next theorems supply one for a second compiler and show that the first has none.

How large the programs are

A compiler is a map on derivations, so it can be measured. Write t_d for the program that item 2 of Theorem 28.49 assigns to d with every parameter read as 0. Is $\|t_d\|$ bounded by a polynomial in $\|d\|$? It is not, and the reason is a step the proof left open: how a realizer that takes a list of arguments is applied, argument by argument, to another one. A realizer in a context of argument types $\Delta = (a_1, \dots, a_L)$ is a closed program of type $a_1 \to \cdots \to a_L \to B$. The direct way to apply u to v pointwise recurses on Δ,

$$\sigma_{()}(u, v) = u\,v, \qquad \sigma_{a::\Delta}(u, v) = \sigma_\Delta\big(\sigma_\Delta(\kappa_\Delta(\mathsf{S}), u), v\big),$$

where κ_Δ lifts a closed program into the context by K's. It calls itself twice for each entry.

Theorem 28.50 (The direct compiler grows exponentially)**.** *Let e_0 be the derivation of $0 = 0$ by reflexivity, and let e_{k+1} derive $0 = 0$ by equality substitution from the reflexivity derivation of $0 = 0$ and from e_k. Let d_n derive $\forall x\,\exists y(y = y)$ from e_n by existential introduction with witness 0 and universal introduction. Then, for every n,*

$$\|d_n\| \le 15(n+1) \qquad \text{and} \qquad \|t_{d_n}\| \ge 4^{n+1} \ge 2^n.$$

Consequently there are no numbers C and k with $\|t_d\| \le C(\|d\| + 1)^k$ for every quantifier-free A and every derivation d of $\forall x\,\exists y\,A$ from no hypotheses. No omniscience principle and no choice principle is used.

Proof. The derivations. The arguments of a substitution node of e_{k+1} are the empty list of hypotheses, the formula $0 = 0$, the terms 0 and 0, the reflexivity derivation and e_k, so $\|e_{k+1}\| = \|e_k\| + 10$ and $\|e_0\| = 3$; the two introductions add 12, and $\|d_n\| = 10n + 15$.

A bound read off a derivation. Since σ calls itself twice for each entry, induction on ℓ gives $\|\sigma_\Delta(u, v)\| \ge \|u\| + \|v\| + 2^\ell$ for a context Δ of length ℓ. At a node of a derivation in Heyting arithmetic, let h be the number of types of the realizers of its hypotheses, counted side by side; the context of the realizer at that node has at least h entries. Define a number $L(e)$ by recursion on the last rule of a derivation e: at an implication elimination with premises e_1

and e_2, $L(e) = \max(2^h, L(e_1), L(e_2))$; at an implication introduction or a universal introduction with premise e', and at an equality substitution whose substituted premise is e', $L(e) = L(e')$; at every other rule, $L(e) = 0$. Two facts follow by induction on e. (i) The first program of the realizer of e has at least $L(e)$ nodes. By the clauses in the proof of Theorem 28.49, at an implication elimination that program applies the first program of the major premise through σ, in a context of at least h entries, to the programs of the minor premise, so it has at least 2^h nodes and contains the first programs of both premises; an implication introduction keeps the programs of its premise; a universal introduction and an equality substitution contain the first program of the premise named. (ii) Adding hypotheses whose realizers have c types in all to every node of e keeps its rules and raises every h by c, so it multiplies $L(e)$ by 2^c.

The translation. Write e^N for the image of e under the map of item 1, the proof of Theorem 20.57 read as a recursion. Reflexivity for a becomes the closed derivation of $a \to \neg\neg a$ applied to it, and that derivation applies the hypothesis $\neg a$ to the hypothesis a, an implication elimination with $h = 2$; so $L(e_0^N) \geq 4$. A substitution from e', deriving $t = u$, and e'', deriving $A(t)$, becomes: under the hypotheses $\neg A(u)$ and then $t = u$, substitute into e''^N, apply $\neg A(u)$, discharge $t = u$, apply e'^N, discharge $\neg A(u)$ and apply stability. There e''^N stands weakened by $\neg A(u)$ and $t = u$, each realized by one program, and every rule below it is an implication elimination, an implication introduction, or the substitution with e''^N as its substituted premise. By (ii), $L(e_{k+1}^N) \geq 4L(e_k^N)$, so $L(e_k^N) \geq 2 \cdot 4^k$ for every k. The derivation of $\forall x \exists y \forall z (y = y)$ that item 2 builds from d_n applies a fixed derivation of $\forall x \exists y (y = y) \to \forall x \exists y \forall z (y = y)$ to d_n. In its image, d_n^N is the minor premise of that implication elimination; it is a universal introduction over the image of the existential introduction, which applies the hypothesis $\forall y \neg (y = y)^N$, instantiated at 0, to e_n^N weakened by that one hypothesis, realized by one program, and discharges it. So the translated derivation has $L \geq 2 \cdot 2 \cdot 4^n = 4^{n+1}$, and by (i) its realizer has at least 4^{n+1} nodes. The functional of item 3 applies that realizer and is closed by bracket abstraction, which keeps it as a part, and dropping the challenge keeps the functional; so $\|t_{d_n}\| \geq 4^{n+1} \geq 2^n$.

No polynomial. Suppose $\|t_d\| \leq C(\|d\| + 1)^k$ for every d. Put $a = C \cdot 16^k$; since $\|d_n\| + 1 \leq 16(n+1)$, we get $2^n \leq a(n+1)^k$ for every n. Let $j = k + a + 4$ and $n = 2^j - 1$. Induction from $4^2 \leq 2^4$ gives $j^2 \leq 2^j$, and $j^2 = jk + ja + 4j \geq jk + a + 16$, so $a + jk \leq 2^j - 1 = n$. Then $a(n+1)^k = a\,2^{jk} < 2^a 2^{jk} \leq 2^n$, a contradiction. □

The blowup is a property of this compiler, not of the derivations. A second compiler with the same specification stays polynomial, of degree five.

Theorem 28.51 (A polynomial compiler of degree five)**.** *There is a map F that assigns to every quantifier-free A and every derivation d of $\forall x \exists y A$ in Peano arithmetic from no hypotheses a closed program $F(d)$ of type $O \to O$ such that, for every number x, A holds at x, at $y = \mathrm{val}(F(d))(x)$ and with every parameter read as 0, and there is a constant C with*

$$\|F(d)\| \leq C(\|d\| + 1)^5 \qquad \text{for all such } A \text{ and } d.$$

No omniscience principle and no choice principle is used.

Proof. The compiler. F follows the proof of Theorem 28.49, changed in three places. Pointwise application calls itself once per entry: $\sigma'_{a::\Delta}(u, v) = \sigma'_\Delta(\mu_\Delta(u), v)$, where μ_Δ composes S with u through one composition combinator $\mathsf{B} = \mathsf{S}\,(\mathsf{K}\,\mathsf{S})\,\mathsf{K}$ per entry. The realizer is built from the

classical derivation in one pass that follows the cases of the negative translation; the realizer of each premise is built once and enters its node once, weakened by composition with a single closed program. Stability of a translated equation, $\neg\neg(s = t) \to s = t$, is realized by the fixed program that applies its argument to the identity: if $s = t$ holds, every number realizes it, and if not, the identity realizes $\neg(s = t)$, so the value realizes $\bot$, that is, satisfies P. No decision of equations is used. Realizability depends on a realizer only through its value, and each new piece has the value of the piece it replaces, so the correctness argument of Theorem 28.49 applies unchanged.

Degree five suffices. Each rule adds to the realizer at most a polynomial in the length of its context and in the sizes of the types it writes, and both are at most linear in $\|d\|$; summed over the nodes this gives $\|d\|$ times a polynomial of degree 4 in $\|d\|$. The functional of item 3 recurses through the matrix: at most linearly many nodes in its shape n, each costing at most $(K+1)^2(W+1)+(W+1)^2$ for a context of length $K = O(n)$ and types of total size $W = O(n^2)$, so $O((n+1)^5)$, plus a part linear in $(n+1)\|A\|$. Every formula concluded in d has shape at most $2\|d\|$, and $\|A\| + 2 \le (\|d\| + 1)^2$. The dummy quantifier of item 2 is bounded node by node in the same way. The sum is at most $C(\|d\| + 1)^5$. □

Degree five is the least, but that rests on a count we report and do not reproduce. Let A_0 be $0 = 0$ and A_{j+1} be $A_j \to 0 = 0$, nested to the left, and let D_j derive $\forall x \exists y\, A_j$ with witness 0; then $\|D_{3n+1}\| = 48n + 35$. On A_j the functional recurses into the antecedents, and its context grows by one entry every second level. A pointwise application in a context of i entries pays, through its composition combinators, the types of all the entries already absorbed, a cubic in i times their size; the recursion at depth $i + 2$ contains the one at depth i in a context one entry longer. A size count of the compiled programs along these lines, which follows the compiler clause by clause, gives $n^5 \le 24\,\|F(D_{3n+1})\|$. Granting it, no bound of degree four holds: if $\|F(d)\| \le C(\|d\| + 1)^4$ for all d, put $E = 24 \cdot C \cdot 48^4 \cdot 16$ and $n = E + 1$; then

$$n^5 \le 24C(48n + 36)^4 \le 24C \cdot 48^4(n + 1)^4 \le En^4 < n^5.$$

A bound of degree m is a bound of every degree $m' \ge m$, since $(\|d\| + 1)^m \le (\|d\| + 1)^{m'}$, so with the count, bounds of degree m exist exactly for $m \ge 5$.

The theorem bounds the degree and says nothing sharp about the constant. The fifth power has a definite source: the functional reading a deeply nested matrix. With the shapes of the input bounded, the degree drops, and the next theorem bounds it in three regimes. For a property R of pairs (A, d), say that F has *degree k on R* when some C has $\|F(d)\| \le C(\|d\|+1)^k$ for every (A, d) with R, and *exponent p/q on R* when some C has $\|F(d)\|^q \le C(\|d\| + 1)^p$ for every such pair. With no restriction degree 5 gives exponent 5, and with the reported count for D_{3n+1} no p/q with $p < 5q$ is an exponent, since such a bound would give $n^{5q} \le 24^q C(48n + 36)^p$ for every n. Table 28.3 collects the answers.

Theorem 28.52 (The degree by the shape of the input)**.** *Let F be the compiler of Theorem 28.51, and let k, c, p, q be natural numbers.*

1. *F has degree* 3 *on the derivations whose matrix has* $\mathrm{sh}(A) \le c$.
2. *F has degree* 3*, and exponent* $5/2$*, on the derivations with* $\mathrm{sh}(d) \le c$.

3. *If $c \geq 3$, then F has degree k on the derivations with $\mathrm{sh}(d) \leq c$ and $\mathrm{cx}(d) \leq c$ exactly when $k \geq 2$; if moreover $p < 2q$, it does not have exponent p/q there.*

No omniscience principle and no choice principle is used.

Proof. Upper bounds. Three estimates carry them. First, for every derivation from no hypotheses the realizer has size at most $C(\|d\| + 1)^3$, whatever its formulas; the fifth power of Theorem 28.51 comes only from the functional. Second, with $\mathrm{sh}(A) \leq c$ the functional has boundedly many nodes and adds at most $C_c(\|d\| + 1)^2$: it evaluates the terms of the matrix, and a substitution multiplies sizes, so $\|A\|$ can be quadratic in $\|d\|$ but no more. These give degree 3 in parts 1 and 2. Third, with $\mathrm{sh}(d) \leq c$ every type in the realizer has bounded size, and the cost concentrates at universal introductions and existential eliminations, which re-project and re-apply the hypotheses of their context: with $M = \mathrm{cx}(d)$ the realizer has size at most $C_c(M+1)^3\|d\|$. A context of M hypotheses is opened by M nodes that each carry the hypotheses opened before them, so $M^2 \leq C\|d\|$, and the realizer is at most $C(\|d\| + 1)^{5/2}$, which with the quadratic functional is the exponent $5/2$. With $\mathrm{cx}(d) \leq c$ as well, every node costs a constant plus the terms it mentions, and the functional is quadratic: degree 2.

Lower bound in part 3. Let u_m be the term $m \cdot 0$, of size $m + 3$. From the axiom instance $m \cdot 0 = 0$ derive $0 = u_m$ by symmetry, and substitute u_m for 0 in a derivation of $\forall y \exists z \big(z = 0+(0+\cdots+(0+0))\big)$ with m summands 0. The result derives $\forall y \exists z \big(z = u_m+(u_m+\cdots+(u_m+0))\big)$; it has size $18m + 51$, empty lists of hypotheses and formulas of shape at most 3, while its matrix has size $m(m + 4) + 3$. The functional contains the decision program χ of the matrix, which contains a program evaluating its terms, so $\|F(d)\| \geq m(m + 3)$, and no exponent p/q with $p < 2q$ survives large m. Degree k is exponent $k/1$, so degrees 0 and 1 fail, and a bound of degree k is one of every larger degree. □

In parts 1 and 2 the degree 3 and the exponent $5/2$ are the least, but again this rests on a count we report and do not reproduce. Let B_K open K hypotheses $0 = 0$ by K implication introductions, each discharged at once by reflexivity, perform K universal introductions and eliminations on reflexivity in that context, and conclude $\forall x \exists y(0 = 0)$ with witness 0. Every formula concluded in B_K is $0 = 0$, $0 = 0 \to 0 = 0$, $\forall(0 = 0)$, $\exists(0 = 0)$ or $\forall\exists(0 = 0)$, so $\mathrm{sh}(B_K) \leq 3$ and the matrix has shape 1, and $\|B_K\| = 14K^2 + 28K + 15$. Each universal introduction pays a cubic in K for each of the $K + 1$ entries of its context, nested K deep; a size count along these lines, clause by clause, gives $5K^5 \leq 6\,\|F(B_K)\|$. Granting it, since $\|B_K\| + 1 \leq 16(K + 1)^2$, a bound $\|F(B_K)\|^q \leq C(\|B_K\| + 1)^p$ with $2p < 5q$ fails for large K. With $p = 2$, $q = 1$ this excludes degree 2 in part 1 for every $c \geq 1$ and in part 2 for every $c \geq 3$, and it excludes every exponent below $5/2$ there.

Conditions and scope.

Base and premises

Peano and Heyting arithmetic; the programs of Definition 28.47 and the size of Definition 28.48; the compilers of Theorems 28.49 and 28.51.

Conclusion

The direct compiler has no polynomial bound (Theorem 28.50, with its count on the page). The second compiler has degree 5 on all inputs (Theorem 28.51); degree 3 with the matrix

Table 28.3. The least degree of a polynomial bound $C(\|d\| + 1)^k$ for the compiler of Theorem 28.51, and the least exponent p/q of a bound $\|F(d)\|^q \le C(\|d\| + 1)^p$, under restrictions on the input d with a constant c. The bounds are Theorems 28.51 and 28.52. That they are least rests, in the first three rows, on the reported size counts of the families in the last column, and in the last row on Theorem 28.52, part 3.

Restriction	Degree	Exponent	Family behind the lower bound
none	5	5	left-nested matrices A_j
matrix of shape at most c, $c \ge 1$	3	between $5/2$ and 3	B_K, whose matrix has shape 1
every formula of shape at most c, $c \ge 3$	3	$5/2$	B_K: size of order K^2, program of order K^5
shape and hypothesis lists at most c, $c \ge 3$	2	2	one substitution at the root makes the matrix quadratic

of bounded shape; degree 3 and exponent $5/2$ with every formula of bounded shape; and exactly degree 2, with no exponent below 2, when shape and hypothesis lists are bounded by some $c \ge 3$ (Theorem 28.52). Free.

Reported computation

That degrees 5 and 3 and the exponents 5 and $5/2$ cannot be lowered rests on the size counts $n^5 \le 24\,\|F(D_{3n+1})\|$ and $5K^5 \le 6\,\|F(B_K)\|$, described above and not reproduced; the arithmetic that turns them into lower bounds is on the page.

Not claimed

Sharp constants.

Under bounded shape the size of the derivation and the program part ways because the derivation pays for each open hypothesis at every node, about K^2 in all, while the program pays about K^4 at each of K nested introductions. The exponent $5/2$ is the ratio of those two growths.

The halving proof, compiled

Theorem 28.53 (The compiled halving proof)**.** *There is a derivation h in Peano arithmetic, from no hypotheses, of*

$$\forall x\, \exists y \big(x = y + y \ \vee \ x = (y + y) + 1 \big)$$

that follows the proof given at the start of this section, with its case split as an instance of excluded middle. Let $t = F(h)$ be its program under the compiler of Theorem 28.51. Then $\mathrm{val}(t)(x)$ *is the integer part of $x/2$ for every number x; in particular* $\mathrm{val}(t)(7) = 3$*, and read on the orbit $\mathbb{N}_\delta$ the program returns the element of* 3 *at the element of* 7*. No omniscience principle and no choice principle is used, although h uses excluded middle.*

Proof. The derivation is the proof written out in the calculus of Definition 8.10: induction on x, an instance of excluded middle for $\exists y(n = y + y)$ in the step, and the laws of addition of

Lemma E.1 for $S(y) + S(y) = S(S(y + y))$. By Theorem 28.51, $y = \mathrm{val}(t)(x)$ satisfies $x = y + y$ or $x = (y + y) + 1$, so $2y \le x \le 2y + 1$, and only the integer part of $x/2$ does that; at 7 it is 3. Item 4 of Theorem 28.49 carries the value to the orbit. □

The theorem reads the value off the specification; it does not look inside the program. A classical proof of a simple fact thus yields a program whose size Theorem 28.51 bounds and a correctness proof that never mentions halving.

The analysis programs, as programs of T

Earlier chapters proved approximate forms of four classical theorems with no principle, each by a finite search over stages of its inputs: the approximate root (Theorem 12.55), the approximate maximum (Lemma 12.57), the approximate mean value point (Theorem 12.86) and the finite Bolzano–Weierstrass theorem (Proposition 12.91). The searches are programs on codes, and they are programs of T.

Definition 28.54 (Stage oracles and four programs)**.** Code the integers by their place in the order $0, -1, 1, -2$, and so on, a ratio by pairing the code of its numerator with its denominator, and a finite list of numbers by repeated pairing, with a pairing of numbers fixed once. A regular sequence s is read through its *stage oracle*, the function that sends m to the code of s_m. A map f of $\mathbb{R}_\delta$ into itself is *presented* on the triadic grid by a family F of regular sequences, $F(n, j)$ naming $f(j/3^n)$, and F is read through the function that sends the code of (n, j, m) to the code of stage m of $F(n, j)$. A *modulus* of uniform continuity of f on $[0, 1]$ is a map ω with $|f(x) - f(y)| \le 1/(k + 1)$ whenever $x, y \in [0, 1]$ and $|x - y| \le 1/(\omega(k) + 1)$. With n and k fixed, let a_j be stage $8k + 7$ of $F(n, j)$.

1. $P_{\mathrm{IVT}}(F, \omega, k)$: with $n = \omega(2k + 1)$, the grid point $j/3^n$ for the first $j \le 3^n$ with $a_j \ge -2/(4k + 4)$, and 1 if there is none.

2. $P_{\mathrm{EVT}}(F, \omega, k)$: with $n = \omega(2k + 1) + 1$, the grid point $j/3^n$ for the first $j \le 3^n$ at which a_j is largest.

3. $P_{\mathrm{MVT}}(G, c, \omega_g, \omega_f, k)$, for a presentation G of a map g and a regular sequence c: with $n = \omega_g(2k + 1) + \omega_f(4k + 4)$, the grid point $i/3^n$ for the first $i < 3^n$ at which stage $16k + 15$ of $G(n, i)$ and stage $16k + 15$ of c differ by at most $6/(8k + 8)$, and 0 if there is none.

4. $P_{\mathrm{BW}}(s, k, M)$, for regular sequences $s_0, s_1, \ldots$: with $m = k + 2$ and $N = (3^m + 1)M + 1$, label each index $i < N$ by $\lfloor 3^m q_i \rfloor$ clamped to $[0, 3^m]$, where q_i is stage $3^m - 1$ of s_i; return the first $M + 1$ indices that carry the first label carried by at least $M + 1$ indices.

Theorem 28.55 (Four analysis programs are programs of T)**.**

1. *Let f be presented by F, with a modulus ω of uniform continuity on $[0, 1]$, and let k be a number.*

   (a) *If $f(0) \le 0 \le f(1)$, then $x = P_{\mathrm{IVT}}(F, \omega, k)$ lies in $[0, 1]$ and $|f(x)| \le 1/(k + 1)$.*

   (b) *The point $x = P_{\mathrm{EVT}}(F, \omega, k)$ lies in $[0, 1]$, and $f(y) \le f(x) + 1/(k + 1)$ for every $y \in [0, 1]$.*

(c) *If g is presented by G with a modulus ω_g, c names $f(1)-f(0)$, and ω_f satisfies, for every j, $|f(y)-f(x)-g(x)(y-x)|\le|y-x|/(j+1)$ whenever $x,y\in[0,1]$ and $|x-y|\le 1/(\omega_f(j)+1)$, then $\xi=P_{\mathrm{MVT}}(G,c,\omega_g,\omega_f,k)$ lies in $[0,1]$ and $|g(\xi)-(f(1)-f(0))|\le 1/(k+1)$.*

(d) *If every s_i names a point of $[0,1]$, then $P_{\mathrm{BW}}(s,k,M)$ lists $M+1$ increasing indices below $(3^{k+2}+1)M+1$ whose terms are pairwise within $1/(k+1)$.*

2. *For each of the four programs there is a closed program of T which, applied to the stage oracles of its inputs, to its moduli and to k (and M), returns the code of the program's output, for all inputs.*

3. *In classical mathematics, for every closed program t of type $(O\to O)\to((O\to O)\to(O\to O))$ there is one partial recursive program that computes $\mathrm{val}(t)(f)(g)(k)$ from k and from codes of programs computing f and g.*

Items 1 and 2 use no omniscience principle and no choice principle.

Proof. 1(a). Put $\varepsilon=1/(4k+4)$, $n=\omega(2k+1)$ and $x_j=j/3^n$. Stage $8k+7$ of a regular sequence lies within $1/(8k+8)\le\varepsilon$ of its point, so $a_j\ge-2\varepsilon$ gives $f(x_j)\ge-3\varepsilon$, and $a_j<-2\varepsilon$ gives $f(x_j)<-\varepsilon$. Since $f(1)\ge0$, $a_{3^n}\ge-\varepsilon$ passes the test, so the first passing j exists. If $j=0$, then $f(x_j)=f(0)\le0$. If $j>0$, then $f(x_{j-1})<-\varepsilon$, and $|x_j-x_{j-1}|=3^{-n}\le1/(n+1)$, so $|f(x_j)-f(x_{j-1})|\le1/(2k+2)=2\varepsilon$ and $f(x_j)<\varepsilon$. In both cases $|f(x_j)|\le3\varepsilon\le1/(k+1)$.

1(b). With $n=\omega(2k+1)+1$, let j be the output index. For $y\in[0,1]$ some grid point x_i lies within 3^{-n} of y, found by a finite search with cotransitivity (proof of Lemma 12.57). Then

$$f(y)\le f(x_i)+2\varepsilon\le a_i+3\varepsilon\le a_j+3\varepsilon\le f(x_j)+4\varepsilon=f(x_j)+\frac{1}{k+1}.$$

1(c). Put $c=f(1)-f(0)$, $\varepsilon=1/(4k+4)$, $\varepsilon'=1/(8k+8)$ and $x_i=i/3^n$; stage $16k+15$ lies within ε' of its point. As $3^{-n}\le1/(n+1)$ and $n\ge\omega_g(2k+1),\omega_f(4k+4)$, the hypotheses on ω_g at $2k+1$ and on ω_f at $j=4k+4$ give $|g(x_{i+1})-g(x_i)|\le2\varepsilon$ and $|f(x_{i+1})-f(x_i)-g(x_i)3^{-n}|\le3^{-n}/(4k+5)<\varepsilon3^{-n}$. Cotransitivity labels each $i<3^n$ with (L) $g(x_i)<c-\varepsilon$, (H) $g(x_i)>c+\varepsilon$ or (M) $|g(x_i)-c|<2\varepsilon$. Neighbours labelled (L) and (H) would be more than 2ε apart, so if no i is (M) all labels agree, and telescoping gives $c<c$ or $c>c$ as in the proof of Theorem 12.86. So some x_i has $|g(x_i)-c|<4\varepsilon'$; its stages differ by at most $6\varepsilon'$, so a passing index exists, and at the first one $|g(x_i)-c|\le6\varepsilon'+2\varepsilon'=1/(k+1)$.

1(d). Of $(3^m+1)M+1$ indices and 3^m+1 labels, some label is carried by at least $M+1$ indices, and a bounded count finds the first. A term with label q lies within $1/(2k+2)$ of $q/3^m$, by the choice of stage and $m=k+2$; two terms with one label are within $1/(k+1)$.

2. Each program reads finitely many stages, compares ratios through cross-multiplied integers, and searches, maximizes or counts over a finite range computed from its inputs. Pairing and unpairing, integer arithmetic on codes, comparison of ratios, bounded search, the running maximum, counting and filtering of a list are defined by primitive recursion, so each is a closed program with R at the types O and $O\to O\to O$. Writing a program as a typed term with these constants and removing its variables by bracket abstraction gives a closed program with the same value, which is the code of the output by induction along the program.

3. This is classical recursion theory. By recursion on types, relate each value of type O to itself and a value of type $A \to B$ to a code e of a partial recursive program when e, run on any code related to a value a, halts with a code related to the value at a. Every constant has a related code, by the parametrization and recursion theorems, and a code related to t, run on codes of f and g and on k, returns $\mathrm{val}(t)(f)(g)(k)$ [6]. □

Item 1 is the agreement with the theorems: each program returns what the corresponding theorem asserts to exist. Item 2 places the programs in T, so they inherit everything said above about programs of T, including their values on the orbit.

The classical display of the line

A last link connects $\mathbb{R}_\delta$ with the classical real numbers. It is a statement of classical mathematics, and it is how a classical theorem about $\mathbb{R}$ is read on $\mathbb{R}_\delta$ value by value.

Theorem 28.56 (In classical mathematics, the line is the real line in exactly one way)**.** *In classical mathematics, with $\mathbb{R}$ the classical real numbers:*

1. *there is a map $D : \mathbb{R}_\delta \to \mathbb{R}$, onto, with $D(0) = 0$, $D(1) = 1$, $D(x + y) = D(x) + D(y)$ and $D(xy) = D(x)D(y)$, and with $x < y$ exactly when $D(x) < D(y)$;*
2. *any two maps with these properties are equal;*
3. *$D(\exp x) = e^{D(x)}$ for every point x, with $\exp$ the exponential of Proposition 12.88;*
4. *for a map F of $\mathbb{R}_\delta$ and a map f of $\mathbb{R}$ with $f(D(x)) = D(F(x))$ for every point x: F is continuous at x (for every k some j has $|F(y) - F(x)| \le 1/(k + 1)$ whenever $|y - x| \le 1/(j + 1)$) exactly when f is continuous at $D(x)$; and d is a derivative of F at x (Definition 28.22) exactly when $D(d)$ is the derivative of f at $D(x)$.*

Proof. 1. The stages of a regular sequence s satisfy $|s_m - s_n| \le 1/(m + 1) + 1/(n + 1)$, so they form a Cauchy sequence of rationals; let D of its point be the limit. Sequences naming the same point have stages within $2/(n + 1)$ (Lemma F.8), hence the same limit, so D is a map on points, and every stage s_n lies within $1/(n + 1)$ of $D(x)$. Sums and products of points are formed from sums and products of stages, so D preserves them, and $D(\eta(q)) = q$. A point is positive when some stage has $s_N > 2/(N + 1)$; then $D(x) \ge s_N - 1/(N + 1) > 0$. Conversely, if $D(x) = L > 0$, then $s_n \ge L - 1/(n + 1) > 2/(n + 1)$ once $3/(n + 1) < L$. So $x < y$ exactly when $D(x) < D(y)$. For $r \in \mathbb{R}$, the ratios $\lfloor 2r(n + 1) \rfloor/(2n + 2)$ lie within $1/(2n + 2)$ of r, so they form a regular sequence, whose point has $D = r$.

2. Let D' have the same properties. Both fix 0 and 1 and preserve sums, differences, products and quotients of ratios, so $D'(\eta(q)) = q = D(\eta(q))$ for every ratio q. If $D'(x) < D(x)$, a rational q between them gives $D'(x) < D'(\eta(q))$ and $D(\eta(q)) < D(x)$, so $x < \eta(q)$ and $\eta(q) < x$, which is impossible; likewise $D(x) < D'(x)$ is impossible, and so $D'(x) = D(x)$.

3. D sends the partial sums $\sum_{m<N} x^m/m!$ to the partial sums of the exponential series at $D(x)$, since it preserves sums, products and ratios. By item 1, $|x - y| \le \eta(q)$ exactly when $|D(x) - D(y)| \le q$, so D carries the limit of the partial sums on $\mathbb{R}_\delta$, which is $\exp(x)$, to their limit in $\mathbb{R}$, which is $e^{D(x)}$.

4. By item 1, $|y - x| \le 1/(j+1)$ exactly when $|D(y) - D(x)| \le 1/(j+1)$, and likewise for the values of F and f, since D preserves sums, products, the order and the ratios. Every real number is $D(y)$ for some point y, and the tolerances $1/(k+1)$ are cofinal among the positive reals. So "for every k some j has $|F(y) - F(x)| \le 1/(k+1)$ whenever $|y - x| \le 1/(j+1)$" says, value by value, that for every k some j has $|f(r) - f(D(x))| \le 1/(k+1)$ whenever $|r - D(x)| \le 1/(j+1)$, which is continuity of f at $D(x)$. The same translation turns the estimate of Definition 28.22 for F, d and x into $|f(r) - f(D(x)) - D(d)(r - D(x))| \le |r - D(x)|/(k+1)$ whenever $|r - D(x)| \le 1/(j+1)$, which is the classical statement that $D(d)$ is the derivative of f at $D(x)$. □

Proposition 28.57 (The determinant is the classical determinant)**.** *In classical mathematics, for every commutative ring R and every $n \times n$ matrix M over R, the determinant of Definition G.2, by expansion along the first column, equals*

$$\sum_{\pi} \operatorname{sgn}(\pi) \prod_{j<n} M_{\pi(j)j},$$

the sum over all permutations π of $\{0, \dots, n-1\}$.

Proof. Induct on n; for $n = 0$ both sides are 1. Group the permutations by $i = \pi(0)$. Deleting row i and column 0 turns the permutations with $\pi(0) = i$ into the permutations τ of the $(n-1) \times (n-1)$ minor $M^{(i,0)}$, with $\prod_{j \ge 1} M_{\pi(j)j}$ the corresponding product for τ and $\operatorname{sgn}(\pi) = (-1)^i \operatorname{sgn}(\tau)$, since moving row i to the top takes i transpositions of neighbours. So the sum is $\sum_{i<n} (-1)^i M_{i0} \det M^{(i,0)}$ by the induction hypothesis, which is the definition. □

Conditions and scope.

Base and premises
Classical mathematics with the classical real numbers $\mathbb{R}$; the line $\mathbb{R}_\delta$, its exponential and its determinant as defined in earlier chapters.

Conclusion
Classically, $\mathbb{R}_\delta$ is the real line in exactly one way, and under that identification its exponential, its pointwise continuity and derivative, and its determinant are the classical ones.

Not claimed
A price: the display reasons classically and identifies objects. The constructive counterpart is Theorem 12.38, which allows at most one structure-preserving map into $\mathbb{R}_\delta$ from every Archimedean ordered field.

28.12 What the mathematics of physics costs

Physics leans on a classical library: the intermediate value theorem for an equilibrium, the spectral theorem for energy levels, Fekete's lemma for a thermodynamic limit, the Picard–Lindelöf theorem for a motion. Appendix Q reads 1467 statements of that library on $\mathbb{R}_\delta$, one to a row, and gives the price of each. Most of the library costs nothing: 1166 rows are free, 14

are proved with a choice principle and no omniscience principle, 265 need an omniscience principle, Ishihara's boundedness principle or the fan theorem, and 22 are open.

The priced rows have a common shape. Each asks for an exact object, such as a root, an eigenvector, the limit of a bounded monotone sequence or a decision, that the physics using it either does not need or gets from information it already has: a gap between energy levels, a slope with a known reciprocal, a surface term of known size. This section proves the rows that show this most clearly. The quantum Ising model in a transverse field has a free energy per site and a ground energy per site in every dimension, with no omniscience principle. The spectral theorem costs exactly LLPO, with countable choice in the converse, and LLPO is the price of an eigenbasis when it is not known whether two levels coincide; when the outcomes of an observable are known to be apart, its eigenbasis is free. Perron's theorem for positive matrices, with the bound of every eigenvalue and the free energy per site of a chain, and the Schmidt decomposition of a real amplitude matrix with positive, pairwise apart coefficients are free with no choice at all.

How the table reads a price. A row is *free* when its proof uses no omniscience principle and no choice beyond the *free allowance*. This is wider than the word in Section 21.1, which allows no choice at all, by one principle: countable choice at pairs of ratios, which makes every located cut the cut of a point (Theorem 12.33) and so makes every Cauchy sequence of points converge (Corollary 12.35). The allowance contains more than it first appears to. A natural number, a yes-or-no answer and a finite table can each be written as a pair of ratios, so it chooses one of them for every n, whether or not the choice is unique; for a unique bit b it chooses the pair (b, b), so countable unique choice of yes-or-no answers is free (row 994). It chooses a regular sequence for every point of a sequence of points, by choosing rational approximations, and so collects points that are each determined uniquely into a sequence. The table applies this one rule to every row. The free rows whose proofs form a limit point say so after their price; 62 of them do. Every theorem below states which choice its proof uses, in the words of Section 21.1.

A price that names AC_ω, countable choice at every type, beside an omniscience principle, such as LLPO with AC_ω, means that the statement implies LLPO with no choice and that LLPO with countable choice proves the statement; the proof uses countable choice, which is an upper bound and is not shown to be needed. We call a price *exact in the table's sense* when both directions hold in this way: the statement implies the principle with no choice, and the principle, with the free allowance and with AC_ω where the price names it, implies the statement. This is weaker than "costs exactly" in Section 21.1, which allows no choice in either direction, so in this section we write "exactly LLPO, with countable choice in the converse". All 1445 rows that are not open are exact in the table's sense. Both directions hold for every one of them. A row with a reference is proved in this book at the result it names; most rows that this book proves give the reference, and most rows without one are not proved in this book.

What the table covers

The rows fall into eleven areas, the sections of the appendix (Table 28.4). The functions a physicist writes every day, the exponential, the logarithm, the trigonometric functions and their complex forms, take 379 rows, and 358 of them are free. Linear algebra takes 326 rows,

Table 28.4. The rows of the price table by area. "Choice" counts rows whose only price is AC_ω; "omniscience" counts rows priced at an omniscience principle, its data form, Ishihara's boundedness principle or the decidable fan theorem.

Area (section of Appendix Q)	Rows	Count	Free	Choice	Omniscience	Open
Conventions at excluded points	1–17	17	5	0	11	1
Limits, continuity and derivatives	18–156	139	113	0	26	0
Exponential, logarithm, trigonometric and complex functions	157–535	379	358	0	21	0
Integrals and measure	536–592	57	49	0	6	2
Differential equations and dynamics	593–638	46	44	0	2	0
Linear algebra and spectra	639–964	326	280	0	45	1
Spectra, lattice models and thermodynamic limits	965–994	30	29	0	1	0
Topology, compactness and fixed points	995–1103	109	34	8	55	12
Probability	1104–1192	89	71	4	14	0
Functional and Fourier analysis	1193–1256	64	40	0	23	1
A first survey across the areas	1257–1467	211	143	2	61	5
All rows		1467	1166	14	265	22

280 free. Apart from the conventions at excluded points, of whose 17 rows only 5 are free, the area with the largest share of prices is topology and compactness: of its 109 rows only 34 are free and 12 are open. The last section of the appendix is a first survey that crosses all the areas, and some of its statements reappear, in other readings, in the areas before it.

Where the prices sit

Table 28.5 shows that each principle has its own place, and that beside a priced row there is usually a free one, which is the statement the physics uses.

Markov's principle sits in hypotheses. The inverse function theorem with a slope $d \neq 0$ costs MP (row 24); with a slope that has a known reciprocal it is free (row 25). Gram–Schmidt for linearly independent vectors costs MP (row 885); for uniformly independent vectors, whose combinations are bounded below by a known $\varepsilon > 0$, it is free (row 884). A measured slope or a measured gap is a number with a known distance from 0, so the free row is the one physics meets.

LLPO sits in exact points. Every LLPO row asks for a point at which something happens, such as a root, a maximizer, an eigenvector or a stationary law, without a separation that locates it. Its approximate form is free. The supremum of a uniformly continuous function on $[0, 1]$ is free (row 1316), while a point attaining it costs exactly LLPO, with countable choice in the converse (row 1313, Theorem 12.60). Every finite Markov chain has, for every tolerance, a law that one step changes by less than the tolerance (row 1131, free), while an exact stationary law of every chain costs LLPO (row 1134); a chain in which some state is entered from every state with probability at least some $\delta > 0$ has exactly one stationary law, for free (row 1130).

Table 28.5. The 279 priced rows by the principle of their price. A row priced at a principle with AC_ω is counted under that principle; the data forms are those of Definition 21.1. Numbers in the last column are rows of Appendix Q.

Principle	Rows	Where it sits, with typical rows
LLPO	72	an exact point where approximate points are free: the intermediate value theorem (1312), a maximum attained (1313), the spectral theorem (1461), a stationary law of every finite Markov chain (1134)
MP	67	a hypothesis $x \neq 0$ used as $x \,\#\, 0$: the inverse function theorem (24), an inverse from a nonzero determinant (650), Gram–Schmidt for independent vectors (885)
LPO	58	limits without a rate, and decisions of order: monotone convergence (1302), Bolzano–Weierstrass (1304), Fekete's lemma (965), the interior point of Rolle's theorem (20), the integer floor (1270)
WLPO	20	decided equality and rank: equality of two numbers (1265), rank–nullity (1456), the least upper bound of a sequence (1307)
$\mathrm{MP}^\vee$	13	L'Hôpital's rule (32), sorting distinct numbers (1070)
FAN	10	covers: Heine–Borel (1418), the Lebesgue number lemma (1102)
LPO function, WLPO function	13	total conventions and stored decisions (rows 1–6, 9, 13–16), the field and the decided order of the line (1429, 1430)
EM, WEM	10	every inhabited set bounded above has a least upper bound (1301), compact sets are closed (1443), projection onto any subspace (1213)
BD-N	2	the uniform boundedness principle for functionals and for operators (1232, 1243)
AC_ω only	14	a member of a set chosen for every n: a supremum as a limit of members (1399), a limit point as a limit of members (1002), a closed interval is connected (1436); Chernoff's optimal constant (1171)

LPO sits in limits without a rate and in decisions of order. Monotone convergence and Bolzano–Weierstrass cost exactly LPO, with countable choice in the converse (Theorems 12.45 and 12.47), and so does Fekete's lemma (Proposition 28.63); a sequence with a known rate needs nothing. The fan theorem sits in compactness by covers: Heine–Borel is exactly the decidable fan theorem (Theorem 13.6), and the supremum of a uniformly continuous function does not need it. The Picard–Lindelöf theorem is free. For a law known only through its values the proof chooses a modulus of continuity in time and a regular sequence for every point the Euler polygons use, and both choices lie inside the free allowance (rows 616 and 617); for a law given by a formula the polygons are computed from the initial data directly (row 620). A choice principle alone is the price of a row when its proof picks a member of a set for every n, as for a supremum written as a limit of members (row 1399); that choice is an upper bound and is not shown to be needed. Baire's theorem and Cantor's intersection theorem (rows 1006 and 1005) choose rational intervals and rational approximations, inside the free allowance, and are free.

These equivalences are known in constructive mathematics, in the setting of Bishop's constructive analysis [14]: the intermediate value theorem and weak König's lemma with LLPO, monotone convergence and Bolzano–Weierstrass with LPO [40, 43], Heine–Borel with the fan theorem [32], the uniform boundedness principle with Ishihara's boundedness principle BD-N [60, 73], and the disjunctive Markov principle $\mathrm{MP}^{\vee}$ in the same order of principles [43, 73]. What the table adds is the reading on $\mathbb{R}_\delta$, the choice each proof uses, and the free statement that stands beside each priced one.

The quantum Ising model in a transverse field

The thermodynamic limit is where physics meets Fekete's lemma, and the quantum Ising model is its standard test. We first fix the levels of a symmetric matrix, the boxes of the lattice and what it means for a family of Hamiltonians to split across a cut.

Definition 28.58 (Levels and partition function)**.** Let A be a symmetric $N \times N$ matrix on $\mathbb{R}_\delta$, $N \geq 1$. A *diagonalization of A at precision* $1/(m+1)$ is an orthonormal list of vectors $v_1, \dots, v_N$ with points $\mu_1, \dots, \mu_N$ such that every coordinate of $Av_c - \mu_c v_c$ is at most $1/(m+1)$ in size. A list $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ is a *list of levels* of A if there is a natural number K such that, for every m and every diagonalization at precision $1/(m+1)$, the jth largest of the μ_c lies within $K/(m+1)$ of λ_j, for each j. For $\beta \geq 0$ the *partition function* of a list of levels is $Z = \sum_j e^{-\beta\lambda_j}$, its *free energy* is $-\log Z$, and its last member λ_N is the *ground energy*.

Every symmetric matrix has a list of levels, and two lists of levels of one matrix agree (rows 971 and 981). So the partition function and the ground energy are fixed by the matrix.

Definition 28.59 (Boxes and splitting across cuts)**.** A *box* in d dimensions is a list $L = (L_1, \dots, L_d)$ of natural numbers, its sides; its volume is $\mathrm{vol}\, L = L_1 \cdots L_d$. We write $L[i : a]$ for L with side i replaced by a, and $A_i(L)$ for the product of the sides other than i, the area of a cut across side i. A function F of boxes is *almost additive with constant C* if

$$\left|F(L[i : a+b]) - F(L[i : a]) - F(L[i : b])\right| \leq C\, A_i(L)$$

for every box L, side i and natural numbers a, b. A family of symmetric matrices H_L, one for each box, *splits across cuts with constant J* if for all L, i, a, b there are a bijection σ from the

basis indices of $H_{L[i:a+b]}$ onto the pairs of basis indices of $H_{L[i:a]}$ and $H_{L[i:b]}$, and a symmetric matrix E with $|x \cdot Ex| \le J A_i(L)\,(x \cdot x)$ for every vector x, such that

$$H_{L[i:a+b]}(p,q) = \big(H_{L[i:a]} \otimes 1 + 1 \otimes H_{L[i:b]}\big)(\sigma p, \sigma q) + E(p,q)$$

for all p, q. Here $A \otimes 1 + 1 \otimes B$ is the Hamiltonian of two systems that do not interact.

Definition 28.60 (The quantum Ising model in a transverse field)**.** Let J and g be points and L a box in d dimensions, whose sites are the integer points x with $0 \le x_i < L_i$; two sites are neighbours when they differ by 1 in one coordinate. A spin configuration is a map s from the sites to $\{+1, -1\}$; there are $2^{\operatorname{vol} L}$ of them. The *transverse-field Ising Hamiltonian* $H_L^{J,g}$ is the symmetric matrix indexed by spin configurations with

$$H_L^{J,g}(s,s) = -J \sum_{\{x,y\}\ \text{neighbours}} s_x s_y, \qquad H_L^{J,g}(s,s') = -g \ \text{ if } s \text{ and } s' \text{ differ at exactly one site,}$$

and all other entries 0. In the usual notation $H = -J\sum_{\langle x,y\rangle} Z_x Z_y - g\sum_x X_x$ with open boundaries, where Z_x reads the spin of site x along the coupling axis and X_x flips it. For $d = 1$ and $L = (n)$ it is the chain on n sites.

Theorem 28.61 (The Ising model splits at every cut)**.** *For all points J and g and every dimension d, the Hamiltonians $H_L^{J,g}$ on the boxes in d dimensions split across cuts with constant $|J|$. No omniscience principle and no choice principle is used.*

Proof. Cut the box $L[i:a+b]$ across side i at a. The sites with $x_i < a$ form a copy of $L[i:a]$, and the others, shifted by a in coordinate i, a copy of $L[i:b]$. A configuration of the large box is a pair of configurations of the two halves; this is σ. A flip at one site of the large box is a flip at one site of one half with the other half unchanged, so the off-diagonal entries of $H^{J,g}_{L[i:a+b]}$ are those of $H^{J,g}_{L[i:a]} \otimes 1 + 1 \otimes H^{J,g}_{L[i:b]}$. A bond inside one half is the same bond of that half. What remains is the diagonal matrix E of the bonds that cross the cut, from the layer $x_i = a-1$ to the layer $x_i = a$: one bond for each site of the cross-section, so $A_i(L)$ bonds, each contributing $+J$ or $-J$. Hence $|E(s,s)| \le |J| A_i(L)$ and $|x \cdot Ex| = \big|\sum_s E(s,s) x_s^2\big| \le |J| A_i(L)(x \cdot x)$. Every step is a finite computation with the entries, and no point is compared with another. □

Theorem 28.62 (The thermodynamic limit over boxes)**.** *Let F be a function of the boxes in d dimensions that is almost additive with constant C. Then there is a point f with*

$$\Big|\frac{F(L)}{\operatorname{vol} L} - f\Big| \le C\Big(\frac{1}{L_1} + \cdots + \frac{1}{L_d}\Big)$$

for every box L whose sides are all at least 1. No omniscience principle is used; countable choice for pairs of ratios is used once, for the limit f.

Proof. Write $x(L) = F(L)/\operatorname{vol} L$. First, multiplying one side ℓ by $q \ge 1$ moves x by at most C/ℓ: cutting the side $q\ell$ into q pieces uses almost additivity $q-1$ times, so $|F(L[i:q\ell]) - qF(L[i:\ell])| \le (q-1)CA_i(L)$, and dividing by $q\ell A_i(L)$ gives the bound. Second, multiplying every side L_i by some $q_i \ge 1$ moves x by at most $C\sum_i 1/L_i$. This is an induction on d: multiply the last side first;

at the new height y the function $M \mapsto F(M, y)$ of boxes in $d-1$ dimensions is almost additive with constant Cy, and dividing the induction bound by y gives the rest. Third, the box with sides $(m+1)L_i$ refines both L and the cube of side $m+1$, so $|x(L) - s_m| \le C\sum_i 1/L_i + Cd/(m+1)$, where s_m is the value of x at that cube. Taking L to be the cube of side $m'+1$ gives $|s_{m'} - s_m| \le Cd/(m'+1) + Cd/(m+1)$, an explicit Cauchy modulus. Corollary 12.35, with countable choice for pairs of ratios, gives the limit f of the s_m, and letting m grow in the third bound gives the theorem. □

For $d = 1$ the theorem is the bounded-defect form of Fekete's lemma. The classical lemma asks for less and pays for it.

Proposition 28.63 (Fekete's lemma and its bounded-defect form)**.**

1. *Consider the assertion: for every sequence of points u with $u(m+n) \le u(m)+u(n)$ for all m, n and with $u(n)/n$ bounded below for $n \ge 1$, the ratios $u(n)/n$ converge to a limit no larger than any of them. The assertion implies LPO, and the proof uses no choice principle. Conversely, LPO and AC_ω imply the assertion. Countable choice is used only through Theorem 12.45, for representatives and once at pairs of ratios.*

2. *If $|u(m+n) - u(m) - u(n)| \le C$ for all m, n, there is a point ℓ with $|u(n)/n - \ell| \le C/n$ for every $n \ge 1$. No omniscience principle is used; countable choice for pairs of ratios is used once, for ℓ.*

Proof. For part 1, let x be a nonincreasing sequence of points bounded below. Put $u(0) = 0$ and $u(k) = k\,x(k-1)$ for $k \ge 1$. Then $u(m+n) = m\,x(m+n-1) + n\,x(m+n-1) \le m\,x(m-1) + n\,x(n-1) = u(m) + u(n)$, and $u(n+1)/(n+1) = x(n)$. So the assertion makes every nonincreasing sequence bounded below converge, which gives LPO by Theorem 12.45 applied to $-x$. Conversely, LPO and AC_ω give monotone convergence (Theorem 12.45). Write $v_n = u(n+1)/(n+1)$; the running minimum $b_n = \min(v_0, \dots, v_n)$ converges to a point ℓ with $\ell \le b_n \le v_n$. For a tolerance ε take M with $b_M \le \ell + \varepsilon$ and, by finitely many comparisons with the tolerance, an index $j \le M$ with $v_j \le b_M + \varepsilon$. Writing $n+1 = q(j+1) + r$ with $r \le j$, subadditivity gives $u(n+1) \le q\,u(j+1) + u(r)$, so $v_n \le v_j + \varepsilon$ for every large n, and $v_n \le \ell + 3\varepsilon$.

For part 2, induction gives $|u(kN) - k\,u(N)| \le (k-1)C$. Comparing $u(p+1)$ and $u(q+1)$ with $u\big((p+1)(q+1)\big)$ gives $|v_p - v_q| \le C/(p+1) + C/(q+1)$, an explicit Cauchy modulus. The limit ℓ exists by Corollary 12.35, and the bound passes to the limit. □

The physics has the bounded defect: the two halves of a chain interact through a boundary term of known size. What remains is to show that the quantum Ising model has it, at every temperature and for the ground state.

Theorem 28.64 (Free and ground energies of quantum lattice models)**.** *Let (H_L) be a family of symmetric matrices on the boxes in d dimensions that splits across cuts with constant J, and let $\beta \ge 0$.*

1. *There is a function F of boxes with $e^{-F(L)} = \sum_j e^{-\beta\lambda_j}$ for every list of levels λ of H_L, and F is almost additive with constant βJ.*

2. *There is a function G of boxes such that $G(L)$ is the last member of every list of levels of H_L, and G is almost additive with constant J.*

No omniscience principle is used; countable choice for pairs of ratios is used to obtain the levels and to collect one value per box into a function of the box.

Proof. The levels of each H_L exist and any two lists of them agree, so the partition function and the ground energy of each box are single points. Countable choice for pairs of ratios, applied once to the boxes coded by natural numbers, makes them functions F and G of the box. The levels of $A \otimes 1 + 1 \otimes B$ are the sums of a level of A and a level of B, sorted (row 981); so the two halves side by side have partition function $Z_a Z_b$ and ground energy $G_a + G_b$, and the relabelling σ changes neither. By Weyl's bound (row 980), a symmetric E with $|x \cdot Ex| \le e\,(x \cdot x)$ moves every level by at most e. With $e = J A_i(L)$, every level of $H_{L[i:a+b]}$ lies within e of the corresponding level of the two halves side by side. So the partition function changes by at most a factor $e^{\beta e}$ either way (row 982), and F by at most $\beta J A_i(L)$; the ground energy changes by at most $J A_i(L)$. □

Corollary 28.65 (The quantum Ising model in the thermodynamic limit)**.** *Let J and g be points and d a dimension. Write F_L for the free energy of $H_L^{J,g}$ at inverse temperature β in the sense of Definition 28.58, $-\log Z$, which is β times the thermodynamic free energy, and G_L for its ground energy. The boxes have open boundaries, as in Definition 28.60.*

1. *For every $\beta \ge 0$ there is a point f, the free energy per site, with $|F_L/\operatorname{vol} L - f| \le \beta|J|\,(1/L_1 + \cdots + 1/L_d)$ for every box whose sides are all at least 1. For the chain on $n \ge 1$ sites, $|F_n/n - f| \le \beta|J|/n$.*

2. *There is a point e_0, the ground energy per site, with $|G_L/\operatorname{vol} L - e_0| \le |J|\,(1/L_1 + \cdots + 1/L_d)$ for every box whose sides are all at least 1. For the chain on $n \ge 1$ sites, $|G_n/n - e_0| \le |J|/n$.*

No omniscience principle is used; countable choice for pairs of ratios is used for the levels, to collect them over the boxes, and for the limits.

Proof. Theorem 28.61 gives the splitting with constant $|J|$. Theorem 28.64 makes F almost additive with constant $\beta|J|$ and G with constant $|J|$, and Theorem 28.62 gives f and e_0. For the chain the cut area is 1, and the bounds read $\beta|J|/n$ and $|J|/n$. □

The Hamiltonian $H_L^{J,g}$ is the quantum Ising magnet: spins on a lattice, coupled along one axis, in a field across that axis; for the chain the quantum phase transition is at $g = J$. Its free energy per site in the infinite lattice, $-\log Z$ per site, which is β times the thermodynamic free energy per site, has a value at every temperature, and so does its ground energy per site, in every dimension. A box of side L in d dimensions is within $\beta|J|d/L$ and $|J|d/L$ of them. Nothing about the spectrum is decided along the way: the bound comes from the surface alone. Fekete's lemma, which uses subadditivity and nothing else, costs LPO; the surface term of known size makes the limit free.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$; the Hamiltonian given by its matrix in the spin basis, on boxes with open boundaries; any points J, g and $\beta \geq 0$.

Conclusion

The splitting at every cut, with no choice; the free energy and the ground energy per site with the finite-size bounds above, with no omniscience principle (rows 987–992 and 993).

Not claimed

Where f fails to be analytic in g, that is, the location of the phase transition, and boxes with periodic boundaries.

The spectral theorem: the price of a coincidence

The spectral theorem in full costs LLPO, and the test that shows it is a matrix whose levels may or may not coincide.

Theorem 28.66 (The spectral theorem costs exactly LLPO, with countable choice in the converse)**.**

1. *Let $n \geq 2$. If every symmetric $n \times n$ matrix on $\mathbb{R}_\delta$ has a unit eigenvector, then LLPO holds. The proof uses no choice principle. In particular the spectral theorem for symmetric $n \times n$ matrices, an orthonormal basis of eigenvectors, implies LLPO.*

2. *LLPO and AC_ω imply that for every n every symmetric $n \times n$ matrix on $\mathbb{R}_\delta$ has an orthonormal basis of eigenvectors. Countable choice is used through weak König's lemma, to collect a finite table for every depth, and to choose rational approximations; both are instances of countable choice at pairs of ratios.*

Proof. For part 1, let x be a point, $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. Let T_x have the diagonal $x^+ \cdot (0, 1, \dots, n-1)$ and x^- in every off-diagonal entry. Let v be a unit eigenvector, $T_x v = \lambda v$, put $s = v_1 + \cdots + v_n$ and $R = (s^2 - n)s^2$. If $x > 0$, the matrix is $x\,\mathrm{diag}(0, 1, \dots, n-1)$; for $i \neq j$ the two eigenvalue equations give $(i - j)x v_i v_j = 0$, so $v_i v_j = 0$, $s^2 = \sum_i v_i^2 = 1$ and $R = 1 - n \leq -1$. If $x < 0$, put $c = -x > 0$; the equations read $c(s - v_i) = \lambda v_i$. Summing them gives $c(n-1)s = \lambda s$, and multiplying each by v_i and summing gives $c(s^2 - 1) = \lambda$; so $cs(n - s^2) = 0$, hence $s(n - s^2) = 0$ and $R = 0$. Since $-1 < -\frac{1}{2}$, cotransitivity gives $-1 < R$ or $R < -\frac{1}{2}$. In the first case $x > 0$ is impossible, so $x \leq 0$; in the second $x < 0$ is impossible, so $x \geq 0$. Thus $x \geq 0$ or $x \leq 0$ for every point, which is LLPO by Corollary 12.18. One matrix and one eigenvector are used.

For part 2, LLPO and AC_ω give weak König's lemma (Theorem 12.72), and weak König's lemma turns approximate zeros into exact ones: a nonnegative Lipschitz function on the box $[-1, 1]^n$ with a point of value below every tolerance vanishes at a point. The tree keeps a string of bits when, at each complete level, a rational approximation of the function at the dyadic point the string names is small; the tree is infinite because the truncations of approximate

zeros pass every test, and along a path the dyadic points converge to a zero. The eigenvectors are found one at a time. With orthonormal eigenvectors $v_1, \dots, v_r$ found, $r < n$, the function

$$h(x) = |x|^2|Ax|^2 - (x \cdot Ax)^2 + \big(|x|^2 - 1\big)^2 + \sum_{c \le r} (x \cdot v_c)^2$$

vanishes exactly at the unit eigenvectors orthogonal to $v_1, \dots, v_r$, by the equality case of the Cauchy–Schwarz inequality. It has approximate zeros with no principle. Let W be the set of vectors orthogonal to $v_1, \dots, v_r$; it is invariant under A, because the v_c are eigenvectors and A is symmetric. The projections of the n coordinate vectors onto W have squared lengths summing to $n - r \ge 1$, so finitely many comparisons find one whose squared length exceeds $1/(2n)$; normalized, it is a unit vector u_0 of W. Let a be a rational bound for the norm of A, read from its entries, take a rational $c \ge a + 1$ and put $B = A + c$. Then $x \cdot Bx \ge (c - a)|x|^2 \ge |x|^2$ for every x, so the iterates $u_{k+1} = Bu_k/|Bu_k|$ are defined, stay in W and are unit vectors. Write $\rho_k = u_k \cdot Bu_k$ and $r_k = |Bu_k - \rho_k u_k|$, so that $|Bu_k|^2 = \rho_k^2 + r_k^2$. The Cauchy–Schwarz inequality for the positive form $(x, y) \mapsto x \cdot By$ gives $|Bu_k|^4 = \big(u_k \cdot B(Bu_k)\big)^2 \le \rho_k \, (Bu_k \cdot B^2 u_k)$, hence

$$\rho_{k+1} = \frac{Bu_k \cdot B^2 u_k}{|Bu_k|^2} \ge \frac{|Bu_k|^2}{\rho_k} = \rho_k + \frac{r_k^2}{\rho_k}.$$

The Rayleigh quotients ρ_k increase and lie between 1 and $c + a$, so $\sum_{k \le K} r_k^2 \le (c + a)^2$ for every K. Given $\varepsilon > 0$, take K with $(K + 1)\varepsilon^2 > (c + a)^2$; comparing each r_k, $k \le K$, with ε and 2ε finds a k with $r_k < 2\varepsilon$. Since $Bu_k - \rho_k u_k = Au_k - (u_k \cdot Au_k)u_k$, the first two terms of h at u_k sum to r_k^2 and the other terms vanish, so $h(u_k) < 4\varepsilon^2$. Every step is a finite computation. Countable choice enters through weak König's lemma and the rational approximations; nothing else uses a choice principle. □

The test matrix shows where LLPO enters. For $x > 0$ the levels of T_x are $0, x, \dots, (n - 1)x$, pairwise apart, and the eigenvectors are the coordinate axes; for $x < 0$ the vector $(1, \dots, 1)$ is an eigenvector; at $x = 0$ every vector is one. A rule that produces an eigenvector for every x decides on which side of 0 the point lies, that is, which family of eigenvectors the coincidence at $x = 0$ breaks into. When the levels are known to be apart there is nothing to decide.

Theorem 28.67 (Apart outcomes: the eigenbasis is free)**.** *Let H be an $n \times n$ Hermitian matrix whose entries are complex numbers with real and imaginary parts on $\mathbb{R}_\delta$, and let $\lambda_1, \dots, \lambda_n$ be points with $\lambda_c - \lambda_d \,\#\, 0$ for $c \ne d$. Suppose that for every c and every k there is a unit vector v such that, for every i, the real and imaginary parts of $(Hv)_i - \lambda_c v_i$ are at most $1/(k + 1)$ in size. Then there is an $n \times n$ matrix V with $V^\dagger V = 1$ and $HV = V\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)$. No omniscience principle and no choice principle is used.*

Proof. The exact eigenvectors are written by algebra from one approximate eigenvector for each value; no limit is taken.

Each λ_c is a root of the characteristic polynomial $p(t) = \det(t - H)$. Put $B = \lambda_c - H$. For a unit vector v the adjugate identity $\mathrm{adj}(B)\,B = \det(B) \cdot 1$ gives $\det B = \langle v, \mathrm{adj}(B)Bv \rangle$, so $|\det B| \le K|Bv|$ with K fixed by the entries of $\mathrm{adj}(B)$. The approximate eigenvectors make $|Bv|$ as small as we like, so $\det B = 0$, an equation between points proved from every precision.

The spectral projector. A monic polynomial of degree n that vanishes at n points whose differences are invertible is the product of the factors $t - \lambda_d$. Hence $q_c(t) = \prod_{d \neq c}(t - \lambda_d)$ satisfies $(t - \lambda_c)q_c(t) = p(t)$, and $q_c(\lambda_c) = \prod_{d \neq c}(\lambda_c - \lambda_d)$ is invertible. By the Cayley–Hamilton theorem $(H - \lambda_c)q_c(H) = p(H) = 0$. Since $q_c(t) - q_c(\lambda_c)$ is divisible by $t - \lambda_c$, the matrix $q_c(H)q_c(\lambda_c)^{-1}$ equals $1 + Q_c(H - \lambda_c)$ for a fixed matrix Q_c. So for every vector v the vector $u = v + Q_c(Hv - \lambda_c v)$ satisfies $Hu = \lambda_c u$.

One precision suffices. Take an approximate unit eigenvector v at a precision, fixed by the entries of Q_c, at which $|\langle v, Q_c(Hv - \lambda_c v)\rangle| \leq \frac{1}{4}$. Then $\langle u, u\rangle \geq \frac{1}{2}$, and $u/\sqrt{\langle u, u\rangle}$ is a unit eigenvector for λ_c. One vector for each of the n values is a finite choice, which needs no choice principle.

Orthogonality. For unit eigenvectors e_c and e_d with $c \neq d$, $\langle He_c, e_d\rangle = \langle e_c, He_d\rangle$ and the values are real, so $(\lambda_c - \lambda_d)\langle e_c, e_d\rangle = 0$; since $\lambda_c - \lambda_d$ is invertible, $\langle e_c, e_d\rangle = 0$. The matrix V with columns e_c is the one required. □

A real symmetric matrix is Hermitian, so the theorem applies to it, and the same proof carried out over $\mathbb{R}_\delta$ gives a real orthonormal basis; row 973 of the appendix states that real form, free with no choice. Everything a measurement uses is free in the same way: every level at every precision (rows 969 and 971), an exact eigenvector of a level apart from the others (row 974), and the projector onto a window of levels separated by gaps from the rest, even when levels inside the window coincide (row 977). An observable whose outcomes are apart has exact eigenstates, each fixed up to a phase, and the probabilities of its outcomes, with no omniscience principle. LLPO is the price of an eigenbasis when it is not known whether two levels coincide.

Conditions and scope.

Base and premises

$\mathbb{R}_\delta$ and the complex numbers over it; a Hermitian matrix given by its entries; values given as points, pairwise apart, each with approximate unit eigenvectors at every precision.

Conclusion

The spectral theorem in full costs exactly LLPO, with countable choice in the converse (row 1461); with apart values the exact unitary eigenbasis is free, with no choice.

Not claimed

An exact eigenbasis computed from the matrix alone when levels may coincide; that is the LLPO case.

Perron's theorem and the Schmidt decomposition

Two more theorems stand beside the table. Each is free with no choice at all, and each has a priced neighbour in the table.

Theorem 28.68 (Perron's theorem for positive matrices). *Let T be an $n \times n$ matrix on $\mathbb{R}_\delta$, $n \geq 1$, whose entries are all positive.*

1. *There are a point $\lambda > 0$ and a vector v with every entry positive and $\sum_i v_i = 1$ such that $Tv = \lambda v$.*
2. *Every eigenvalue is at most λ in size: if w is a complex vector with some entry apart from 0 and $Tw = \mu w$ for a complex number μ, then $|\mu| \leq \lambda$.*
3. *With $\mathbf{1}$ the vector of ones and $s_m = \sum_i (T^m \mathbf{1})_i$, there is a rational $q \geq 1$, read from the entries, with $0 \leq \frac{1}{m} \log s_m - \log \lambda \leq \frac{\log q}{m}$ for every $m \geq 1$; so $\frac{1}{m} \log s_m$ converges to $\log \lambda$.*

No omniscience principle and no choice principle is used.

Proof. A finite search over the entries gives rational bounds $1/(k+1) \leq T_{ij} \leq D$. With $\theta = 1/((k+1)(D+1))$ we have $0 < \theta \leq 1$ and $\theta D \leq 1/(k+1)$, and Birkhoff's contraction holds in ratio form: if $w \geq 0$, $\sum_i w_i = 1$ and $cw \leq u \leq dw$ entrywise, then

$$\big(c + \theta(\textstyle\sum_i u_i - c)\big)\, Tw \leq Tu \leq \big(d - \theta(d - \sum_i u_i)\big)\, Tw,$$

so the gap $d - c$ shrinks by the factor $1 - \theta$. The normalized iterates $x_m = T^m \mathbf{1} / \sum_i (T^m \mathbf{1})_i$ therefore satisfy $c_m x_m \leq T x_m \leq d_m x_m$ with $d_m - c_m \leq G(1-\theta)^m$ for a fixed G, and $|x_m(i) - x_{m+1}(i)| \leq (k+1)G(1-\theta)^m$. The iterates are computed from regular sequences naming the entries, by sums, products and the inverse of a sum known to be positive, so each coordinate of the limit $v = \lim x_m$ is the diagonal of computed sequences, as in part 2 of Theorem 12.93, and no choice is used. Put $\lambda = \sum_i (Tv)_i$; then $Tv = \lambda v$ by continuity. Since $\sum_i v_i = 1$ and $v \geq 0$, every $(Tv)_i \geq 1/(k+1)$, so $\lambda > 0$ and $v_i = (Tv)_i / \lambda > 0$. This is part 1.

For part 2, every $(Tv)_i$ is at most D, so $\lambda \leq nD$ and $v_i \geq 1/((k+1)nD)$. Take a rational M with $|w_i| \leq M$ for every i, which exists since every point has a rational bound, and put $c = (k+1)nD\,M$; then $|w_i| \leq M \leq c\, v_i$ for every i. Take an index i with $|w_i| > 0$. The entries of T are positive, so $|(Tx)_j| \leq (T|x|)_j$ for every complex vector x, where $|x|$ has the entries $|x_j|$; by induction $|\mu|^m |w| = |T^m w| \leq T^m |w| \leq c\, T^m v = c\lambda^m v$ entrywise. Hence $(|\mu|/\lambda)^m \leq c\, v_i / |w_i|$ for every m. If $|\mu| > \lambda$, the left side would exceed every bound; so $|\mu| \leq \lambda$.

For part 3, put $q = (k+1)nD$, a rational with $q \geq 1$, so that $v_i \geq 1/q$ and $v_i \leq 1$ for every i. Then $v \leq \mathbf{1} \leq q\, v$ entrywise, and since T^m has nonnegative entries, $\lambda^m v = T^m v \leq T^m \mathbf{1} \leq q\, T^m v = q\lambda^m v$. Summing over i gives $\lambda^m \leq s_m \leq q\lambda^m$, and taking logarithms gives the bound. □

Theorem 28.69 (The Schmidt decomposition with apart coefficients). *Let M be a real $n \times n$ matrix on $\mathbb{R}_\delta$, and let $\sigma_1, \ldots, \sigma_n$ be points with $\sigma_c > 0$ for every c and $\sigma_c - \sigma_d \# 0$ for $c \neq d$. Suppose that for every c and every k there are unit vectors u and v with $|(Mv)_i - \sigma_c u_i| \leq 1/(k+1)$ and $|(M^{\mathsf{T}} u)_i - \sigma_c v_i| \leq 1/(k+1)$ for every i. Then there are $n \times n$ matrices U and V with orthonormal columns such that $M = U \operatorname{diag}(\sigma_1, \ldots, \sigma_n) V^{\mathsf{T}}$. No omniscience principle and no choice principle is used.*

Proof. The matrix $A = M^{\mathsf{T}} M$ is symmetric, and an approximate pair (u, v) for σ_c gives an approximate eigenvector of A for σ_c^2:

$$Av - \sigma_c^2 v = M^{\mathsf{T}}(Mv - \sigma_c u) + \sigma_c (M^{\mathsf{T}} u - \sigma_c v).$$

The differences $\sigma_c^2 - \sigma_d^2 = (\sigma_c - \sigma_d)(\sigma_c + \sigma_d)$ are invertible. The four steps of the proof of Theorem 28.67, carried out over $\mathbb{R}_\delta$, give unit vectors e_c with $Ae_c = \sigma_c^2 e_c$, pairwise orthogonal. Let V have the columns e_c and U the columns Me_c/σ_c. Since $(Me_c)\cdot(Me_d) = e_c \cdot Ae_d$, the columns of U are orthonormal; and $MV = U\,\mathrm{diag}(\sigma)$ with $VV^{\mathsf{T}} = 1$ gives $M = U\,\mathrm{diag}(\sigma)V^{\mathsf{T}}$. □

A positive matrix is the transfer matrix of a one-dimensional lattice model with positive Boltzmann weights. For a chain of $m + 1$ sites with free ends the partition function is $s_m = \sum_i (T^m \mathbf{1})_i$, so by part 3 the free energy per site, $-\log Z$ per site, tends to $-\log \lambda$, with an explicit bound at every length. Perron's theorem gives the positive eigenvector v and its eigenvalue λ, which no other eigenvalue exceeds in size, all computed from the weights with no choice of approximations. Whether v is the only positive eigenvector up to scale is not claimed here. The table prices the neighbouring statement for Markov chains: every finite chain has a stationary law at exactly LLPO, with countable choice in the converse (row 1134), and a chain whose states are each entered from every state with positive probability has one for free (rows 1130 and 1135). A pure state of two n-level systems with a real amplitude matrix M is $\sum_c \sigma_c\, a_c \otimes b_c$ with orthonormal a_c and b_c. When its Schmidt coefficients σ_c are known to be positive and pairwise apart, the Schmidt bases exist with no omniscience principle and no choice, and with them the reduced density matrices and the entanglement entropy. The general Schmidt decomposition and the singular value decomposition cost exactly LLPO, with countable choice in the converse (rows 846 and 855): the price is again the coincidence of two coefficients.

Rows a course teaches

Table 28.6 lists the rows a first course in analysis teaches, with the place in this book that proves each one.

The prices LLPO and LPO in the table are exact in the table's sense: the theorems cited prove the converse with countable choice, and the choices they use, of representatives, of yes-or-no answers, of finite tables and at pairs of ratios, all lie inside the free allowance.

Two of the rows without a printed proof have short arguments. The Gaussian integral of a positive-definite form, written as $d_1 y_1^2 + d_2(y_2 + m_{21}y_1)^2 + \cdots$ with every weight d_i positive, is taken one coordinate at a time over the whole line; each step is the one-dimensional integral of Theorem 12.95 after a shift, and the value is the product of the square roots of π/d_i. For the Picard–Lindelöf theorem the Euler polygons of a Lipschitz law converge uniformly at a rate fixed by the Lipschitz constant and the bound on the law, so the motion is a limit with a known modulus. When the law is known only through its values at points, the proof chooses a modulus of continuity in time and a regular sequence for every point the polygons use; both are choices of ratios or natural numbers, inside the free allowance. When the law is a formula in the operations of the line, the polygons are computed on regular sequences of the initial data. Banach's theorem in Theorem 12.93 is the case of $[0, 1]$ with a rational factor; row 1335 is the whole line with any factor q, $0 \le q < 1$.

Table 28.6. Rows a first course teaches. Every price is exact in the table's sense. A dash means that the proof is not printed in this book; "on $[0,1]$ only" means that the theorem cited proves a special case of the row.

Statement	Row	Price	Proved in
Gaussian integral over the line	1136	free	Theorem 12.95
Gaussian integral of a positive-definite form	1159	free	–
Intermediate value theorem on $[0,1]$	1312	LLPO	Theorem 12.54
Heine–Borel on $[0,1]$	1418	FAN	Theorem 13.6
Monotone convergence	1302	LPO	Theorem 12.45
Bolzano–Weierstrass	1304	LPO	Theorem 12.47
Banach's fixed point theorem	1335	free	on $[0,1]$ only: Theorem 12.93
Picard–Lindelöf, law known through its values	616	free	–
Picard–Lindelöf, law given by a formula	620	free	–
Fekete's lemma	965	LPO	Proposition 28.63
Fekete's lemma, bounded defect	966	free	Proposition 28.63
Spectral theorem, symmetric matrices	1461	LLPO	Theorem 28.66

The open rows

Twenty-two rows have no exact price among the letters of the table. Thirteen concern compactness and continuity at each point: the Heine–Cantor theorem on compact sets, squares and cubes (rows 1440, 1441, 1057, 1088, 1089 and 1094), closed and bounded sets, closed subsets and continuous images of compact sets (rows 1052–1054, 1447 and 1450), sequential compactness (row 1444) and Brouwer's theorem on the square for maps continuous at each point (row 1058). Four concern closed and connected sets on the line (rows 1019, 1023, 1060 and 1079), two the convergence theorems of integration (rows 552 and 553), one functional analysis (row 1221), one linear algebra (row 684), and one is the derivative convention (row 10).

Five of them have an exact price in the key's sense in a principle that is not one of the table's letters. The derivative convention costs exactly MP together with a differentiability decision, and MP with the WEM function gives it (Theorem 28.24). Dini's theorem and the bounded convergence theorem for integrals (rows 552 and 553) are equivalent over countable choice to the measure form of the Heine–Borel theorem and to continuity of measure from above, the direction from the theorem using no choice. The Heine–Cantor theorem on $[0,1]$ (row 1441) is equivalent with no choice to the decidable fan theorem together with a presentation of the local control data by countable covers, and the continuous image of a compact set (row 1447) to a margin property of covers. Two more rows have an equivalence that uses choice in the direction from the statement, so it is not exact in the key's sense: Parseval's identity from a trivial orthogonal complement (row 1221) is MP together with the double negation of LPO for sequences, taken one instance at a time, with the choice of representatives, and a complete set

has an open complement (row 1023) exactly when two principles hold, over countable choice. The rest lie between a lower and an upper bound, which each row states after its price.

Three rows that sit next to these have exact prices. The uniform boundedness principles for functionals and for operators (rows 1232 and 1243) are exactly BD-N, Ishihara's boundedness principle, which the key of the appendix defines. Every complex square matrix has a unit eigenvector at exactly LLPO, with countable choice in the converse (row 933), and so does Schur triangularization (row 934): with LLPO and countable choice the eigenvector statement holds exactly when the approximate fundamental theorem of algebra does, and the fundamental theorem of algebra itself is free (row 482). A sequence that is bounded at infinitely many indices has a convergent subsequence at exactly LPO (row 1388), by the Bolzano–Weierstrass theorem.

What this section proves

This section proves its headline rows: rows 970, 983, 987–992 and 993 through Theorems 28.61, 28.62 and 28.64 and Corollary 28.65; rows 965 and 966 through Proposition 28.63; row 1461 and the free form for apart outcomes, row 973, through Theorems 28.66 and 28.67; and, beside the table, Theorems 28.68 and 28.69. Most rows that this book proves name the result in the appendix, and the rows of Table 28.6 with a theorem in their last column are proved there. For most other rows the proofs are not printed in this book: for those rows "exact" means that both directions hold, and the book records the statement and the price.

Questions to work through

1. (Theorem 28.9, items 1 and 4.) For the sign map S built from the WLPO function in the proof of item 4, show that $S(x) = 1$ exactly when $\neg(x \leq 0)$. Then assume MP and show that $|S(y)| > 0$ implies $y \# 0$, so that the jump of S carries a test. What does Theorem 28.8 then give, and why does the argument fail without MP?

2. (Theorem 28.45.) Over the strict hypotheses $0 < a$ and $a < b$, write a cone certificate for $a \cdot a \cdot a < b \cdot b \cdot b$, and check the identity it rests on. Which part of Theorem 28.42 would apply to this statement, and what weaker conclusion would it give?

3. (Theorem 28.50.) Using $\|d_n\| = 10n + 15$, find the least n with $\|d_n\| + 1 < 2^n$, and compare $\|d_n\|$ with the lower bound for $\|t_{d_n}\|$ there. For given C and k, check that the n chosen in the proof satisfies $C(\|d_n\| + 1)^k < 2^n$.

4. (Theorem 28.52, part 2, and the count after it.) From $\|B_K\| = 14K^2 + 28K + 15$ and the reported count $5K^5 \leq 6\,\|F(B_K)\|$, show directly that F has no exponent p/q with $2p < 5q$ on the derivations of shape at most 3. Why does the same family give no information about part 3?

5. (Theorem 28.53.) Show that a function t on the numbers with $x = t(x) + t(x)$ or $x = t(x) + t(x) + 1$ for every x is $x \mapsto \lfloor x/2 \rfloor$. Which results does the proof of the theorem use, and why does it need no evaluation of the program?

6. (Theorem 28.56, Proposition 28.57.) Show, classically, that every map φ of $\mathbb{R}$ into itself that preserves 0, 1, sums and products is the identity, and deduce item 2 of Theorem 28.56 from it. Then check Proposition 28.57 for $n = 2$.

Chapter 29

Time, chance and learning

A detector has been examined a thousand times and has never fired. What does that record let us conclude? Not that it never will: the next examination may fire, and no record of zeros refutes that (Section 17.7). A record of past events confirms some questions, refutes others, and settles more in the limit if its verdict may be revised as it grows. This chapter asks two further questions. What logic do records obey, and what does it cost to learn a law from them?

Here is a small case of the second question. Four laws label the points 0, 1 and 2; we write each law as its three labels in order:

$$h_1 = 000, \qquad h_2 = 011, \qquad h_3 = 101, \qquad h_4 = 110.$$

One of them is the true law. A learner is shown the points in order, predicts each label, and is then told the true one. The *halving learner* predicts the label that at least half of the laws still consistent with what it has been told predict, and 1 at a tie. Suppose the truth is h_1. At point 0 two laws say 1 and two say 0; the learner says 1 and errs, and h_3 and h_4 are refuted. At point 1, h_1 says 0 and h_2 says 1; the learner says 1 and errs again, and only h_1 remains, so point 2 is predicted correctly. Two mistakes, and no learner of these four laws is safe from two. Whatever a learner predicts at point 0, let the truth be the opposite label: two laws remain with that label, and they disagree at point 1, so the same move costs the learner a second mistake. The halving learner is also never worse than this: each of its mistakes refutes at least half of the surviving laws, the true law always survives, and so m mistakes need $2^m \leq 4$.

Finite learning of this kind costs nothing: every step is a count. Prices appear when the candidate laws are infinitely many, or when a law is a whole infinite record. Identifying which of two worlds known to differ is the true one costs $\mathrm{MP}_{\mathrm{bin}}$, whether the learner must commit once or may settle in the limit; identifying which of two worlds that may be equal is the true one costs LPO either way (Theorem 29.24). For the worlds computed by programs of Gödel's system T, learning every one in the limit costs exactly LPO for the sequences those programs compute (Theorem 29.45). Over arbitrary systems of programs, the method that runs the least program reproducing the record costs exactly Σ_2-LPO, the form of LPO for questions one quantifier higher (Theorem 29.51); the same principle decides every question that verdicts revised without a bound confirm (Proposition 29.50). How often a learner of any kind must revise is read off its candidate laws alone (Theorem 29.58); a computable learner can need more (Theorem 29.59). The stage at which a learner held to few revisions commits is a certificate: every deeply imitated rival law already differs before it (Theorem 29.61). On a catalogue built from running times, three revisions make that stage, with a table of answers to two-quantifier questions, settle a question with three alternating quantifiers (Theorem 29.62), and for every $r \geq 1$ a catalogue built from nested sums has r revisions too few and $r+2$ enough (Theorem 29.64).

The chapter has two halves. The first concerns time and chance. Section 29.1 shows that one recorded history has exactly the logic of Gödel and Dummett and the tree of possible

records exactly intuitionistic logic, and prices the laws that hold at the start of a history. Section 29.2 derives chance from exchangeable records: finitely many are close to a mixture of coins, infinitely many are exactly one, and on $\mathbb{R}_\delta$ deciding whether the mixture is a single coin costs WLPO. Section 29.3 shows that the recognition history of Chapter 21 is the least varied record that never repeats, finds the places of δ's ratios, and reads dimension off the growth of comparisons. Section 29.4 proves Brouwer's fixed-point theorem on a triangle in $\mathbb{R}_\delta^2$: approximately for free, exactly at the price of LLPO when countable choice is assumed.

The second half concerns learning. Section 29.5 prices identification of a law from its records and the revision of verdicts; Section 29.6 counts mistakes, samples and predictions; Section 29.7 treats learners that are programs; Section 29.8 learns with the cost J of Chapter 21; Section 29.9 prices learning on the programs of Gödel's T; Section 29.10 prices the standard methods and their bounds over every system of programs, and separates decision tables from propositions; and Section 29.11 counts the changes of mind that a catalogue of laws forces.

Maps are read as in Section 21.1: rules given as data, without unique choice. A result is *free* when its proof uses no omniscience principle and no choice principle; a proof *with no choice principle* may use an omniscience principle named as a hypothesis; and a result *proved in classical mathematics* is said to be so before its statement. Worlds, records $w\restriction n$ and guessing rules are those of Section 17.7, and the principles are those of Definition 7.8. A ratio written among points of $\mathbb{R}_\delta$ stands for its point.

29.1 The logic of one recorded history

A record grows, and what it has confirmed stays confirmed: a later record keeps everything an earlier one recorded. Kripke's semantics of intuitionistic logic is built on exactly this persistence [53]. We use it to ask which logical laws records obey, and what the shape of time, a single history or a tree of possible histories, does to that logic.

Definition 29.1 (Record frames and forcing)**.**

1. A *record frame* is a set W of *stages* with a preorder $\leq$, a reflexive and transitive relation; $w \leq u$ says that u is a later record keeping everything w recorded. *One history* is $(\mathbb{N}, \leq)$. The *record tree* is the set of finite binary words, with $u \leq v$ when u is an initial segment of v.

2. *Formulas* are built from atoms $p_0, p_1, \ldots$ and $\bot$ by $\wedge$, $\vee$ and $\to$; $\neg a$ abbreviates $a \to \bot$. We name three:

$$\mathrm{lin} = (p_0 \to p_1) \vee (p_1 \to p_0), \qquad \mathrm{wem} = \neg p_0 \vee \neg\neg p_0, \qquad \mathrm{em} = p_0 \vee \neg p_0.$$

3. A *persistent valuation* V assigns to every atom p a set $V(p)$ of stages closed upward: $w \in V(p)$ and $w \leq u$ give $u \in V(p)$.

4. *Forcing* is defined by recursion on formulas: $w \Vdash p$ when $w \in V(p)$; never $w \Vdash \bot$; $w \Vdash a \wedge b$ when $w \Vdash a$ and $w \Vdash b$; $w \Vdash a \vee b$ when $w \Vdash a$ or $w \Vdash b$; and $w \Vdash a \to b$ when every $u \geq w$ with $u \Vdash a$ has $u \Vdash b$. A formula is *valid on* W when it is forced at every stage under every persistent valuation.

5. *W never branches* when $w \le u$ and $w \le v$ give $u \le v$ or $v \le u$; it is *confluent* when $w \le u$ and $w \le v$ give some z with $u \le z$ and $v \le z$; and it is *static* when $w \le u$ gives $u \le w$.

Forcing persists: if $w \le u$ and $w \Vdash a$, then $u \Vdash a$. The proof is an induction on a that uses transitivity at $\to$ and no principle.

Theorem 29.2 (Three shapes of time and their logical laws)**.** *For every record frame W:*

1. lin *is valid on W if and only if W never branches;*
2. wem *is valid on W if and only if W is confluent;*
3. em *is valid on W if and only if W is static; in particular, if some $w \le u$ has $u \not\le w$, so that a later record adds something, then* em *is not valid on W;*
4. *one history never branches, is confluent and is not static; the record tree branches and is not confluent.*

The implications from validity to shape in parts 1 and 3, the clause on a later record, and part 4 use no principle. The other implications are proved in classical mathematics.

Proof. For stages u and v let $V_{u,v}$ be the valuation in which p_0 holds exactly at the stages above u and every other atom exactly at the stages above v.

(1) Let $w \le u$ and $w \le v$, and force lin at w under $V_{u,v}$. If $p_0 \to p_1$ is forced at w, then u, which forces p_0, forces p_1, so $v \le u$. If $p_1 \to p_0$ is forced, then $u \le v$. Conversely let W never branch, and suppose $p_0 \to p_1$ is not forced at w: some $u \ge w$ forces p_0 and not p_1. Every $v \ge w$ forcing p_1 is comparable with u. If $u \le v$, then v forces p_0 by persistence; $v \le u$ would make u force p_1. So $p_1 \to p_0$ is forced at w.

(2) Let $w \le u$ and $w \le v$, under $V_{u,u}$. The stage $u \ge w$ forces p_0, so $\neg p_0$ is not forced at w, and validity gives $w \Vdash \neg\neg p_0$. So v does not force $\neg p_0$: some $y \ge v$ forces p_0, that is, $u \le y$. Conversely, let W be confluent. If some $u \ge w$ forces p_0, then every $x \ge w$ has a common upper bound with u, which forces p_0; so no $x \ge w$ forces $\neg p_0$, and $w \Vdash \neg\neg p_0$. Otherwise no stage above w forces p_0, and $w \Vdash \neg p_0$.

(3) Let $w \le u$, under $V_{u,u}$. The stage u refutes $\neg p_0$ at w, so validity gives $w \Vdash p_0$, that is, $u \le w$. If $u \not\le w$, then neither p_0 nor $\neg p_0$ is forced at w, which is the clause on a later record. Conversely, in a static frame, if p_0 is not forced at w, any $u \ge w$ forcing it has $u \le w$ and would make w force it; so $w \Vdash \neg p_0$.

(4) $\mathbb{N}$ is totally ordered, $\max(u, v)$ bounds u and v, and $0 \le 1$ while $1 \not\le 0$. In the tree, the words 0 and 1 extend the empty word, are incomparable, and have no common extension. □

A history can now be asked about events. Let an event be a binary sequence α, with $\alpha(n) = 1$ recording that it was seen at examination n.

Definition 29.3 (The history of two events)**.** For binary sequences α and β, the *history* $H(\alpha, \beta)$ is the valuation on one history in which p_0 holds at stage n when $\alpha(k) = 1$ for some $k \leq n$, and every other atom holds at n when $\beta(k) = 1$ for some $k \leq n$. We write $H(\alpha)$ for $H(\alpha, \alpha)$, and $\Sigma_1(\alpha)$ for the statement that $\alpha(n) = 1$ for some n.

At stage 0 a history has seen one examination of each event. Which laws are forced there, in every history? Each answer is an omniscience principle.

Theorem 29.4 (The prices of a history's laws at its start)**.**

1. lin *is forced at stage* 0 *of* $H(\alpha, \beta)$ *for all binary* α *and* β *if and only if LLPO holds.*
2. wem *is forced at stage* 0 *of* $H(\alpha)$ *for every binary* α *if and only if WLPO holds.*
3. *Not every* $H(\alpha)$ *forces* em *at stage* 0*.*
4. *"If* $\neg\neg p_0$ *is forced at stage* 0 *of* $H(\alpha)$*, then* $\Sigma_1(\alpha)$*" holds for every binary* α *if and only if* $\mathrm{MP}_{\mathrm{bin}}$ *holds.*
5. *The statement that* $(\Sigma_1(\alpha) \to \Sigma_1(\beta)) \vee (\Sigma_1(\beta) \to \Sigma_1(\alpha))$ *for all binary* α *and* β *is equivalent to LLPO; the statement that* $\neg\Sigma_1(\alpha) \vee \neg\neg\Sigma_1(\alpha)$ *for every binary* α*, to WLPO; and the statement that* $\Sigma_1(\alpha) \vee \neg\Sigma_1(\alpha)$ *for every binary* α*, to LPO.*

Every implication is proved with no choice principle, and part 3 uses no principle at all.

Conditions and scope.

Base and premises
One history $(\mathbb{N}, \leq)$ with the valuations $H(\alpha, \beta)$; binary sequences as events.

Conclusion
Prelinearity at the start of every history is exactly LLPO, weak excluded middle exactly WLPO, and turning "the event cannot stay unseen" into "the event is seen" exactly $\mathrm{MP}_{\mathrm{bin}}$; excluded middle fails outright, at no price.

Not claimed
Prices of formulas with atoms beyond two events, or of laws forced at later stages.

Proof. For binary α and β, call stage k *quiet* when $\alpha(i) = \beta(i) = 0$ for every $i < k$, and define the *race*

$$r(2k) = 1 \iff \alpha(k) = 1,\ \beta(k) = 0 \text{ and } k \text{ is quiet},$$
$$r(2k+1) = 1 \iff \beta(k) = 1,\ \alpha(k) = 0 \text{ and } k \text{ is quiet},$$

and $r = 0$ elsewhere: $r(2k) = 1$ says that α is seen strictly first, at stage k. The race has at most one entry 1. Two entries at stages $k \leq k'$ have $k = k'$, since k' quiet makes $\alpha(k) = \beta(k) = 0$ when $k < k'$; and at one stage only one of the two cases can hold. For every u, a search through the stages up to u either shows them all quiet or finds the first stage $m \leq u$ with $\alpha(m) = 1$ or $\beta(m) = 1$.

(1) Assume LLPO and apply it to the race. Suppose $r(2k) = 0$ for every k, and let stage u force p_0, so $\alpha(i) = 1$ for some $i \le u$. The search finds the first stage $m \le u$ at which α or β is 1; m is quiet, and $\beta(m) = 0$ would give $r(2m) = 1$. So $\beta(m) = 1$ and u forces p_1: the formula $p_0 \to p_1$ is forced at 0. If instead $r(2k+1) = 0$ for every k, then $p_1 \to p_0$ is forced. Conversely, let α have at most one entry 1, and put $e(k) = \alpha(2k)$ and $o(k) = \alpha(2k+1)$. In $H(e, o)$, suppose $p_0 \to p_1$ is forced at 0. If $\alpha(2k) = 1$, stage k forces p_0, hence p_1, so $\alpha(2i+1) = 1$ for some $i \le k$, a second entry 1. So $\alpha(2k) = 0$ for every k. The other disjunct gives the odd places.

(2) In $H(\alpha)$, stage 0 forces $\neg p_0$ exactly when α is all zero: an entry $\alpha(n) = 1$ makes stage n force p_0. Stage 0 forces $\neg\neg p_0$ exactly when α is not all zero: if α is all zero, stage 0 forces $\neg p_0$; and if α is not all zero, no stage x forces $\neg p_0$, for that would make α all zero, since $\alpha(n) = 1$ makes stage $x + n$ force p_0. So wem at stage 0 of $H(\alpha)$ says that α is all zero or not, and the equivalence with WLPO follows.

(3) Let $\alpha(n) = 1$ exactly when $n = 1$. Stage 0 does not force p_0, as $\alpha(0) = 0$, and does not force $\neg p_0$, as stage 1 forces p_0.

(4) By (2), $\neg\neg p_0$ is forced at stage 0 exactly when α is not all zero. So the statement is $\mathrm{MP}_{\mathrm{bin}}$ itself.

(5) If $p_0 \to p_1$ is forced at stage 0 of $H(\alpha, \beta)$ and $\alpha(i) = 1$, then stage i forces p_1, so $\Sigma_1(\beta)$; hence (1) gives the first equivalence from LLPO, and the argument of (1) with e and o gives LLPO back. Since $\neg\Sigma_1(\alpha)$ says that α is all zero, the other two statements are WLPO and LPO as written in Definition 7.8. □

Read with stages as the records of successive examinations, the theorem says that one linear time costs LLPO, convergent futures cost WLPO, a closed future costs LPO, and waiting for an event that cannot stay unseen costs $\mathrm{MP}_{\mathrm{bin}}$, the price that Theorem 17.17 found for a recorder. The reading is ours; the theorem is about frames and binary sequences.

The complete logics of the two frames

Which formulas are valid on one history, and which on the record tree? Two calculi answer.

Definition 29.5 (Intuitionistic logic and Gödel–Dummett logic)**.** *Intuitionistic propositional logic*, IPC, derives formulas from the axiom schemas

$$a \to (b \to a), \qquad (a \to (b \to c)) \to ((a \to b) \to (a \to c)),$$
$$a \wedge b \to a, \qquad a \wedge b \to b, \qquad a \to (b \to a \wedge b), \qquad a \to a \vee b, \qquad b \to a \vee b,$$
$$(a \to c) \to ((b \to c) \to (a \vee b \to c)), \qquad \bot \to a$$

by modus ponens. *Gödel–Dummett logic*, LC, adds every instance of $(a \to b) \vee (b \to a)$ as an axiom.

Theorem 29.6 (The tree has intuitionistic logic and one history has LC)**.**

1. *A formula is valid on the record tree if and only if IPC derives it.*
2. *A formula is valid on one history if and only if LC derives it.*
3. *IPC does not derive $p_0 \vee \neg p_0$; LC derives $(p_0 \to p_1) \vee (p_1 \to p_0)$ and IPC does not.*

The theorem is proved in classical mathematics: Lindenbaum's lemma uses Zorn's lemma, and the transfer between frames splits on whether a stage forces a formula. That every formula IPC derives is valid on every frame uses no principle.

Conditions and scope.

Base and premises
Propositional formulas, persistent valuations of atoms, the two calculi of Definition 29.5; classical metatheory.

Conclusion
The tree of possible records is a complete frame for intuitionistic logic, and one history a complete frame for Gödel–Dummett logic: no further law holds of either.

Not claimed
First-order logic, or that the propositional rules of Chapter 5 form exactly the calculus IPC; that identification is not proved here.

Proof. Soundness. Each axiom of IPC is forced at every stage of every frame, by persistence for the first two schemas and directly for the others, and modus ponens keeps a formula forced at every stage. On a frame that never branches every instance of $(a \to b) \vee (b \to a)$ is forced: the proof of Theorem 29.2(1) applies with a and b in place of p_0 and p_1.

Completeness for all frames. This is Kripke's theorem [53]; we recall the canonical model. Write $\Gamma \vdash \varphi$ for derivability from hypotheses Γ in IPC, or in LC. The first two schemas give the deduction theorem. If $\Gamma \nvdash \varphi$, Zorn's lemma gives a maximal $\Delta \supseteq \Gamma$ with $\Delta \nvdash \varphi$, since a derivation uses finitely many hypotheses. Such a Δ is closed under derivation, omits $\bot$, and is *prime*: if $a \vee b \in \Delta$ but $a, b \notin \Delta$, maximality gives $\Delta, a \vdash \varphi$ and $\Delta, b \vdash \varphi$, so $\Delta \vdash \varphi$ by the deduction theorem and the schema for $\vee$. The prime theories, ordered by inclusion, form a frame; let an atom hold at T when it belongs to T. Then $T \Vdash \psi$ exactly when $\psi \in T$, by induction on ψ; at $a \to b \notin T$, the deduction theorem gives $T \cup \{a\} \nvdash b$, and a prime $U \supseteq T \cup \{a\}$ omitting b refutes $a \to b$ at T. A formula valid on every frame therefore lies in every prime theory, and so is derivable: otherwise a prime theory omits it. For LC the canonical frame never branches. If $T \subseteq U$ and $T \subseteq V$ with $a \in U \setminus V$ and $b \in V \setminus U$, then $(a \to b) \vee (b \to a) \in T$; primeness puts $a \to b$ in $T \subseteq U$, so $b \in U$, or $b \to a$ in $T \subseteq V$, so $a \in V$. Hence a formula valid on every frame that never branches is derivable in LC.

Transfer. It remains to show that validity on the tree gives validity on every frame, and validity on one history gives validity on every frame that never branches. Fix φ and its finite set of subformulas. For a frame W with a valuation, the *type* $\mathrm{typ}(z)$ of a stage z is the set of subformulas of φ forced at z; by persistence $z \le z'$ gives $\mathrm{typ}(z) \subseteq \mathrm{typ}(z')$. Let A be a frame and $f : A \to W$ any function with

forth $x \le y$ in A gives $\mathrm{typ}(f(x)) \subseteq \mathrm{typ}(f(y))$;

back for every $x \in A$ and every $z \ge f(x)$ in W there is $y \ge x$ with $\mathrm{typ}(f(y)) = \mathrm{typ}(z)$.

Let an atom hold at $x \in A$ when it lies in $\mathrm{typ}(f(x))$; by forth this valuation is persistent. Then $x \Vdash_A \psi$ exactly when $f(x) \Vdash_W \psi$, for every subformula ψ. The induction is immediate except at $a \to b$. If $f(x)$ forces it and $y \ge x$ forces a, then $f(y)$ forces a by induction and $a \to b$ by forth, so it forces b, and so does y. If x forces it and $z \ge f(x)$ forces a, back gives $y \ge x$ of the type of z; then y forces a, hence b, and b lies in the type of z. So if φ is valid on A, it is forced at every $f(x)$.

The tree. Fix W, a valuation and a stage w_0. List the sets of subformulas as $S_0, \dots, S_{K-1}$, and for each j realized as a type pick a stage c_j of type S_j. Walk a binary word from the state $(w_0, 0)$, a stage and a counter. A letter 1 raises the counter. A letter 0 at counter $2j$ moves to $(c_j, 0)$ when c_j exists and the current type is contained in S_j, and otherwise resets the counter to 0; at an odd counter it resets. Let f send a word to the stage it reaches. No move loses a forced subformula, which gives forth. For back, let the word x end at (c, i) and let $z \ge c$ have type S_j, which contains the type of c. The word $x1^{i+1}0$ ends at $(c, 0)$, and $x1^{i+1}01^{2j}0$ ends at c_j. So the tree transfers to every frame, at every stage $w_0 = f(\varnothing)$.

One history. Let W never branch, and fix a valuation and a stage w_0. Any two stages above w_0 are comparable, so their types form a chain. Put $g(0) = w_0$, and let $g(n+1)$ be a stage above w_0 whose type is the least type strictly containing that of $g(n)$, or $g(n)$ if there is none. Types grow along g, which gives forth; and g passes through every type of a stage above $g(n)$, one at a time, which gives back. So one history transfers to every frame that never branches.

Putting the pieces together, validity on the tree gives validity on every frame, hence derivability in IPC, hence validity on the tree; and validity on one history gives validity on every frame that never branches, hence derivability in LC, hence validity on one history, which never branches. For part 3, em fails on the frame $0 < 1$ by Theorem 29.2(3), and lin, an axiom of LC, fails on the tree by Theorem 29.2(1) and (4). □

Gödel introduced the finite-valued logics from which LC descends [128], and Dummett gave LC its axiom and its semantics [125]. The new point is the transfer: two particular frames, the possible records and a single history, already carry the whole logic.

29.2 Chance from exchangeable records

A coin is tossed n times and the outcomes are recorded. Suppose nothing distinguishes the order of the tosses: every rearrangement of a record is as probable as the record. De Finetti's theorem says that such records behave as if a coin of unknown bias had been tossed. Chance is then not posited; it is derived from a symmetry of the record. We prove a finite form with an explicit bound, the exact form for unbounded runs, and then price on $\mathbb{R}_\delta$ the question whether the coin is a single coin.

Definition 29.7 (Exchangeable laws of records)**.**

1. A *law of n binary records* is a map P from $\{0,1\}^n$ to the nonnegative rationals with $\sum_x P(x) = 1$. It is *exchangeable* when $P(x \circ \sigma) = P(x)$ for every record x and every permutation σ of the positions $0, \dots, n-1$.

2. For $m \le n$, the law of the first m records is $P_m(y) = \sum P(x)$, summed over the x whose first m entries are y.

3. Write $k(x)$ for the number of entries 1 of a record x. The *coin law* of bias p on $\{0,1\}^m$ is $c_p(y) = p^{k(y)}(1-p)^{m-k(y)}$, and the *coin mixture* of P is

$$Q_m(y) = \sum_{x \in \{0,1\}^n} P(x)\, c_{k(x)/n}(y) \ :$$

   toss m times a coin whose bias is the frequency of 1 in a record drawn from P.

4. $(n)_m = n(n-1)\cdots(n-m+1)$ counts the ways to read m distinct positions of n in order.

Theorem 29.8 (Finitely many exchangeable records are nearly a mixture of coins)**.** *Let $m \le n$ and let P be an exchangeable law of n binary records. Then the sum of the absolute differences satisfies*

$$\sum_{y \in \{0,1\}^m} \left| P_m(y) - Q_m(y) \right| \ \le \ 2\Big(1 - \frac{(n)_m}{n^m}\Big) \ \le \ \frac{m(m-1)}{n}.$$

The theorem is proved in classical mathematics; the argument is finite counting over the rationals.

Both P_m and Q_m are probability laws on $\{0,1\}^m$, so the largest difference between the probabilities they give one set of records, their total variation distance, is half this sum: at most $1 - (n)_m/n^m$, hence at most $m(m-1)/(2n)$. For ten records of which two are read, $(10)_2/10^2 = 0.9$, and the sum of the absolute differences is at most 0.2.

Proof. (1) Every injective map i from $\{0, \dots, m-1\}$ to the positions extends to a permutation of the positions. By exchangeability, the records read at the positions $i(0), \dots, i(m-1)$ have the law P_m: the sum of $P(x)$ over the x with $x \circ i = y$ is $P_m(y)$.

(2) Average (1) over the $(n)_m$ injective maps. With $\mathrm{inj}(x,y)$ the number of injective i with $x \circ i = y$,

$$P_m(y) = \sum_x P(x)\, \frac{\mathrm{inj}(x,y)}{(n)_m} \ :$$

P_m is the P-mixture of reading m positions of x without repetition.

(3) Reading with repetition gives the coin law. A map i with $x \circ i = y$ sends each of the $k(y)$ entries 1 of y to one of the $k(x)$ entries 1 of x, and each entry 0 to one of the $n - k(x)$ entries 0. So the number $\mathrm{all}(x,y)$ of all maps i with $x \circ i = y$ is $k(x)^{k(y)}(n-k(x))^{m-k(y)}$, and $\mathrm{all}(x,y)/n^m = c_{k(x)/n}(y)$. Hence $Q_m(y) = \sum_x P(x)\,\mathrm{all}(x,y)/n^m$.

(4) Fix x and write all $=$ inj $+$ rep, where $\mathrm{rep}(x,y)$ counts the maps that repeat a position. Since $\sum_y \mathrm{inj}(x,y) = (n)_m$ and $\sum_y \mathrm{rep}(x,y) = n^m - (n)_m$,

$$\sum_y \Big|\frac{\mathrm{inj}}{(n)_m} - \frac{\mathrm{all}}{n^m}\Big| \le \sum_y \mathrm{inj}\Big(\frac{1}{(n)_m} - \frac{1}{n^m}\Big) + \sum_y \frac{\mathrm{rep}}{n^m} = 2\Big(1 - \frac{(n)_m}{n^m}\Big).$$

Weighting by $P(x)$ and using the triangle inequality gives the first bound.

(5) $(n)_m/n^m = \prod_{i<m}(1 - i/n)$, and $\prod_i (1-a_i) \ge 1 - \sum_i a_i$ for numbers a_i in $[0,1]$, by induction on their number. So $(n)_m/n^m \ge 1 - m(m-1)/(2n)$, which is the second bound. □

Finite forms of de Finetti's theorem with bounds of this order are due to Diaconis and Freedman [124]. When the records go on without end, the approximation becomes an identity.

Theorem 29.9 (De Finetti's theorem for binary records)**.** *Work in classical mathematics, with the classical real numbers and Borel measures. Let (P_n) be an* exchangeable run*: for every n, P_n is an exchangeable map from $\{0,1\}^n$ to the nonnegative reals with sum* 1*, and the laws are* consistent*, $P_n(x) = P_{n+1}(x0) + P_{n+1}(x1)$. Then there is exactly one Borel probability measure ν on $[0,1]$ with*

$$P_n(x) = \int_0^1 p^{k(x)}(1-p)^{n-k(x)}\, d\nu(p) \qquad \text{for every } n \text{ and every } x \in \{0,1\}^n.$$

Its jth moment is $m_j = P_j(1\cdots 1)$, and for $j \le n$ the frequency moments satisfy

$$\Big|\sum_x P_n(x)\Big(\frac{k(x)}{n}\Big)^j - m_j\Big| \le \frac{j(j-1)}{n}.$$

Proof. Frequency moments. In the notation of Theorem 29.8 for the law P_n and $m = j$, the frequency moment $\sum_x P_n(x)(k(x)/n)^j$ is $Q_j(1\cdots 1)$, and by consistency the law of the first j records of P_n is P_j, so m_j is $P_j(1\cdots 1)$. The theorem bounds their difference by $j(j-1)/n$.

Existence. Let ν_n be the law of the frequency $k(x)/n$ under P_n. The probability measures on $[0,1]$ are compact under weak convergence (Helly's selection theorem), so a subsequence of (ν_n) converges weakly to some ν. Moments are integrals of continuous functions, so the jth moment of ν is the limit of the frequency moments, which is m_j.

Representation. The family $Q_n(x) = \int p^{k(x)}(1-p)^{n-k(x)} d\nu$ is exchangeable and consistent, since $p + (1-p) = 1$, and $Q_n(1\cdots 1) = m_n = P_n(1\cdots 1)$. Two exchangeable consistent families that agree on the records of 1s agree everywhere. Indeed, by exchangeability each depends only on the length n and the number k of entries 1; write $q(n,k)$. Consistency, after a permutation moves the new entry, reads $q(n,k) = q(n+1,k) + q(n+1,k+1)$. So $q(n+1,k) = q(n,k) - q(n+1,k+1)$, and induction on the number $n-k$ of entries 0 fixes every value from the values $q(n,n)$.

Uniqueness. Two probability measures on $[0,1]$ with the same moments integrate every polynomial alike, hence every continuous function, by Weierstrass's approximation theorem, and so they are equal. □

The theorem is de Finetti's [122, 123], proved here by the route through moments [126, Chapter VII]. It speaks of the classical real numbers. On $\mathbb{R}_\delta$ we ask a sharper question: is the mixing law a single coin?

Definition 29.10 (Runs on the line)**.** A *run on* $\mathbb{R}_\delta$ assigns to every binary record x of every length n a point $P_n(x) \geq 0$ of $\mathbb{R}_\delta$, with $P_0 = 1$ on the empty record, each P_n exchangeable and the family consistent as in Theorem 29.9. Its *moments* are $m_j = P_j(1 \cdots 1)$, and its *variance* is $v = m_2 - m_1^2$. The run has a *fixed chance* when $P_n(x) = m_1^{k(x)}(1 - m_1)^{n-k(x)}$ for every record x.

Consistency gives $m_0 = 1$ and $0 \leq m_{j+1} \leq m_j$, since $m_j = P_{j+1}(1 \cdots 10) + m_{j+1}$.

Theorem 29.11 (The price of a fixed chance)**.**

1. *A run on* $\mathbb{R}_\delta$ *has a fixed chance if and only if its second moment is the square of its first,* $m_2 = m_1^2$.
2. *"Every run on* $\mathbb{R}_\delta$ *has a fixed chance or does not" is equivalent to WLPO.*
3. *"Every run on* $\mathbb{R}_\delta$ *without a fixed chance has a variance apart from* 0*" is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$.

Part 1 is free, and the equivalences of parts 2 and 3 are proved with no choice principle.

Conditions and scope.

Base and premises

Runs whose probabilities are points of $\mathbb{R}_\delta$, exchangeable and consistent; nothing else is assumed of them.

Conclusion

Second moment equal to the squared first moment is a single coin, with no omniscience. Deciding whether a run is a single coin costs WLPO; certifying, from its failure, a visible spread costs $\mathrm{MP}_{\mathrm{bin}}$.

Not claimed

The price of deciding whether the classical mixing law of Theorem 29.9 has an atom, or of comparing it with a ratio; or that a physical source is such a run, which is an identification.

Proof. (1) A fixed chance gives $m_2 = m_1^2$ at $x = 11$. For the converse, assume $m_2 = m_1^2$. Fix N and let c be the number of entries 1 of a record of length N. The expected value of the binomial coefficient $\binom{c}{r}$ under P_N is

$$\sum_x P_N(x)\binom{k(x)}{r} = \binom{N}{r} m_r \ :$$

$\binom{k(x)}{r}$ counts the sets of r positions at which x is 1, and for a fixed set the total probability of the records that are 1 there is m_r, by a permutation that moves the set to the front and by consistency. We write $E[\cdot]$ for such expected values. From $c^2 = 2\binom{c}{2} + c$ and $c\binom{c}{j} = (j+1)\binom{c}{j+1} + j\binom{c}{j}$,

$$E[(c - Nm_1)^2] = Nm_1(1 - m_1), \qquad E\Big[\binom{c}{j}(c - Nm_1)\Big] = \binom{N}{j}\big(Nd_j - j(m_{j+1} - m_j)\big),$$

where $d_j = m_{j+1} - m_1 m_j$; the first uses $m_2 = m_1^2$. Let $j \le N$, $M \ge 1$ and $e = \pm 1$. For $0 \le c \le N$ the quantity

$$V(c) = \Big(\binom{N}{j}(c - Nm_1) - eM\binom{c}{j}\Big)^2 + M^2\binom{c}{j}\Big(\binom{N}{j} - \binom{c}{j}\Big)$$

is at least 0, since $\binom{c}{j} \le \binom{N}{j}$. Expanding, the terms in $\binom{c}{j}^2$ cancel, and its expected value divided by $\binom{N}{j}^2$ gives

$$0 \le Nm_1(1 - m_1) - 2eM\big(Nd_j - j(m_{j+1} - m_j)\big) + M^2 m_j.$$

Take $N = M^2$ with $M \ge j + 1$, and divide by $2M^2$:

$$eMd_j \le \tfrac{1}{2}\big(m_1(1 - m_1) + m_j\big) + \frac{ej(m_{j+1} - m_j)}{M} \le 1 + j,$$

because $m_1(1-m_1) \le \frac{1}{4}$, $0 \le m_{j+1} \le m_j \le 1$ and $M \ge 1$. With $e = 1$ and $e = -1$, $|d_j| \le (1+j)/M$ for every $M \ge j + 1$. A point bounded so is neither positive nor negative, by the Archimedean property (Proposition 12.19), so $d_j = 0$ and $m_{j+1} = m_1 m_j$. By induction $m_j = m_1^j$. The family $m_1^k(1 - m_1)^{n-k}$ is exchangeable and consistent with the same values on records of 1s, so the induction on entries 0 in the proof of Theorem 29.9, which is algebra and uses no principle, shows that the run has the fixed chance m_1.

(2) For a point s with $0 \le s \le \frac{1}{2}$, the *two-coin run* mixes the coins of bias $\frac{1}{2} + s$ and $\frac{1}{2} - s$ with weight $\frac{1}{2}$ each. It is a run on $\mathbb{R}_\delta$, with $m_1 = \frac{1}{2}$ and $m_2 = \frac{1}{4} + s^2$, so its variance is s^2, and by (1) it has a fixed chance exactly when $s = 0$. For a binary sequence α, let s_α be half the jump point of α (Lemma 12.5, all signs +1): it is 2^{-N-1} if the first entry 1 of α is at N, and 0 if α is all zero. Deciding whether the two-coin run of s_α has a fixed chance decides whether α is all zero: that is WLPO. Conversely, WLPO decides $x = 0$ or $x \ne 0$ for every point x (Corollary 12.18); applied to $m_2 - m_1^2$, it decides the fixed chance by (1).

(3) $\mathrm{MP}_{\mathrm{bin}}$ is equivalent to "every point $x \ne 0$ is apart from 0" (Theorem 12.11), and by (1) a run without a fixed chance has $v \ne 0$. Conversely, let α be not all zero. Then the two-coin run of s_α has no fixed chance, since that would give $s_\alpha = 0$ and α all zero. So its variance s_α^2 is apart from 0, hence positive, so $s_\alpha > 0$, and a positive jump point exhibits an entry 1 of α: the regular sequence naming it has a stage above its tolerance, and that stage has found the first event. □

Theorem 29.9 provides a mixing law in classical mathematics. On $\mathbb{R}_\delta$, deciding whether it is a single coin costs WLPO, and certifying that it is not costs $\mathrm{MP}_{\mathrm{bin}}$; the criterion itself, two moments, costs nothing.

The cost of telling two frequency profiles apart

How far apart are two profiles of frequencies? The cost J of Definition 21.11 gives an answer with the properties a measure of distinguishability should have.

Theorem 29.12 (The J-cost of two profiles is half Neyman's χ^2)**.** *Work in classical mathematics. For vectors p and q with positive entries indexed by finitely many outcomes, put*

$$D_J(p\,\|\,q) = \sum_i q_i\, J\Big(\frac{p_i}{q_i}\Big), \qquad J(x) = \tfrac{1}{2}(x + x^{-1}) - 1.$$

1. *If $\sum_i p_i = \sum_i q_i = 1$, then $D_J(p\,\|\,q) = \frac{1}{2}\sum_i (q_i - p_i)^2/p_i = \frac{1}{2}\big(\sum_i q_i^2/p_i - 1\big)$, half of Neyman's χ^2 with p as the observed frequencies, equivalently half of Pearson's χ^2 with p as the expected law.*
2. *Merging outcomes never increases it: for every map g of the outcomes onto a set of cells, with g_*p summing p over each cell, $D_J(g_*p\,\|\,g_*q) \le D_J(p\,\|\,q)$.*
3. *$D_J(p + \varepsilon h\,\|\,p) = \frac{\varepsilon^2}{2}\sum_i h_i^2/(p_i + \varepsilon h_i)$ exactly, whenever every $p_i + \varepsilon h_i$ is positive.*
4. *As $\varepsilon \to 0$ with $\varepsilon \neq 0$, $D_J(p + \varepsilon h\,\|\,p)/\varepsilon^2 \to \frac{1}{2}\sum_i h_i^2/p_i$, half the Fisher information of p in the direction h.*
5. *$D_J(p\,\|\,q) \ge 0$, with equality exactly when $p = q$.*

Proof. $J(x) = (x-1)^2/(2x)$, so $J \ge 0$ with equality only at $x = 1$, which is (5). For (1), $q_i J(p_i/q_i) = \frac{1}{2}p_i + \frac{1}{2}q_i^2/p_i - q_i$; sum, with $\sum p = \sum q = 1$. For (2), the same identity writes D_J as the sum of $\frac{1}{2}p_i - q_i$, which merging does not change, and $\frac{1}{2}q_i^2/p_i$. On one cell A, the Cauchy–Schwarz inequality gives $\sum_{i\in A} q_i^2/p_i \ge (\sum_{i\in A} q_i)^2/\sum_{i\in A} p_i$. For (3), with $x_i = (p_i + \varepsilon h_i)/p_i$, $p_i J(x_i) = p_i(\varepsilon h_i/p_i)^2 p_i/(2(p_i + \varepsilon h_i))$. Part (4) follows from (3) by continuity. □

D_J is the divergence of Csiszár's family [172] generated by J, which is convex with $J(1) = 0$; data processing, part 2, is the property that family shares. What is particular to J is part 4. The calibration of the cost, unit curvature at the reference, makes its local form half the Fisher metric, which Čencov showed to be, up to a constant factor, the only metric invariant under the Markov embeddings of one set of outcomes in another [119]; part 1 identifies its global form with a classical statistic [134]. Whether J itself is forced by such an invariance is not settled here.

29.3 The golden record, the places of ratios and growth

A record that never repeats must be varied. How varied? Count the different stretches it shows. This section proves that the recognition history of Section 21.5 is as little varied as a record that never repeats can be. It then asks which notions of size δ's ratios carry, and how the dimension of a space of comparisons shows in the growth of their records.

Definition 29.13 (Windows of a record)**.** Let s be an infinite binary record, a binary sequence. Its *window* of length n at i is $(s(i), \dots, s(i+n-1))$, and $p_s(n)$ is the number of different windows of length n. The record is *eventually periodic* when some N and $p > 0$ give $s(i + p) = s(i)$ for every $i \ge N$. It is *balanced* when any two windows of the same length differ by at most one in their numbers of entries 1. A window w is *right special* when both $w0$ and $w1$ are windows.

Theorem 29.14 (A record that never repeats shows more than n windows of length n)**.** *Work in classical mathematics. For every infinite binary record s:*

1. *s is eventually periodic if and only if $p_s(n) \le n$ for some n; so a record that never repeats has $p_s(n) \ge n+1$ for every n;*
2. *if s is balanced and not eventually periodic, then $p_s(n) = n+1$ for every n.*

Proof. (1) $p_s(0) = 1$, and $p_s(n) \le p_s(n+1)$, since every window of length n is the start of one of length $n+1$. If $p_s(n) \le n$, then p_s does not increase at every step from 0 to n, so $p_s(m+1) = p_s(m)$ for some $m < n$. Then every window of length m has exactly one continuation among the windows, so the window of length m at $i+1$ is determined by the one at i. Two of the windows at $0, 1, \dots, p_s(m)$ are equal, say at $i < j$, and then $s(k+j-i) = s(k)$ for every $k \ge i$. Conversely, if s is periodic with period p from N on, the window at $i \ge N$ equals the window at $N + ((i-N) \bmod p)$, so $p_s(n) \le N + p$ for every n, and $p_s(N+p) \le N+p$.

(2) Every window of length n has one or two continuations, so $p_s(n+1) = p_s(n) + r(n)$, where $r(n)$ is the number of right-special windows of length n. In a balanced record $r(n) \le 1$: if $u \ne v$ were right special, write $u = u'az$ and $v = v'bz$ with $a \ne b$, say $a = 0$ and $b = 1$; then $0z0$ and $1z1$ are windows of the same length whose numbers of entries 1 differ by 2. In a record that never repeats $r(n) \ge 1$, since $r(n) = 0$ would give $p_s(n+1) = p_s(n)$ and, by the argument of (1), a period. So $r(n) = 1$, and $p_s(n) = n+1$ by induction. □

Part 1 is the theorem of Morse and Hedlund [132]. Part 2 goes back to Morse and Hedlund [133], and its converse, that a record with exactly $n+1$ windows of every length is balanced, is due to Coven and Hedlund [121]; these are the Sturmian records. See also [100, Chapter 2].

The recognition history h of Definition 21.24 posts debits and credits. Read a credit as 1 and a debit as 0, and call the resulting record the *golden record* g:

$$h = \mathsf{dcddcdcddcddc}\cdots, \qquad g = 0100101001001\cdots.$$

Let $c(N)$ be the number of entries 1 among its first N entries.

Theorem 29.15 (The golden record is the least varied record that never repeats)**.** *Work in classical mathematics. The golden record satisfies*

$$(N+1)(2-\varphi) - 1 < c(N) < (N+1)(2-\varphi) \qquad \text{for every } N.$$

It is balanced and not eventually periodic, and it has exactly $n+1$ windows of every length n. By Theorem 29.14, no record that never repeats has fewer windows of any length.

Proof. Section 21.5 proves, by sign tests in $\mathbb{Z}[\varphi]$, that the discrepancy $D_N = N\varphi^{-3} - b_N$ satisfies $-2/\varphi < D_N < 2/\varphi^2$, where b_N is the number of debits minus the number of credits among the first N postings. Here $b_N = N - 2c(N)$, and $\varphi^{-3} = 2\varphi - 3$, so $D_N = 2c(N) - 2N(2-\varphi)$. With $1/\varphi = \varphi - 1$ and $1/\varphi^2 = 2 - \varphi$, the bounds read $N(2-\varphi) - (\varphi - 1) < c(N) < (N+1)(2-\varphi)$, and $N(2-\varphi) - (\varphi - 1) = (N+1)(2-\varphi) - 1$.

The window of length m at a has $c(a+m) - c(a)$ entries 1, and the bounds place this number strictly between $m(2-\varphi) - 1$ and $m(2-\varphi) + 1$. Two integers in an open interval of length 2

differ by at most 1: the record is balanced. Suppose it were periodic from N on with period $p > 0$ and q entries 1 in a period. Then $c(N + kp) = c(N) + kq$, and the bounds at N and at $N + kp$ give $|k(q - p(2 - \varphi))| < 2$ for every k. But $q - p(2 - \varphi) = (q - 2p) + p\varphi$ is not 0, since $p \neq 0$ and an element $a + b\varphi$ of $\mathbb{Z}[\varphi]$ is 0 only when $a = b = 0$; so large k give a contradiction. Theorem 29.14(2) now gives $n + 1$ windows of every length. □

The frequency of credits is $2 - \varphi = \varphi^{-2} \approx 0.382$, an irrational number, and that is what forbids a period. Reading the postings as the events of the recognition hierarchy is the identification of Definition 21.24; under it, the hierarchy posts the least varied record that never repeats.

The places of δ's ratios

An *absolute value* on $\mathbb{Q}_\delta$ is a map f to the classical nonnegative reals with $f(q) = 0$ exactly when $q = 0$, $f(xy) = f(x)f(y)$ and $f(x + y) \leq f(x) + f(y)$; it is *nontrivial* when some $f(q)$ is neither 0 nor 1. Its *place cost* is $C_f = J \circ f$ on the nonzero ratios.

Theorem 29.16 (The places of δ's ratios and their costs)**.** *Work in classical mathematics.*

1. *Every nontrivial absolute value on $\mathbb{Q}_\delta$ is $q \mapsto |q|^c$ with $0 < c \leq 1$, or $q \mapsto |q|_p^c$ for a prime p and some $c > 0$, where q is read in $\mathbb{Q}$ along the unique isomorphism.*
2. *For every absolute value f and all nonzero x and y, the place cost obeys the composition law $C_f(xy) + C_f(x/y) = 2C_f(x)C_f(y) + 2C_f(x) + 2C_f(y)$.*
3. *If f is nontrivial and $C_f(2) = \frac{1}{4}$, then f is the real absolute value or the 2-adic one.*
4. *Both occur: the real and the 2-adic absolute values have $C(2) = \frac{1}{4}$, and the 2-adic one has $C(3) = 0$.*
5. *If f is nontrivial, $C_f(2) = \frac{1}{4}$ and $C_f(3) \neq 0$, then $C_f(q) = J(q)$ for every positive ratio q.*

Proof. (1) $\mathbb{Q}_\delta$ maps uniquely and isomorphically onto $\mathbb{Q}$ (Theorem 4.15), so an absolute value on one is an absolute value on the other, and Ostrowski's theorem [135] classifies those on $\mathbb{Q}$. (2) f is positive and multiplicative on nonzero ratios, and $J(ab) + J(a/b) = 2J(a)J(b) + 2J(a) + 2J(b)$ for $a, b > 0$: with $a = e^s$ and $b = e^t$ this is $\cosh(s + t) + \cosh(s - t) = 2\cosh s \cosh t$, since $J(e^t) = \cosh t - 1$. (3) $J(z) = \frac{1}{4}$ with $z \geq 0$ gives $2z^2 - 5z + 2 = (2z - 1)(z - 2) = 0$, so $f(2)$ is 2 or $\frac{1}{2}$. In the real case $2^c = 2$ gives $c = 1$. In a p-adic case $|2|_p = 1$ for $p \neq 2$ gives $C(2) = 0$, and for $p = 2$, $2^{-c} = \frac{1}{2}$ gives $c = 1$. (4) $J(2) = J(\frac{1}{2}) = \frac{1}{4}$ and $|3|_2 = 1$. (5) $C_f(3) \neq 0$ excludes the 2-adic place, so f is the real absolute value. □

The reference value $J(2) = \frac{1}{4}$ of Chapter 21 (Theorem 21.22, part 3) therefore leaves exactly two places of the ratios, the real one and the 2-adic one, and one nonzero value at 3 selects the real place, on which the place cost is J itself. Which physical datum supplies $C(3) \neq 0$ is not addressed here.

Dimension as growth

Take D independent comparisons, each of which can step forward or back. After at most n steps the record of where they stand is a point of $\mathbb{Z}^D$ with $|v_1| + \cdots + |v_D| \le n$. How many such points are there, and does their number reveal D?

Definition 29.17 (Balls of steps)**.** For a finite set $S \subseteq \mathbb{Z}^D$ of steps, $W_S(n)$ is the set of sums of at most n elements of S, repetitions allowed; S *generates* when every point of $\mathbb{Z}^D$ lies in some $W_S(n)$. The ball $B(D,n)$ is the set of $v \in \mathbb{Z}^D$ with $\sum_i |v_i| \le n$. A function f from the natural numbers to the reals has *growth degree* D when there are $c, C > 0$ with $cn^D \le f(n) \le Cn^D$ for every $n \ge 1$.

Theorem 29.18 (Dimension is read off the growth of comparisons)**.** *Work in classical mathematics.*

1. $|B(D,n)| = \sum_{k=0}^{D} 2^k \binom{D}{k}\binom{n}{k}$*, the Delannoy number of* D *and* n*.*
2. *For every finite generating* $S \subseteq \mathbb{Z}^D$*,* $n \mapsto |W_S(n)|$ *has growth degree* D*.*
3. *A function has at most one growth degree.*

Proof. (1) Write $F(D,n)$ for the sum. Splitting off the first coordinate t, the points of $B(D+1,n)$ with first coordinate t form a copy of $B(D, n-|t|)$, so

$$|B(D+1,n)| = |B(D,n)| + 2\sum_{s<n} |B(D,s)|,$$

and subtracting this at n from this at $n+1$ gives $|B(D+1,n+1)| = |B(D+1,n)| + |B(D,n+1)| + |B(D,n)|$. Both $|B(0,n)|$ and $|B(D,0)|$ are 1. F has the same boundary values, and Pascal's rule $\binom{D+1}{k} = \binom{D}{k} + \binom{D}{k-1}$, with the same rule for n, gives $F(D+1,n+1) - F(D+1,n) - F(D,n+1) = \sum_k 2^k \binom{D}{k-1}\binom{n}{k-1} - F(D,n) = F(D,n)$.

(2) Let L bound the sums $\sum_i |s_i|$ for $s \in S$. Then $W_S(n) \subseteq B(D,nL)$, which lies in a cube with $(2nL+1)^D \le (3L)^D n^D$ points, since $L \ge 1$ when $D \ge 1$; for $D = 0$ there is one point. For the lower bound with $D \ge 1$, let each unit vector e_i be a sum of at most $K \ge 1$ steps. A point with coordinates in $\{0, \dots, a\}$ is a sum of at most KDa steps, so the cube with $a = \lfloor n/(KD) \rfloor$ lies in $W_S(n)$, and it has $(a+1)^D \ge (KD)^{-D} n^D$ points.

(3) If $cn^D \le f(n) \le C' n^E$ for every $n \ge 1$ with $D > E$, then $cn^{D-E} \le C'$ for every n, which fails for large n. □

With the steps $\pm e_i$, $W_S(n)$ is the ball itself, and D independent reversible comparisons reach exactly the Delannoy number of states within n steps [113]. Its leading term $2^D n^D / D!$ is the volume of the cross-polytope. Dimension is therefore an invariant of the growth of comparison records. The theorem forces no value of D; the topological reason for $D = 3$ is Chapter 23.

29.4 Fixed points on δ's triangle

Stir a cup of coffee: some point of the surface ends where it began. That is Brouwer's fixed-point theorem, and Chapter 12 priced its one-dimensional case, the intermediate value theorem: an approximate root costs nothing, and an exact root costs LLPO, given countable choice: the root gives LLPO with no choice principle, and LLPO gives the root when countable choice is assumed (Theorems 12.55 and 12.54). The same holds on a triangle in the plane of $\mathbb{R}_\delta$.

Definition 29.19 (The triangle and its maps)**.** The *triangle* T is the set of $(x, y) \in \mathbb{R}_\delta^2$ with $x \geq 0$, $y \geq 0$ and $x + y \leq 1$. Points p and q are *k-near* when $|p_1 - q_1| \leq 1/(k+1)$ and $|p_2 - q_2| \leq 1/(k+1)$. A map F from $\mathbb{R}_\delta^2$ to itself is *uniformly continuous on T with modulus ω* when points of T that are $\omega(k)$-near have k-near images, and it is a *self-map of T* when it sends T into T.

The proofs label a grid. Write $b_0 = 1 - x - y$, $b_1 = x$ and $b_2 = y$ for the *barycentric coordinates* of (x, y); on T they are at least 0 and sum to 1. The grid of side N has the vertices $(i/N, j/N)$ with $i, j \geq 0$ and $i + j \leq N$, and its *small triangles* have the vertices (i, j), $(i+1, j)$, $(i, j+1)$ with $i + j + 1 \leq N$, or $(i+1, j)$, $(i, j+1)$, $(i+1, j+1)$ with $i + j + 2 \leq N$, in grid units.

Lemma 29.20 (Sperner's lemma for the triangle)**.** *Let $N \geq 1$, and label every vertex of the grid of side N with 0, 1 or 2, so that no vertex of the bottom row $j = 0$ has label 2, no vertex of the left edge $i = 0$ has label 1, and no vertex of the hypotenuse $i + j = N$ has label 0. Then an odd number of small triangles carry all three labels; in particular one does. The proof uses no principle.*

Proof. Call an edge between adjacent vertices labelled 0 and 1 a *door*. A small triangle with all three labels has exactly one door; any other has none or two. So the number of fully labelled small triangles has the parity of the total number of doors counted triangle by triangle. An inner edge lies in two small triangles and a boundary edge in one, so that total has the parity of the number of doors on the boundary. The left edge has no label 1 and the hypotenuse no label 0, so their edges are not doors. The bottom row has labels 0 and 1 only; its first vertex has label 0, being on the left edge, and its last has label 1, being on the hypotenuse. So the labels change an odd number of times along it. □

Theorem 29.21 (Approximate fixed points on the triangle are free)**.** *Let F be a self-map of T, uniformly continuous on T with modulus ω. For every k there is a point $p \in T$ such that $F(p)$ and p are k-near. Moreover, for every $n \geq \max(4k + 3, \omega(4k + 3))$ such a p can be found among the vertices of the grid of side $n + 1$. The proof uses no omniscience principle and no choice principle.*

Proof. Put $t = 1/(4k + 4)$ and fix n as stated. At a vertex v and for each c, cotransitivity, applied to $b_c(v) < b_c(v) + t$ at the point $b_c(F(v))$, gives *good*, $b_c(F(v)) < b_c(v) + t$, or *raised*, $b_c(v) < b_c(F(v))$. Exclude the labels c with $b_c(v) = 0$: 0 on the hypotenuse, 1 on the left edge and 2 on the bottom row. If every remaining coordinate were raised, then, since the excluded ones have $b_c(v) = 0 \leq b_c(F(v))$, the sums would give $1 = \sum_c b_c(v) < \sum_c b_c(F(v)) = 1$. So some remaining coordinate is good; label v with the first. The grid is finite, so the labels are collected by induction over the vertices, and no choice principle enters. They meet the conditions of

Lemma 29.20, which gives vertices P_0, P_1, P_2 with labels 0, 1, 2, pairwise within $1/(n+1) \le t$ in each coordinate. Since $n \ge \omega(4k+3)$, their images are pairwise within t. Good labels give

$$F(P_0)_1 \le F(P_1)_1 + t < (P_1)_1 + 2t \le (P_0)_1 + 3t, \qquad F(P_0)_2 < (P_0)_2 + 3t,$$
$$(P_0)_1 + (P_0)_2 < F(P_0)_1 + F(P_0)_2 + t,$$

the last from label 0 at P_0. Hence $(P_0)_1 < F(P_0)_1 + (F(P_0)_2 - (P_0)_2) + t < F(P_0)_1 + 4t$, and likewise for the second coordinate. With $4t = 1/(k+1)$, $F(P_0)$ and P_0 are k-near. Every step compares given points with ratio margins, and no exact sign is decided. □

Theorem 29.22 (An exact fixed point costs LLPO, given countable choice)**.**

1. *If every uniformly continuous self-map of T has a fixed point, then every uniformly continuous f on $[0,1]$ with $f(0) \le 0 \le f(1)$ has a root, and LLPO holds. No choice principle is used.*
2. *Assume countable choice AC_ω (Definition 12.21). If LLPO holds, every uniformly continuous self-map of T has a fixed point. No further choice is used.*

So, assuming countable choice, the exact fixed-point theorem on T is equivalent to LLPO.

Conditions and scope.

Base and premises
The triangle in $\mathbb{R}^2_\delta$; maps uniformly continuous on it with a modulus. Part 2 assumes AC_ω.

Conclusion
Approximate fixed points are free (Theorem 29.21). An exact fixed point implies LLPO with no choice, and LLPO gives it at the rung of countable choice, as for the intermediate value theorem.

Not claimed
That LLPO alone, without countable choice, gives exact fixed points; that question is open here.

Proof. (1) Let $\mathrm{cl}(z) = \max(0, \min(z, 1))$ and $F(x, y) = (\mathrm{cl}(x - f(\mathrm{cl}(x))), 0)$. It maps the plane into T and is uniformly continuous, since cl moves points by no more than their distance. A fixed point is $(x, 0)$ with $x \in [0, 1]$ and $x = \mathrm{cl}(x - f(x))$. If $f(x) > 0$, then $x > 0$ is impossible, since $\min(x - f(x), 1) < x$ and $0 < x$ would make $\mathrm{cl}(x - f(x)) < x$; so $x = 0$ and $f(0) > 0$, against the hypothesis. So $f(x) \le 0$, and symmetrically $f(x) \ge 0$. Theorem 12.54(1) turns roots into LLPO.

(2) Let F have modulus ω. *Tests.* For a tolerance t and a corner $v = (a/2^d, b/2^d)$ in T, cotransitivity in each coordinate shows that there is a bit that is 1 only if $F(v)$ and v are t-near and 0 only if they are not $(2t+1)$-near. Which bit it finds depends on the regular sequences that name $F(v)$, so countable choice is what collects these bits into one table $B(t, v)$. Say that t is *open* at depth d when $2^d \ge \max(\omega(6t+5) + 1, 6t + 6)$.

The tree. A node of depth d is a word of d letters from $\{0,1,2,3\}$; each letter halves the current dyadic square in both directions, and the node names the lower left corner $(a/2^d, b/2^d)$ of its square. Keep a node when, at every depth $d' \le d$ along it, the corner lies in T $(a+b \le 2^{d'})$ and passes $B(t,\cdot) = 1$ for every $t \le d'$ open at d'. The tree is decided by finitely many integer comparisons and table entries, and each node has at most four children.

It has nodes at every depth. Given D_0, apply Theorem 29.21 with $k = 6D_0 + 5$ on the grid of side 2^M, M large, to get a vertex $p = (i/2^M, j/2^M)$ of T with $F(p)$ and p $(6D_0+5)$-near. Its binary digits give the nested dyadic squares containing it, whose corners at depth $d' \le D_0$ lie in T and within $2^{-d'}$ of p. If $t \le d'$ is open at d', such a corner v has $|F(v) - F(p)| \le 1/(6t+6)$ by the modulus, $|F(p) - p| \le 1/(6t+6)$ and $|p - v| \le 1/(6t+6)$ in each coordinate. So $F(v)$ and v are $(2t+1)$-near, the bit 0 is impossible, and $B(t,v) = 1$. The word of p's digits up to D_0 is a node.

A path and its limit. Writing each letter as two binary digits turns the tree into a decidable binary tree with members of every length. LLPO with AC_ω gives weak König's lemma (Theorem 12.72), hence a path. Its corners c_d are ratios with $|c_d - c_{d'}| \le 2^{-d}$ for $d \le d'$, so they form a regular sequence and name a point q, with no choice; $q \in T$ since T is closed. For each t, the tolerance t is open at every large depth, where the path's corner passes the test, so $F(c_d)$ and c_d are t-near. By the modulus, $|F(q) - q| \le 1/(t+1) + 1/(k+1) + 2^{-d}$ in each coordinate for every k and all large d. So $|F(q) - q| \le 1/(t+1)$ for every t, and $F(q) = q$. □

Sperner's lemma [136] and the approximate fixed point it gives are the standard constructive route to Brouwer's theorem [117]. On $\mathbb{R}_\delta$ the route ends at an approximate fixed point; the last step, to an exact one, is a compactness step, and it costs exactly what weak König's lemma costs.

29.5 What identifying a law costs

Section 17.7 priced the decision of a question from records. A learner faces a different task: it must name the law that produces the record. Its candidates are listed in advance, and it must end up naming the right one.

Definition 29.23 (Learners and identification)**.** Worlds are binary runs, read by themselves. A *prior* is a sequence $h = (h_0, h_1, \ldots)$ of worlds.

1. A *learner* is a rule M that assigns an index to every finite binary record. It *identifies ω in the limit*, relative to h, when some j and N give $h_j = \omega$ and $M(\omega{\restriction}n) = j$ for every $n \ge N$. A class of worlds is *learnable* when some prior and some learner identify each of its members in the limit.

2. A *certain learner* assigns to every finite record an index or the mark $\circ$, "silent". It *identifies ω with certainty* when it announces an index at some record of ω, and every index j it announces at a record of ω has $h_j = \omega$.

3. The *pair prior* of two worlds h_0 and h_1 lists h_0 at index 0 and h_1 at every other index.

4. The *zero world* is the run of zeros. The *flag world* of a binary sequence α has entry 1 at n when $\alpha(k) = 1$ for some $k \le n$.

Table 29.1. The price of identifying one of two worlds (Theorem 29.24).

	Worlds known to differ	Worlds that may be equal
With certainty	$\mathrm{MP}_{\mathrm{bin}}$	LPO
In the limit	$\mathrm{MP}_{\mathrm{bin}}$	LPO

Like recorders, learners are arbitrary rules; nothing requires them to be computable. The flag world of α is the zero world exactly when α is all zero, and the two have the same record of length n exactly when $\alpha(k) = 0$ for every $k < n$. Comparing the two worlds through a learner is the engine of every lower bound below.

Theorem 29.24 (Learning a pair costs $\mathrm{MP}_{\mathrm{bin}}$ or LPO)**.** *Each of the following equivalences is proved in both directions with no choice principle.*

1. *Every pair of different worlds is identified with certainty, relative to its pair prior, by some certain learner, if and only if* $\mathrm{MP}_{\mathrm{bin}}$ *holds.*
2. *Every pair of different worlds is identified in the limit if and only if* $\mathrm{MP}_{\mathrm{bin}}$ *holds.*
3. *Every pair of worlds, equal or not, is identified with certainty if and only if LPO holds.*
4. *Every pair of worlds is identified in the limit if and only if LPO holds.*
5. *Every prior has a learner that identifies each of its members in the limit if and only if LPO holds.*

Conditions and scope.

Base and premises
Binary worlds read by themselves; learners are arbitrary rules on finite records, not required to be computable.

Conclusion
The price of naming the true one of two candidate laws is $\mathrm{MP}_{\mathrm{bin}}$ when they are known to differ and LPO when they may be equal, for certain and for convergent identification alike.

Not claimed
Prices for classes of more than two worlds other than part 5, or for learners restricted to programs (Sections 29.7 and 29.9).

Table 29.1 arranges the first four. What costs is not certainty as against convergence. It is whether the two candidate worlds are known to differ.

Proof. Upper bounds. Let $h_0 \neq h_1$. The sequence that is 1 where they differ is not all zero, so $\mathrm{MP}_{\mathrm{bin}}$ finds a place k with $h_0(k) \neq h_1(k)$. The learner that stays silent on records shorter than $k + 1$ and then names 0 or 1 by the entry at k identifies both worlds with certainty, and, naming 0 while silent, in the limit. Under LPO two worlds are equal or differ at a place found;

for equal worlds announce 0 at once. For a prior h, the *enumeration learner* names, at a record σ, the least $i \le |\sigma|$ with $h_i{\restriction}|\sigma| = \sigma$, and $|\sigma| + 1$ if there is none. On ω, LPO finds the least i with $h_i = \omega$ and, for each $j < i$, a place where h_j and ω differ; past these finitely many places the learner names i.

Lower bounds. For (3), let α be binary and take the pair of the zero world, at index 0, and the flag world of α. The certain learner announces on the zero world at some record of length n, an index j. If $j \neq 0$, the announcement is correct, so the flag world is the zero world and α is all zero. If $j = 0$, inspect $\alpha(0), \dots, \alpha(n-1)$: an entry 1 settles the matter; if there is none, the flag world has the same record of length n, the learner announces 0 there too, and correctness makes the flag world the zero world. This is LPO. For (1), let α be not all zero; the two worlds then differ, the first and last cases are contradictions, and an entry 1 remains: $\mathrm{MP}_{\mathrm{bin}}$.

For (4), take the pair of the flag world of α, at index 0, and the zero world. The learner settles on the zero world at j from stage N and on the flag world at j' from stage N'. If $j = 0$, the index 0 names the zero world, and if $j' \neq 0$, the index j' names the flag world; either way the two worlds are equal and α is all zero. Otherwise $j \neq 0 = j'$; inspect α below $N + N'$. An entry 1 settles the matter, and if there is none the two records of length $N + N'$ coincide, so the learner names j and j' at one record and $j = j'$, a contradiction. For (2) the first two cases are contradictions when α is not all zero. Part 5 contains part 4, since a pair prior is a prior. Every case split is on natural numbers and bits. □

Theorem 29.25 (What the limit buys and what no prior covers)**.**

1. *Let s_n be the world whose only entry* 1 *is at n. The enumeration learner of the prior $(0, s_0, s_1, \dots)$ identifies each of its members in the limit, with no principle; no certain learner identifies all of them with certainty relative to this prior.*
2. *For every prior and every learner, some binary world is not identified in the limit.*
3. *Every learnable class lies in some prior, and, under LPO, every class that lies in a prior is learnable. In classical mathematics a class of worlds is learnable exactly when it injects into the natural numbers.*

All but the last statement are proved with no choice principle; the last is classical.

Proof. (1) Any two members differ at a known place: s_n and 0 at n, s_n and s_m at n. On s_n, the record of length $n + 1$ refutes the indices $0, \dots, n$, so from then on the learner names $n + 1$; on the zero world it names 0 throughout. A certain learner that announces an index on the zero world at a record of length n announces the same index on s_n, whose record of length n is the same; that index cannot name both worlds. (2) A world identified relative to h is some h_j, and the world $\omega(n) = 1 - h_n(n)$ is none of them. (3) The first statement is (2)'s observation, and the second is the enumeration learner of Theorem 29.24. Classically LPO holds, and a class lies in a prior exactly when it injects into the natural numbers: send a member to its least index, or list the members along the inverse of an injection. □

The spike class separates the two notions on a single class: convergence is free there, while certainty is impossible at any price. The four cells of Table 29.1 measure something else, the

knowledge whether two candidates differ. Gold's identification in the limit [63] is the notion priced here, and the enumeration learner is his identification by enumeration.

The price of revising a verdict

Theorem 17.20 showed that verdicts revised as the record grows confirm exactly the Σ_2 questions. What does it cost to know, now, what the verdicts will settle on? It depends on how often they may change.

Definition 29.26 (Revisions and Σ_2 excluded middle)**.** A guessing rule g (Definition 17.19) *changes at most k times* on w when there are at most k numbers m with $g(w{\restriction}m) \neq g(w{\restriction}(m+1))$. Σ_2 *excluded middle*, Σ_2-LEM, states that for every binary double sequence β, either some m has $\beta(m,n) = 0$ for every n, or no m does.

Theorem 29.27 (The prices of revisable verdicts)**.**

1. *If a guessing rule confirms Q in the limit and changes at most k times on every world, then LPO decides Q at every world.*
2. *On binary runs, "some entry is 1" is confirmed in the limit by a rule that changes at most once, and deciding it for every run is LPO.*
3. *For each k, "the number of entries 1, capped at $k+1$, is eventually odd" is confirmed in the limit by a rule that changes at most $k+1$ times, and by no rule that changes at most k times on every run.*
4. *Σ_2-LEM decides every question that a guessing rule confirms in the limit.*
5. *On binary runs, finiteness, $\exists N \forall n \geq N\, \alpha(n) = 0$, is confirmed in the limit, and deciding it for every run is equivalent to Σ_2-LEM.*

The negative half of part 3 is proved in classical mathematics; everything else is proved with no choice principle.

Proof. (1) Fix w. At a stage s, apply LPO to the sequence that is 1 at i when $g(w{\restriction}(s+i)) \neq g(w{\restriction}s)$. If it is all zero, the verdicts have settled on $g(w{\restriction}s)$, which answers Q. Otherwise move to the stage $s+i$ found, having used one change. After at most $k+1$ steps no change remains, and the first case answers.

(2) "Yes once a 1 has appeared" changes at most once, and Proposition 17.18(1) prices the decision.

(3) Saying *yes* when the capped count so far is odd changes only when the capped count rises, at most $k+1$ times. Against a rule with at most k changes, start from the zero world: its capped count is even, so the verdicts settle on *no* from some stage. Place an entry 1 after that stage: the count becomes odd, and the verdicts must settle on *yes*, after a change. Repeating $k+1$ times forces $k+1$ changes along one run.

(4) If g confirms Q in the limit, $Q(w)$ says that some m has $g(w{\restriction}(m+n)) =$ yes for every n. Apply Σ_2-LEM to $\beta(m,n) = 1$ exactly when $g(w{\restriction}(m+n))$ is *no*.

(5) "Yes when the record is empty or ends in 0" confirms finiteness, and (4) gives the decision from Σ_2-LEM. Conversely, let β be a binary double sequence. Let cand(s) be the least $m \leq s$ with $\beta(m, n) = 0$ for every $n \leq s$, or $s + 1$ if there is none, and let $c(s) = 1$ exactly when cand$(s + 1) \neq$ cand(s). If c has finitely many entries 1, the candidate is eventually constant, and since a missing candidate changes at every stage it is a real one, whose row is 0 everywhere. If some row m is 0 everywhere, then cand$(s) \leq m$ for $s \geq m$, and it never decreases from there; a bounded sequence of numbers that never decreases cannot fail forever to become constant, so c cannot fail to be finite. So a decision of finiteness for c decides whether some row of β is 0 everywhere. □

Part 1 and part 4 refine Theorem 17.20: a Σ_2 question confirmed with a bounded number of revisions costs at most LPO, and the event question shows that LPO is needed; with no bound the price is Σ_2 excluded middle, attained by finiteness. Σ_2-LEM is equivalent, with no choice principle, to the principle Σ_2-LPO of Section 29.10 (Proposition 29.50), the price of Harrington's learning method.

29.6 Mistakes, dimensions and samples

The opening case counted mistakes. Here the count is made exact for every finite class of laws, and the same counting answers how large a sample must be and how well a law can be predicted from its program.

Definition 29.28 (Online learning)**.** Let X be a set of points; a *law* is a map $f : X \to \{0, 1\}$. An *online learner* L maps a finite history of labelled points and a new point to a predicted label. On a list of points labelled by f, it predicts each label from the history of the earlier points with their true labels, and mist$(L, f, x_1 \dots x_r)$ is the number of wrong predictions. For a finite list H of laws, repetitions allowed, with length $|H|$, the *version space* of a history is the list of laws in H consistent with it. The *halving learner* of H predicts 1 exactly when at least half of the version space predicts 1. The *labellings* of n points are the 2^n laws on $\mathbb{N}$ that are 0 from n on.

Theorem 29.29 (Mistakes are distinctions)**.**

1. *If f is in H, the halving learner of H makes m mistakes on a list of points only if $2^m \leq |H|$.*
2. *For every online learner and every n, some labelling of the points $0, \dots, n-1$ makes it err at each of them.*
3. *The halving learner of the 2^n labellings of n points errs at most n times on every list of points; the labellings correspond one-to-one with the binary words of length n.*
4. *Let* dec *send some binary words to laws, and let $f =$ dec(w) with $|w| \leq b$. The halving learner of the laws* dec(w')*, $|w'| \leq b$, errs at most b times on every list of points.*

The theorem is free.

Proof. (1) When the halving learner errs, the true label is given by at most half of the version space, so at most half survives; f always survives. After m mistakes, $1 \le |H|/2^m$. (2) Label each point with the opposite of the prediction. (3) is (1) with $|H| = 2^n$, and a labelling is its word of first n values. (4) There are $2^{b+1} - 1$ words of length at most b, so $2^m \le |H| < 2^{b+1}$. □

Part 4 is a mistake-bound form of Occam's razor, whose sample-size form is due to Blumer, Ehrenfeucht, Haussler and Warmuth [116]: a law with a b-bit description costs at most b mistakes. A halving learner can be far from optimal, and the exact worst case is a dimension.

Definition 29.30 (Littlestone dimension)**.** Let X_0 be a finite list of points. A point x *splits* a list H of laws when both labels occur at x; $H_{x,b}$ is the list of laws in H with label b at x. The *dimension* $\mathrm{ldim}(H)$ is 0 when no point of X_0 splits H, and otherwise the largest, over splitting points x of X_0, of $1 + \min(\mathrm{ldim}(H_{x,0}), \mathrm{ldim}(H_{x,1}))$. The *rank* of a list is 0 when it is empty and ldim + 1 otherwise. The *standard optimal learner* predicts at x the label b whose part of the version space has the larger rank, and 1 at a tie.

Theorem 29.31 (The Littlestone dimension is the exact worst case)**.** *Let H be a finite list of laws and X_0 a finite list of points.*

1. *If $f \in H$, the standard optimal learner makes at most* $\mathrm{ldim}(H)$ *mistakes on every list of points from X_0.*
2. *If H is not empty, every online learner makes at least* $\mathrm{ldim}(H)$ *mistakes on some law of H and some list of points from X_0.*
3. *If H is not empty, $2^{\mathrm{ldim}(H)} \le |H|$.*
4. *The 2^n labellings of the points $0, \dots, n-1$ have dimension n.*

The theorem is free.

Proof. (1) A mistake at x means that the true label's part of the version space has the smaller rank, or the equal rank when the prediction was 1. If x splits the version space V, that part has dimension at most $\mathrm{ldim}(V) - 1$, by the definition of ldim. If it does not, the true part is all of V and the other is empty, with rank 0, so the prediction was right. So each mistake lowers the dimension of the version space, which never becomes empty. (2) Answer each prediction at a point x that realizes the dimension with the label of the other part, and continue inside that part, whose dimension is at least $\mathrm{ldim}(H) - 1$. (3) By induction on $|H|$: at a point realizing the dimension both parts have dimension at least $\mathrm{ldim}(H) - 1$, hence at least $2^{\mathrm{ldim}(H)-1}$ laws each. (4) By (3) the dimension is at most n, and splitting at $0, 1, \dots$ in turn reaches n. □

The theorem is Littlestone's [131]. For the four laws of the opening the dimension is 2, equal to $\log_2 4$; for the n laws with a single 1 among n points it is 1, far below $\log_2 n$, since one mistake reveals the position of the 1.

Samples, counted

How many examples must a learner see to be probably nearly right? The standard answer is a statement about probability. It is, underneath, a count.

Theorem 29.32 (Probably approximately correct learning is counting)**.** *Let X_0 be a list of N points, repetitions allowed, f a law and H a finite list of laws. A* sample *is a list of m points of X_0; there are N^m. Write* $\mathrm{err}(h)$ *for the number of points of X_0 at which h differs from f.*

1. *A law h agrees with f on exactly $(N - \mathrm{err}(h))^m$ samples.*
2. *At most $|H|(N - e - 1)^m$ samples have a law in H with more than e errors that agrees with f on the sample.*
3. *Let a learner return, for every sample labelled by f, a law of H consistent with it. Then its law has more than e errors on at most $|H|(N - e - 1)^m$ samples.*

The theorem is free.

Proof. A sample on which h agrees with f is a list of m points chosen among the $N - \mathrm{err}(h)$ points of agreement, which is (1). Summing (1) over the laws with more than e errors gives (2), and (3) follows since the returned law agrees with f on its sample. □

Dividing by N^m, a sample drawn uniformly from X_0 misleads a consistent learner by at least $\varepsilon = (e + 1)/N$ with frequency at most $|H|(1 - \varepsilon)^m$, the bound of Valiant's model [137]. Read as frequencies over all samples, the bound needs no posit of probability.

Predicting from programs

A *predictor* sees the record $\omega{\restriction}k$ and predicts $\omega(k)$; its mistakes up to n are the $k < n$ at which it is wrong. A *system of programs* is a partial evaluation $\mathrm{Run}(c, k)$, defined at some pairs, with the bits encoded injectively as values; program c *computes* ω when $\mathrm{Run}(c, k)$ is defined and encodes $\omega(k)$ for every k. A *halting table* for it is a map d with $d(c, k) = 1$ exactly when $\mathrm{Run}(c, k)$ is defined. The *elimination predictor* follows the least program $c \le k$ that agrees with the record and halts at k, read through d. The *phased halving predictor* works in phases: in phase q it takes the majority vote, 1 at a tie, of the programs below 2^q that agree with the record and halt at k, and the phase is the least $q \le k$ for which some program below 2^q agrees with the record.

Theorem 29.33 (Universal prediction needs a halting table)**.**

1. *For every computable predictor there is a computable binary world on which it errs at every examination.*
2. *For any system of programs with a halting table given as data, the elimination predictor makes at most $2j$ mistakes on every world that program j computes, at every stage.*
3. *The phased halving predictor makes at most $p(p + 1)/2 + 2p + 1$ mistakes there, where p is the number of binary digits of j.*

Part 1 is proved in classical mathematics; parts 2 and 3 are free.

Proof. (1) Let $\omega(k)$ be the opposite of the prediction at $\omega{\restriction}k$; the recursion is computable. (2) Before stage j there are at most j mistakes. From stage j on, program j is a candidate that agrees and halts, so the followed program c is at most j; if $c = j$ the prediction is right, and if $c < j$ errs, c disagrees with the record from then on. So at most j further mistakes occur. (3) Program j lies below 2^p and always agrees, so the phase is at most p once $k \geq p$, and it never decreases, since a program that disagrees with a record disagrees with its extensions. Before stage p there are at most p mistakes. A program that does not halt at k disagrees with every longer record, and at a mistake the majority of those that halt were wrong; so each mistake in phase q at least halves the agreeing programs below 2^q, of which there are at most 2^q. Phase q therefore sees at most $q + 1$ mistakes, and $\sum_{q \leq p}(q + 1) = p(p + 1)/2 + p + 1$. □

No computable predictor is universal, by part 1; with a halting table, even a crude one is, with a mistake bound set by the size of the program. The size of the shortest program is itself out of reach: the length of the shortest description of a number is not computable, and a consistent formal system certifies lower bounds on it only up to a constant [130, 120].

29.7 Learners that are programs

So far learners were arbitrary rules. Suppose instead that the learner must be a program, and the laws to be learned are the computable binary worlds. Gold proved that no single program then identifies every law in the limit [63]. Weaker criteria of success were studied by Bārzdiņš and by Case and Smith [114, 118]. This section prices them. Recognition Science's processes enter through an identification: if its native processes are the computable functions, which we do not prove here, the theorems below say what a native learner can learn. Section 29.9 takes a narrower reading, under which every program halts.

Definition 29.34 (Criteria of success)**.** Fix a standard numbering $\varphi_0, \varphi_1, \ldots$ of the partial computable functions [19]. A *code e computes* a binary world ω when $\varphi_e(x)$ is defined and equals $\omega(x)$ for every x. A learner M assigns a code to every finite binary record, and its *conjecture* at stage n on ω is $M(\omega{\restriction}n)$. M identifies ω

EX when its conjectures settle on one code that computes ω;

BC when from some stage on every conjecture computes ω;

BC* when from some stage on every conjecture computes ω at all but finitely many inputs;

BC^k when from some stage on every conjecture computes ω at all but at most k inputs;

EX* when its conjectures settle on one code that computes ω at all but finitely many inputs.

EX implies BC and EX*, and each of BC, BC^k and EX* implies BC*. The lower bounds below all use the zero world and the flag worlds, and they do not use the particular numbering.

Theorem 29.35 (A learner of the zero world and its flags yields LPO)**.**

1. *Let G be a class of binary worlds that contains the zero world and the flag world of each of its members. Let hypotheses be drawn from any set, with any relation "h computes ω" under which two worlds computed by one hypothesis agree from some input on. If a learner M, a rule from records to hypotheses, has for every $\omega \in G$ a stage from which every conjecture computes ω, then every $\alpha \in G$ is all zero or has an entry* 1*.*

2. *Hence every learner, computable or not, that EX-, BC- or BC*-identifies every computable binary world yields LPO for computable binary sequences.*

3. *Conversely, for any system of programs* Run *with a halting table given as data, LPO makes the learner that names the least code consistent with the record, consistency read through the table, EX-identify every world that some program computes.*

Parts 1 and 3 are proved with no choice principle, for every system of hypotheses or programs; part 2 is their instance for the codes φ_e, stated in classical mathematics.

Proof. (1) Let the conjectures compute the zero world from stage N and the flag world of α from stage N'. Inspect α below $N + N'$. If an entry 1 appears, we are done. If not, the two worlds have the same record of length $N + N'$, so the conjecture there computes both, and they agree from some input B on. Then α is all zero: $\alpha(k) = 1$ would put a 1 into the flag world at $\max(k, B)$. (2) The computable worlds contain the zero world and the flag world of each member. A code computes at most one world, and a code that computes two worlds at all but finitely many inputs makes them agree from some input on; EX gives BC, and BC gives BC*. (3) LPO decides, for each code below the least code of the target, whether it reproduces the target everywhere, reading definedness through the table, or finds a place where it does not; past these finitely many places the learner settles, as the enumeration learner does in Theorem 29.24. □

So identifying every computable law, by any of these criteria and by any means, is at least as strong as deciding every computable binary sequence, and with a halting table in hand LPO suffices for the strongest criterion. Which programs can do it is classical computability theory.

Theorem 29.36 (What a computable learner can do)**.** *Work in classical mathematics.*

1. *No computable learner EX-identifies every computable binary world.*

2. *No computable learner BC-identifies every computable binary world.*

3. *No computable learner EX*-identifies every computable binary world, and for each k none BC^k-identifies every computable binary world.*

4. *One computable learner BC*-identifies every computable binary world: at a record σ its conjecture is the program that, on input x, finds the least $e \le |\sigma|$ whose computation, run for x steps, reproduces every recorded value, and returns $\varphi_e(x)$.*

Proof. (1) Let M be computable and EX-identify every computable world. For a finite record σ, M settles on a code of $\sigma 0^\infty$, say from $\sigma 0^K$ on, and on a different code of $\sigma 0^K 10^\infty$, a different world. So a search finds k and m with $M(\sigma 0^k 10^m) \neq M(\sigma 0^k)$. Iterating from the empty record builds a computable world along which M changes its conjecture infinitely often. (2) From σ, search for k and t such that the conjecture at $\sigma 0^k$ returns 0 at input $|\sigma| + k$ within t steps; the search succeeds because $\sigma 0^\infty$ is computable and identified. Extend by $\sigma 0^k 1$. Along the limit world infinitely many conjectures are wrong at their next input. (3) The same method, with blocks: for BC^k, find a stage at which the conjecture returns values at $k+1$ of the next inputs, and let the world contradict all of them; for EX^*, run the construction of (2). Its search still succeeds, since from some stage on the conjectures on $\sigma 0^\infty$ are one code, which returns 0 at all large inputs; and once M has settled along the limit world on one code, every later round makes that code wrong at a new input. (4) Let E be the least code of ω. Each $e < E$ fails to reproduce ω at some input, where it is wrong or undefined, so once the record contains all these inputs no $e < E$ is selected at any x. The finitely many recorded values of φ_E are computed within some number of steps, so for all large x the selection is E, and the conjecture returns $\varphi_E(x) = \omega(x)$. □

Part 1 is Gold's theorem [63], part 2 is due to Bārzdiņš and to Case and Smith [114, 118], and part 4 is Harrington's theorem as reported by Case and Smith [118]. Of the standard criteria only the weakest, finitely many errors with changing codes, is within reach of one computable learner, and by Theorem 29.35 any learner that meets it yields LPO for computable sequences. Harrington's learner therefore cannot be justified without that much omniscience; Section 29.10 prices his method exactly.

Learning from examples alone

A learner may see only members of a language $L \subseteq \mathbb{N}$, never non-members. A *text* for L is a sequence of members of L and pauses in which every member appears. A family (L_i) of languages, each decided by a map $F(i, x)$, is *learnable from text* when some learner, given the finite pieces of any text for any L_i, settles on an index j with $L_j = L_i$. A *tell-tale* for L_i is a finite set $T \subseteq L_i$ such that no L_j with $T \subseteq L_j$ is a proper subset of L_i.

Theorem 29.37 (Positive data need finite tell-tales)**.**

1. *No learner learns from text the family consisting of* $\mathbb{N}$*, at index* 0*, and the segments* $\{0, \dots, s\}$*, at index* $s+1$*. The proof uses no choice principle.*
2. *In classical mathematics, a family of decided languages is learnable from text if and only if each of its members has a finite tell-tale.*

Proof. (1) Build a text for $\mathbb{N}$ in stages. Having written a text piece listing $0, \dots, s$, repeat s until the learner names $s+1$; it must, since the continuation that repeats s forever is a text for $\{0, \dots, s\}$, whose only index is $s+1$, and whether it names $s+1$ is decided. Then write $s+1$. The result is a text for $\mathbb{N}$ on which the learner names $1, 2, 3, \dots$ in turn, so it never settles on 0. (2) If the family is learnable, every member L_i has a *locking sequence*, a finite text piece τ of L_i after which the learner names a code of L_i on every extension within L_i, by an argument by contradiction; the members of τ form a tell-tale. Conversely, the learner that names the least

$i \le n$ whose tell-tale has appeared and whose language contains everything seen learns the family. □

Part 1 is Gold's theorem on superfinite classes [63], part 2 is Angluin's [112], and locking sequences are due to Blum and Blum [115].

29.8 Learning with the cost J

A system compares each new record with a frame, a positive point m against which records are read, and pays $J(x/m)$ for a record x. Which frame makes a history cheapest, and what does a learner pay that must choose its frame before each record arrives? Everything in this section takes place on $\mathbb{R}_\delta$ and uses no choice principle; the frame, the regret and Heron's update are free, and only exact tests carry a price.

Definition 29.38 (The cost of a frame and the J-centre)**.** Let records $x_0, \dots, x_{n-1} > 0$ carry weights $w_i > 0$, with $S = \sum_i w_i x_i$ and $T = \sum_i w_i/x_i$. The *cost* of a frame $m > 0$ is $C(m) = \sum_i w_i J(x_i/m)$. A *$J$-centre* is a point $m_* > 0$ with $m_*^2 T = S$, and we put $\tau = m_* T$.

The centre is the geometric mean of the weighted arithmetic mean $S/\sum w$ and the weighted harmonic mean $\sum w/T$ of the records.

Theorem 29.39 (The J-centre, and regret as the cost of re-framing)**.**

1. *The J-centre of a nonempty weighted history exists on $\mathbb{R}_\delta$ and is unique.*
2. *For every frame $m > 0$, $C(m) = C(m_*) + \tau\, J(m/m_*)$, and $C(m_*) = \tau - \sum_i w_i$.*
3. *Hence $C(m_*) \le C(m)$, with equality only at $m = m_*$.*
4. *Let a learner start from any frame $f_0 > 0$ and use, for record n, the centre f_n of the records before it, and write T_{n+1} for the sum T of the first $n+1$ records. Then*

$$\sum_{n<N} w_n\, J\Big(\frac{x_n}{f_n}\Big) = C_N(f_N) + \sum_{n<N} f_{n+1} T_{n+1}\, J\Big(\frac{f_n}{f_{n+1}}\Big),$$

where C_N is the cost of the first N records.

The theorem is free; the identities of parts 2 and 4 hold in every commutative ring in which 2, the records and the frames are invertible.

Proof. (2) $C(m) = \frac{1}{2}\sum_i w_i(x_i/m + m/x_i) - \sum_i w_i = S/(2m) + mT/2 - \sum_i w_i$. With $S = m_*^2 T$, $S/(2m) + mT/2 = \frac{m_* T}{2}(m_*/m + m/m_*) = \tau(J(m/m_*) + 1)$. (1) $S/T > 0$ has a positive square root (Lemma F.12). If two positive points have $m_1^2 T = m_2^2 T$, then $(m_1 - m_2)(m_1 + m_2)T = 0$ with $(m_1 + m_2)T > 0$, so $m_1 = m_2$. (3) $2aJ(a) = (a-1)^2$ is a square, so $J \ge 0$, and $J(a) = 0$ gives $(a-1)^2 = 0$, hence $a = 1$; and $\tau > 0$. (4) Write C_n for the cost of the first n records, so $C_{n+1}(m) = C_n(m) + w_n J(x_n/m)$ and $C_0 = 0$. Part 2 for the first $n+1$ records gives $C_{n+1}(f_n) = C_{n+1}(f_{n+1}) + f_{n+1}T_{n+1}J(f_n/f_{n+1})$, so $w_n J(x_n/f_n) = C_{n+1}(f_{n+1}) - C_n(f_n) + f_{n+1}T_{n+1}J(f_n/f_{n+1})$, and the sum telescopes. □

By part 3, $C_N(f_N)$ is the least cost of the whole history, so the learner's *regret*, its cost minus the least cost, is exactly the sum of the J-costs of its own changes of frame. How large can that be?

Theorem 29.40 (Logarithmic regret, and no better)**.** *Write $H_N = \sum_{n<N} 1/(n+1)$.*

1. *Let $R \geq 1$, let unit-weight records x_n and the start f_0 lie in $[1/R, R]$, and let the learner follow the centre as in Theorem 29.39(4). Its regret after N records satisfies $0 \leq \mathrm{Reg}_N \leq R^{10} H_N$.*
2. *On the alternating records $2, \frac{1}{2}, 2, \frac{1}{2}, \ldots$ the same learner has $H_K \leq 20\,\mathrm{Reg}_{2K} + 1$.*
3. *There are numbers K_c and K_d ($K_c = 12$, $K_d = 48$ will do) such that for every learner, a rule that gives a positive frame for each finite record, and every horizon N, some records $y_0, \ldots, y_{N-1} \in \{2, \frac{1}{2}\}$ and some frame $m > 0$ satisfy*

$$H_N \leq K_c\Big(\sum_{n<N} J\big(y_n / L(y_0 \ldots y_{n-1})\big) - C_N(m)\Big) + K_d.$$

By Theorem 29.39(3) the same holds with the least cost in place of $C_N(m)$, so every learner has regret at least $(H_N - 48)/12$ on some history.

The theorem is free.

Proof. (1) By Theorem 29.39(4) the regret is $\sum_{n<N} r_n$ with $r_n = f_{n+1} T_{n+1} J(f_n/f_{n+1}) = T_{n+1}(f_n - f_{n+1})^2/(2f_n) \geq 0$. The centres lie in $[1/R, R]$. For $n \geq 1$, $f_n^2 T_n = S_n$ gives $f_{n+1}^2 - f_n^2 = (x_n^2 - f_n^2)/(x_n T_{n+1})$, so

$$r_n = \frac{(x_n^2 - f_n^2)^2}{2 f_n x_n^2 T_{n+1}(f_n + f_{n+1})^2} \leq \frac{R^4}{2 \cdot R^{-1} \cdot R^{-2} \cdot (n+1) R^{-1} \cdot 4R^{-2}} = \frac{R^{10}}{8(n+1)},$$

and $r_0 = (f_0 - x_0)^2/(2 f_0 x_0) \leq R^4/2$, since $f_1 = x_0$.

(2) After $2k$ records the centre is 1, and after $2k+1$ records $f^2 = (5k+4)/(5k+1)$, so for $k \geq 1$, $r_{2k} = \frac{5k+1}{2}(f-1)^2/2 = 9/(4(5k+1)(f+1)^2) \geq 1/(4(5k+1)) \geq 1/(20(k+1))$, using $f \leq 2$. Summing over $1 \leq k < K$ gives $\mathrm{Reg}_{2K} \geq (H_K - 1)/20$.

(3) Count one extra record of each kind: after n records with k twos, let $K = k+1$, $L = n-k+1$, $D = K + L = n+2$, $U = 2K + L/2$ and $V = K/2 + 2L$, the sums of the records and of their inverses plus $\frac{5}{2}$; let $G = \sqrt{UV}$ and $e = K - L$. Put $c = \frac{1}{8}$,

$$\alpha_n = \frac{c}{(n+2)(n+3)} \sum_{n \leq j < N} \frac{1}{j+2}, \qquad \Phi_n = \mathrm{Loss}_n + n - G - \alpha_n e^2,$$

where Loss_n is the learner's cost on the first n records. Let f be the learner's frame for record n, and X_+ and X_- the increase of Φ if the record is 2 or $\frac{1}{2}$. We show $K X_+ + L X_- \geq D Y$ with $Y = c/(n+3) - \alpha_{n+1}$. Write $G_\pm$ for the new G. Then $K X_+ + L X_- = t_1 + t_2 + (D\alpha_n - (D+2)\alpha_{n+1}) e^2 - c\,D/(D+1) + DY$, where $t_1 = \frac{1}{2}(U/f + Vf) - G$ and $t_2 = (D+1)G - K G_+ - L G_-$, because $K(e+1)^2 + L(e-1)^2 = (D+2)e^2 + D$ and the learner's step costs, weighted by K and L, add up to $\frac{1}{2}(U/f + Vf) - D$. Three identities, each with $G^2 = UV$, finish:

$$t_1 \cdot Vf = \tfrac{1}{2}(Vf - G)^2, \qquad \tfrac{25}{4} D^2 - 4G^2 = \tfrac{9}{4}(K-L)^2,$$
$$2D(D+1) G t_2 = \tfrac{9}{4} DKL + K\big(DG_+ - (D+1)G\big)^2 + L\big(DG_- - (D+1)G\big)^2.$$

The first gives $t_1 \geq 0$. The other two give $G \leq \frac{5}{4}D$ and then $D(D+1)t_2 \geq \frac{9DKL}{8G} \geq \frac{9KL}{10} \geq 4cKL$. By the choice of α, $D\alpha_n - (D+2)\alpha_{n+1} = c/(D(D+1))$, and $4KL + e^2 = D^2$; so $KX_+ + LX_- \geq DY$. Since $K, L > 0$, cotransitivity gives $X_+ \geq Y - \varepsilon$ or $X_- \geq Y - \varepsilon$ for $\varepsilon = 1/(N+1)$, with no decision of a sign; choose that record. From $\Phi_0 = -\frac{5}{2}$, after N steps $\Phi_N \geq \sum_{n<N}(c/(n+3) - \alpha_{n+1}) - \frac{5}{2} - 1 \geq \frac{H_N}{12} - \frac{15}{4}$, using $\sum_n \alpha_{n+1} \leq cH_N/3$. Finally $\alpha_N = 0$, and the frame $m = G/V$ costs at most $G - N$ on the records, since their sums are $U - \frac{5}{2}$ and $V - \frac{5}{2}$; so $\mathrm{Loss}_N - C_N(m) \geq \Phi_N$. □

Parts 1 and 3 together say that the minimax regret of learning a frame under J is of order $\log N$: following the centre achieves it, and no learner does better than a constant fraction of it. This is the logarithmic regret known for strongly convex losses [129], here with J as the loss and with no averaging over sequences.

The centre can be reached without a square root. Heron's update $m \mapsto \frac{1}{2}(m + m_*^2/m)$ is Newton's method for $m^2 T = S$, and also the gradient step on C with step size m/T, since $C'(m) = \frac{T}{2}(1 - m_*^2/m^2)$; reading it as the natural gradient step of the frame is an interpretation.

Theorem 29.41 (Heron's update is free and converges)**.**

1. *For $m, m_* > 0$, $\frac{1}{2}(m + m_*^2/m) - m_* = m_* J(m/m_*) = (m - m_*)^2/(2m)$: the overshoot of Heron's update is the J-cost of the frame it leaves.*
2. *On $\mathbb{R}_\delta$, for $q \geq 0$ and $h_0 > 0$ the iterates $h_{k+1} = \frac{1}{2}(h_k + q/h_k)$ exist and are positive. If $m > 0$ and $m^2 = q$, then $0 \leq h_{k+1} - m \leq 2^{-k}(h_1 - m)$ and $h_{k+2} - m \leq (h_{k+1} - m)^2/(2m)$ for every k.*

The theorem is free.

Proof. (1) is algebra. (2) A positive point has an inverse, so the iterates exist, and $h_{k+1} - m = (h_k - m)^2/(2h_k) \geq 0$ by (1). For $k \geq 1$, $0 \leq (h_k - m)/h_k \leq 1$, so the error at least halves; and $1/h_{k+1} \leq 1/m$ gives the quadratic bound. □

The error falls geometrically at first and then quadratically, and the limit m is not needed to run the iteration, only to measure it.

When to stop, and what exact steps cost

An optimizer stops when its residual is small. A residual is a point of $\mathbb{R}_\delta$, and an exact test of smallness is a comparison of points. Theorem 29.42 prices the tests, and Table 29.2 the step rules.

Theorem 29.42 (The prices of stopping)**.** *Let $r_0, r_1, \ldots$ be points of $\mathbb{R}_\delta$, the residuals of a run.*

1. *If $a < b$ and some $r_k < a$, there is a k with $r_k < b$ and $a < r_j$ for every $j < k$: a test with a tolerance band stops, and stops in time. This is free.*
2. *"Whenever some $r_k < b$, there is a least k with $r_k < b$", for all runs and points b, is equivalent to LPO.*

Table 29.2. Prices of the step rules of an optimizer on $\mathbb{R}_\delta$. Each row asks for every g (or a, b) a choice with the stated guarantees; the tolerance $\varepsilon > 0$ is a point.

Step rule	Price	Source
Sign step: a bit, 0 only if $g \geq 0$, 1 only if $g \leq 0$	LLPO	Corollary 12.18
Three-way step: down if $g > 0$, stay if $g = 0$, up if $g < 0$	LPO	Corollary 12.18
Exact stop: 1 only if $g = 0$, 0 only if $g \neq 0$	WLPO	Corollary 12.18
Exact acceptance: 1 only if $a < b$, 0 only if not	LPO	Table 12.2
Tolerance step: 0 only if $g > -\varepsilon$, 1 only if $g < \varepsilon$	free	cotransitivity
Tolerance acceptance: 1 only if $a < b + \varepsilon$, 0 only if $b < a$	free	cotransitivity

3. *For runs given by regular sequences naming the residuals, "$\neg\neg\exists k\, r_k < b$ implies $\exists k\, r_k < b$", for all such runs and points b, is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$.

The equivalences are proved with no choice principle.

Proof. (1) For $j = 0, 1, \ldots$ apply cotransitivity to $a < b$ at r_j: $r_j < b$ or $a < r_j$. Stop at the first j with $r_j < b$; at the given k the second case is impossible, so the search ends by k. (2) LPO decides $r_j < b$ for each j (Table 12.2), and a search below the given k finds the least. Conversely, for points x and y take the run $x,\ y - 1$ and $b = y$: the least k is 0 exactly when $x < y$, so the rule decides $x < y$, which is LPO. (3) A strict inequality $r_k < b$ holds exactly when, at some stage, the regular sequences show a gap larger than their tolerance, a decided condition on the pair of the residual's index and the stage; $\mathrm{MP}_{\mathrm{bin}}$ applied to the pairs finds one. Conversely, for binary α, the run $r_k = -1$ if $\alpha(j) = 1$ for some $j \leq k$ and $r_k = 1$ otherwise, with $b = 0$, turns $\neg(\alpha$ all zero) into an entry 1. □

For residuals that are bare points, not given by their sequences, the rule of part 3 still implies $\mathrm{MP}_{\mathrm{bin}}$, and whether the converse holds is not settled here. Steps that branch on exact signs carry the prices of Chapter 12; Table 29.2 lists them, together with the free rules that use a tolerance instead.

An optimizer on exact points therefore uses overlapping tolerance guards, as Theorem 17.22 found for tests of a law: every exact branch costs an omniscience principle, and every guarded branch is free.

29.9 Learning on the programs of Gödel's T

Section 29.7 read Recognition Science's processes as the computable functions. A narrower reading takes its execution layer to be the closed programs of Gödel's system T with their evaluation, the system of Chapter 28. Either reading is an identification, and the two are different: the closed programs of T compute only some of the total computable functions, and the computable functions include partial ones. The theorems of each section are about its own programs, and this section's do not apply to partial programs. The programs here halt on every input, and this changes the prices of Section 29.7.

Recall the system. Its programs are built by application from the constants K, S, 0, Succ and the recursors R at every finite type, and a program t of type $O \to O$ has a value $\mathrm{val}(t)$, a function from numbers to numbers defined at every input. Programs are finite trees labelled by constants and types, so they can be coded by numbers. Decoding a number is a bounded computation, and it lists the programs of type $O \to O$ as $t_0, t_1, \ldots$ with every program occurring.

Definition 29.43 (Worlds of the layer and learning on it)**.** A binary world ω is a *world of the layer* when some program t of type $O \to O$ *computes* it: $\mathrm{val}(t)(x)$ is 1 when $\omega(x) = 1$ and 0 otherwise. LPO_T and MP_T are LPO and $\mathrm{MP}_{\mathrm{bin}}$ for the binary sequences that are worlds of the layer. A *learner on the layer* maps finite binary records to programs; EX and BC^* are as in Definition 29.34, with programs in place of codes. It *changes its mind* at k on ω when its conjectures at the records of lengths k and $k+1$ differ. A learner *commits once* on ω when, for some stage N, it gives no program at the records shorter than N and one fixed program computing ω at all the others. The *fixed learner* E names, at a record σ, the program t_i for the least $i \le |\sigma|$ whose values reproduce σ, and $t_{|\sigma|+1}$ if there is none.

Theorem 29.44 (No program of the layer evaluates the layer)**.** *No program U of type $O \to (O \to O)$ satisfies $\mathrm{val}(U)(j)(y) = \mathrm{val}(t_j)(y)$ for all j and y. This is free. In classical mathematics, $(j, y) \mapsto \mathrm{val}(t_j)(y)$ is a computable function, and the fixed learner E is computable through the indices it names.*

Proof. Bracket abstraction turns $y \mapsto \mathrm{val}(U)(y)(y) + 1$ into a program, which is t_j for some j; at $y = j$ it would give $\mathrm{val}(U)(j)(j) = \mathrm{val}(U)(j)(j) + 1$. For the classical statement, a code a *realizes* a value v of type O when $a = v$, and one of type $\sigma \to \rho$ when $\varphi_a(b)$ is defined and realizes $v(w)$ for every b realizing a value w of type σ. One code realizes K at every type, one S, one Succ, 0 realizes 0, and one code realizes every recursor R, its run on a numerical argument n halting by induction on n. A realizer of v applied to one of w realizes $v(w)$, so the code assembled along a program's tree, computed from its index, realizes its value. Hence $(j, y) \mapsto \mathrm{val}(t_j)(y)$ is computable, and so is E's search through $i \le |\sigma|$. □

So the layer cannot run its own interpreter, while a program outside it can. That is why the prices below stop at LPO for the layer's own sequences.

Theorem 29.45 (Learning the layer costs exactly LPO for its own sequences)**.** *Each of the following is equivalent to LPO_T, with no choice principle in either direction:*

1. *some learner EX-identifies every world of the layer;*
2. *some learner BC^*-identifies every world of the layer;*
3. *the fixed learner E EX-identifies every world of the layer;*
4. *some learner EX-identifies every world of the layer while changing its mind at most $c + 1$ times on every world that t_c computes.*

With no hypothesis at all, E changes its mind at most c times on every world that t_c computes.

Conditions and scope.

Base and premises
The closed programs of Gödel's T of type $O \to O$ and the binary worlds they compute; learners of any kind.

Conclusion
Learning every world of the layer in the limit costs exactly LPO for the layer's own sequences, by one fixed learner, and costs no more with a count of revisions attached.

Identification
That these programs are Recognition Science's execution layer is an identification; the theorem is about the programs.

Proof. Upper bounds. Let ω be computed by t. For each j, the world that is 1 where $\mathrm{val}(t_j)$ and $\mathrm{val}(t)$ disagree is a world of the layer, since comparing two numbers is primitive recursive. So LPO_T decides whether t_j computes ω or finds an input where it does not. It therefore finds the least i with t_i computing ω and, for each $j < i$, a refuting input; past these finitely many inputs E names t_i. *The count.* The index named by E never decreases, since a refuted index stays refuted; it never exceeds c when t_c computes ω, since t_c is never refuted; and it rises at every change. A nondecreasing sequence of numbers bounded by c rises at most c times. This gives (4) from (3). *Lower bounds.* The worlds of the layer contain the zero world and the flag world of each member, by primitive recursion, and two worlds that one program computes at all but finitely many inputs agree from some input on. So Theorem 29.35(1) turns a BC^* learner into LPO_T, and each of (1), (3) and (4) gives (2). □

Why is the layer priced by LPO_T alone, while the learners of Section 29.7 use a halting table or, for Harrington's method, Σ_2-LPO (Section 29.10)? Whether a program of the layer computes a given world of the layer is a question about one decided sequence: LPO answers it. Whether a code computes a world asks also whether it halts, a question one level higher.

Theorem 29.46 (Revisions on the layer)**.**

1. *For every m, no learner EX-identifies every world of the layer while changing its mind at most m times on each.*

2. *Some learner EX-identifies, with at most one change of mind, every world of the layer that has at most one entry* 1*, if and only if* LPO_T *holds.*

3. *For a finite list s of numbers, let ω_s have its entries* 1 *exactly at the members of s. One learner EX-identifies every ω_s with at most $|s|$ changes of mind, and for every k no learner identifies every ω_s with $|s| \le k + 1$ within k changes.*

Part 2 is proved with no choice principle, and parts 1 and 3 are free.

Proof. (1) Run the learner on the zero world until it has settled on a program, then let the world change from a later stage on; worlds that differ need different programs, so the learner must change its mind. The worlds built are eventually constant, hence worlds of the layer, and

$m + 1$ rounds force $m + 1$ changes. (2) The learner that names the program of the first entry 1 seen, and the zero program before, changes its mind at most once and identifies the zero world and every world with a single entry 1. Under LPO_T every world of the layer with at most one entry 1 is one of these. Conversely, let a learner identify such worlds, and let α be a world of the layer. The world that has its only entry 1 at the first entry 1 of α is a world of the layer with at most one entry 1. The learner has settled on the zero world by some stage N and on this world by N'; if α is 0 below $N + N'$, the two records agree there, the shared conjecture computes both worlds, and α is all zero. (3) Name the program of the set of positions of the entries 1 seen so far; this changes only at a new entry 1. Against a learner with k changes, add a new entry 1 after each settling, $k + 1$ times. □

Committing once

A learner that commits once must know when it has seen enough. Here the price sits in knowing what form a world has.

Theorem 29.47 (Committing once on the layer)**.**

1. *No learner commits once correctly on a class that contains the zero world and, for every N, a world that is* 0 *below N and has an entry* 1*. In particular no learner commits once correctly on every world of the layer.*
2. *One learner, defined with no hypothesis, commits once correctly on every* headed *world:* 0 *below some j,* 1 *at j, and from there on the values of t_j, shifted to begin at $j + 1$.*
3. *For a world ω of the layer, some learner commits once correctly on both the zero world and ω if and only if ω is all zero or has an entry* 1*. Hence committing once on the zero world and the flag world of α, for every world α of the layer, is equivalent to* LPO_T*, and on the zero world and ω, for every world ω of the layer that is not all zero, to* MP_T*.*

Parts 1 and 2 are free, and part 3 is proved with no choice principle.

Proof. (1) On the zero world the learner commits at some stage N to a program z. A member that is 0 below N has the same record of length N, so the learner commits to z there as well, and z cannot compute a world with an entry 1. (2) Stay silent until the first entry 1, at j; then commit to the program that writes j zeros, a 1, and then the values of t_j. (3) If ω is all zero, commit to the zero program at once; if $\omega(m) = 1$, stay silent below $m + 1$ and commit by the entry at m. Conversely, let the learner commit on the zero world at N_0 and on ω at N_1. If ω is 0 below $N_0 + N_1$, the records agree there, the learner commits to one program on both, and ω is the zero world. The flag world of α is all zero or has an entry 1 exactly when α is. □

Theorem 29.48 (Catalogues, limits and locking functions)**.** *Let $W_0, W_1, \ldots$ be a* catalogue*: worlds given with programs P_i computing each W_i.*

1. *Some learner EX-identifies every W_i if and only if any two members are equal or differ at some place; some learner BC^*-identifies every W_i if and only if any two members agree from some place on or differ at some place. The two conditions are equivalent.*

2. *A finite catalogue admits a learner that commits once correctly on each member if and only if any two members are equal or differ at some place.*

3. *A catalogue admits a learner that commits once correctly on each member if and only if it has a* locking function*: a map g such that whenever W_j agrees with W_i below $g(i)$, $W_j = W_i$.*

4. *The zero world with the worlds that have a single entry* 1 *form a catalogue that a learner defined with no hypothesis EX-identifies with at most one change of mind, and on which no learner commits once correctly.*

5. *Let W_i be* 0 *below $d_i = \mathrm{val}(t_i)(i)$ and* 1 *from d_i on. This catalogue has the locking function $i \mapsto d_i + 1$, so a learner commits once correctly on it, but no program of the layer computes a locking function for it.*

Each statement is proved with no choice principle.

Proof. (1) Let a learner identify W_i and W_j, settled from stages N and N'. If the members differ below $N + N'$, a search finds the place; if not, the conjecture at the shared record of length $N + N'$ computes both, exactly or from some input on. Agreement from B on, with a search below B, decides equality, which gives the equivalence of the two conditions. Conversely, the learner that names the least index consistent with the record identifies every member, as in Theorem 29.24, the decisions refuting smaller indices. (2) Two members learned by committing once are decided, by the argument of Theorem 29.47(3). Conversely, choose a stage m beyond one place of difference of every pair that differs; stay silent below m, then commit to the program of the first listed member that agrees with the record below m. (3) Given such a learner, let $g(i)$ be the stage at which it commits on W_i, found by search; if W_j agrees with W_i below $g(i)$, the learner commits on W_j to the program of W_i, so $W_j = W_i$. Given g, commit at the first stage n at which some $i \le n$ has $g(i) \le n$ and agrees with the record below $g(i)$, to P_i. (4) Name the program of the first entry 1 seen, and the zero program before; Theorem 29.47(1) excludes committing once. (5) If W_j agrees with W_i below $d_i + 1$, its step is at d_i, so $W_j = W_i$. Every program $c \mapsto v$ with a constant value is some t_j, so for each i some member steps at $d_i + 1$ and agrees with W_i below d_i; a locking function g therefore has $g(i) > d_i$. If t_k computed one, then $\mathrm{val}(t_k)(k) > d_k = \mathrm{val}(t_k)(k)$. □

Part 4 separates convergence from commitment on one catalogue at no price: the learner may change its mind once, and nobody can know when. The omniscience sits in knowing which form a world has: LPO_T is equivalent to every world of the layer with at most one entry 1 belonging to that catalogue. Part 5 shows that a locking function can exist although the layer cannot compute it. Locking sequences go back to Blum and Blum [115], and counted mind changes to Case and Smith [118], with ordinal counts due to Freivalds and Smith [127].

29.10 Methods, bounds and decision tables

Harrington's learner (Theorem 29.36(4)) and the learner that reads a halting table (Theorem 29.35(3)) are methods, and a method makes sense for any system of programs whose runs may fail to stop. This section prices each method over all such systems at once, then prices the bounds that say when a method has become right.

Definition 29.49 (Systems of programs, the two methods and decision tables)**.**

1. A *system of programs* is an evaluation $\mathrm{ev}(s, e, y)$, the value of program e at input y after s steps, or none yet; a value once given stays given at later steps. Program e *computes* ω when for every y some s gives the encoded value $\omega(y)$.

2. The *selection method* is any rule $\mathrm{sel}(\sigma, x)$ that returns the least $e \le |\sigma|$ whose computation, run x steps, reproduces the record σ, whenever there is one. Its conjecture at σ answers input x with the value of program $\mathrm{sel}(\sigma, x)$ at x. It *succeeds* on ω when, from some record length on, each conjecture answers correctly at all large inputs.

3. The *oracle method*, for a partial evaluation with a halting table, names the least $c \le |\sigma|$ whose values, read through the table, reproduce σ; it succeeds when it settles on a program computing ω.

4. Σ_2*-LPO* states that for every binary double sequence R, either every row y has some s with $R(y, s) = 1$, or some row y has $R(y, s) = 0$ for every s.

5. A Σ_1 *table* exists when every binary double sequence R has a map D with $D(r) = 1$ exactly when $R(r, s) = 1$ for some s. A Σ_2 *table* exists when every binary triple sequence R has a map T with $T(r) = 1$ exactly when every y has some s with $R(r, y, s) = 1$.

6. A method has a *bound* on a system when some map F gives, for every program c and every world ω that c computes, a record length $F(c)$ from which the method is right on ω.

Σ_2-LPO decides, for a whole array, whether every row eventually fires or some row never does. It is the same principle as Σ_2-LEM of Definition 29.26, which asks only whether some row never fires.

Proposition 29.50 (Σ_2 excluded middle is Σ_2-LPO)**.** Σ_2*-LEM and* Σ_2*-LPO are equivalent. Both directions are proved with no choice principle.*

Proof. Let Σ_2-LPO hold, and let β be a binary double sequence. Applied to $R = \beta$, it says either that every row of β has an entry 1, and then no row is 0 everywhere, or that some row is 0 everywhere. Either way the alternative of Σ_2-LEM is decided.

Conversely, let Σ_2-LEM hold. For a binary sequence α, apply it to $\beta(m, n) = 1 - \alpha(m)$: row m of β is 0 everywhere exactly when $\alpha(m) = 1$, so either α has an entry 1 or none, which is LPO. Now let R be a binary double sequence, and apply Σ_2-LEM to R itself. If some row is 0 everywhere, that is the second alternative of Σ_2-LPO. Otherwise no row is 0 everywhere. Then, for each row y, LPO applied to the row finds an entry 1 or shows that the row is 0 everywhere, and the second case is excluded. So every row has an entry 1, the first alternative. □

Theorem 29.51 (The price of the two methods)**.**

1. *The selection method succeeds on every world that some program computes, for every system of programs and every rule meeting its specification, if and only if Σ_2-LPO holds.*
2. *The oracle method succeeds on every world that some program computes, for every system with a halting table, if and only if LPO holds.*
3. *Σ_2-LPO implies LPO.*

Each implication is proved with no choice principle.

Proof. *Σ_2-LPO buys selection.* Let c compute ω. For each $e \le c$, Σ_2-LPO applied to $R(y, s) = 1$ when $\mathrm{ev}(s, e, y)$ is the value $\omega(y)$ says that e computes ω or names an input y_e at which e never gives $\omega(y_e)$. Let E be the least program that computes ω. Once the record is longer than every y_e with $e < E$, no $e < E$ is selected at any x, and once x exceeds the steps E needs on the recorded inputs, E is selected; its value at x is $\omega(x)$.

Selection gives LPO. Use the system in which program 0 computes the flag world of α and every other program answers 0 at once. The method succeeds on the zero world from some record length N. If α has an entry 1 below N, we are done. Otherwise program 0 reproduces the zero record of length N at every x, so it is selected, and its values must be 0 at all large x: the flag world is eventually 0, and α is all zero.

Selection gives Σ_2-LPO. Given R, let program 0 answer 0 at y within s steps once every row $y' \le y$ has a witness up to s, and let the other programs answer 0 at once. The method succeeds on the zero world from some length N. By the previous step, LPO decides each of the rows up to N. If one has no witness, that is the answer. Otherwise all have witnesses up to some S, program 0 reproduces the record of length N within S steps, and it is selected at every $x \ge S$; its value must then exist at all large x, so every row has a witness.

The oracle method. LPO buys it by Theorem 29.35(3). Conversely, let program 0 compute the flag world of α, the others the zero world, with the table saying that every program halts. The method names a program of the zero world from some length N. If α is 0 below N, program 0 is consistent and least at that record, so it computes the zero world, and α is all zero.

(3) Apply Σ_2-LPO to $R(y, s) = \alpha(s)$. □

Harrington's method thus costs Σ_2-LPO, the Σ_2 form of LPO. It implies LPO, the price of the halting-table method. Whether it is strictly stronger under this chapter's reading of maps is open; with countable choice LPO gives it, since LPO then decides every row and the answers can be collected, and under the default reading, which contains unique choice, LPO gives it as well (Theorem 29.53). Both prices imply LPO_T, the price on the layer of Section 29.9, where programs always halt. By Proposition 29.50, Σ_2-LPO is also the price of deciding finiteness of every run (Theorem 29.27(5)): the selection method and verdicts revised without a bound cost one principle.

Theorem 29.52 (Bounds on the methods are decision tables)**.**

1. *The selection method has a bound on every system of programs if and only if Σ_2 tables exist.*
2. *The oracle method has a bound on every system with a halting table if and only if Σ_1 tables exist.*
3. *Σ_2 tables give Σ_1 tables, and Σ_1 tables give LPO and $\mathrm{MP}_{\mathrm{bin}}$; so a bound for either method gives LPO and $\mathrm{MP}_{\mathrm{bin}}$.*

Each implication is proved with no choice principle.

Proof. Bounds give Σ_1 tables. Given R, take the system in which program 0 computes the zero world and program $r+1$ the flag world of row r, every program halting at once. Let F be a bound, for either method, and put $D(r) = 1$ when row r has a witness below $F(0) + F(r+1)$. If it has none, the zero world and the flag world of row r have the same record of that length, past both bounds, so one conjecture is right on both from some input on, as in Theorem 29.35(1), and row r has no witness at all. With Theorem 29.53(4) below, a bound for the selection method also gives Σ_2 tables.

Tables give bounds. For the oracle method, a Σ_1 table on the pairs (j, c) decides whether program j differs from program c somewhere, reading definedness through the halting table; let $F(c)$ exceed c and the first such place of every $j < c$ that differs. For the selection method, a Σ_2 table decides for $e < c$ whether e computes the world of c, and the Σ_1 table it contains, with the LPO that it gives, finds the first input up to which e fails to agree; let $F(c)$ exceed c and these inputs. In both cases the argument of Theorem 29.51 then runs from $F(c)$ on.

(3) A Σ_1 question is a Σ_2 question with an idle universal quantifier. A Σ_1 table for the family whose every row is α decides, at row 0, whether α has an entry 1: that is LPO, and LPO gives $\mathrm{MP}_{\mathrm{bin}}$. □

A decision table answers a whole family of questions at once and is data; LPO answers one question at a time. Without unique choice the two differ, and by exactly unique choice.

Theorem 29.53 (Decision tables and propositions differ by unique choice)**.** *Say that* unique choice holds for Σ_1 families *when every binary double sequence R whose rows are each decided, $\exists s\, R(r,s) = 1$ or not, has a map D with $D(r) = 1$ exactly when $R(r,s) = 1$ for some s; define it for Σ_2 families in the same way.* Boolean unique choice *says: if for every n exactly one bit b has $P(n,b)$, then some map f has $P(n, f(n))$ for every n.*

1. *Σ_1 tables exist if and only if LPO holds and unique choice holds for Σ_1 families.*
2. *Σ_2 tables exist if and only if Σ_2-LPO holds and unique choice holds for Σ_2 families.*
3. *Boolean unique choice gives unique choice for Σ_1 and for Σ_2 families.*
4. *Σ_1 tables give Σ_2-LPO, and Σ_1 tables exist if and only if Σ_2 tables do.*

The theorem uses no choice principle; parts 1 to 3 use no principle of any kind.

Conditions and scope.

Base and premises
The narrower reading of maps of Section 21.1, in which a map is a rule given as data.

Conclusion
A decision table is exactly its principle together with unique choice for the family; Σ_1 and Σ_2 tables are equivalent.

Not claimed
Any difference between tables and principles under the default reading of maps, which contains unique choice and makes them coincide.

Proof. (1) A Σ_1 table gives LPO by Theorem 29.52(3), and it is itself the map that unique choice asks for. Conversely, LPO decides every row, and unique choice collects the decided rows into a map. (2) is the same argument, with Σ_2-LPO obtained from the table as in (4). (3) For a decided row, the relation "$b = 1$ exactly when the row has a witness" holds of exactly one bit. (4) Given R, a Σ_1 table D on its rows and LPO applied to $y \mapsto 1 - D(y)$ show that every row has a witness or name a row y with $D(y) = 0$, which has none. For Σ_2 tables, a Σ_1 table on the pairs (r, y) gives $E(r, y) = 1$ exactly when row (r, y) has a witness, and a second Σ_1 table on "some y has $E(r, y) = 0$" gives the complement of the Σ_2 table. □

This is the separation that Theorem 16.27 exhibits in δ's calculus, where LPO decides each question but cannot collect the answers. Under the default reading of maps of this book, which contains unique choice (Appendix C and Section 21.1), the tables and the principles coincide: Σ_1 tables are LPO, and Σ_2 tables are Σ_2-LPO. The distinction is a statement about the narrower reading of Section 21.1.

Bounds for learners of computable laws

For the codes of Section 29.7, a *bound* for a learner M is a map F from codes to record lengths such that, whenever c computes ω and $n \geq F(c)$, the conjecture $M(\omega{\restriction}n)$ computes ω at all but finitely many inputs. A *halting table* is a map d with $d(c, k) = 1$ exactly when $\varphi_c(k)$ is defined, and a *totality table* a map T with $T(c) = 1$ exactly when $\varphi_c(k)$ is defined for every k. A *catalogue* is a sequence of worlds $W_0, W_1, \ldots$, and a locking function is as in Theorem 29.48.

Theorem 29.54 (Bounds on learners, and halting questions as the price)**.**

1. *Let W be any catalogue with an index z of the zero world and, for each i, an index $f(i)$ of the flag world of W_i, and let hypotheses be as in Theorem 29.35(1). If a learner has a map F such that from stage $F(i)$ on its conjectures compute W_i, then some map D has $D(i) = 1$ exactly when W_i has an entry* 1*. Hence, in classical mathematics, every learner that BC*-identifies every computable world with a bound given as a map on codes yields a halting table.*

2. *In classical mathematics, a map F on codes is the bound of some BC* learner of every computable world if and only if it* separates*: whenever codes c and c' compute worlds that agree below* $\max(F(c), F(c'))$*, those worlds agree from some input on. The implication from bound to separation holds for every system of hypotheses with no choice principle.*

3. *In classical mathematics, for every catalogue that is computable and has a locking function, a computable learner commits once correctly on every member of it, naming indices, by asking at each record of length n some $n+1$ halting questions fixed by n; neither the questions nor the learner read the locking function. And there is a computable catalogue of worlds of the layer of Section 29.9 that has a locking function given a halting table, on which no computable learner commits once correctly naming indices.*

Proof. (1) Put $D(i) = 1$ when W_i has an entry 1 below $F(z) + F(f(i))$. If it has none, the records of W_i's flag world and of the zero world agree up to that length, past both bounds, so one conjecture computes both from some input on, and W_i is all zero. For the instance, list as W the worlds that record when the computation of $\varphi_c(k)$ halts; its flag worlds and zero world are computable worlds with codes found from c and k, and D is a halting table. (2) If F is a bound and the worlds of c and c' agree below $\max(F(c), F(c'))$, the conjecture at that shared record computes both from some input on. Conversely, let F separate. At σ the learner names the least code c with $F(c) \leq |\sigma|$ that computes a world whose record of length $|\sigma|$ is σ, and 0 if there is none; classical mathematics allows this test. If c_0 computes ω and $n \geq F(c_0)$, then c_0 qualifies at $\omega{\restriction}n$, so the named code c_1 computes a world that agrees with ω below $n \geq \max(F(c_1), F(c_0))$, and separation makes the two agree from some input on. (3) The learner asks, for each place $\langle m, i\rangle \leq n$, whether the search for a member agreeing with W_i below m and differing later halts. Its answers tell, for each i, whether m locks W_i, and the learner commits to the first i that the record and a locking stage confirm. For the second statement, list for each code e the world that is 1 below e, and, when $\varphi_e(0)$ halts in exactly $s+1$ steps, the same world with one more entry 1 at $e+s+1$. A halting table bounds the steps and gives a locking function, while a computable learner committing once on every member would make non-halting a computably enumerable property, against Turing's theorem [21]. □

Theorem 29.54(2) says which maps are bounds but not what a bound costs. The cheapest costs exactly a halting table: every bound, of any learner, gives one back by a finite check, and a halting table builds a separating map, which is the bound of some learner.

Theorem 29.55 (The cheapest bound costs exactly a halting decision)**.** *Work in classical mathematics. The* halting trace *of code c at input k is the world $\tau_{c,k}$ with $\tau_{c,k}(s) = 1$ exactly when the computation of $\varphi_c(k)$ halts within s steps; codes* $\mathrm{tc}(c,k)$ *of these worlds exist.*

1. *Let d be a halting table. For codes $c' < c$, consider the search through the pairs (x, s) for the first at which $\varphi_c(x)$ and $\varphi_{c'}(x)$ have both returned within s steps with different values, returning x. Let $F(c)$ be 1 plus the largest input returned by these searches over the $c' < c$ for which d says the search halts. Then two worlds computed by codes c and c' that agree below $\max(F(c), F(c'))$ agree everywhere; in particular F separates.*

2. *Let F separate, let z be a code of the zero world, and let* tc *give codes of the halting traces. Then $\varphi_c(k)$ is defined if and only if $\tau_{c,k}(s) = 1$ for some $s < \max(F(z), F(\mathrm{tc}(c,k)))$.*

3. *Let W be a catalogue with $(i, x) \mapsto W_i(x)$ computable. The learner that names, at a record σ, the least place $i \leq |\sigma|$ whose world agrees with σ, and 0 if there is none, is computable. Given a halting table, let $b(k)$ be $k+1$ plus the largest first place of difference between W_i and W_k over the $i < k$ for which the table says that the search for such a place halts. On W_k, from stage $b(k)$ on, the learner names a place holding W_k.*

Once the searches and the table's answers are given, each argument uses no choice principle.

Proof. (1) Let c and c' compute ω and ω', which agree below $\max(F(c), F(c'))$. If $c = c'$ the worlds are equal. Otherwise let $c' < c$, say. Were the worlds to differ somewhere, the search for (c, c') would halt, both codes being defined everywhere; d would say so, and the search would return an input at which they differ and which lies below $F(c)$, against the agreement. So they differ nowhere, and since bits are decided, $\omega = \omega'$.

(2) If $\tau_{c,k}(s) = 1$ the computation halts. Conversely, let it halt within s_0 steps, so that $\tau_{c,k}(s) = 1$ for every $s \geq s_0$, and inspect the s below $m = \max(F(z), F(\mathrm{tc}(c, k)))$. If none has $\tau_{c,k}(s) = 1$, the trace agrees with the zero world below m, so the two agree from some input B on, and $\tau_{c,k}(\max(s_0, B))$ would be both 1 and 0.

(3) Agreement of W_i with a finite record is a finite computation, so the learner is computable. Let $n \geq b(k)$. Then $k \leq n$ and W_k agrees with its own record, so the least agreeing place is at most k. A place $i < k$ with $W_i \neq W_k$ makes the search halt, since both worlds are total, and their first difference lies below $b(k) \leq n$; so W_i does not agree with the record. The least agreeing place therefore holds W_k. □

With Theorem 29.54(2), a halting table builds the bound of some learner, of any kind, that BC*-identifies every computable world, and the bound of any such learner yields a halting table. Part 3 builds, in the same way from a halting table, a bound for a computable learner on catalogues of total worlds. A stronger demand on the bound, or a particular learner, can ask for more.

Theorem 29.56 (Stronger bounds carry a totality decision)**.** *Work in classical mathematics. A map F on codes is a* strong bound *when, whenever c computes ω, every code $e < c$ that reproduces ω at the inputs below $F(c)$ computes ω at all but finitely many inputs.*

1. *For each code p there are, found from p, a code $\mathrm{tr}(p)$ of the program that at input x waits until φ_p has halted at every input $y \leq x$ and then returns 0, and a code $\mathrm{za}(p) > \mathrm{tr}(p)$ of the zero world. If F is a strong bound and d a halting table, then φ_p is total if and only if $d(p, y) = 1$ for every $y < F(\mathrm{za}(p))$.*

2. *If F is a strong bound, then $c \mapsto \max(F(c), c + 1)$ is a bound for Harrington's learner of Theorem 29.36(4).*

3. *A totality table, together with $\mathrm{MP}_{\mathrm{bin}}$, yields a strong bound.*

4. *Fix a pairing of numbers, and let $\mathrm{col}_{p,j}$ be the world that is 1 exactly at the inputs whose first coordinate is the pair (p, j). The* trap (p, j) *is the program that at input x waits until φ_p has halted at every input $y \leq x$ and then returns $\mathrm{col}_{p,j}(x)$. List the programs in blocks $0, 1, 2, \ldots$, block p being program p followed by the traps (p, j) for $j < p + 2$. The* trap-first learner *conjectures at σ the program that at input x runs the first entry of blocks $0, \ldots, |\sigma|$ whose computation, run x steps, reproduces σ, and diverges if there is none. It is computable and BC*-identifies every computable world. If F is a bound for it and d a halting table, then φ_p is total if and only if $d(p, y) = 1$ for every $j < p + 2$ and every $y < F(w_{p,j})$, where $w_{p,j}$ is a code of $\mathrm{col}_{p,j}$ found from p and j.*

The construction of part 3 uses no choice principle once the table is given.

Proof. (1) If φ_p is total, $d(p, y) = 1$ for every y. Conversely, let $d(p, y) = 1$ for every $y < N = F(\mathrm{za}(p))$. Then φ_p halts at every $y < N$, so $\mathrm{tr}(p)$ returns 0 at every $x < N$ and reproduces the zero world below N. Since $\mathrm{tr}(p) < \mathrm{za}(p)$ and $\mathrm{za}(p)$ computes the zero world, $\mathrm{tr}(p)$ returns 0 at all large x. So φ_p halts at every $y \leq x$ for arbitrarily large x: it is total.

(2) Let c compute ω and $n \geq \max(F(c), c + 1)$. The code $c \leq n$ reproduces ω below n, so the least code $e \leq n$ that does is at most c; if $e < c$, it reproduces ω below $F(c)$, and by strength it computes ω at all but finitely many inputs. Every code below e is wrong or undefined at some recorded input and is never selected, while e is selected once x exceeds the steps it needs on the record. So the conjecture returns $\varphi_e(x) = \omega(x)$ at all large x.

(3) For codes e and c, let $A(e, c)$ say that $\varphi_e(y)$ and $\varphi_c(y)$ are defined and equal for every y, and $A_y(e, c)$ that they are at y. Each is a totality question, the second about a program that ignores its input, so the table decides them. If $A(e, c)$ fails, not every $A_y(e, c)$ holds, and $\mathrm{MP}_{\mathrm{bin}}$ gives a y at which $A_y(e, c)$ fails; let $y(e, c)$ be the least, and $y(e, c) = 0$ when $A(e, c)$ holds. Put $F(c) = \max_{e<c}(y(e, c) + 1)$. If c computes ω and $e < c$ reproduces ω below $F(c)$, then $A_y(e, c)$ holds at every $y < F(c)$, in particular at $y(e, c)$; so $A(e, c)$ holds, and e computes ω.

(4) The learner is computable because the list up to block $|\sigma|$ and the runs are finite computations. Fitting a record within x steps persists as x grows, so for a fixed record the selected entry is, for all large x, the first entry that fits within some number of steps. Let ω be computable; program c computing it heads block c. Let q be the first entry of the whole list that computes ω. Each earlier entry is wrong or undefined at some input; once the record contains these inputs and has length at least c, no earlier entry is selected and q is, at all large x. So every late conjecture computes ω at all large inputs.

For the totality test, the forward direction is clear. Conversely, let $d(p, y) = 1$ for all $j < p + 2$ and $y < F(w_{p,j})$. Two different column worlds differ at infinitely many inputs, so each of the $p + 1$ programs $0, \ldots, p$ computes, at all but finitely many inputs, at most one of the $p + 2$ worlds $\mathrm{col}_{p,j}$. Choose j such that none of them does so for $\mathrm{col}_{p,j}$, and let $n = F(w_{p,j})$ and σ the record of $\mathrm{col}_{p,j}$ of length n. The conjecture at σ computes $\mathrm{col}_{p,j}$ at all large inputs, so for all large x it runs one entry q of blocks $0, \ldots, n$, and q computes $\mathrm{col}_{p,j}$ at all but finitely many inputs. A trap of another pair computes another column or is undefined from some input on, so q is not one, and by the choice of j it is none of the programs $0, \ldots, p$. If $n < p$, every entry of blocks $0, \ldots, n$ is excluded, which is impossible; so $n \geq p$, and the trap (p, j) is in the list. Since φ_p halts at every input below n, the trap reproduces σ, so q is the trap or an earlier entry. The earlier entries are programs $0, \ldots, p$ and traps of other pairs, all excluded; so q is the trap (p, j). It is defined at all large inputs, so φ_p is total. □

Harrington's learner sits between the two prices. A bound for it is exactly a map past which the least code consistent with the record computes the world from some input on; it separates, and a strong bound raised above the code is one. A learner that differs from Harrington's only at records of a narrow kind has bounds worth exactly a totality table, given $\mathrm{MP}_{\mathrm{bin}}$, and the bounds of the trap-first learner, which runs Harrington's rule on another list, carry a totality decision. Whether Harrington's own bounds, in the standard order of codes, carry one is open.

29.11 How many revisions a catalogue forces

The catalogue of Theorem 29.48(4) needs one change of mind because every record of the zero world is also a record of a world with a late entry 1. This section turns that observation into an exact count for every catalogue, and prices one or two changes on catalogues built from the halting problem.

Definition 29.57 (Place learners and derived catalogues)**.** Let $W = (W_0, W_1, \ldots)$ be a catalogue of binary worlds, and write $W_i \neq W_{i'}$ when the two worlds differ at some input.

1. A *place learner* assigns a place $M(\sigma)$ to every finite binary record. It *identifies* ω when some j with $W_j = \omega$ and some N give $M(\omega{\restriction}n) = j$ for every $n \geq N$. It *changes its mind at* m on ω when $M(\omega{\restriction}(m+1)) \neq M(\omega{\restriction}m)$. It *learns* W *with at most* k *changes* when it identifies every W_i and changes its mind at no more than k stages on each.

2. The *derived catalogues* are sets of places: D^0 holds every place, and $i \in D^{j+1}$ when $i \in D^j$ and, for every N, some $i' \in D^j$ with $W_{i'} \neq W_i$ agrees with W_i below N.

3. The *halting catalogue* H uses a pairing $\langle a, b\rangle$ of numbers. Place $\langle e, 0\rangle$ holds the *header world* A_e, which is 1 below e and 0 from e on. Place $\langle e, s+1\rangle$ holds A_e with one more entry 1, at $e + s + 1$, when $\varphi_e(0)$ halts in exactly $s+1$ steps, and A_e otherwise.

By induction on j, places with the same world lie in the same derived catalogues, and $D^{j+1} \subseteq D^j$. For the catalogue of the opening paragraph, D^1 holds the places of the zero world, since the world with its only 1 at N agrees with it below N; a world with its only 1 at n is not in D^1, since only that world agrees with it below $n+1$; and D^2 is empty. The halting catalogue is the one of Theorem 29.54(3).

Theorem 29.58 (Revision bounds are read off derived catalogues)**.** *Let W be a catalogue and k a number.*

1. *Some place learner learns W with at most k changes if and only if D^{k+1} is empty and all places in D^k hold one world.*

2. *Let each W_i be a world of the layer of Section 29.9, given with a program P_i computing it. Then some learner on the layer EX-identifies every W_i while changing its mind at most k times on each if and only if D^{k+1} is empty and all places in D^k hold one world.*

3. *The halting catalogue has D^1 empty. So some place learner learns it with at most one change, and none with no change.*

In parts 1 and 2, the implication from a learner to the condition on the derived catalogues is proved with no choice principle; the converse, which builds the learner, and part 3 are proved in classical mathematics.

Conditions and scope.

Base and premises

Any catalogue of binary worlds; learners of any kind that name places, or, in part 2, programs of the layer, read as in Section 29.9.

Conclusion

The fewest changes of mind that suffice for a catalogue, when some number does, is the least k with D^{k+1} empty and D^k holding one world: how deeply the candidate laws accumulate on one another.

Not claimed

That the learner built in the proof is computable; membership in a derived catalogue is not decided by a record.

Proof. The adversary. Let M identify every W_i. Suppose $a \in D^{j+1}$ and M changes its mind on W_a at the stages of a finite set S, all below n. Let M name t on W_a from stage N on, and take $a' \in D^j$ with $W_{a'} \neq W_a$ that agrees with W_a below $N + n$. On $W_{a'}$ the records up to length $N+n$ are those of W_a, so M changes its mind at the stages of S and names t at stage $N+n$. From some stage it names a place t' holding $W_{a'}$, and $t' \neq t$ because the two places hold different worlds. So M changes its mind at some stage at or after $N + n$, a stage not in S. One step thus passes from a place of D^{j+1} to a place of D^j and adds a change; j steps from a place of D^j add j changes.

From a learner to the condition. Let M learn W with at most k changes. A place in D^{k+1}, with S empty, would give $k + 1$ changes on one world. If two places i, i' of D^k held worlds that differ at some input, the place M names on the empty record would hold at most one of them, say not W_i; then M changes its mind on W_i before it settles, and k steps from i give $k + 1$ changes. Each case split compares numbers or bits. For part 2 the same argument runs with programs in place of places: programs of the layer are finite trees, so whether two are equal is decided, and two programs computing different worlds differ.

From the condition to a learner. At a record σ, let its *level* be the largest $j \leq k$ for which some place of D^j holds a world extending σ, and call σ *unambiguous* when all such places of that level hold one world. The learner names the least such place at an unambiguous record, and otherwise what it named at the record one shorter (place 0 at an ambiguous empty record); in part 2 it names that place's program. Along a world of W the level never rises, since a world extending a record extends its shorter records; while the level stays, an unambiguous record stays unambiguous with the same least place. So the potential, the level plus 1 at an ambiguous record, never rises, and it falls at every change of mind. At the empty record it is at most k: the level is at most k, and at level k the record is unambiguous because D^k holds one world. Hence at most k changes.

For identification, let r be the largest j with $i \in D^j$; it is at most k, since D^{k+1} is empty. As $i \notin D^{r+1}$, some N makes every place of D^r whose world agrees with W_i below N hold W_i. From stage N on, a place of a higher level extending the record would lie in D^r, hold W_i, and put i into D^{r+1}; so the level is r, the record is unambiguous, and the learner names the least place of D^r holding W_i, from then on.

(3) A world of H agreeing with A_e below $e+1$ has header e. Of these, at most one differs from A_e, the one with its extra 1 at $e+s+1$ for the s with $\varphi_e(0)$ halting in exactly $s+1$ steps, and agreement below $e+s+2$ excludes it. That world, in turn, is the only world of H that agrees with it below $e+s+2$. So no place is in D^1. Part 1 with $k=1$ gives the learner; with $k=0$ it would need all places to hold one world, but $A_0 \neq A_1$. □

For a computable learner on the halting catalogue the count is two, and a halting table buys back one.

Theorem 29.59 (One change on the halting catalogue costs a halting decision)**.** *Work in classical mathematics.*

1. *A computable place learner learns the halting catalogue with at most two changes.*
2. *A halting table yields a place learner that learns it with at most one change.*
3. *No computable place learner learns it with at most one change.*
4. *Let a place learner learn the halting catalogue with at most one change, let the place it names at the empty record not hold A_e, and let it name a place holding A_e at the record of A_e of length n. Then $\varphi_e(0)$ is defined if and only if it halts within fewer than n steps.*
5. *Let W be a catalogue with $(i,x) \mapsto W_i(x)$ computable. If some place learner learns W with at most one change, then there are computable maps Q, giving a finite list of codes for each record, and L such that, for every halting table d, the learner that names $L(\sigma, b)$, where b lists the answers $d(c,0)$ for the codes c in $Q(\sigma)$, learns W with at most one change.*

The counting in parts 4 and 5 uses no choice principle.

Proof. (1) Name place 0 until the record shows an entry 0. With the first 0 at e, name $\langle e, 0\rangle$; once an entry 1 appears at some $p > e$, name $\langle e, p-e\rangle$. An extra 1 at p occurs only in the world of $\langle e, p-e\rangle$, so the last place named holds the world, and the learner names at most three places in turn.

(2) With the table, find for each e the number k_e of steps in which $\varphi_e(0)$ halts exactly, and $k_e = 0$ if it does not halt. Name place 0 until the record shows its first 0, at e, and its entry at $e+k_e$; then name $\langle e, k_e\rangle$ if that entry is 1 and $\langle e, 0\rangle$ if it is 0. An extra 1 can sit only at $e+k_e$, so the place is right, and it is named after at most one change.

(4) If $\varphi_e(0)$ halts within fewer than n steps, it halts. Conversely, let it halt in exactly $k \geq n$ steps. The world of $\langle e, k\rangle$ has its extra 1 at $e+k \geq n$, so its record of length n is that of A_e. On it the learner has left its opening place by stage n, which is one change, and it names a place holding A_e; it must end at a place holding the world of $\langle e, k\rangle$, a second change, which the bound forbids. So $\varphi_e(0)$ halts in fewer than n steps.

(3) Let the learner be computable and j_0 its place at the empty record; at most one header world is held by j_0. Whether a place holds A_e is decided by running $\varphi_{e'}(0)$ for at most $s+1$ steps, for place $\langle e', s+1\rangle$. For every other e, part 4 and the convergence of the learner on A_e give: $\varphi_e(0)$ diverges if and only if some n has the learner naming a place holding A_e at the

record of length n while $\varphi_e(0)$ has not halted within n steps. That makes divergence on input 0 a property that a computable search confirms, for all codes but one; with halting also so confirmed, halting would be decided, against Turing's theorem [21].

(5) Let M be the given learner and j_0 the place it names at the empty record. Every place i with $W_i \neq W_{j_0}$ has a *locking length* m: every place agreeing with W_i below m holds W_i. Indeed, let M name t on W_i from stage m on; $t \neq j_0$, since $W_t = W_i \neq W_{j_0}$. A place agreeing with W_i below m has the same record of length m, so M names t there after one change, and it cannot change again; it ends at t, and the world is $W_t = W_i$. For a pair $\langle m, i\rangle$, the search for a place agreeing with W_i below m and differing from it somewhere is a program found from m and i; it fails to halt exactly when m locks W_i. Let $Q(\sigma)$ list these programs for the pairs up to $|\sigma|$, and let L name i for the least pair $\langle m, i\rangle \leq |\sigma|$ such that W_i agrees with σ below m and the answer is "does not halt", and j_0 if there is none. A pair that qualifies keeps qualifying as the record grows, and later pairs are larger, so the learner names j_0 and then one place. If a pair $\langle m, i\rangle$ qualifies on W_{i_0}, then W_{i_0} agrees with W_i below m, so $W_{i_0} = W_i$. If none ever qualifies, then $W_{i_0} = W_{j_0}$, since otherwise W_{i_0} has a locking length m and the pair $\langle m, i_0\rangle$ qualifies. □

A second level of the same construction moves the price from halting to totality.

Theorem 29.60 (Running times and revision counts)**.** *Work in classical mathematics. Let $L_{e,k}$ be A_e with one more entry* 1*, at $e+1+k$, and $P_{e,k,s}$ be $L_{e,k}$ with one more entry* 1*, at $e+2+k+s$. The catalogue C holds at place $\langle e, 0\rangle$ the world A_e; at $\langle e, \langle k, 0\rangle+1\rangle$ the world $L_{e,k}$; and at $\langle e, \langle k, s+1\rangle+1\rangle$ the world $P_{e,k,s}$ when, within s steps, φ_e has halted at every input below k and not at k, and $L_{e,k}$ otherwise.*

1. *A computable place learner learns C with at most three changes.*
2. *Some place learner learns C with at most two changes, and none with at most one.*
3. *Let a place learner learn C with at most two changes, let the place it names at the empty record not hold A_e, and let it name a place holding A_e at the record of A_e of length n. Then φ_e is total if and only if it is defined at every input below n.*

The counting in part 3 uses no choice principle.

Proof. (1) Name place 0 until an entry 0 appears, at e; then $\langle e, 0\rangle$; once an entry 1 appears at some $p_1 > e$, the place $\langle e, \langle p_1 - e - 1, 0\rangle + 1\rangle$ of a line; once a further 1 appears at some $p_2 > p_1$, the place $\langle e, \langle p_1 - e - 1, p_2 - p_1\rangle + 1\rangle$ of a point. A second 1 past the header occurs only in a point, and a point is a member only when its condition holds, so the last place named holds the world; at most four places are named in turn.

(2) Every header world is in D^1, since $L_{e,k}$ agrees with A_e below $e + 1 + k$. No point is: a world agreeing with $P_{e,k,s}$ below $e + 3 + k + s$ is that point. A line $L_{e,k}$ is in D^1 only when its point condition holds for arbitrarily large s, that is, when φ_e halts at every input below k and never at k; this happens for at most one k, since for $k < k'$ the condition for k' makes φ_e halt at k. So each header has at most one line in D^1. Then D^2 is empty. Among the worlds held in D^1, a line $L_{e,k}$ is the only one agreeing with it below $e + 2 + k$, and A_e is the only one agreeing with A_e below $e + 2 + k$, for its line $L_{e,k}$ in D^1, or below $e + 1$ if it has none. Theorem 29.58(1)

with $k = 2$ gives the learner; with $k = 1$ it would need the header worlds $A_0 \neq A_1$ of D^1 to coincide.

(3) If φ_e is total, it is defined below n. Conversely, let it be defined below n but not everywhere, and let $y_0 \geq n$ be the least input at which it is undefined. For all large s, within s steps φ_e has halted below y_0 and not at y_0. The line $L = L_{e,y_0}$ agrees with A_e below n, so on L the learner has left its opening place by stage n and names a place holding A_e; from some stage $N \geq n$ on it names a place holding $L \neq A_e$, so it changes its mind a second time in between. Take s large with $e + 2 + y_0 + s \geq N$. The point $P_{e,y_0,s}$ is a member that agrees with L below N, so on it the learner makes both changes and then a third, to a place holding the point. □

Only $\langle e, 0\rangle$ holds A_e, and the opening place holds at most one header world; so a learner of C with two changes, with a halting table, decides by a finite check which programs halt on every input. For the catalogue $C_2(\mathbf{t}_0)$ of Section 29.12, which is C with each input read as a coded list, a halting table alone does not buy two changes, even for learners that choose each question from the answers so far; the answers to "does some row of φ_e never halt?" do (Theorem 29.68). The read-offs in Theorems 29.59(4) and 29.60(3) are two cases of one fact about every catalogue.

Theorem 29.61 (Committing early certifies the deep laws)**.** *Let W be a catalogue and k a number, and let a place learner learn W with at most $k + 1$ changes. Let the place it names at the empty record not hold W_a, and let it name a place holding W_a at the record of W_a of length n. Then every $b \in D^k$ with $W_b \neq W_a$ has $W_b(x) \neq W_a(x)$ for some $x < n$. The theorem is free.*

Proof. Comparing the entries below n one at a time, W_b and W_a differ below n or agree below n. If they agree, then on W_b, whose records up to length n are those of W_a, the learner leaves its opening place before stage n and names a place holding $W_a \neq W_b$ at stage n; it ends at a place holding W_b, a second change. The adversary in the proof of Theorem 29.58, applied k times from $b \in D^k$, then gives a world with $k + 2$ changes, more than $k + 1$. □

With $k = 2$, a catalogue one level deeper than C reads off a question one quantifier higher.

Theorem 29.62 (Three revisions on running times read off a third-level question)**.** *Work in classical mathematics. Row y of φ_e is* total *when φ_e halts at every input $\langle y, j\rangle$. With A_e and $L_{e,y}$ as in Theorem 29.60, let $S_{e,y,m}$ be $L_{e,y}$ with one more entry* 1*, at $e+2+y+m$, and $P_{e,y,m,s}$ be $S_{e,y,m}$ with one more entry* 1*, at $e+3+y+m+s$. The catalogue C_3 holds at $\langle e, 0\rangle$ the world A_e; at $\langle e, \langle y, 0\rangle + 1\rangle$ the world $L_{e,y}$; at $\langle e, \langle y, \langle m, 0\rangle + 1\rangle + 1\rangle$ the world $S_{e,y,m}$; and at $\langle e, \langle y, \langle m, s + 1\rangle + 1\rangle + 1\rangle$ the world $P_{e,y,m,s}$ when the condition $c(e, y, m, s)$ holds, and $S_{e,y,m}$ otherwise: within s steps, φ_e has halted at every input $\langle y, j\rangle$ with $j \leq m$, and for every $y' < y$ some $j \leq m$ has φ_e not halted at $\langle y', j\rangle$.*

1. *A computable place learner learns C_3 with at most four changes.*
2. *Some place learner learns C_3 with at most three changes, and none with at most two.*
3. *If row y of φ_e is total and no row $y' < y$ is, then $L_{e,y} \in D^2$.*
4. *Let a place learner learn C_3 with at most three changes, let the place it names at the empty record not hold A_e, and let it name a place holding A_e at the record of A_e of length n. Then no row of φ_e is total if and only if no row $y < n$ is.*

Conditions and scope.

Base and premises

The catalogue C_3, built from the running times of programs on rows of inputs; place learners of any kind, computable in part 1; classical mathematics. Reading computable learners as native ones is the identification of Section 29.7.

Conclusion

Four changes computably; three, not two, for some learner; a three-change learner's commitment to A_e bounds the search for a total row.

Not claimed here

That "no row of φ_e is total" is complete at the third level for the coding of this theorem. For the coded rows of Section 29.12 the decided question is complete at every level (Theorem 29.66), and no learner computable from a table of the row answers, listed or adaptive, manages three changes on $C_3(\mathbf{t}_1)$ (Theorem 29.68).

Proof. (1) Name place 0 until an entry 0 appears, at e; then $\langle e, 0\rangle$; at an entry 1 at some $p_1 > e$, the place of $L_{e,y}$ with $y = p_1 - e - 1$; at a further 1 at $p_2 > p_1$, the place of $S_{e,y,m}$ with $m = p_2 - p_1 - 1$; at a further 1 at $p_3 > p_2$, the place of $P_{e,y,m,s}$ with $s = p_3 - p_2 - 1$. A third 1 past the header occurs only in a point $P_{e,y,m,s}$, a member only when $c(e, y, m, s)$ holds, so the last place named holds the world, after at most four changes.

(2) Every world is 1 below its header e and 0 at e. Header worlds and lines lie in D^1: $L_{e,N}$ agrees with A_e below N, and $S_{e,y,N}$ with $L_{e,y}$. No point does, since a world agreeing with $P_{e,y,m,s}$ below $e + 4 + y + m + s$ is that point. The members agreeing with $S_{e,y,m}$ below $e + 3 + y + m$ are $S_{e,y,m}$ and its points, so $S_{e,y,m} \in D^1$ only when $c(e, y, m, s)$ holds for arbitrarily large s. Hence every A_e lies in D^2; no $S_{e,y,m}$ does; and $L_{e,y} \in D^2$ only when, for arbitrarily large m, $c(e, y, m, s)$ holds for arbitrarily large s. Two rows $y_1 < y_2$ cannot both have this property: take such an m for y_2, then $m_1 \geq m$ and s_1 with $c(e, y_1, m_1, s_1)$, then $s \geq s_1$ with $c(e, y_2, m, s)$; the last gives $j \leq m$ with φ_e not halted at $\langle y_1, j\rangle$ within $s \geq s_1$ steps, against the second. So D^2 holds the header worlds and at most one line per header. A line $L_{e,y}$ in D^2 is the only world of D^2 agreeing with it below $e + 2 + y$, and A_e the only one agreeing with A_e below $e + 2 + y$, for its line $L_{e,y}$ in D^2, or below $e + 1$. So D^3 is empty, and Theorem 29.58(1) with $k = 3$ gives the learner; with $k = 2$ it would need $A_0 \neq A_1$, both in D^2, to coincide.

(3) $L_{e,y} \in D^1$. Each row $y' < y$ has an input $\langle y', d_{y'}\rangle$ at which φ_e never halts; let M_0 exceed every $d_{y'}$. For $m \geq M_0$, φ_e halts at the inputs $\langle y, j\rangle$ with $j \leq m$ within some number of steps, and from there on $c(e, y, m, s)$ holds; so the points $P_{e,y,m,s}$ put $S_{e,y,m}$ in D^1, and S_{e,y,M_0+N} differs from $L_{e,y}$ and agrees with it below N.

(4) If no row is total, none below n is. Conversely, let some row be total and y_0 the least. By part 3, $L_{e,y_0} \in D^2$, and it differs from A_e only at $e + 1 + y_0$. Theorem 29.61 with $k = 2$ gives $e + 1 + y_0 < n$, so row $y_0 < n$ is total. □

"No row of φ_e is total" has three alternating quantifiers: every y has a j at which, for every s, φ_e has not halted within s steps. As only $\langle e, 0\rangle$ holds A_e, a learner of C_3 with three changes and a table of the total rows decide it by a finite check, for every e but one.

Definition 29.63 (Sums and nested sums)**.** A *test* b assigns 0 or 1 to every finite list of numbers; b_u is b with u put in front of each list, and $[u]$ is the list with the single entry u. The *shift* $\mathrm{sh}_u w$ of a world w is 0 below u, 1 at u, and $w(z-u-1)$ at $z>u$. Given catalogues K_u and a set E of enabled parts, the *sum* holds the zero world, its *head*, at place 0, and at place $\langle u,i\rangle+1$ the shift sh_u of the world at place i of K_u if $u\in E$, and the zero world otherwise. The *nested sums*: $G_0(b)$ holds only the zero world, and $G_{j+1}(b)$ is the sum of the $K_u=G_j(b_u)$, with every part enabled when $j\geq 1$ and the parts u with $b([u])=1$ when $j=0$. $\mathrm{IO}^0(b)$ says that b gives 1 on the empty list, and $\mathrm{IO}^{j+1}(b)$ that for every N some $u\geq N$ has $\mathrm{IO}^j(b_u)$. For $k\geq 2$ and a family $\mathbf{b}=(b^{(x)})_x$ of tests, the catalogue $C_k(\mathbf{b})$ holds at place $\langle x,i\rangle$ the world that is 1 below x, 0 at x, and from $x+1$ on the world at place i of $G_k(b^{(x)})$.

So $\mathrm{IO}^1(b)$ says $b([u])=1$ for arbitrarily large u, IO^j has $2j$ alternating quantifiers, and places $\langle x,0\rangle$ and $\langle x,\langle y,0\rangle+1\rangle$ of $C_k(\mathbf{b})$ hold A_x and $L_{x,y}$.

Theorem 29.64 (Catalogues of every depth)**.** *Let b be a test and $\mathbf{b}=(b^{(x)})_x$ a family of tests.*

1. *For every j, no place of $G_j(b)$ lies in D^{j+1}, and every place of $G_j(b)$ in D^j holds the zero world. The head of $G_{j+1}(b)$ lies in D^j, and it lies in D^{j+1} if and only if $\mathrm{IO}^{j+1}(b)$.*
2. *Let $k\geq 2$. In $C_k(\mathbf{b})$, D^{k+1} is empty, so some place learner learns it with at most $k+1$ changes; every A_x lies in D^{k-1}, so none learns it with at most $k-1$; and $L_{x,y}\in D^{k-1}$ if and only if $\mathrm{IO}^{k-1}(b^{(x)}_y)$. With at most one such y for each x, some place learner manages k changes.*
3. *Let a place learner learn $C_k(\mathbf{b})$ with at most k changes, let the place it names at the empty record not hold A_x, and let it name a place holding A_x at the record of A_x of length n. Then no y has $\mathrm{IO}^{k-1}(b^{(x)}_y)$ if and only if no $y<n$ does.*

Part 1 is free; parts 2 and 3 are proved in classical mathematics.

Proof. (1) In a sum, a world of an enabled part u is 0 below u and 1 at u; the zero world is 0 at u, and a world of part $u'\neq u$ differs from it at the smaller of u,u'. So a member agreeing with a world of part u up to u lies in part u, and two worlds of part u agree below $u+1+N$, or differ, exactly when the worlds of K_u they shift do. By induction on r, for enabled u, place $\langle u,i\rangle+1$ lies in D^r of the sum exactly when place i lies in $D^r(K_u)$ (part transfer). A member agreeing with the zero world below N and differing from it lies in an enabled part $u\geq N$, and every such member agrees with it below u. By part transfer, the head lies in D^{r+1} exactly when it lies in D^r and for every N some enabled $u\geq N$ has a place of $D^r(K_u)$.

Now induct on j. $G_0(b)$ holds one world. In $G_{j+1}(b)$, a place of an enabled part in D^{j+1} would give by part transfer a place of $G_j(b_u)$ in D^{j+1}; so the places in D^{j+1} hold the zero world, and none lies in D^{j+2}. Next, the head of $G_{j+1}(b)$ lies in D^r for $r\leq j$, by induction on j and then r: for $r+1\leq j$ every part is enabled, and the head of $G_j(b_u)$ lies in D^r, for every u. Last, by the head criterion, the head of $G_{j+1}(b)$ lies in D^{j+1} exactly when for every N some enabled $u\geq N$ has a place of $G_j(b_u)$ in D^j. For $j=0$ this says $b([u])=1$ for arbitrarily large u, which is $\mathrm{IO}^1(b)$. For $j\geq 1$ such a place holds the zero world, as the head does, so it may be taken to be the head, and by induction the condition reads $\mathrm{IO}^j(b_u)$ for arbitrarily large u, which is $\mathrm{IO}^{j+1}(b)$.

(2) Worlds of different columns differ at the smaller index, so columns transfer like parts: place $\langle x, i\rangle$ lies in $D^r(C_k(\mathbf{b}))$ exactly when place i lies in $D^r(G_k(b^{(x)}))$. Part 1 then gives D^{k+1} empty and $A_x \in D^{k-1}$; and as part y of $G_k(b^{(x)})$ is enabled, $L_{x,y} \in D^{k-1}$ exactly when the head of $G_{k-1}(b_y^{(x)})$ lies in D^{k-1}, which is $\mathrm{IO}^{k-1}(b_y^{(x)})$. Theorem 29.58(1) gives the learner with $k+1$ changes, and excludes $k-1$ as $A_0 \neq A_1$ both lie in D^{k-1}. Under the uniqueness condition D^k is empty: a place of D^k in column x holds A_x (part 1), so $A_x \in D^k$, which is $\mathrm{IO}^k(b^{(x)})$ and gives two values of y. Theorem 29.58(1) with k gives the learner. (3) If $\mathrm{IO}^{k-1}(b_y^{(x)})$, then $L_{x,y} \in D^{k-1}$ differs from A_x only at $x+1+y$, and Theorem 29.61 with $k-1$ gives $y < n$. □

So for every $r \geq 1$, r changes are too few on $C_{r+1}(\mathbf{b})$ and $r+2$ suffice. Section 29.12 answers the two questions this raises: the tests built from the rows of programs make the question of part 3 complete at every level, and one computable learner manages $k+1$ changes on every $C_k(\mathbf{b})$ (Theorems 29.66 and 29.67).

So, in this half of the chapter: identifying one of two laws costs $\mathrm{MP}_{\mathrm{bin}}$ or LPO; learning a layer whose programs always halt costs LPO for its own sequences; a method that must survive programs that run forever costs Σ_2-LPO, which also prices what verdicts revised without a bound confirm; a guarantee of when a learner is right costs a decision table: for some learner of every computable law exactly a halting table, for the trap-first learner a decision of which programs halt on every input. The changes of mind a catalogue forces on learners of any kind are read off its derived catalogues; computable learners can need more; and a learner held to few changes commits only after every deep rival has differed.

29.12 Catalogues at every level of the arithmetical hierarchy

Theorem 29.64 built, for every family $\mathbf{b}$ of tests, a catalogue $C_k(\mathbf{b})$ that needs at least k changes of mind and admits $k+1$, and read off from every learner with k changes the answer to a question with $2k-2$ alternating quantifiers. Three questions were left: which tests make that question as hard as its quantifiers allow, whether a computable learner manages $k+1$ changes, and what information a learner needs for k. This section answers all three for the tests built from programs. The answer is exact: a learner computable from an oracle joined with the row table manages the least number of changes exactly when the joined oracle computes one decided question of the arithmetical hierarchy. All results of the section are proved in classical mathematics, as in Section 29.11; programs are those of Section 29.7, with φ_e the function computed by program e.

Definition 29.65 (Row questions, decided questions and their tests)**.** Code finite lists of numbers by $\ulcorner[\,]\urcorner = 0$ and $\ulcorner a :: l\urcorner = \langle a, \ulcorner l\urcorner\rangle + 1$. For a program e, a list l and $n \geq 0$ put

$$R_0(e,l) \;:\; \varphi_e(\ulcorner l\urcorner) \text{ never halts}, \qquad R_{n+1}(e,l) \;:\; \forall m\, \neg R_n(e, m :: l).$$

The *row question* $R_n(e,y)$ is $R_n(e,[y])$, a Π_{n+1} question about row y of φ_e; $R_1(e,y)$ says that φ_e halts at every input $\ulcorner[m,y]\urcorner$. The *decided question* at level n is $\mathrm{Dc}_n(e) \;:\; \forall y \neg R_n(e,y)$, a Π_{n+2} question, and $U_n(e)$ is its negation, "some row of φ_e satisfies R_n". The bounded form $R_n(e,l;u_1,\dots,u_{n+1})$ replaces each quantifier, from the outside in, by one bounded by u_i, the

innermost bound being a number of steps; it is decided by running programs. The *test* $t_n(e)$ holds of a list that begins with $y, u_1, \dots, u_{n+1}$ when row y satisfies the bounded form with these bounds and no row $y' < y$ does, and fails on every list with fewer than $n+2$ entries. The catalogue $C_{n+2}(\mathbf{t}_n)$ is the catalogue of Definition 29.63 for the family $\mathbf{t}_n = (t_n(x))_x$.

Row questions are complete, at every level

Theorem 29.66 (The row questions and the hierarchy)**.**

1. Limits. *For all n, e and y, $\mathrm{IO}^{n+1}(t_n(e)_y)$ holds exactly when y is the first row of φ_e with $R_n(e, y)$.*

2. Completeness. *Call a question $P(x)$ Σ_{k+1}* in normal form *when it reads*

$$\exists m_1 \, \neg \exists m_2 \, \neg \cdots \exists s \; R(x, \dots),$$

with $k+1$ quantifiers over a computable Boolean matrix R. For every n and every computable matrix R there is a computable f with $\neg R_n(f(x), y)$ exactly when the Σ_{n+1} normal form of R holds at (x, y). Hence every Π_{n+2} question in normal form is Dc_n at a computable program.

3. Normal forms. *At every level the questions in normal form contain the decidable ones and are closed under substitution of a computable map, conjunction, disjunction, an existential quantifier, a bounded universal quantifier with a computable bound, and passage to the next level; and a Σ_{k+2} question is exactly $\exists m \, \neg Q(\langle x, m \rangle)$ for a Σ_{k+1} question Q.*

4. Every formula. *Let an arithmetical formula be built from computable Boolean atoms by negation, conjunction, disjunction and the two quantifiers, and classified by its shape as Σ_{k+1} or Π_{k+1}. A Σ_{k+1} formula, and the negation of a Π_{k+1} formula, are Σ_{k+1} in normal form; every Π_{n+2} formula is $\mathrm{Dc}_n(f(x))$ for a computable f; and Dc_n is itself a Π_{n+2} formula with computable atoms.*

5. The hierarchy. *There is a computable h such that program $h(x)$ runs program x on input x and then runs the program it returns. For every n and every such computable h, the question $\mathrm{Dc}_n(h(x))$ is not Σ_{n+2} in normal form. Similarly, "program x halts on every input $\langle x, m \rangle$" is not Σ_2 in normal form.*

6. Neither enumerable. *For every n, neither the set of e with $\mathrm{Dc}_n(e)$ nor its complement is computably enumerable.*

Proof. (1) Induct on n. Running longer never undoes a halt, so "not halted within s steps" is antitone in s, and it holds for arbitrarily large s exactly when the program never halts: this is the case $n = 0$ for one row. At each level the bounded form of row y is antitone in the outer bound, and "no row $y' < y$ satisfies the bounded form" is monotone in it. For such a pair the conjunction $A(u) \wedge B(u)$ holds for arbitrarily large u exactly when A holds for every u and B for some u. Applied level by level, this turns "infinitely often, $n+1$ times over" into the unbounded alternating question for row y and its failure for every earlier row.

(2) Let program $f(x)$, on input $\ulcorner l \urcorner$, search for an s with $R(x, s :: l)$ true and halt when it finds one; f is computable by the parameter theorem [19]. Then $\neg R_0(f(x), l)$ holds exactly when

some s has $R(x, s{:}:l)$, and by induction on n, $\neg R_n(f(x), l)$ holds exactly when the Σ_{n+1} normal form of R holds at x and l: both sides add the alternation $\exists m\,\neg$ in front. For the last sentence, a Π_{n+2} question in normal form is the negation of $\exists y \neg \Sigma_{n+1}$, which is $\neg\exists y\, R_n(f(x), y)$, that is, $\mathrm{Dc}_n(f(x))$.

(3) and (4) are induction on the level and on the formula. Conjunction and disjunction interleave the witnesses of the two normal forms by pairing, an existential is absorbed into the outermost one by pairing, and a bounded universal is moved inside by collecting the finitely many witnesses into one number, which the computable bound permits. Each step produces a computable matrix from the given ones.

(5) The program $h(x)$ exists by composition of programs. Suppose $\mathrm{Dc}_n(h(x))$ held exactly when a Σ_{n+2} normal form $S(x)$ does. By (2) there is a computable g with $\mathrm{Dc}_n(g(x))$ exactly when $\neg S(x)$. By the recursion theorem [19] some program c returns $g(c)$ on input c. Then $h(c)$ runs $g(c)$, so it computes the same function, and $\mathrm{Dc}_n(h(c))$ holds exactly when $\mathrm{Dc}_n(g(c))$ does, that is, exactly when $\neg S(c)$, that is, exactly when $\neg\mathrm{Dc}_n(h(c))$. That is impossible: Kleene's diagonal argument, at every level. The last sentence of (5) is the same argument at $n = 0$ with another program: if "program x halts on every input $\langle x, m\rangle$" held exactly when a Σ_2 normal form $S(x)$ does, take g from (2) with $\mathrm{Dc}_0(g(x))$ exactly when $\neg S(x)$, and by the recursion theorem a program c that on every input $\langle z, m\rangle$ runs $g(c)$ on $\ulcorner[m]\urcorner$. Then c halts on every input $\langle c, m\rangle$ exactly when $g(c)$ halts on every $\ulcorner[m]\urcorner$, that is, exactly when $\mathrm{Dc}_0(g(c))$, that is, exactly when $\neg S(c)$.

(6) A computably enumerable set is Σ_1 in normal form, and the complement of one is Π_1, which is Σ_2 in normal form; by (3) both are Σ_{n+2} in normal form. So if the set of e with $\mathrm{Dc}_n(e)$, or its complement, were enumerable, Dc_n would be Σ_{n+2} in normal form, and so would $\mathrm{Dc}_n(h(x))$ by substitution, against (5). □

The catalogue of each level

Theorem 29.67 (The construction at every level)**.** *Let $n \geq 0$.*

1. *Some place learner learns $C_{n+2}(\mathbf{t}_n)$ with at most $n + 2$ changes, and none with at most $n + 1$.*
2. *Let a place learner learn $C_{n+2}(\mathbf{t}_n)$ with at most $n + 2$ changes, let the place it names at the empty record not hold A_e, and let it name a place holding A_e at the record of A_e of length N. Then $\mathrm{Dc}_n(e)$ holds exactly when no row $y < N$ of φ_e satisfies R_n.*
3. One computable learner for every catalogue. *Let* lrn *name place* 0 *until the record shows an entry* 0*, and from then on the place $\langle x, \ulcorner g\urcorner\rangle$, where x is the position of the first* 0 *and g lists the gaps between the entries* 1 *seen after it. Then* lrn *is computable, and for every k and every family $\mathbf{b}$ of tests it learns every member of $C_k(\mathbf{b})$ with at most $k + 1$ changes.*

Proof. (1) and (2). By Theorem 29.66(1), at most one y has $\mathrm{IO}^{n+1}(t_n(x)_y)$ for each x: the first row with R_n. So Theorem 29.64(2) gives a learner with $n + 2$ changes and excludes $n + 1$, and Theorem 29.64(3) gives the read-off, since some y has $\mathrm{IO}^{n+1}(t_n(e)_y)$ exactly when some row satisfies R_n.

(3) A world of $C_k(\mathbf{b})$ is 1 below its column x, 0 at x, and from $x + 1$ on a nested shift: each level of the nested sum puts one entry 1 at the start of its part, so the world after the header

is determined by the gaps between its entries 1, and there are at most k of them. The place $\langle x, \ulcorner g \urcorner \rangle$ is a place of that world once all its entries 1 have appeared. Along a world the learner moves from place 0 once, at the header, and once at each further entry 1: at most $k+1$ changes. The rule reads a finite record, so it is computable. □

Part 3 settles the second question left after Theorem 29.64: computable learners do manage $k+1$ changes, on every catalogue of that kind. Part 2 says what $n+2$ changes know: a commitment to the header world A_e at stage N answers the Π_{n+2} question $\mathrm{Dc}_n(e)$ by a check of the rows below N. By Theorem 29.66(5) that question is not Σ_{n+2}, and this is what makes learners with $n+2$ changes expensive.

What a learner needs for the least number of changes

A learner *computable from an oracle O*, a map from numbers to bits, is given by one procedure that reads the record and a finite prefix of O, answers or not, answers on every long enough prefix, and never answers wrongly. A learner *asking a table* asks at each record a computable list of questions chosen from the record alone and names a place computed from the record and the answers; an *adaptive* learner chooses each question from the record and the answers so far, by an oracle machine whose run ends on every record. The *row table* at level n answers every row question $R_n(e, y)$; the *decided answers* U_n answer $U_n(e)$ for every e.

Theorem 29.68 (The least number of changes costs the decided question)**.** *Let $n \geq 0$, and let learning mean learning every member of $C_{n+2}(\mathbf{t}_n)$ with at most $n+2$ changes.*

1. Computable learners fail. *No computable place learner learns $C_2(\mathbf{t}_0)$ with at most two changes, and none computable learns $C_{n+2}(\mathbf{t}_n)$ at any level.*
2. The row table does not suffice. *No learner asking the row table at level n, and no adaptive learner with the row table, learns $C_{n+2}(\mathbf{t}_n)$.*
3. The decided answers suffice. *There are computable rules such that the learner asking the decided answers U_n learns $C_{n+2}(\mathbf{t}_n)$; it suffices to give it the row table and, at each record, the one decided answer about the record's column. At level two the learner that asks U_0 at its column and the row answers below the record's length learns $C_2(\mathbf{t}_0)$.*
4. Exactly. *For every oracle T, joined with the row table, some learner computable from the joined oracle learns $C_{n+2}(\mathbf{t}_n)$ if and only if U_n is computable from it. One procedure, given one bit, the learner's first guess, a program e, the learner's places on the records of A_e and the row answers of e, decides $U_n(e)$ for every learner with $n+2$ changes.*
5. Prefixes are oracle machines. *A function is computable from an oracle through finite prefixes, as above, exactly when an oracle machine that asks one value at a time computes it.*

Conditions and scope.

Base and premises

The catalogues $C_{n+2}(\mathbf{t}_n)$ of Definition 29.65; learners computable, asking a table, adaptive, or computable from an oracle; classical mathematics.

Conclusion

Among learners computable from an oracle joined with the row table, the least number of changes, $n + 2$, is bought exactly by the joined oracles that compute the Σ_{n+2} question U_n: the decided question, not the row questions below it.

Not claimed

The same for a bare oracle, not joined with the row table (Remark 29.70).

Proof. Lower bounds. Let M learn with $n + 2$ changes. For every program e except at most one (the column of the place named at the empty record), Theorem 29.67(2) gives: $\mathrm{Dc}_n(e)$ holds exactly when there is an N at which M names a place holding A_e on the record of A_e of length N and no row $y < N$ satisfies R_n. Whether a place holds A_e is decided, since only $\langle e, 0\rangle$ holds it. If M is computable, the condition is Σ_{n+2} in normal form by Theorem 29.66(3): an existential over N of a computable check and a bounded universal over Π_{n+1} failures. If M asks the row table, or adaptively uses it, its answers are Π_{n+1} and their negations Σ_{n+1}, and writing "these answers are the true ones" into the condition keeps it Σ_{n+2}. Either way Dc_n would be Σ_{n+2} at all programs but one, hence at the composed programs $h(x)$ after a finite patch, against Theorem 29.66(5).

Upper bounds. By Theorem 29.66(1) and Theorem 29.64, column x has a line in D^{n+1} exactly when some row of φ_x satisfies R_n, and then it is the line of the first such row. The learner of Theorem 29.58(1) with $k = n + 2$ needs, at a record of column x, only this information. With the decided answer $U_n(x)$ it knows whether the line exists: if not, it names A_x at once; if so, it keeps its opening place until the row answers show the first row with R_n and the record has passed that row, and from then on it follows the worlds of the column as the learner of Theorem 29.67(3) does, with one change saved at the header. Its questions and its place are computable from the record and the answers, and the potential argument of Theorem 29.58 bounds its changes by $n + 2$.

Exactness. If U_n is computable from the oracle, the upper-bound learner is. Conversely, given a learner computable from the oracle, the lower-bound condition, now read with the oracle, decides $U_n(e)$ by a search that ends: wait until the learner names A_e on the record of A_e and the row answers below that stage are in. The single exceptional column is handled by one built-in bit, the answer at the column of the first guess; that bit is why the procedure is uniform in everything except one bit.

(5) A prefix procedure is simulated by a machine that asks for the next value whenever the procedure has not answered on the prefix so far; conversely a machine's run asks finitely many values, which a long enough prefix contains. □

A column's own rows are not enough above level two

At level two the learner of Theorem 29.68(3) asks only about the program of the record's own column. From level three on that fails, even with the decided answer at the column.

Theorem 29.69 (Own rows at levels three, four and five)**.** *For $n = 1, 2, 3$, consider learners that receive at each record the decided answer U_n about the record's column and row answers R_n about the program of that column only: either a computable list of rows chosen from the record (*listed*), or rows chosen one at a time from the record and the answers so far by a computable oracle machine whose run ends on every record (*adaptive*). No such learner, listed or adaptive, learns $C_{n+2}(\mathbf{t}_n)$ with at most $n + 2$ changes. Moreover, for $n = 1, 2, 3$, a learner that, on every record showing an entry 0 whose column's program has no row with R_n, names a value of a fixed partial computable function at that record does not learn $C_{n+2}(\mathbf{t}_n)$ with at most $n + 2$ changes.*

Proof. We give the argument at level three for listed learners; the other cases change it as described at the end. Fix the rules of a learner of the stated kind, and let M' be the computable map that names, at a record, the place the learner would name if every answer it receives were "no". By the recursion theorem [19] build a program x that, on the input $\ulcorner[m, y]\urcorner$, searches for the least stage $N_0 > x$ at which M' names the place of A_x on the record of A_x; then, if $y = N_0 - x - 1$, searches for the least stage N_1 at which M' names the place of the line $L_{x,y}$ on the record of that line; and halts exactly when $m \leq N_1 - (x + 2 + y)$. Every row of x halts on at most finitely many inputs, so no row satisfies R_1: the true decided answer at column x and every true row answer about x are "no", and on the records of column x the learner acts as M'. Two such programs that differ at the input 0 exist, and one of them is not the column of the place the learner names at the empty record; take x to be that one.

Suppose the learner learns $C_3(\mathbf{t}_1)$ with three changes. To learn A_x it names its place from some stage on, so the search for N_0 ends; to learn the line $L_{x,y}$ with $y = N_0 - x - 1$ it names that line's place from some stage on, and the search for N_1 ends. The point world of that line whose next entry 1 comes after N_1 is a member, so the learner leaves the line's place at some stage $N_2 > N_1$; and a full world of column x whose last entry 1 comes after N_2 is a member too, because row y halts below the bound and the earlier rows never halt. On that world the learner changes its mind in $[0, N_0)$, in $[N_0, N_1)$, in $[N_1, N_2)$ and after N_2: four changes.

An adaptive learner is handled in the same way: every question it asks is about a row of x, the answer is "no", and its run on all-"no" answers is a computable map M'.

At level four the program x is built one quantifier higher. On the input $\ulcorner[m_2, m_1, y]\urcorner$ it runs the two searches of level three, each bounded by m_2: for the least stage $N_0 > x$ at which M' names the place of A_x on the record of A_x, and, if $y = N_0 - x - 1$, for the least stage N_1 at which M' names the place of the line $L_{x,y}$ on the record of that line. It fails to halt exactly when both stages are found, $N_1 \leq m_2$ and $m_1 \leq N_1$, and halts otherwise; with the searches bounded, this is decided by running M'. Every row y has an m_1, any $m_1 > N_1$, or any m_1 at all when the stages do not exist, at which every input $\ulcorner[m_2, m_1, y]\urcorner$ halts; so no row satisfies R_2, the true answers about x are all "no", and on the records of column x the learner acts as M'. Suppose it learns $C_4(\mathbf{t}_2)$ with four changes. Then N_0 and N_1 exist, as at level three. The point world of the line whose second entry 1 comes at N_1 is a member, so the learner leaves the line's place for that world's place at some stage $N_2 > N_1$; the world of the next kind whose third gap is N_2 is a member, so it leaves that place at some stage $N_3 > N_2$; and the full world of column x whose fourth entry 1 comes after N_3, far enough for every row below y to have halted at the inputs the test reads, is a member too, because row y fails to halt at the inputs $\ulcorner[m_2, m_1, y]\urcorner$ with $m_1 \leq N_1 \leq m_2$. On it the learner changes its mind in $[0, N_0)$, in $[N_0, N_1)$, in $[N_1, N_2)$, in $[N_2, N_3)$ and after N_3: five changes. At level five the program is

built one quantifier higher again, with one more search and one more world, so that no row satisfies R_3. For a learner that agrees, on the columns without a row satisfying R_n, with a fixed partial computable function, take M' to be that function: the searches of x run it, and they end because the learner, which equals it there, must name the places in question. □

Learner on $C_{n+2}(\mathbf{t}_n)$	With at most $n+2$ changes
any place learner	yes; $n+1$ changes never (Theorem 29.67)
computable	no, at every level; $n+3$ changes yes (Theorem 29.67(3))
asking the row table, listed or adaptive	no, at every level
asking the decided answers U_n	yes, at every level
row table and one decided answer per record	yes, at every level
computable from an oracle joined with the row table	exactly when U_n is computable from the joined oracle
own column's rows and its decided answer	yes at $n = 0$; no at $n = 1, 2, 3$, listed or adaptive

Table 29.3. What buys the least number of changes on the catalogue of level $n+2$ (Theorems 29.67, 29.68 and 29.69). All entries are proved in classical mathematics.

Remark 29.70 (What is not settled)**.** Theorem 29.69 is proved at the levels $n = 1, 2, 3$; at $n \geq 4$ the statement for own rows is not proved here. The read-off of Theorem 29.68(4) uses one built-in bit, so the reduction from a learner to U_n is not uniform in the learner. And part (4) is proved for oracles joined with the row table. For a bare oracle O the lower-bound argument still needs the row answers below the stage at which the learner commits, so it shows only that U_n is computable from O joined with the row table; whether every bare oracle from which some learner with $n+2$ changes is computable must itself compute U_n is not settled here.

Questions to work through

1. In the history $H(\alpha)$ of Definition 29.3, let α have its first entry 1 at 3. At which stages are p_0, $\neg p_0$ and $\neg\neg p_0$ forced? Use the answer to explain why em fails at stage 0 while wem holds there for this α (Theorem 29.4).

2. Let P be the law of three binary records that gives probability $\frac{1}{3}$ to each record with exactly one entry 1. Compute P_2 and the coin mixture Q_2, and check the bound of Theorem 29.8 for $n = 3$, $m = 2$.

3. Write the first 21 entries of the golden record and list its windows of length 3. How many are there, and which window of length 2 is right special (Theorems 29.14 and 29.15)?

4. Let H consist of the four laws on the points $0, 1, 2, 3$ whose labels are 1 at exactly one point. Compute $\mathrm{ldim}(H)$, run the standard optimal learner on the target that is 1 at 3, with the points in the order $0, 1, 2, 3$, and compare with the halving learner (Theorems 29.29 and 29.31).

5. For the records $1, 4$ with unit weights, find the J-centre, the least cost, and the regret of the learner that starts at $f_0 = 1$ and follows the centre; check the identity of Theorem 29.39(4).

6. For each world ω of the layer with a single entry 1, Theorem 29.47(3) gives a learner that commits once correctly on the zero world and on ω. Show that no single learner does so for all of them at once, and that the catalogue of these worlds has no locking function (Theorems 29.47 and 29.48).

Chapter 30

Conjectures and refutations

Chapter 21 showed that, at zero holding, two golden comparisons in a row never change a record with certainty. With the reference action set to one cycle, a comparison at ratio x drives the register by $2\pi J(x)$, and a golden comparison drives $2\pi J(\varphi) = 2\pi(\varphi - \frac{3}{2}) \approx 0.742$. Two of them drive $2\pi\varphi^{-3} \approx 1.483$, short of the quarter turn $\pi/2 \approx 1.571$ that every change with certainty needs when the register does not hold. Do three?

Work it out on one channel of Recognition Science's register, a channel with eight states. Its native operations rotate three states at a time, a *triad* e_i, e_j, e_7, about the axis $e_i + e_j + e_7$. A rotation by one third of a turn is the exact cycle $e_i \mapsto e_j \mapsto e_7 \mapsto e_i$, and it moves e_i to e_j with certainty. Three golden comparisons drive ≈ 2.225, more than $2\pi/3 \approx 2.094$, and they can spend it exactly. The inequality $J(\varphi) \geq \frac{1}{9}$ says $\varphi \geq \frac{29}{18}$, and that holds because

$$\left(\frac{29}{18}\right)^2 - \frac{29}{18} = \frac{319}{324} < 1 = \varphi^2 - \varphi,$$

while $t^2 - t$ increases for $t > \frac{1}{2}$. So each golden comparison can carry a rotation by $2\pi/9$ about one triad's axis and idle for the rest of its drive, and three of them compose to the cycle. At zero holding, three golden comparisons change a record with certainty; two never do; one never does (Theorem 30.21).

The example has the shape of every result in this chapter: a question about recognition, ledgers or certain change with a definite answer, settled by a proof. Each statement below has one of four outcomes. It is *proved.* It is *refuted*: we prove its negation, give the counterexample and prove the corrected statement. It is *reduced*: we prove that it is equivalent to a sharper statement that we name exactly and cannot yet settle. Or it is *open.* Section 30.5 lists every open statement.

The chapter is organised by theme. Section 30.1 asks how well readings at golden ratios fix the calibration of a recorded phase: the answer is the Diophantine type of φ^{-3}, and adding probes leads to classical open problems of simultaneous approximation. Section 30.2 finds when two recognizers of one history agree on a common count, and what it costs in memory to tell records apart. Section 30.3 classifies the long-run rates of self-similar double-entry ledgers and decides when they keep balanced books. Section 30.4 prices a change with certainty on the native register: exactly $2\pi/3$ for one channel, $\pi/2$ per channel when a low-bit flip suffices, π for two channels at once, floors under every holding, and the phases that a returned helper can deliver.

Everything in this chapter is proved in classical mathematics, over the classical real and complex numbers. We claim no price on $\mathbb{R}_\delta$ for any of it; several statements are about counts and finite words, but we prove them classically and claim nothing more. Two readings are identifications and stay separate from what is proved: that a recorded phase equals a calibration times J of the ratio, and that the tokens below are the operations of Recognition Science's register. Two proofs end in finite computations in exact rational arithmetic. Lemma 30.23

rests on eight checks of positive definiteness whose data Table 30.2 prints, so a reader can repeat them. The exact range in Theorem 30.18 rests on an exhaustive search over 1573 cases, which its proof describes step by step but does not print; its scope box records this. Numerical exploration appears only outside the numbered results, in the text and in the open problems; we label it as evidence, and no result depends on it.

30.1 Calibrating a phase with golden probes

A recognition at ratio $x > 0$ leaves a record whose phase, relative to a reference, is $a\,J(x)$ for an unknown *calibration* $a > 0$. Recognition Science reads a as the ratio of the reference action to $\hbar$; that reading is an identification, and the theorems below use only the form $aJ(x)$. A phase is read through its cosine and sine, so it is known only modulo 2π, and the calibration is ambiguous: at the ratio φ^2, where $J(\varphi^2) = \frac{1}{2}$, the calibrations a and $a + 4\pi k$ give the same phase for every integer k. A second reading at a ratio x separates the two by $4\pi k J(x)$ modulo 2π. At $x = \varphi$ that separation is $2\pi k(\sqrt{5} - 2)$, because $2J(\varphi) = 2\varphi - 3 = \varphi^{-3} = \sqrt{5} - 2$, and it is never close to a multiple of 2π for small k. How precisely, then, do two golden readings fix the calibration?

Definition 30.1 (Readings and consistent calibrations)**.** In classical mathematics: a *quadrature reading* at a ratio $x > 0$ is a triple of reals (c, s, ε), estimates of a cosine and a sine with error ε. A phase Δ is *consistent* with it when $|\cos\Delta - c| \le \varepsilon$ and $|\sin\Delta - s| \le \varepsilon$. A calibration $a > 0$ is consistent with a reading at x when the phase $aJ(x)$ is. For a real t, $\|t\|$ is its distance to the nearest integer.

Lemma 30.2 (Two phases in one reading)**.** *If the phases α and β are both consistent with one reading of error ε, then $|\alpha - \beta - 2\pi k| \le 6\varepsilon$ for some integer k.*

Proof. Each coordinate of $e^{i\alpha} - e^{i\beta}$ is at most 2ε in absolute value, so $|e^{i\alpha} - e^{i\beta}| \le 2\sqrt{2}\,\varepsilon$. Choose k with $t = \alpha - \beta - 2\pi k \in [-\pi, \pi]$. Then $|e^{i\alpha} - e^{i\beta}| = 2|\sin(t/2)| \ge 2|t|/\pi$, because sin is concave on $[0, \pi/2]$. Hence $|t| \le \sqrt{2}\,\pi\varepsilon \le 6\varepsilon$. □

The number $\varphi^{-3} = \sqrt{5} - 2$ is a quadratic irrational. Its continued fraction is $[0; 4, 4, 4, \ldots]$, since $\sqrt{5} - 2 = 1/(4 + (\sqrt{5} - 2))$, and a real number whose partial quotients are bounded is *badly approximable*: $\|k\alpha\| \ge c/|k|$ for some $c > 0$ and every integer $k \ne 0$ [83]. Hurwitz's theorem shows that for every irrational α such a constant is at most $1/\sqrt{5}$ [82]. We need one explicit constant.

Lemma 30.3 (φ^{-3} is badly approximable)**.** *For all integers $k \ne 0$ and m,*

$$1 \le 6|k|\,\big|k(\sqrt{5} - 2) - m\big|.$$

Proof. Put $M = m + 2k$, so $k(\sqrt{5} - 2) - m = k\sqrt{5} - M$. The integer $5k^2 - M^2$ is not 0, since $\sqrt{5}$ is irrational, so $|k\sqrt{5} - M| \cdot |k\sqrt{5} + M| = |5k^2 - M^2| \ge 1$. If $|k\sqrt{5} - M| \ge 1$ the claim is clear. Otherwise $|M| < \sqrt{5}|k| + 1$, so $|k\sqrt{5} + M| < 2\sqrt{5}|k| + 1 < 6|k|$, and $|k\sqrt{5} - M| > 1/(6|k|)$. □

Theorem 30.4 (Two golden probes fix a calibration with exponent exactly 2)**.** *In classical mathematics: let $B \geq 1$, let $a, b \in (0, B]$, and let r_1 and r_2 be readings at the ratios φ and φ^2, with errors ε_1 and ε_2. If both calibrations a and b are consistent with both readings, then*

$$|a - b| \leq 12\,(1 + B)^2\,(\varepsilon_1 + \varepsilon_2).$$

Moreover, for every real p the following are equivalent:

(i) there is $C > 0$ such that $|a - b| \leq C(1 + B)^p(\varepsilon_1 + \varepsilon_2)$ for all $B \geq 1$, all $a, b \in (0, B]$ and all readings r_1 at φ and r_2 at φ^2 with which both a and b are consistent;

(ii) $p \geq 2$.

Proof. The bound. Lemma 30.2 at r_2, where the phases are $a/2$ and $b/2$, gives an integer k with $\frac{a-b}{2} = 2\pi k + e_2$ and $|e_2| \leq 6\varepsilon_2$. At r_1 the phases are $a(\sqrt{5} - 2)/2$ and $b(\sqrt{5} - 2)/2$, whose difference is $(\sqrt{5} - 2)(2\pi k + e_2)$; the lemma gives an integer m with

$$2\pi\big|k(\sqrt{5} - 2) - m\big| \leq 6\varepsilon_1 + 6(\sqrt{5} - 2)\varepsilon_2 \leq 6(\varepsilon_1 + \varepsilon_2).$$

Note $|a - b| < B$. If $\varepsilon_1 + \varepsilon_2 \geq B/12$, then $|a - b| < 12(\varepsilon_1 + \varepsilon_2)$. Otherwise $12\varepsilon_2 < B$. If $k = 0$, then $|a - b| = 2|e_2| \leq 12\varepsilon_2$. If $k \neq 0$, Lemma 30.3 gives $1 \leq 6|k| \cdot 6(\varepsilon_1 + \varepsilon_2)/(2\pi)$, so $\varepsilon_1 + \varepsilon_2 \geq \pi/(18|k|)$; and $4\pi|k| \leq |a-b| + 2|e_2| < 2B$, so $|k| < B/(2\pi)$ and $\varepsilon_1 + \varepsilon_2 > \pi^2/(9B) > 1/B$. Then $|a - b| < B < B^2(\varepsilon_1 + \varepsilon_2) \leq (1 + B)^2(\varepsilon_1 + \varepsilon_2)$. In every case the bound holds.

(ii) implies (i). Take $C = 12$; for $p \geq 2$, $(1 + B)^p \geq (1 + B)^2$.

(i) implies (ii). Let $p < 2$ and suppose C works. By Dirichlet's theorem [81], for every $N \geq 1$ there are integers $1 \leq k \leq N$ and m with $|k(\sqrt{5} - 2) - m| \leq 1/(N + 1)$, and Lemma 30.3 forces $k \geq (N + 1)/6$. Put $a = 1 + 4\pi k$, $b = 1$ and $B = a$. The phases at φ^2 differ by $2\pi k$, so the exact reading $(\cos\frac{a}{2}, \sin\frac{a}{2}, 0)$ is consistent with both. The phases at φ differ by $2\pi k(\sqrt{5}-2) = 2\pi m + \delta$ with $|\delta| \leq 2\pi/(N+1)$; the reading centred halfway, at $b(\sqrt{5}-2)/2 + \delta/2$, with error $\varepsilon_1 = |\delta|/2 \leq \pi/(N + 1) < \pi/k$ (as $k \leq N$), is consistent with both, because cos and sin are 1-Lipschitz. Then (i) gives

$$4\pi k \leq C(2 + 4\pi k)^p\,\frac{\pi}{k}, \qquad \text{so} \qquad 4k^2 \leq C\,(15k)^{\max(p,0)},$$

using $2 + 4\pi k \leq 15k$. As N grows, $k \geq (N + 1)/6$ grows without bound, and $2 - \max(p, 0) > 0$: a contradiction. □

Conditions and scope.

Base and premises

Classical real numbers; readings and consistency as in Definition 30.1; the phase form $aJ(x)$ at the two ratios φ and φ^2.

Conclusion

Two calibrations that both fit golden readings of errors $\varepsilon_1, \varepsilon_2$ differ by at most $12(1+B)^2(\varepsilon_1 + \varepsilon_2)$, and no exponent below 2 works for any constant.

Identification
That a recorded relative phase is $aJ(x)$, with a the reference action over $\hbar$, is the admissible phase law of Recognition Science, used as stated. No value of a is supplied.

Not claimed
That the golden pair is the best pair of probes; the proof uses only the Diophantine type of φ^{-3}, and other quadratic probes are not treated.

So the reading precision a calibration range B needs scales as B^{-2}: one factor of B comes from the size of the ambiguity, the other from how close $k\varphi^{-3}$ can come to an integer for $|k| \le B$. The exponent is the Diophantine exponent of φ^{-3}.

A third probe can only help, and the question is by how much. Add a reading r_3 at the ratio $\sqrt{2}$, where $J(\sqrt{2}) = \frac{3\sqrt{2}}{4} - 1$; the alias $a + 4\pi k$ now also turns by $2\pi k \cdot \frac{3\sqrt{2}}{2}$ modulo 2π. A wrong branch hides exactly as well as $k(\sqrt{5}-2)$ and $k \cdot \frac{3\sqrt{2}}{2}$ are simultaneously close to integers.

Theorem 30.5 (A third probe and simultaneous approximation)**.** *In classical mathematics, put $\alpha = \sqrt{5}-2$ and $\beta = \frac{3\sqrt{2}}{2}$, and consider three readings r_1, r_2, r_3 at the ratios φ, φ^2 and $\sqrt{2}$, with errors $\varepsilon_1, \varepsilon_2, \varepsilon_3$. For a real p, say that p* works *if some $C > 0$ gives $|a-b| \le C(1+B)^p(\varepsilon_1+\varepsilon_2+\varepsilon_3)$ for all $B \ge 1$, all $a, b \in (0, B]$ and all such readings with which both a and b are consistent.*

(a) No $p < \frac{3}{2}$ works.

(b) The following are equivalent: (i) a real p works exactly when $p \ge \frac{3}{2}$; (ii) the pair (α, β) is simultaneously badly approximable*: there is $c > 0$ with*

$$\sqrt{|k|}\,\max\big(|k\alpha - m|,\ |k\beta - n|\big) \ge c \qquad \text{for all integers } k \ne 0,\ m,\ n.$$

Proof. (a) Let $p < \frac{3}{2}$ and suppose C works. For $Q \ge 1$ the Q^2+1 points $(\{k\alpha\}, \{k\beta\})$, $0 \le k \le Q^2$, of the unit square lie in Q^2 boxes of side $1/Q$; two share a box, and their difference gives $1 \le k \le Q^2$ and integers m, n with $|k\alpha - m| \le 1/Q$ and $|k\beta - n| \le 1/Q$. For fixed k this cannot hold for all large Q, since $k\alpha \notin \mathbb{Z}$; so $k \to \infty$ as $Q \to \infty$. Put $a = 1 + 4\pi k$, $b = 1$, $B = a$. The phases differ by $2\pi k$ at φ^2, by $2\pi k\alpha$ at φ, and by $4\pi k J(\sqrt{2}) = 2\pi k\beta - 4\pi k$ at $\sqrt{2}$. As in the proof of Theorem 30.4, readings centred halfway have $\varepsilon_2 = 0$ and $\varepsilon_1, \varepsilon_3 \le \pi/Q \le \pi/\sqrt{k}$. So $4\pi k \le C(15k)^{\max(p,0)} \cdot 2\pi/\sqrt{k}$, that is $2k^{3/2-\max(p,0)} \le C\,15^{\max(p,0)}$, which fails for large k.

(b) *(ii) implies (i).* By (a) it suffices to show that $C = 12 + 1/c$ works for $p \ge \frac{3}{2}$. Lemma 30.2 at r_2 gives $\frac{a-b}{2} = 2\pi k + e_2$ with $|e_2| \le 6\varepsilon_2$. At r_1 the phase difference is $2\pi k\alpha + \alpha e_2$ and at r_3 it is $2\pi k\beta - 4\pi k + (\frac{3\sqrt{2}}{2} - 2)e_2$; the lemma gives integers m, n with

$$2\pi|k\alpha - m| \le 6\varepsilon_1 + 6\alpha\varepsilon_2, \qquad 2\pi|k\beta - n| \le 6\varepsilon_3 + 6\big(\tfrac{3\sqrt{2}}{2} - 2\big)\varepsilon_2.$$

Both coefficients α and $\frac{3\sqrt{2}}{2} - 2$ lie in $(0, 1)$, so with $E = \varepsilon_1+\varepsilon_2+\varepsilon_3$ we get $\max(|k\alpha - m|, |k\beta - n|) \le 6E/(2\pi) < E$. If $E \ge B/12$, then $|a-b| < B \le 12E$. Otherwise $12\varepsilon_2 \le B$; if $k = 0$, $|a-b| \le 12\varepsilon_2$; if $k \ne 0$, then $|k| < B/(2\pi) < B$ as before, and (ii) gives $E > c/\sqrt{|k|} > c/\sqrt{B}$, so $|a-b| < B = B^{3/2}/\sqrt{B} < B^{3/2}E/c$. In each case $|a-b| \le (12 + 1/c)(1+B)^p E$.

(i) implies (ii). We show that $c = 1/(30C)$ satisfies (ii) when C works at $p = \frac{3}{2}$. Let $k \neq 0$, m and n be integers. Replacing (k, m, n) by $(-k, -m, -n)$ we may take $k > 0$, and it suffices to treat the integers m^*, n^* nearest to $k\alpha$ and $k\beta$. With $a = 1 + 4\pi k$, $b = 1$, $B = a$ and readings centred halfway, $\varepsilon_2 = 0$, $\varepsilon_1 = \pi|k\alpha - m^*|$ and $\varepsilon_3 = \pi|k\beta - n^*|$. The bound gives $4\pi k \leq C(2+4\pi k)^{3/2} \cdot 2\pi \max(|k\alpha - m^*|, |k\beta - n^*|)$, and $2 + 4\pi k \leq 15k$, so $\sqrt{k}\max(|k\alpha - m^*|, |k\beta - n^*|) \geq 2/(15^{3/2}C) > 1/(30C)$. $\square$

Both numbers are badly approximable on their own, being quadratic irrationals; the question is whether one k can bring both close to integers at once. The two numbers come from different quadratic fields, and we know no proof that the pair is simultaneously badly approximable; it is Problem 5 of Section 30.5. Schmidt's theorem on simultaneous approximation to algebraic numbers [84] gives $\max(\|k\alpha\|, \|k\beta\|) \geq c_\eta |k|^{-1/2-\eta}$ for every $\eta > 0$, since $1, \alpha, \beta$ are linearly independent over $\mathbb{Q}$; the argument above then makes every exponent $p > \frac{3}{2}$ work. Only the endpoint is open.

Can a better pair of probes do more? Fix the reference probe φ^2. The calibrations a and $a + 4\pi q$ agree there, and a probe at ratio x turns them apart by $4\pi q J(x)$. A pair of probes separates the remote aliases well when, for every large q, one of the two turns is far from a multiple of 2π.

Definition 30.6 (Separation score and the simultaneous approximation constant)**.** In classical mathematics: for $x > 0$ and $q \in \mathbb{N}$ let $\chi_q(x) = (\cos(4\pi q J(x)) - 1)^2 + \sin^2(4\pi q J(x))$, the squared chord between the two phase points of the alias q at the probe x. The *separation score* of probes $x, y > 0$ is

$$\mathrm{sc}(x, y) = \liminf_{q \to \infty} \frac{q}{4} \max\big(\chi_q(x), \chi_q(y)\big).$$

For real α, β, the *simultaneous approximation constant* is

$$S(\alpha, \beta) = \liminf_{q \to \infty} q \max\big(\|q\alpha\|^2, \|q\beta\|^2\big).$$

Adams proved that the supremum of $S(\alpha, \beta)$ over all pairs for which $1, \alpha, \beta$ is a basis of a real cubic field is $\frac{2}{7}$ [78]. Cusick studied the constant for real cubic fields and conjectured that $\frac{2}{7}$ bounds it for every real pair [79]; the supremum over all real pairs is not known.

Theorem 30.7 (Probe pairs and the constant $\frac{2}{7}$)**.** *In classical mathematics:*

(a) $\mathrm{sc}(x, y) = \pi^2 \, S\big(2J(x), 2J(y)\big)$ *for all* $x, y > 0$*;*

(b) the gain $x \mapsto 2J(x)$ *maps* $(1, \varphi^2)$ *bijectively onto* $(0, 1)$*, with inverse*

$$\alpha \mapsto \frac{1}{2}\big(2 + \alpha + \sqrt{\alpha(4 + \alpha)}\big);$$

(c) $\mathrm{sc}(x, y) \leq \pi^2$*;*

(d) some $x, y \in (1, \varphi^2)$ *have* $\mathrm{sc}(x, y) > 2\pi^2/7$ *if and only if some real* α, β *have* $S(\alpha, \beta) > \frac{2}{7}$*.*

Proof. (a) Put $\alpha = 2J(x)$ and $\beta = 2J(y)$. Since $(\cos\theta - 1)^2 + \sin^2\theta = 4\sin^2(\theta/2)$ and $\theta = 2\pi q\alpha$, $\frac{q}{4}\chi_q(x) = q\sin^2(\pi q\alpha)$. For every real t, $4\|t\|^2 \le \sin^2(\pi t) \le \pi^2\|t\|^2$, and the ratio $\sin^2(\pi t)/\|t\|^2$ tends to π^2 as $\|t\| \to 0$; moreover $\sin^2(\pi t)$ increases with $\|t\|$. Write $T_q = q\max(\|q\alpha\|^2, \|q\beta\|^2)$ and $T_q' = q\max(\sin^2(\pi q\alpha), \sin^2(\pi q\beta))$, so $4T_q \le T_q' \le \pi^2 T_q$. By the box argument in the proof of Theorem 30.5(a), with an irrational gain in place of $\sqrt{5}-2$, $T_q \le 1$ for infinitely many q; if both gains are rational, $T_q = 0$ infinitely often. So both lower limits are finite. Along any sequence of q on which T_q or T_q' stays bounded, both $\|q\alpha\|$ and $\|q\beta\|$ tend to 0, so $T_q' = (\pi^2 + o(1))\,T_q$. Taking sequences that realise each lower limit gives $\liminf T_q' = \pi^2 \liminf T_q$.

(b) $2J(x) = x + x^{-1} - 2$ increases on $(1,\infty)$, vanishes at 1, and equals 1 at φ^2 because $\varphi^2 + \varphi^{-2} = 3$. Solving $x^2 - (2+\alpha)x + 1 = 0$ for the root above 1 gives the inverse.

(c) $\liminf T_q \le 1$, as in (a).

(d) If x, y exist, (a) gives the pair $2J(x), 2J(y)$. Conversely, S does not change when α or β changes by an integer. If α is an integer, then $S(\alpha,\beta) = \liminf q\|q\beta\|^2 = 0$: for irrational β, Dirichlet's theorem gives infinitely many q with $q\|q\beta\|^2 \le 1/q$, and for rational β, $\|q\beta\| = 0$ infinitely often. So a pair with $S > \frac{2}{7}$ has fractional parts in $(0,1)$, and (b) turns them into probes in $(1,\varphi^2)$ whose score is $\pi^2 S > 2\pi^2/7$ by (a). □

Conditions and scope.

Base and premises

Classical real numbers; the score and the constant of Definition 30.6; the reference probe φ^2 with $J(\varphi^2) = \frac{1}{2}$.

Conclusion

Separating remote aliases with two probes is exactly the classical problem of simultaneous approximation in the maximum norm: the score is π^2 times the constant of the pair of gains, never more than π^2, and a pair beating $2\pi^2/7$ exists exactly when Cusick's bound $S \le \frac{2}{7}$ fails for some real pair.

Not claimed

The value of the supremum. It is Problem 6.

The value $\frac{2}{7}$ arises in cubic fields. The field of $\vartheta = 2\cos(2\pi/7)$, a root of $\vartheta^3 + \vartheta^2 - 2\vartheta - 1 = 0$, is the real cubic field of discriminant 49, and its integers form the ring $\mathbb{Z}[\vartheta]$. For integers p, q let $\gamma(p,q)$ be the probe whose gain is the fractional part of $p\vartheta + q\vartheta^2$, by Theorem 30.7(b).

Proposition 30.8 (Integral cubic probe designs)**.** *In classical mathematics, $\vartheta = 2\cos(2\pi/7)$ satisfies $\vartheta^3 + \vartheta^2 - 2\vartheta - 1 = 0$, and the following are equivalent:*

(i) for every $\eta > 0$ there are integers p, q, r, s with $ps - qr = 1$ such that $\gamma(p,q)$ and $\gamma(r,s)$ lie in $(1,\varphi^2)$ and $2\pi^2/7 - \eta < \mathrm{sc}(\gamma(p,q), \gamma(r,s)) < 2\pi^2/7$;

(ii) for every $\eta > 0$ there are integers p, q, r, s with $ps - qr = 1$ and

$$\frac{2}{7} - \eta < S(p\vartheta + q\vartheta^2,\ r\vartheta + s\vartheta^2) < \frac{2}{7}.$$

Proof. With $c = \cos(2\pi/7)$, the identity follows from $\cos(6\pi/7) = \cos(8\pi/7)$, that is $4c^3 - 3c = 2(2c^2-1)^2 - 1$, that is $(c-1)(8c^3+4c^2-4c-1) = 0$; and $c \neq 1$. The cubic has no rational root (± 1 give -1 and 1), so $1, \vartheta, \vartheta^2$ are linearly independent over $\mathbb{Q}$, and $p\vartheta + q\vartheta^2$ is not an integer unless $p = q = 0$. When $ps - qr = 1$, neither (p, q) nor (r, s) is $(0, 0)$, so both fractional parts lie in $(0, 1)$ and both probes lie in $(1, \varphi^2)$, automatically. By Theorem 30.7(a) and the invariance of S under integer shifts, the score of the two probes is $\pi^2 S(p\vartheta + q\vartheta^2, r\vartheta + s\vartheta^2)$; so (i) and (ii) say the same thing, with η scaled by π^2. □

Whether unimodular integral bases of $\mathbb{Z}[\vartheta]$ come arbitrarily close to $\frac{2}{7}$ from below is open (Problem 7), and so is the corresponding question for coordinates of index two, where the conjectured ceiling is $\frac{1}{14}$ (Problem 8).

30.2 Two recognizers, and the memory that recognition needs

A recognizer of a distinction history reads the count of distinctions up to an index–period relation $\sim_{i,p}$ (Definition 7.2): below the index i it tells every count apart, and from i on it reads the count modulo the period p. Chapter 7 catalogues the single recognizers. Here two of them read one history. Their common coarse-graining identifies what either identifies. When two readings agree after that coarse-graining, did they come from one count?

Take $\sim_{0,2}$ and $\sim_{1,3}$. Together they identify any two counts, since $\gcd(2, 3) = 1$; in particular they identify 0 and 1. Is there a count n that the first reads as 0 and the second reads as 1? The second recognizer tells 0 apart from everything, but it reads 1 as $4, 7, 10, \ldots$; and the first reads 0 as every even count. So $n = 4$ works. Now swap the roles: a count read as 1 by $\sim_{0,2}$ and as 0 by $\sim_{1,3}$ would have to be 0, which is even. The transient of the later-starting recognizer is a genuine obstruction.

The *join* $\approx_1 \vee \approx_2$ of two additive congruences on counts is the least additive congruence containing both. It is the transitive closure of their union, because the union is reflexive, symmetric and preserved by adding a count to both sides, and the transitive closure keeps these properties.

Theorem 30.9 (Synchronized recognition)**.** *In classical mathematics: let i, j be counts, $p, q > 0$ and $g = \gcd(p, q)$.*

(a) The join of two recognizers is a recognizer: $\sim_{i,p} \vee \sim_{j,q} = \sim_{\min(i,j),\, g}$.

(b) The following are equivalent: (i) whenever $a (\sim_{i,p} \vee \sim_{j,q}) b$, some count n satisfies $n \sim_{i,p} a$ and $n \sim_{j,q} b$; (ii) $i = j$, or $i < j$ and $p \mid q$, or $j < i$ and $q \mid p$.

Proof. (a) Write $m = \min(i, j)$. Each of the two relations lies below $\sim_{m,g}$: a related pair is equal, or consists of counts at least i (or j), hence at least m, with equal remainders modulo p (or q), hence modulo g. As $\sim_{m,g}$ is an additive congruence (Proposition 7.3), the join lies below it. Conversely, the join is transitive, so it suffices to show $t \simeq t + g$ in the join for every $t \geq m$; repeating the step then relates any two counts at least m with equal remainders modulo g.

Say $m = i$ (the other case is symmetric). By Bézout's identity there is a count M with $pM \equiv g \pmod{q}$. Then

$$t \sim_{i,p} t + p(M + j) \sim_{j,q} t + g + pj \sim_{i,p} t + g :$$

the first and last steps relate counts at least i that differ by a multiple of p; the middle step relates counts at least $pj \geq j$ whose difference $pM - g$ is a multiple of q.

(b) Suppose (ii). If $i = j$ and $a \neq b$ are related by the join, both are at least i and $a \equiv b \pmod{g}$; by Bézout's identity $b - a = ps - qt$ for some integers s, t, so $n = a + ps + Mpq$, with M large, is a count $n \geq i$ with $n \equiv a \pmod{p}$ and $n \equiv b \pmod{q}$, and it is the required count. If $i < j$ and $p \mid q$, the join is $\sim_{i,p}$ by (a), and $n = b$ works: $b \sim_{i,p} a$ and $b \sim_{j,q} b$. The case $j < i$, $q \mid p$ is the mirror image, with $n = a$.

Suppose (ii) fails, say $i < j$ and $p \nmid q$ (the case $j < i$, $q \nmid p$ is the mirror image). Then $g < p$, and the join, which is $\sim_{i,g}$, relates $a = i + g$ with $b = i$. A count n with $n \sim_{j,q} i$ is i itself, since $i < j$. Then $i \sim_{i,p} i + g$ would need $p \mid g$, which fails because $0 < g < p$. □

So agreement after coarse-graining is not enough for a common origin unless the recognizers start together or the later one refines the earlier one's period. Part (a) closes the catalogue of index–period recognizers under joins.

The second question concerns memory. A record can be real, in the sense that any two of its possible values can be told apart by some later examination, and yet be expensive to measure, in the sense of one fixed procedure that tells all values apart at once. We measure the cost in bits kept within reach.

Definition 30.10 (Asking histories and the window world)**.** A *record* is a finite binary word. An *asking history* is a finite list of acts: *post r* appends the record r to the list of records; *retain* changes nothing but the stage; *compare* (x, y), when positions x and y of the list hold records, appends to the transcript whether the two records are equal, and otherwise records nothing. Running a history from the empty list gives a list of records and a transcript of outcomes. The *result* of an act after a history is the outcome of a comparison whose positions both hold records, and no result otherwise.

1. The *window cost* of a list of records $r_0, \dots, r_{\ell-1}$ is $\sum_t (|r_t| + 1)$: each record costs its bits and one separator. Position x is in the *window of capacity* N when $x < \ell$ and the records from position x on cost at most N.

2. The *window world of capacity* N permits an asking history when every comparison in it addresses, at the moment it is made, two positions in the window of capacity N.

3. A *recognition observable* with values in a set X assigns to each $x \in X$ a permitted history $o(x)$, such that for distinct x, y there are a history c and an act a for which $o(x)\,c\,a$ and $o(y)\,c\,a$ are permitted and the results of a after $o(x)\,c$ and after $o(y)\,c$ differ.

A recognition observable is the operational reading of "the values are real": each value can occur, and any two can be told apart by some later examination, which may depend on the pair.

Theorem 30.11 (Pairwise recognition is cheap; common measurement is not)**.** *Let N and L be counts, $n \geq 2$, and let $m_1, \dots, m_n$ be distinct records of length L.*

(a) There is a recognition observable in the window world of capacity N with values $1, \dots, n$ and $o(i) = \text{post } m_i$ if and only if $N \geq 2(L+1)$.

(b) There is one asking history c such that every history $(\text{post } m_i)\, c$ is permitted in the window world of capacity N and the transcripts of these n runs are pairwise distinct if and only if $N \geq n(L+1)$.

Proof. In both parts the runs differ only in the record at position 0: after $(\text{post } m_i)\, c$ the list is m_i followed by the records that c posts, the same for every i. A comparison of two positions other than 0, or of 0 with itself, therefore has the same result in every run; a comparison of 0 with a position $y \neq 0$ holding a record r has the result "$m_i = r$". Whether a position holds a record does not depend on i.

(a) If $N \geq 2(L+1)$, take for $x \neq y$ the history $c = \text{post } m_x$ and the act $a = \text{compare } (0, 1)$. After $o(x)\, c$ it finds two equal records, after $o(y)\, c$ two different ones, and the list costs $2(L+1) \leq N$, so both positions are in the window. Conversely, let $x \neq y$, with c and a as in the definition. The act a must be a comparison of 0 with some $y' \neq 0$ holding a record r, and "$m_x = r$" and "$m_y = r$" must differ, so r is m_x or m_y and has length L. Position 0 is in the window, so the whole list costs at most N, and it contains two records of length L: $2(L+1) \leq N$.

(b) If $N \geq n(L+1)$, let c post m_1, compare $(0, 1)$, post m_2, compare $(0, 2)$, and so on up to m_{n-1}. When position t is compared the list costs $(t+1)(L+1) \leq N$, so every comparison is permitted. Run i has transcript $(m_i = m_1, \dots, m_i = m_{n-1})$, which is the indicator of i for $i < n$ and all "false" for $i = n$; these are distinct. Conversely, let c work, and let R be the set of records that c compares with position 0. If two indices $i \neq i'$ had $m_i, m_{i'} \notin R$, every result in their runs would agree. So at least $n - 1$ of the m_i lie in R. At the last comparison of position 0 with a member of R, all members of R have been posted, and position 0 is in the window: the whole list, which holds m_i and at least $n - 1$ further distinct records of length L, costs at most N. So $n(L+1) \leq N$. □

For all records of L bits, $n = 2^L$: being real costs a window of $2(L+1)$ bits, while one common measurement costs $2^L(L+1)$. The gap is exponential, and it comes from one fact visible in the proof: a comparison learns only whether two records are equal, so a fixed procedure must keep every candidate it rules out within reach of the unknown record.

30.3 Self-similar ledgers: rates, balance and fluctuation

Chapter 21 built Recognition Science's ledger from one rule: every debit is posted as a debit followed by a credit, every credit as a debit. The resulting history is the Fibonacci word [100]; its running rate tends to φ^{-3}, and its books stay within 2 of the line of that slope (Chapter 21). Any rule that replaces each posting by a fixed word defines a self-similar ledger in the same way. Which rates can such ledgers have, and which of them keep balanced books?

Definition 30.12 (Joins, rates and balanced books)**.** In classical mathematics:

1. A *join* σ assigns a posting word to each of the letters d (debit) and c (credit) and acts on words letter by letter. Its *words* are $W_k = \sigma^k(\mathsf{d})$. Its *incidence counts* are the numbers DD and CD of debits and credits in $\sigma(\mathsf{d})$ and DC and CC of debits and credits in $\sigma(\mathsf{c})$; its trace is $\mathrm{tr} = DD + CC$, its determinant $\det = DD \cdot CC - DC \cdot CD$, its discriminant $\mathrm{tr}^2 - 4\det$, its *Perron root* $\lambda = \frac{1}{2}(\mathrm{tr} + \sqrt{\mathrm{tr}^2 - 4\det})$ and its *second mode* $\mu = \mathrm{tr} - \lambda$.
2. The join is *primitive* when for some $k \geq 1$ both letters occur in $\sigma^k(\mathsf{d})$ and in $\sigma^k(\mathsf{c})$.
3. It has *rate* r when the rates $\mathrm{bal}(W_k)/|W_k|$ tend to r, where bal counts debits minus credits.
4. The *discrepancy* of a word at rate r is $\delta_r(w) = \mathrm{bal}(w) - r|w|$. The join *keeps balanced books* at rate r when some C bounds $|\delta_r(p)|$ for every prefix p of every word W_k.

The golden join of Chapter 21, $\mathsf{d} \mapsto \mathsf{dc}$ and $\mathsf{c} \mapsto \mathsf{d}$, has trace 1, determinant -1, discriminant 5 and rate $\sqrt{5} - 2 = \varphi^{-3}$.

Which rates occur

Theorem 30.13 (The rates of self-similar ledgers)**.** *In classical mathematics: a real number r is the rate of some primitive join if and only if $-1 < r < 1$ and r is a root of an integer polynomial $at^2 + bt + c$ with $a \neq 0$ whose other root, $-b/a - r$, lies outside $[-1, 1]$. Every rational number in $(-1, 1)$ and φ^{-3} are rates; $\sqrt{2}/2$ is not, although it lies in $(-1, 1)$ and is a root of $2t^2 - 1$. So not every quadratic number in $(-1, 1)$ is a rate.*

Proof. Counts. Let (x_k, y_k) be the numbers of debits and credits in W_k. Each debit of W_k becomes $\sigma(\mathsf{d})$ and each credit $\sigma(\mathsf{c})$, so $(x_{k+1}, y_{k+1})^{\mathsf{T}} = M(x_k, y_k)^{\mathsf{T}}$ with $M = \left(\begin{smallmatrix} DD & DC \\ CD & CC \end{smallmatrix}\right)$. If σ is primitive, then $DC \geq 1$ (otherwise no power of σ puts a debit into the image of a credit), $CD \geq 1$ likewise, and $\mathrm{tr} \geq 1$ (if $DD = CC = 0$, every $\sigma^k(\mathsf{d})$ consists of one letter only). The discriminant $(DD - CC)^2 + 4\,DC \cdot CD$ is positive, so $\lambda > \mu$, and $\lambda + \mu = \mathrm{tr} > 0$ gives $\lambda > |\mu|$. The vector $\xi = (DC, \lambda - DD)$ has positive entries and satisfies $M\xi = \lambda\xi$; the vector $\zeta = (DC, \mu - DD)$ satisfies $M\zeta = \mu\zeta$ and has entries of opposite signs. Writing $(1, 0) = A\xi + B\zeta$ gives $A = (DD - \mu)/(DC(\lambda - \mu)) > 0$, so $(x_k, y_k) = A\lambda^k\xi + B\mu^k\zeta$ and the rates of W_k tend to $r = (\beta - 1)/(\beta + 1) \in (-1, 1)$, where $\beta = DC/(\lambda - DD) > 0$ is the ratio of the entries of ξ. The two rows of $M\xi = \lambda\xi$ give $CD\,\beta^2 + (CC - DD)\beta - DC = 0$; this is the theorem of Perron and Frobenius in dimension two [89].

The polynomial. Substitute $\beta = (1 + r)/(1 - r)$ and multiply by $(1 - r)^2$: r is a root of

$$P(t) = CD\,(1 + t)^2 + q(1 + t)(1 - t) - DC\,(1 - t)^2, \qquad q = CC - DD,$$

an integer polynomial with $P(1) = 4\,CD > 0$ and $P(-1) = -4\,DC < 0$. If P has degree two, with other root r', then $P(1)P(-1) = a^2(1 - r^2)(1 - r'^2) < 0$ for its leading coefficient a, so $|r'| > 1$. If P has degree one, r is rational, and $(t - 2)P(t)$ is an integer quadratic with roots r and 2.

Realization. Let $r \in (-1, 1)$ be a root of $P_0(t) = at^2 + bt + c$, $a \neq 0$, with other root r', $|r'| > 1$. As above $P_0(1)P_0(-1) < 0$; changing the sign of P_0 if necessary, $P_0(1) > 0 > P_0(-1)$. Put

$CD = a + b + c \geq 1$, $DC = -(a - b + c) \geq 1$ and $q = 2(c - a)$; a direct expansion shows that the polynomial P built from these numbers is $4P_0$. Choose $DD = |q| + 1 - q$ and $CC = |q| + 1$, both at least 1 with $CC - DD = q$, and let $\sigma(\mathsf{d}) = \mathsf{d}^{DD}\mathsf{c}^{CD}$ and $\sigma(\mathsf{c}) = \mathsf{d}^{DC}\mathsf{c}^{CC}$. Both images contain both letters, so σ is primitive, and its rate is the root of $4P_0$ in $(-1, 1)$, which is r.

Examples. A rational $m/n \in (-1, 1)$ is a root of $(nt - m)(t - 2)$. The rate $\sqrt{5} - 2$ is a root of $t^2 + 4t - 1$, whose other root is $-2 - \sqrt{5}$. An integer quadratic with root $\sqrt{2}/2$ is a rational multiple of $2t^2 - 1$, because $\sqrt{2}/2$ is irrational; its other root is $-\sqrt{2}/2 \in [-1, 1]$. □

The proof explains the shape of the answer. A rate is a ratio of entries of a positive eigenvector of a positive integer matrix, and the conjugate eigenvector of such a matrix has entries of opposite signs; in the rate coordinate this puts the conjugate root outside $[-1, 1]$.

Which ledgers keep balanced books

The discrepancy adds over concatenation, and on the two letters it is $\delta_r(\mathsf{d}) = 1 - r$ and $\delta_r(\mathsf{c}) = -1 - r$. The row vector $\ell = (1 - r, -1 - r)$ is orthogonal to the Perron vector $\xi \propto (1 + r, 1 - r)$, so it is a left eigenvector of M for the second mode: $\ell M = \mu\ell$. In words,

$$\delta_r(\sigma(w)) = \mu\,\delta_r(w) \quad \text{for every word } w, \qquad \delta_r(W_k) = \mu^k(1 - r). \tag{30.1}$$

We also need the prefix decomposition of Dumont and Thomas [80]: every proper prefix p of $\sigma^k(a)$ is

$$p = \sigma^{k-1}(u_1)\,\sigma^{k-2}(u_2)\cdots\sigma^0(u_k), \tag{30.2}$$

where $b_1 = a$, each u_j is a proper prefix of $\sigma(b_j)$, and b_{j+1} is the letter of $\sigma(b_j)$ that follows u_j. It follows by induction on k, since $\sigma^k(a)$ is the concatenation of the words $\sigma^{k-1}(y)$ over the letters y of $\sigma(a)$. By (30.1), $\delta_r(p) = \sum_j \mu^{k-j}\delta_r(u_j)$.

Theorem 30.14 (The balance law of two-letter ledgers)**.** *In classical mathematics, let σ be a primitive join with rate r.*

(a) σ keeps balanced books if and only if either $|\det| < \lambda$, or $|\det| = \lambda$ and the prefix discrepancies of σ^2 form a coboundary*: some π on the two letters satisfies*

$$\delta_r\big(\textit{first } m \textit{ letters of } \sigma^2(b)\big) = \pi\big(\textit{letter } m \textit{ of } \sigma^2(b)\big) - \pi(b)$$

for both letters b and every position $m < |\sigma^2(b)|$, counted from 0*.*

(b) If the discriminant is not a perfect square, σ keeps balanced books if and only if $|\det| < \lambda$.

Proof. Since $\mu\lambda = \det$, the condition $|\det| < \lambda$ says $|\mu| < 1$, and $|\det| = \lambda$ says $|\mu| = 1$.

If $|\mu| < 1$, let m_0 bound $|\delta_r(u)|$ over the proper prefixes u of $\sigma(\mathsf{d})$ and $\sigma(\mathsf{c})$. Every proper prefix of W_k has $|\delta_r(p)| \leq \sum_j |\mu|^{k-j} m_0 \leq m_0/(1 - |\mu|)$, and the whole words have $|\delta_r(W_k)| \leq 1 - r$; the books are balanced. If $|\mu| > 1$, the whole words alone have $|\delta_r(W_k)| = |\mu|^k(1 - r) \to \infty$.

Let $|\mu| = 1$ and put $\tau = \sigma^2$, whose second mode is $\mu^2 = 1$, so τ preserves discrepancies. Draw the *prefix graph* of τ: for every letter b and position m of $\tau(b)$ an edge from b to the letter at m,

weighted by the discrepancy of the first m letters of $\tau(b)$. By (30.2) for τ, the proper prefixes of $\tau^k(b)$ correspond to the paths of length k from b, and the discrepancy of the prefix is the weight of its path.

Suppose the coboundary π exists. The edge weights are $\pi(\text{end}) - \pi(\text{start})$, so every path weight telescopes to a difference of two values of π, and every prefix of $W_{2k} = \tau^k(\mathsf{d})$ has discrepancy at most $2\max|\pi|$ in absolute value. A prefix of $W_{2k+1} = \tau^k(\sigma(\mathsf{d}))$ is $\tau^k(v)p'$ with v a prefix of $\sigma(\mathsf{d})$ and p' a prefix of some $\tau^k(y)$, and $\delta_r(\tau^k(v)) = \delta_r(v)$; so those are bounded as well.

Conversely, suppose the books are balanced, with bound C. Every path from d has weight at most C in absolute value. Since σ is primitive, so is τ, and both letters can be reached from each letter. If some cycle had nonzero weight, going round it repeatedly after reaching it from d would give paths from d of unbounded weight. So every cycle has weight 0, and $\pi(a) =$ the weight of any path from d to a is well defined: two such paths, each followed by one path back to d, are cycles. For an edge from b to a, a path to b followed by the edge is a path to a, so its weight is $\pi(a) - \pi(b)$. This is the coboundary.

(b) If $\mu = \pm 1$, then $\lambda = \mathrm{tr} \mp 1$ and the discriminant $(\lambda - \mu)^2 = (\mathrm{tr} \mp 2)^2$ is a perfect square. So a nonsquare discriminant excludes $|\mu| = 1$, and (a) reduces to $|\det| < \lambda$. □

Conditions and scope.

Base and premises
Classical real numbers; primitive two-letter joins of Definition 30.12; the rate of Theorem 30.13.

Conclusion
A self-similar ledger keeps its books within a fixed distance of the line of its rate exactly when its second mode contracts, $|\det| < \lambda$; at the neutral boundary $|\det| = \lambda$ it does so exactly when the prefix discrepancies of the doubled rule are a coboundary.

Not claimed
Ledgers with three or more letters; their balance law is not settled here.

Every unimodular join, $|\det| = 1$, keeps balanced books, the golden join among them. The join $\mathsf{d} \mapsto \mathsf{dcc}$, $\mathsf{c} \mapsto \mathsf{dddc}$ has trace 2, determinant -5 and $\lambda = 1 + \sqrt{6} \approx 3.449 < 5$: its whole words drift like $(\sqrt{6} - 1)^k$. At the neutral boundary the incidence counts no longer decide. The joins $\mathsf{d} \mapsto \mathsf{dcd}$, $\mathsf{c} \mapsto \mathsf{cdc}$ and $\mathsf{d} \mapsto \mathsf{ddc}$, $\mathsf{c} \mapsto \mathsf{ccd}$ have the same counts, with $\lambda = 3 = |\det|$ and rate 0. The first is balanced, with $\pi(\mathsf{d}) = 0$ and $\pi(\mathsf{c}) = 1$; the second is not, since its words begin $\mathsf{ddc\,ddc\,ccd}\cdots$ and the prefix balance climbs by one at each level.

How an unbalanced neutral ledger wanders

At the neutral boundary an unbalanced ledger does not drift: its whole words have discrepancy $\pm(1 - r)$. It spreads instead, and the spread has an exact law. For a primitive join with rate r and $|\mu| = 1$, let V_k be the variance of the $|W_k| + 1$ numbers $\delta_r(p)$, p running over all prefixes of W_k, the empty prefix and W_k included, each with weight $1/(|W_k| + 1)$.

Theorem 30.15 (Variance at the neutral boundary)**.** *In classical mathematics: for every primitive join with rate r and second mode $\mu = \pm 1$, the limit $v = \lim_k V_k/k$ exists, $v \geq 0$, and $v = 0$ if and only if the join keeps balanced books.*

Proof. The limit. The proof computes it. For letters y put $\alpha_y = (|\sigma(y)| - \mu)/(\lambda - \mu)$, so that $|\sigma^k(y)| = \alpha_y \lambda^k + (1 - \alpha_y)\mu^k$ (both sides satisfy the recurrence with roots λ, μ and agree at $k = 0, 1$). Primitivity makes $|\sigma^k(y)|$ unbounded, so $\alpha_y > 0$. For a letter b and a position m of $\sigma(b)$, carrying the letter y_m, let δ_m be the discrepancy of the first m letters of $\sigma(b)$ and $q_m = \alpha_{y_m}/(\lambda \alpha_b)$; these weights sum to 1 over m. Put $\omega(b) = \sum_m q_m \delta_m$, $\eta = \mu\lambda(\omega(\mathsf{d}) - \omega(\mathsf{c}))/(\lambda - 1)$, the *block values* $v_m = \delta_m + \eta\,[y_m = \mathsf{d}]$, their weighted mean $\kappa(b) = \sum_m q_m v_m$ and

$$\tau(b) = \sum_m q_m \big(v_m - \kappa(b)\big)^2, \qquad v = \tfrac{1}{2}(1 + r)\alpha_{\mathsf{d}}\,\tau(\mathsf{d}) + \tfrac{1}{2}(1 - r)\alpha_{\mathsf{c}}\,\tau(\mathsf{c}).$$

For a letter b, let $m_k(b)$ and $s_k(b)$ be the mean and the variance of the discrepancies of the $|\sigma^k(b)|$ proper prefixes of $\sigma^k(b)$. The word $\sigma^{k+1}(b)$ is the concatenation of the blocks $\sigma^k(y_m)$, and a proper prefix ending in block m has discrepancy $\mu^k \delta_m$ plus that of a proper prefix of $\sigma^k(y_m)$. The law of total variance gives

$$s_{k+1}(b) = \sum_m w_{k,m}\, s_k(y_m) + \tau_k(b),$$

with block weights $w_{k,m} = |\sigma^k(y_m)|/|\sigma^{k+1}(b)| \to q_m$ and $\tau_k(b)$ the weighted variance of the block values $\mu^k \delta_m + m_k(y_m)$. Since $\mu^2 = 1$, these values are $\mu^k(\delta_m + \eta_k[y_m = \mathsf{d}])$ plus a constant, with $\eta_k = \mu^k(m_k(\mathsf{d}) - m_k(\mathsf{c}))$. Let κ_k be the weight that the blocks of $\sigma(\mathsf{d})$ carrying a debit receive, minus the same for $\sigma(\mathsf{c})$. The block decomposition of the means gives $\eta_{k+1} = \mu(\omega_k(\mathsf{d}) - \omega_k(\mathsf{c})) + \mu\kappa_k\eta_k$, with ω_k formed like ω from the weights $w_{k,m}$. The vector (α_y) is a left eigenvector of M for λ, which gives $\lambda\kappa_k \to DD - DC\,\alpha_{\mathsf{d}}/\alpha_{\mathsf{c}} = DD + CC - \lambda = \mu$; so $\mu\kappa_k \to 1/\lambda < 1$ and $\eta_k \to \mu\lambda(\omega(\mathsf{d}) - \omega(\mathsf{c}))/(\lambda - 1) = \eta$. Hence $\tau_k(b) \to \tau(b)$.

Weight the letters by $\rho_k(b) = f(b)|\sigma^k(b)|/(2\lambda^k)$ with $f = (1 + r, 1 - r)$. Because f is proportional to the Perron vector, $Mf = \lambda f$, and a direct substitution gives $\sum_b \rho_{k+1}(b) \sum_{m\,:\,y_m = y} w_{k,m} = \rho_k(y)$. So $\sum_b \rho_k(b) = \sum_b \rho_0(b) = 1$, and $Y_k = \sum_b \rho_k(b) s_k(b)$ satisfies $Y_{k+1} = Y_k + \sum_b \rho_{k+1}(b)\tau_k(b)$, whose increment tends to $\sum_b \frac{1}{2} f(b)\alpha_b \tau(b) = v$. By Cesàro's lemma, $Y_k/k \to v$. The difference $g_k = s_k(\mathsf{d}) - s_k(\mathsf{c})$ satisfies $g_{k+1} = \kappa_k g_k + O(1)$ with $|\kappa_k| \to 1/\lambda < 1$, so g_k is bounded; since $Y_k = s_k(\mathsf{d}) - \rho_k(\mathsf{c}) g_k$, also $s_k(\mathsf{d})/k \to v$. Finally V_k adds the whole word, of discrepancy $\pm(1 - r)$, to the $|W_k|$ proper prefixes; the means grow at most linearly in k and $|W_k|$ exponentially, so $V_k - s_k(\mathsf{d}) \to 0$.

Zero limit. If the books are balanced with bound C, then $V_k \leq C^2$ and $v = 0$. Conversely let $v = 0$. The weights $\frac{1}{2}(1 \pm r)\alpha$ are positive, so $\tau(\mathsf{d}) = \tau(\mathsf{c}) = 0$, and all blocks of $\sigma(b)$ carry one value $V(b)$. Write $e(\mathsf{d}) = 1 - r$ and $e(\mathsf{c}) = -1 - r$, both nonzero. Three facts follow. Consecutive blocks give $e(y_m) = \eta([y_m = \mathsf{d}] - [y_{m+1} = \mathsf{d}])$, so no image contains two equal adjacent letters, and a step dc inside an image forces $\eta = 1 - r$, a step cd forces $\eta = 1 + r$. The first block gives $V(b) = \eta[y_0 = \mathsf{d}]$. The whole image gives, by (30.1), $\mu e(b) = V(b) - \eta[y_{\text{last}} = \mathsf{d}] + e(y_{\text{last}})$. We look for a potential h with $\delta_m + h(y_m) = \mu h(b)$ for every block m of every image $\sigma(b)$; with $h(\mathsf{d}) = \eta + t$ and $h(\mathsf{c}) = t$ this asks $V(b) + t = \mu(\eta[b = \mathsf{d}] + t)$ for both letters.

If $\mu = 1$ it asks $V(\mathsf{d}) = \eta$ and $V(\mathsf{c}) = 0$. Should $\sigma(\mathsf{c})$ end with d, the last-block relation gives $V(\mathsf{c}) = \eta - 2$, and the first-block relation then forces $\sigma(\mathsf{c})$ to begin with c and $\eta = 2$; but the

step cd inside it forces $\eta = 1 + r < 2$. So $\sigma(\mathsf{c})$ ends with c and $V(\mathsf{c}) = 0$; symmetrically $\sigma(\mathsf{d})$ ends with d and $V(\mathsf{d}) = \eta$.

If $\mu = -1$ it asks $V(\mathsf{c}) - V(\mathsf{d}) = \eta$, and then $t = -(\eta + V(\mathsf{d}))/2$. The last-block relation gives $V(\mathsf{d}) = \eta - 2 + 2r$ or $2r$ as $\sigma(\mathsf{d})$ ends with d or c, and $V(\mathsf{c}) = 2 + 2r$ or $2r + \eta$ as $\sigma(\mathsf{c})$ ends with c or d. If $\sigma(\mathsf{d})$ ends with d, the first-block relation excludes $V(\mathsf{d}) = \eta$ (it would give $r = 1$), so $\sigma(\mathsf{d})$ begins with c, $\eta = 2 - 2r$, and its step cd gives $\eta = 1 + r$, so $r = \frac{1}{3}$ and $\eta = \frac{4}{3}$; then $V(\mathsf{c}) \in \{\frac{8}{3}, 2\}$ is neither 0 nor η, a contradiction. If both images end with c, the first-block relations give $r = 0$ and $\eta = 2$, while the step dc in $\sigma(\mathsf{c})$ gives $\eta = 1$, a contradiction. So $\sigma(\mathsf{d})$ ends with c and $\sigma(\mathsf{c})$ with d, and then $V(\mathsf{c}) - V(\mathsf{d}) = \eta$.

In every case the potential exists. Along (30.2), $\delta_r(u_j) = \mu h(b_j) - h(b_{j+1})$, and $\delta_r(p) = \sum_j \mu^{k-j}\delta_r(u_j) = \mu^k h(b_1) - h(b_{k+1})$ telescopes; the books are balanced. □

So at the neutral boundary there is no middle ground: either the books stay bounded, or the prefix discrepancies spread like a random walk, with variance close to νk after k generations. For instance $\mathsf{d} \mapsto \mathsf{ddc}$, $\mathsf{c} \mapsto \mathsf{ccd}$ has $\nu = \frac{2}{3}$, and $\mathsf{d} \mapsto \mathsf{ccd}$, $\mathsf{c} \mapsto \mathsf{cdd}$, with $\mu = -1$, has $\nu = \frac{2}{9}$, by the formula in the proof. How large can ν be? A neutral ledger with Perron root λ and rate r cannot wander faster, it seems, than $(\lambda^2 - 1)(1 + |r|)^2/12$, which at rate 0 is the variance of a uniform choice among λ positions; the bound is open (Problem 2), but at rate 0 it is attained at every integer.

Proposition 30.16 (The neutral slope $(n^2 - 1)/12$ is attained)**.** *In classical mathematics: for every integer $n \geq 2$ there is a primitive join with rate 0, $|\mu| = 1$ and $\lambda = n$ whose variances satisfy $V_k/k \to (n^2 - 1)/12$. For $n \geq 3$ one is $\mathsf{d} \mapsto \mathsf{cdd}$, $\mathsf{c} \mapsto \mathsf{d}^{n-2}\mathsf{c}^{n-1}$; for $n = 2$ one is $\mathsf{d} \mapsto \mathsf{c}$, $\mathsf{c} \mapsto \mathsf{cdd}$.*

Proof. Let $n \geq 3$. The incidence counts are $DD = 2$, $CD = 1$, $DC = n - 2$, $CC = n - 1$; so $\mathrm{tr} = n + 1$, $\det = n$, $\lambda = n$, $\mu = 1$, both images contain both letters, and the Perron vector $(1, 1)$ gives rate 0. By the proof of Theorem 30.15 the limit is the explicit ν. Here $\alpha_{\mathsf{d}} = \frac{2}{n-1}$ and $\alpha_{\mathsf{c}} = \frac{2(n-2)}{n-1}$. In $\sigma(\mathsf{d}) = \mathsf{cdd}$ the blocks have weights $\frac{n-2}{n}, \frac{1}{n}, \frac{1}{n}$ and prefix discrepancies $0, -1, 0$, so $\omega(\mathsf{d}) = -\frac{1}{n}$. In $\sigma(\mathsf{c})$ the $n - 2$ debit blocks have weight $\frac{1}{n(n-2)}$ and prefix discrepancies $0, 1, \dots, n-3$, and the $n-1$ credit blocks have weight $\frac{1}{n}$ and prefix discrepancies $n-2, n-3, \dots, 0$, so $\omega(\mathsf{c}) = \frac{n-3}{2n} + \frac{(n-2)(n-1)}{2n} = \frac{n^2-2n-1}{2n}$. Hence $\eta = n(\omega(\mathsf{d}) - \omega(\mathsf{c}))/(n-1) = -\frac{n-1}{2}$. The block values of $\sigma(\mathsf{d})$ are 0, $\eta - 1$, η with mean -1, so

$$\tau(\mathsf{d}) = \tfrac{n-2}{n} + \tfrac{1}{n}\eta^2 + \tfrac{1}{n}(\eta + 1)^2 = \frac{(n-1)^2}{2n}.$$

The block values of $\sigma(\mathsf{c})$ are $j - \frac{n-1}{2}$ for $j = 0, \dots, n-3$ and $0, 1, \dots, n-2$; their weighted mean is $\frac{n-3}{2}$, and

$$\tau(\mathsf{c}) = \frac{1}{n(n-2)}\sum_{j=1}^{n-2} j^2 + \frac{1}{n}\sum_{t=0}^{n-2}\Big(t - \frac{n-3}{2}\Big)^2 = \frac{(n-1)^2(n+3)}{12n}.$$

With rate 0 the weights are $\frac{1}{2}\alpha_{\mathsf{d}} = \frac{1}{n-1}$ and $\frac{1}{2}\alpha_{\mathsf{c}} = \frac{n-2}{n-1}$, and

$$\nu = \frac{n-1}{2n} + \frac{(n-2)(n-1)(n+3)}{12n} = \frac{(n-1)(n^2+n)}{12n} = \frac{n^2-1}{12}.$$

For $n = 2$: $DD = 0$, $CD = 1$, $DC = 2$, $CC = 1$, so $\lambda = 2$, $\mu = -1$, rate 0, and σ^2 puts both letters into both images. Here $\alpha_{\mathsf{d}} = \frac{2}{3}$, $\alpha_{\mathsf{c}} = \frac{4}{3}$; $\sigma(\mathsf{d})$ is one block, so $\tau(\mathsf{d}) = 0$ and $\omega(\mathsf{d}) = 0$; $\sigma(\mathsf{c}) = \mathsf{cdd}$ has weights $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ and prefix discrepancies $0, -1, 0$, so $\omega(\mathsf{c}) = -\frac{1}{4}$ and $\eta = -2 \cdot \frac{1}{4} = -\frac{1}{2}$. Its block values $0, -\frac{3}{2}, -\frac{1}{2}$ have mean $-\frac{1}{2}$ and $\tau(\mathsf{c}) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$; so $\nu = \frac{2}{3} \cdot \frac{3}{8} = \frac{1}{4}$. □

A hold letter, and a ledger that wanders cubically

Give the ledger a third letter h, a *hold*, that posts neither a debit nor a credit. Joins, primitivity, words $W_k = \sigma^k(\mathsf{d})$ and rates are defined as before, the rate still being debits minus credits over all letters.

Theorem 30.17 (A hold letter strictly enlarges the rates)**.** *In classical mathematics:*

(a) the three-letter join $\mathsf{d} \mapsto \mathsf{d}^9\mathsf{ch}$, $\mathsf{c} \mapsto \mathsf{dh}$, $\mathsf{h} \mapsto \mathsf{dc}$ *is primitive with rate* $\sqrt{2}/2$*, which is the rate of no primitive two-letter join;*

(b) every rational number in $(-1, 1)$ *is the rate of a primitive three-letter join;*

(c) the statement "the rates of primitive three-letter joins are exactly the numbers in $(-1, 1)$ *that are algebraic of degree at most* 3*" is equivalent to the conjunction of two statements about the joins* $\sigma(a) = \mathsf{d}^{A(a)}\mathsf{c}^{B(a)}\mathsf{h}^{H(a)}$ *given by nine counts: that every rate of such a primitive join lies in* $(-1, 1)$ *and is algebraic of degree at most* 3*, and that every irrational algebraic number of degree at most* 3 *in* $(-1, 1)$ *other than* $\sqrt{2}/2$ *is the rate of one.*

Proof. (a) Let x_k, y_k, z_k count debits, credits and holds in W_k. Then $x_{k+1} = 9x_k + y_k + z_k$, $y_{k+1} = x_k + z_k$ and $z_{k+1} = x_k + y_k$, so $y_{k+1} - z_{k+1} = z_k - y_k$, and from $y_0 = z_0 = 0$ the credits and holds stay equal. The pair (x_k, y_k) follows the matrix $\left(\begin{smallmatrix} 9 & 2 \\ 1 & 1 \end{smallmatrix}\right)$, with Perron root $5 + 3\sqrt{2}$ and Perron vector $(4 + 3\sqrt{2}, 1)$; as in the proof of Theorem 30.13, $x_k/y_k \to 4 + 3\sqrt{2}$, and the rate $(x_k - y_k)/(x_k + 2y_k)$ tends to $(3 + 3\sqrt{2})/(6 + 3\sqrt{2}) = \sqrt{2}/2$. All three letters occur in σ^2 of each letter. Theorem 30.13 excludes $\sqrt{2}/2$ for two letters.

(b) Write $q = m/n$ with $|m| < n$. The join whose three images all equal $\mathsf{d}^{2(n+m)-1}\mathsf{c}^{2(n-m)-1}\mathsf{h}^2$ is primitive, since each exponent is at least 1; each image has length $4n$ and balance $4m$, and every W_k with $k \geq 1$ is a concatenation of images, so its rate is exactly q.

(c) The incidence counts alone determine the letter counts of every W_k, primitivity and every whole-word rate, and the join with images $\mathsf{d}^A\mathsf{c}^B\mathsf{h}^H$ has the same counts as any join; so restricting to these normal forms loses nothing. Rationals and $\sqrt{2}/2$ are covered by (a) and (b), and the remaining clauses are the two named statements. □

Whether the two statements of (c) hold is Problem 1. The eigenvector reasoning behind Theorem 30.13 suggests degree at most 3; what is missing is a cubic analogue of the realization step.

A hold letter also lets the second mode form a Jordan block, and then the ledger wanders faster than any random walk. Consider the census with images of length 7 in which $\sigma(\mathsf{d})$ has 4 debits, 1 credit and 2 holds, $\sigma(\mathsf{c})$ has 3, 2, 2 and $\sigma(\mathsf{h})$ has 0, 4, 3. Its *witness* ordering is

$$\sigma_* : \quad \mathsf{d} \mapsto \mathsf{cddddhh}, \qquad \mathsf{c} \mapsto \mathsf{hhccddd}, \qquad \mathsf{h} \mapsto \mathsf{hhhcccc}.$$

Weight the letters by $u = (1, 1, -2)$ and write $f_\sigma(a, i)$ for the u-weight of the first i letters of $\sigma(a)$, $i = 0, \dots, 6$. Let $\bar{f} = \frac{1}{21}\sum_{a,i} f_\sigma(a, i)$, let P be the census average, $(P\psi)(a) = \frac{1}{7}\sum_{c\in\sigma(a)} \psi(c)$ over the seven letters of $\sigma(a)$, and let z_σ solve $z - Pz = h$ with $z(\mathsf{d}) + z(\mathsf{c}) + z(\mathsf{h}) = 0$, where $h(a) = \frac{1}{7}\sum_i f_\sigma(a, i) - \bar{f}$. Put

$$g_\sigma(a, i) = f_\sigma(a, i) - \bar{f} + z_\sigma(\text{letter } i \text{ of } \sigma(a)) - z_\sigma(a), \qquad \nu(\sigma) = \frac{4}{63}\sum_a \sum_{i=0}^{6} g_\sigma(a, i)^2.$$

Theorem 30.18 (A Jordan census: cubic fluctuation and its exact range)**.** *In classical mathematics: the witness σ_* is primitive with rate 0. For every join σ with the same nine counts, the variance V_k of the balances of all prefixes of W_k satisfies $V_k/k^3 \to \nu(\sigma)$, and*

$$\frac{100}{441} \le \nu(\sigma) \le \frac{7408}{441}.$$

The upper value is attained exactly by σ_* and by σ_* with all three images reversed; the lower value is attained, for instance by $\mathsf{d} \mapsto \mathsf{dchdhdd}$, $\mathsf{c} \mapsto \mathsf{dhdchdc}$, $\mathsf{h} \mapsto \mathsf{hcchcch}$.

Proof. The Jordan block. Let M be the incidence matrix, column a counting the letters of $\sigma(a)$. With the balance weights $\mathrm{b} = (1, -1, 0)$ one checks $\mathrm{b}M = \mathrm{b} + 2u$ and $uM = u$; so the balance of $\sigma^j(w)$ is $\mathrm{bal}(w) + 2j\,u(w)$, with $u(w)$ the u-weight of w. Each letter occurs seven times among the three images, so the uniform law on letters is stationary for P, and the Perron vector is $(1, 1, 1)$: the rate is 0, and the witness is primitive because σ_*^2 of every letter contains all three.

Prefixes as a Markov chain. All images have length 7, so a prefix of W_k of length $N < 7^k$ corresponds to the base-7 digits $(m_1, \dots, m_k)$ of N, and (30.2) reads $p = \sigma^{k-1}(u_1)\cdots\sigma^0(u_k)$ with u_j the first m_j letters of $\sigma(b_j)$, $b_1 = \mathsf{d}$ and b_{j+1} the letter at m_j. A uniformly chosen N makes the digits independent and uniform, and the letters b_j a Markov chain with kernel P. The balance of p is $\sum_j\big(\mathrm{bal}(u_j) + 2(k-j)F_j\big)$ with $F_j = f_\sigma(b_j, m_j)$. By the Poisson equation, $g_j = g_\sigma(b_j, m_j)$ has conditional mean 0 given the past, and $F_j - \bar{f} = g_j - (z_\sigma(b_{j+1}) - z_\sigma(b_j))$. Summation by parts shows that the terms z_σ, $\mathrm{bal}(u_j)$ and the mean $\bar{f}$ contribute $O(k)$ to the centred balance in mean square, while the martingale part $\sum_j 2(k-j)g_j$ has variance $\sum_j 4(k-j)^2\,\mathbb{E}[g_j^2]$. As P is primitive, $\mathbb{E}[g_j^2]$ tends geometrically to the stationary value $\frac{1}{21}\sum_{a,i} g_\sigma(a, i)^2$, and $\sum_j 4(k-j)^2 \sim \frac{4}{3}k^3$. So the variance over the proper prefixes, divided by k^3, tends to $\nu(\sigma)$; the whole word adds one value $1 + 2k$ and changes nothing in the limit.

The range. The coefficient $\nu(\sigma)$ is a function of the three orderings; there are 105, 210 and 35 orderings of the three images. Both $\bar{f}$ and z_σ depend on them only through the three image sums $\sum_i f_\sigma(a, i)$, which take 11, 11 and 13 values. Once the sums are fixed, each image's contribution to $\nu(\sigma)$ depends on its own ordering alone. The range is therefore a finite exhaustive computation in exact rationals, which we describe but do not reproduce: for each of the $11 \cdot 11 \cdot 13 = 1573$ triples of sums, solve for z_σ, take the largest and the smallest contribution of each image among its orderings with the given sum, and compare. It gives the minimum $100/441$ and the maximum $7408/441$, the maximum reached only at image sums $(18, -12, -24)$ and $(-12, 18, 12)$, where the orderings that reach it are exactly σ_* and its reversal. That the two displayed joins evaluate to $7408/441$ and $100/441$ can be checked by hand from the formula for $\nu(\sigma)$. □

Conditions and scope.

Base and premises
Classical real numbers and finite probability; the census with images of length 7 and the nine counts above; all orderings of its three images.

Conclusion
Every ordering fluctuates at the cubic scale, $V_k/k^3 \to \nu(\sigma)$, and $\nu(\sigma)$ lies between $100/441$ and $7408/441$; the upper value is attained exactly by σ_* and its reversal.

Finite computation
The primitivity, the rate and the limit are proved on the page, and the values at the two displayed joins can be checked by hand. The bounds $100/441 \le \nu(\sigma) \le 7408/441$ and the uniqueness of the maximizers rest on the exhaustive computation in exact rationals over the 1573 triples of image sums that the proof describes and does not print.

The order of postings alone moves the cubic coefficient by a factor of 74.08: holds spread evenly keep the ledger close to its line, holds bunched at the ends of the images push it away. Which orderings are extremal for other neutral Jordan censuses, and when a zero coefficient is possible, are open (Problem 4); so is the positivity of the linear coefficient for unbalanced neutral three-letter ledgers (Problem 3). Two facts there are proved: balanced books have slope 0, and an affine prefix potential, $\delta(\text{prefix}) + H(b) = \mu H(a) + d$ before every letter b of every image $\sigma(a)$, keeps the books balanced.

30.4 The price of a certain change

Chapters 21 and 24 proved the quarter-turn floor: a word that moves a configuration of a register to a different one with certainty drives at least $\pi/2$, less its holding drift, and two entangling tokens attain it. That floor uses only the norms of the generators. The native register has more structure, and it raises the price in exact amounts. Table 30.1 collects the answers.

Definition 30.19 (The native register)**.** In classical mathematics:

1. A *channel* has the eight labels $0, \dots, 7$, read as three bits; the *low-bit partner* of a label flips its last bit, so the low-bit pairs are $\{0,1\},\{2,3\},\{4,5\},\{6,7\}$. A register of n channels has the configurations $u = (u_1, \dots, u_n)$ as orthonormal basis e_u.

2. For labels $i < j \le 6$ the *triad* $(i, j, 7)$ has the generator B with $Be_i = (e_j - e_7)/\sqrt{3}$, $Be_j = (e_7 - e_i)/\sqrt{3}$, $Be_7 = (e_i - e_j)/\sqrt{3}$ and $Be_a = 0$ for the other labels. It is real and antisymmetric, kills $e_i + e_j + e_7$, and has eigenvalues $0, \pm\mathrm{i}$ on the triad, so $\|B\| = 1$ and $e^{\theta B}$ rotates the triad by the angle θ about $(e_i + e_j + e_7)/\sqrt{3}$; $e^{2\pi B/3}$ is the cycle $e_i \mapsto e_j \mapsto e_7 \mapsto e_i$.

3. A *braid token* on channel k has generator θB acting on that channel, for a triad and an angle θ; with $\theta = 0$ it is *idle*. An *entangling token* on channels $k \ne l$ has generator $\frac{\pi}{4}\mathrm{i}S_kF_l$, where S_k multiplies e_u by $(-1)^{\text{number of bits 1 in } u_k}$ and F_l replaces u_l by its partner; channel l is its *target*.

Table 30.1. Prices of a certain change on the native register (drive in units of $\hbar$; B braid drive, N entangling tokens, D holding drift).

Change	Holding	Price or floor
three golden comparisons in a row	none	suffice; two never do
any change of one channel	none	exactly $2\pi/3$
channels changed to low-bit partners	none	exactly $\pi/2$ per channel
both channels of two	none	at least π; exactly π for the cheapest (two low-bit flips), also with entanglement
a channel leaving its low-bit pair	none	braid drive $\geq (\pi/2)\sqrt{3/2}$ there; $\geq 2\pi/3$ with one non-idle braid there
any change, any register	any	$\sqrt{2/3}\,B + N\pi/4 + D \geq \pi/2$
any change of one channel	diagonal	$\geq (\pi/2)\sqrt{3/2} \approx 1.924$
any change of one channel	scalar	exactly $2\pi/3$
one pulse, or repeated equal pulses	label seven	$\geq 2\pi/3$; exact pulses exist
any change	drift $\pi/2$	drive 0 suffices

4. A *word* w is a finite list of tokens. Its *drive* is $g(w) = \sum_t \|G_t\| = B(w) + \frac{\pi}{4}N(w)$, where $B(w)$ sums $|\theta|$ over the braid tokens and $N(w)$ counts the entangling tokens. Its *ideal matrix* is the ordered product of the e^{G_t}. Given a self-adjoint *holding* H and $\hbar, \varepsilon > 0$, the *paid word* W is the product of the paid tokens $e^{-\mathrm{i}\|G_t\|}\exp(G_t - \mathrm{i}\varepsilon H/\hbar)$, as in Chapter 24, and its *drift* is $D = |w|\,\varepsilon\|H\|/\hbar$. At zero holding W is the ideal matrix times a unit phase.

5. The word makes a *certain change* $u \to v$, $u \neq v$, when $|\langle e_v, We_u\rangle| = 1$, that is $We_u = ce_v$ with $|c| = 1$.

Replacing H by $(\varepsilon/\hbar)H$ changes neither W nor D, so statements made in units with $\hbar = \varepsilon = 1$ hold for all $\hbar$ and ε.

Every floor in this section comes from one estimate: an orthogonal projection can gain weight only through the part of the generator that crosses it.

Lemma 30.20 (Speed through a subspace)**.** *Let R be an orthogonal projection, $X_1, \dots, X_m$ skew-Hermitian matrices and ψ a unit vector, and put $\psi_m = e^{X_m}\cdots e^{X_1}\psi$. If $R\psi = 0$ and $\|R\psi_m\| = 1$, then*

$$\sum_{t=1}^{m} \|RX_t(1-R)\| \geq \frac{\pi}{2}.$$

Proof. Follow $\psi(s) = e^{sX_t}\psi_{t-1}$ for $s \in [0,1]$, token after token, and write $p = R\psi(s)$, $q = (1-R)\psi(s)$, $\kappa_t = \|RX_t(1-R)\|$. Since RX_tR is skew-Hermitian, $\frac{d}{ds}\|p\|^2 = 2\,\mathrm{Re}\langle p, RX_t(1-R)q\rangle$, so $|\frac{d}{ds}\|p\|^2| \leq 2\kappa_t\|p\|\|q\|$ with $\|p\|^2 + \|q\|^2 = 1$. Where $0 < \|p\| < 1$, the angle $A = \arcsin\|p\|$ therefore moves at rate at most κ_t. Let s_0 be the last time with $A = 0$ and s_1 the first later time with $A = \pi/2$; between them A climbs from 0 to $\pi/2$ at rate at most κ_t, so $\pi/2 \leq \sum_t \kappa_t$. □

A phase factor does not move $\|R\psi\|$, so the lemma applies to paid words with $X_t = G_t - \mathrm{i}\varepsilon H/\hbar$. With R the projection onto one basis vector e_v, $\|RX(1-R)\|$ is the norm of the off-diagonal part of row v of X. For a braid generator θB that row has two entries of size $|\theta|/\sqrt{3}$ when v lies in the triad, and norm $\sqrt{2/3}\,|\theta|$; for an entangling generator, a signed permutation without fixed points times $\frac{\pi}{4}$, it has one entry of size $\frac{\pi}{4}$.

Golden steps

Theorem 30.21 (Three golden steps change a record with certainty; two never do)**.** *In classical mathematics, consider a register of $n \geq 1$ channels at zero holding, any $\hbar, \varepsilon > 0$, and comparison controllers with reference action $2\pi\hbar$, as in Chapter 21, selecting words of this register. There are three controllers whose words at the ratio φ, run one after another, make a certain change. No two controllers' golden words in a row, and no single golden word, make a certain change.*

Proof. By the drive law of Chapter 21, every controller's word at ratio x has drive $2\pi J(x)$. Two golden words drive $2\pi\varphi^{-3} < \pi/2$, because $\varphi^{-3} = \sqrt{5} - 2 < \frac{1}{4}$, and one drives less; at zero holding the quarter-turn floor of Chapter 24, or Lemma 30.20 with the norms of the generators, excludes a certain change.

For the three, recall $J(\varphi) \geq \frac{1}{9}$ from the opening of this chapter. Let w_0 be the word: a braid token of angle $2\pi/9$ on a triad $(i, j, 7)$ of channel 1, then the braid tokens of angles t and $-t$ on the same triad, with $2t = 2\pi J(\varphi) - 2\pi/9 \geq 0$. It drives $2\pi J(\varphi)$ and its ideal matrix is $e^{2\pi B/9}$. The controller that selects, for records (d, r), two opposite braid tokens of angles $\pm\pi d/r$ satisfies both laws of a controller; change it at the golden records only, selecting w_0 there, which keeps every drive and hence both laws. Three of its golden words multiply to $e^{2\pi B/3}$, the cycle, which moves e_u with $u_1 = i$ to e_v, where v has $v_1 = j$ and agrees with u elsewhere. At zero holding the paid word is this matrix times a phase. □

One channel at zero holding: exactly a third of a turn

On a single channel the only tokens are braids. Each $e^{\theta B}$ is a real rotation fixing $\mathbb{1} = \sum_a e_a$, so the states $x = Ue_a$ reached from a basis vector are real and lie on the sphere

$$\mathcal{S} = \Big\{ x \in \mathbb{R}^8 \;:\; \sum_m x_m = 1,\ \sum_m x_m^2 = 1 \Big\},$$

and a certain change $a \to b$ ends exactly at e_b: its sign is fixed by $\sum_m (Ue_a)_m = \langle \mathbb{1}, Ue_a \rangle = \langle U^{\mathsf{T}}\mathbb{1}, e_a \rangle = 1$. For $x \in \mathcal{S}$ and a label a put

$$h_a(x) = \frac{8x_a - 1}{7} \in [-1, 1],$$

the cosine of the angle between $x - \mathbb{1}/8$ and $e_a - \mathbb{1}/8$, two vectors of squared length $\frac{7}{8}$.

Theorem 30.22 (Floors for one channel)**.** *In classical mathematics, on one channel at zero holding:*

(a) every word that makes a certain change drives at least $\arccos(-\frac{1}{7}) \approx 1.7141$*, and in fact at least* $\frac{\sqrt{21}}{4}\arccos(-\frac{1}{7}) \approx 1.9638$*;*

(b) a word whose nonidle braid tokens all use one triad, and a word of two braid tokens, drive at least $2\pi/3$ *if they make a certain change;*

(c) one braid token of angle $2\pi/3$ *makes a certain change.*

Proof. (a) The vectors $e_a - \mathbb{1}/8$ and $e_b - \mathbb{1}/8$ have inner product $-\frac{1}{8}$, so the angle between them is $\arccos(-\frac{1}{7})$. A rotation $e^{\theta B}$ moves a real vector y by an angle at most $|\theta|$, since $\langle y, e^{\theta B}y\rangle = \|y_0\|^2 + \cos\theta\|y_1\|^2$ with y_0 in the kernel of B and y_1 in its rotation plane, and angles between vectors obey the triangle inequality. That gives the first floor. For the second, let a unit drive turn the triad $(i, j, 7)$. If a is not in the triad, x_a does not move; if it is, $\dot{x}_a = \pm(x_p - x_q)/\sqrt{3}$ with p, q the other two triad labels. The seven amplitudes other than x_a have sum $1 - x_a$ and sum of squares $1 - x_a^2$; the square of a difference of two of them is at most twice their variance sum, so $(x_p - x_q)^2 \le \frac{4}{7}(1 - x_a)(3 + 4x_a)$, while $1 - h_a^2 = \frac{16}{49}(1 - x_a)(3 + 4x_a)$. Hence $|\dot{h}_a| \le \frac{8}{7\sqrt{3}}\sqrt{\frac{4}{7}(1 - x_a)(3 + 4x_a)} = \frac{4}{\sqrt{21}}\sqrt{1 - h_a^2}$, and $\arcsin h_a$ moves at rate at most $4/\sqrt{21}$ per unit drive. Along a certain change $a \to b$ it falls from $\arcsin 1 = \frac{\pi}{2}$ to $\arcsin(-\frac{1}{7})$, a total of $\arccos(-\frac{1}{7})$.

(b) Tokens on one triad commute, so the word is $e^{\Theta B}$ with $|\Theta|$ at most its drive. The diagonal entry of this rotation at a triad label is $(1 + 2\cos\Theta)/3$, and a certain change out of a label needs a zero diagonal entry there (outside the triad nothing moves), so $\cos\Theta = -\frac{1}{2}$ and $|\Theta| \ge 2\pi/3$. For two tokens on different triads, the intermediate state $x = e^{\alpha B'}e_a = e^{-\beta B''}e_b$ is supported both in $\{a\}\cup$ the first triad and in $\{b\}\cup$ the second. If a is outside the first triad, $x = e_a$ and the second token alone makes the change; similarly for b. Otherwise x is supported in the intersection of two distinct triads, at most two labels, and $x_1 + x_2 = 1 = x_1^2 + x_2^2$ forces x to be a basis vector: one of the two tokens alone makes a certain change, and pays $2\pi/3$.

(c) The cycle $e^{2\pi B/3}$ moves e_i to e_j. □

The gap between 1.964 and 2.094 is closed by a second potential, a phase. For labels $a \ne b$ other than 7 put

$$\ell(x) = \frac{x_a + x_b}{2} - x_7 + \mathrm{i}\,\frac{\sqrt{3}}{2}\,(x_b - x_a),$$

the complex coordinate of the projection of x onto the plane of the triad $(a, b, 7)$ orthogonal to its axis. It equals $e^{-\mathrm{i}\pi/3}$ at e_a and $e^{\mathrm{i}\pi/3}$ at e_b, and the rotation of that triad turns it at exactly unit rate. The question is whether other triads can turn it faster.

Lemma 30.23 (The phase turns no faster than its own triad)**.** *In classical mathematics, let $a \neq b$, put $K = \sqrt{21}/4$, $\vartheta_c(x) = \arccos h_c(x)$, and let $\mathscr{E}_{ab}$ be the set of $x \in \mathcal{S}$ with $K(\vartheta_a(x) + \vartheta_b(x)) < 2\pi/3$.*

(a) If $a, b \neq 7$, then on $\mathscr{E}_{ab}$ the real part of ℓ is positive and $|\mathrm{Im}(\ell(Bx)\,\overline{\ell(x)})| \leq |\ell(x)|^2$ for the generator B of every triad.

(b) If $b = 7$, the same holds for $\ell_k(x) = \frac{x_a + x_7}{2} - x_k + \mathrm{i}\frac{\sqrt{3}}{2}(x_7 - x_a)$ whenever x_k is the least amplitude among the labels other than a and 7.

Proof. Write $P = x_a$ and $Q = x_b$. Put $c_0 = -\frac{129}{500}$ and $L = 8\pi/(3\sqrt{21}) \approx 1.828$. On $\mathscr{E}_{ab}$, $\vartheta_a + \vartheta_b < L < \frac{\pi}{2} + 0.258$, so $\cos(\vartheta_a + \vartheta_b) > c_0$, and $\vartheta_b < \pi - \vartheta_a$ gives $h_a + h_b > 0$, that is $4(P + Q) > 1$. As $\cos(\vartheta_a + \vartheta_b) = h_a h_b - \sin\vartheta_a \sin\vartheta_b$, this gives $0 \leq \sin\vartheta_a \sin\vartheta_b < h_a h_b - c_0$; squaring, $h_a^2 + h_b^2 - 2c_0 h_a h_b > 1 - c_0^2$, which with $h = (8x - 1)/7$ reads

$$\Omega = 250000\big((8P-1)^2 + (8Q-1)^2\big) + 129000(8P-1)(8Q-1) - 11434591 > 0.$$

The other labels enter through the Cauchy–Schwarz bound $(\sum_{m\in U} x_m)^2 \leq |U| \sum_{m\in U} x_m^2$ for the set U of labels outside the form and the generator; with the two sums read off the sphere it says $C \geq 0$, where

$$C = |U|\Big(1 - \sum_{m\notin U} x_m^2\Big) - \Big(1 - \sum_{m\notin U} x_m\Big)^2;$$

in (b) the minimality $x_m \geq x_k$ enters as well.

For the generator of the triad $(i, j, 7)$ put $w = -\sqrt{3}\,Bx$, so $w_i = x_j - x_7$, $w_j = x_7 - x_i$, $w_7 = x_i - x_j$ and w vanishes elsewhere. After the factors $\sqrt{3}$ cancel, the rate inequality for the form with labels (a, b, c) reads

$$\big|(w_b - w_a)(x_a + x_b - 2x_c) - (w_a + w_b - 2w_c)(x_b - x_a)\big| \leq (x_a + x_b - 2x_c)^2 + 3(x_b - x_a)^2.$$

When the generator's triad is the form's own, the two sides are equal. Otherwise let F_+ and F_- be the right side plus and minus the expression inside the absolute value; we need $F_\pm \geq 0$, quadratic inequalities in at most five amplitudes. Exchanging a and b in (a) conjugates ℓ and changes neither side; reversing a generator changes the sign of w; exchanging the two outside labels of a disjoint generator changes the sign inside the absolute value. So (a) needs both signs for a generator sharing a, one sign for a disjoint generator, and the positivity of the real part; (b) needs both signs for a generator sharing a, both for one sharing k, and one for a disjoint generator. Write P, Q, R for x_a, x_b, x_c, so that $R = x_7$ in (a) and $Q = x_7$, $R = x_k$ in (b), and S, T for the amplitudes at the generator's labels outside the form.

Two of these are immediate. In (b) the real part is positive because the six amplitudes other than x_a, x_7 sum to $1 - P - Q \geq 6R$ and $4(P + Q) > 1$. For a generator sharing a in (b), one sign reads $F = 2(P - Q)^2 + 2(S - R)(P + Q - 2R) \geq 0$, which holds by minimality, $S \geq R$, and the positive real part.

Each of the others has the certificate of Table 30.2. In each row the quadratic polynomial $F - \lambda C - \kappa\Omega$ has a positive definite matrix in the basis formed by 1 and the amplitudes that occur,

Table 30.2. Certificates for Lemma 30.23. Each row gives an inequality $F \ge 0$, the Cauchy–Schwarz bound C (on $|U|$ outside labels) and multipliers λ, κ such that $F - \lambda C - \kappa\Omega$ has a positive definite matrix; the last column is its least Gaussian pivot, rounded down. The vector $w = (w_a, w_b, w_c)$ is that of the generator, read on the form's labels.

Row	Case	F	$\lvert U\rvert$	λ	κ	Pivot
1	(a), real part	$2g\big(\frac{P+Q}{2}\big) - \frac{1}{50}$	—	0	$8 \cdot 10^{-7}$	0.62
2	(a), shares a	F_-, $w = (S-R,\ 0,\ P-S)$	4	1.5	$4.6 \cdot 10^{-7}$	0.10
3	(a), shares a	F_+, $w = (S-R,\ 0,\ P-S)$	4	1.5	$4.6 \cdot 10^{-7}$	0.47
4	(a), disjoint	F_+, $w = (0,\ 0,\ S-T)$	3	3	$7 \cdot 10^{-7}$	1.56
5	(b), shares a	F_+, $w = (S-Q,\ P-S,\ 0)$	4	1.5	$4.6 \cdot 10^{-7}$	0.47
6	(b), shares k	F_+, $w = (0,\ R-S,\ S-Q)$	4	1.5	$4.6 \cdot 10^{-7}$	0.47
7	(b), shares k	F_-, $w = (0,\ R-S,\ S-Q)$	4	1.5	$4.6 \cdot 10^{-7}$	0.10
8	(b), disjoint	F_-, $w = (0,\ S-T,\ 0)$	3	3	$7 \cdot 10^{-7}$	1.56

in the order $1, P, Q, R, S, T$: Gaussian elimination in that order, in exact rational arithmetic, gives positive pivots, the least of which the table lists. Hence $F > \lambda C + \kappa\Omega > 0$ on $\mathscr{E}_{ab}$. Row 1 serves the real part in (a). Cauchy–Schwarz on the five labels outside $a, b, 7$ says $g(x_7) \le 0$, where

$$g(t) = 6t^2 - 2(1-P-Q)\,t + (1-P-Q)^2 - 5(1-P^2-Q^2).$$

The quadratic g increases for $t \ge (1-P-Q)/6$, a point below $(P+Q)/2$ because $4(P+Q) > 1$, and row 1 gives $2g\big(\frac{P+Q}{2}\big) > \frac{1}{50}$. So $g > 0$ on $[\frac{P+Q}{2}, \infty)$, and $x_7 < \frac{P+Q}{2}$, which says $\operatorname{Re}\ell > 0$. The narrowest margins are in rows 2 and 7, generators sharing one label with the form; there the bound on the two end angles, through Ω, carries the proof. □

Theorem 30.24 (One channel: a certain change costs exactly a third of a turn)**.** *In classical mathematics: on one channel at zero holding, every word that makes a certain change drives at least $2\pi/3$, and one braid token of angle $2\pi/3$ attains it. Consequently there is no certain change of one channel below $2\pi/3$, in particular none that switches between triads.*

Proof. Suppose a word of drive $g < 2\pi/3$ makes a certain change $a \to b$. Reversing it (the inverse tokens in reverse order) gives a change $b \to a$ of the same drive, so we may take $a \ne 7$. By the proof of Theorem 30.22(a), $K\vartheta_a$ moves at rate at most 1 per unit drive; so at every intermediate state x, inside tokens included, $K\vartheta_a(x)$ is at most the drive spent and $K\vartheta_b(x)$ at most the drive remaining. Every state therefore lies in $\mathscr{E}_{ab}$.

If $b \ne 7$, the phase $\Phi = \arg\ell$ is continuous on $\mathscr{E}_{ab}$ by Lemma 30.23(a), and along a braid token $\dot{x} = \theta Bx$ it moves at rate $\theta\operatorname{Im}(\ell(Bx)\overline{\ell(x)})/|\ell(x)|^2$, at most $|\theta|$. So Φ changes by at most g; but it runs from $-\pi/3$ at e_a to $\pi/3$ at e_b. If $b = 7$, use $\Phi(x) = \arg\big(\frac{x_a+x_7}{2} - \min_k x_k + \mathrm{i}\frac{\sqrt{3}}{2}(x_7 - x_a)\big)$, the minimum over the labels other than a and 7. At every point some k attains the minimum, the right derivative of Φ there is that of $\arg\ell_k$ for such a k, and Lemma 30.23(b) bounds it by $|\theta|$; a continuous function whose right derivatives are so bounded moves at rate at most $|\theta|$. The endpoint values are again $\mp\pi/3$, because the minimum is 0 at e_a and at e_7. Either way $g \ge 2\pi/3$, a contradiction. Theorem 30.22(c) gives attainment. □

Conditions and scope.

Base and premises
Classical real and complex matrices; one channel of Definition 30.19; zero holding (or any scalar holding, which only adds a phase).

Conclusion
The cheapest certain change of one eight-state channel costs exactly $2\pi/3$, a third of a turn, and the triad cycle is optimal: in the region that a cheaper word could reach, no triad turns the phase faster than the cycle's own.

Identification
That these tokens are the operations of Recognition Science's register, and that drive measures action in units of $\hbar$.

Not claimed
The price under a nonscalar holding; see below and Section 30.5.

Two conjectures fall with this theorem: that some unentangled word changes a channel with certainty below $2\pi/3$, and that a word switching between triads, with drive in $[\pi/2, 2\pi/3)$, does so. Both are false. The cycle plays the role of a geodesic: inside the region a cheaper word would have to cross, nothing turns faster.

Many channels: a quarter turn per channel, and what it buys

On several channels the entangling tokens become available. Assign each braid token to its channel and each entangling token to its target; the *channel drive* of channel k is the drive of the tokens assigned to it, and the channel drives add up to the drive of the word.

Theorem 30.25 (Drive is additive over changed channels)**.** *In classical mathematics, at zero holding on a register of n channels:*

(a) every word that makes a certain change $u \to v$ has channel drive at least $\pi/2$ on every channel k with $u_k \neq v_k$; hence its drive is at least $\frac{\pi}{2} \cdot \#\{k : u_k \neq v_k\}$;

(b) on two channels, a certain change of both labels drives at least π, and π is the least drive of a certain change of both labels by a word containing an entangling token. So entanglement does not bring a double change below π.

Proof. (a) Apply Lemma 30.20 with R the projection onto the configurations whose channel k shows a label other than u_k: it starts at 0 and ends at 1. A braid token on another channel commutes with R. An entangling token whose target is not k commutes with R as well, because its sign operator is diagonal and its flip acts on its target alone. So only the tokens assigned to channel k contribute, each at most its norm. Summing over the changed channels counts every token at most once.

(b) The floor is (a) with two changed channels. Two entangling tokens on $(1, 2)$ have ideal product $\mathrm{i}S_1F_2$, which flips the low bit of channel 2; two more on $(2, 1)$ flip channel 1. The four tokens drive π and move every configuration to the one with both low bits flipped, up to a phase. □

Theorem 30.26 (Exactly the low-bit masks attain the quarter turn)**.** *In classical mathematics, at zero holding on a register of $n \geq 2$ channels: a certain change $u \to v$ can be made with drive exactly $\frac{\pi}{2} \cdot \#\{k : u_k \neq v_k\}$ if and only if every channel either keeps its label or moves to its low-bit partner. If channel k moves to any other label, its channel drive is at least $\frac{\pi}{2} + \big(1 - \sqrt{2/3}\big)\frac{\pi}{2} \approx 1.859$.*

Proof. Let channel k move to a label v_k outside the low-bit pair $\{u_k, u_k'\}$. Let β_k be the braid drive on channel k and ϵ_k the drive of the entangling tokens targeting k. First, take R the projection onto "channel k shows a label outside $\{u_k, u_k'\}$". Entangling tokens commute with it, since F_k preserves the pair; so Lemma 30.20 charges only braid tokens on k, each at most its norm, and $\beta_k \geq \pi/2$. Second, take R the projection onto "channel k shows v_k". An entangling token targeting k contributes at most $\frac{\pi}{4}$; a braid token θB on k contributes at most the norm of the off-diagonal part of row v_k of θB, which is $\sqrt{2/3}\,|\theta|$. So $\epsilon_k + \sqrt{2/3}\,\beta_k \geq \pi/2$, and

$$\epsilon_k + \beta_k \geq \frac{\pi}{2} + \big(1 - \sqrt{2/3}\big)\beta_k \geq \frac{\pi}{2} + \big(1 - \sqrt{2/3}\big)\frac{\pi}{2}.$$

If the drive equals $\frac{\pi}{2}$ times the number of changed channels, then by Theorem 30.25(a) every changed channel has channel drive exactly $\pi/2$, which this excludes. Conversely, if every changed channel k moves to its partner, pick a channel $l \neq k$ and two entangling tokens on (l, k); the product over the changed channels is a signed permutation times a phase that moves e_u to e_v, with drive $\pi/2$ per changed channel. □

A changed channel that leaves its low-bit pair pays even more in its own braids, whatever the entangling tokens do.

Proposition 30.27 (Leaving a low-bit pair)**.** *In classical mathematics, at zero holding on any register: if a word makes a certain change $u \to v$ in which v_k is neither u_k nor its low-bit partner, its braid tokens on channel k drive at least $\frac{\pi}{2}\sqrt{3/2} \approx 1.924$.*

Proof. Take R as in the first step of the previous proof. For a braid generator B on channel k, the block $RB(1 - R)$ maps the pair into the rest of its triad. If the triad meets the pair in one label p, the block is the off-diagonal part of column p, of norm $\sqrt{2/3}$; if the triad contains both labels of the pair, the block has rank one, sending e_p and $e_{p'}$ to $\pm\frac{1}{\sqrt{3}}$ times the third label, again of norm $\sqrt{2/3}$; otherwise it is 0. Lemma 30.20 gives $\sqrt{2/3}\,\beta_k \geq \pi/2$. □

So a non-partner change costs its channel at least $\frac{\pi}{2}\sqrt{3/2}$, an excess of $(\sqrt{3/2} - 1)\frac{\pi}{2} \approx 0.353$ over the quarter turn. The natural guess is that it costs $2\pi/3$, the one-channel price.

The rate argument forgets that a token turns through a finite angle. One braid token cannot carry a state further from the pair than the angle between the pair and its image. Put

$$\beta(\theta) = \arccos\left|\frac{1 + 2\cos\theta}{3}\right| \in \left[0, \frac{\pi}{2}\right],$$

the angle whose cosine is the modulus of a diagonal entry of $e^{\theta B}$ at a label of its triad.

Theorem 30.28 (Leaving a low-bit pair: an angle budget)**.** *In classical mathematics, at zero holding on any register, let a word make a certain change $u \to v$ in which v_k is neither u_k nor its low-bit partner.*

(a) The angles θ of its braid tokens on channel k satisfy $\sum \beta(\theta) \geq \pi/2$.

(b) $\beta(\theta) \leq \sqrt{2/3}\,|\theta|$ for every real θ; so (a) implies Proposition 30.27.

(c) If all braid tokens on channel k but one are idle, the remaining one has $|\theta| \geq 2\pi/3$: the braid drive on channel k is at least $2\pi/3$.

Proof. (a) Let Q be the projection onto the configurations whose channel k shows u_k or its partner, and for a unit vector ψ let $\alpha(\psi) \in [0, \pi/2]$ be the angle with $\cos\alpha(\psi) = \|Q\psi\|$. Entangling tokens and tokens on other channels commute with Q and leave α unchanged. Let $T = e^{\theta B}$ act on channel k, with $r = |1 + 2\cos\theta|/3$ the modulus of its diagonal entries on the triad and $s = \sqrt{1 - r^2} = \sin\beta(\theta)$. Both blocks $(1-Q)TQ$ and $QT(1-Q)$ have norm at most s. If the triad misses the pair they vanish. If it contains one label p of the pair, T fixes the partner, and both blocks are carried by label p: their norms are those of column and row p of T without the diagonal entry, $\sqrt{1 - r^2}$, since the rows and columns of T are unit vectors. If it contains both labels, the third label c of the triad carries both blocks in the same way, with norm $\sqrt{1 - r^2}$. Hence for a unit ψ with $\alpha(\psi) = \alpha$,

$$\|QTQ\psi\|^2 = \|Q\psi\|^2 - \|(1-Q)TQ\psi\|^2 \geq (1 - s^2)\cos^2\alpha, \qquad \|QT(1-Q)\psi\| \leq s\sin\alpha,$$

so $\|QT\psi\| \geq \cos\beta\cos\alpha - \sin\beta\sin\alpha = \cos(\alpha + \beta)$ when $\alpha + \beta \leq \pi/2$: the token moves α by at most $\beta(\theta)$. The word starts at $\alpha(e_u) = 0$ and ends at $\alpha = \pi/2$, because $QWe_u = cQe_v = 0$. If the angles summed to less than $\pi/2$, induction over the tokens would give $\|QWe_u\| \geq \cos\sum\beta(\theta) > 0$.

(b) If $\sqrt{2/3}\,|\theta| \geq \pi/2$ there is nothing to show. Otherwise $x = |\theta| < \pi$ and it suffices that $g(x) = 1 + 2\cos x - 3\cos(\sqrt{2/3}\,x) \geq 0$. Now $g(0) = 0$ and $g'(x) = \sqrt{6}\sin(\sqrt{2/3}\,x) - 2\sin x \geq 0$ on $[0, \pi]$, because sin is concave there, so $\sin(\lambda x) \geq \lambda\sin x$ for $\lambda \in [0, 1]$, and $\sqrt{6}\sqrt{2/3} = 2$.

(c) An idle token has $\beta(0) = 0$, so by (a) the remaining token has $\beta(\theta) \geq \pi/2$, that is $\cos\theta = -\frac{1}{2}$, and $|\theta| \geq 2\pi/3$. □

So one braid token cannot take a channel off its pair below the one-channel price, whatever else the word does; a cheaper departure interleaves several.

Corollary 30.29 (What a sharp pair floor needs)**.** *In classical mathematics, the following are equivalent:*

(i) for every $\eta > 0$ some register of $n \geq 2$ channels has a word at zero holding that makes a certain change moving channel k off its low-bit pair, with braid drive on channel k below $\frac{\pi}{2}\sqrt{3/2} + \eta$;

(ii) the same, where moreover the word has braid drive below $2\pi/3$ on channel k, at least two braid tokens there that are not idle, and angles with $\frac{\pi}{2} \leq \sum\beta(\theta) \leq \sqrt{2/3}\sum|\theta|$.

Proof. (ii) gives (i) at once. Given (i) and $\eta > 0$, apply (i) with $\eta' = \min\bigl(\eta, \frac{1}{2}(\frac{2\pi}{3} - \frac{\pi}{2}\sqrt{3/2})\bigr) > 0$. The braid drive of the word on channel k is below $2\pi/3$. By Theorem 30.28(a) its angles there sum to at least $\pi/2$, so some braid token on channel k is not idle, and by part (c) not only one; so at least two are not idle, and parts (a) and (b) give the angle inequalities. □

Proposition 30.30 (Every channel price is attained)**.** *In classical mathematics, let $\varpi(a, b)$ be 0 if $a = b$, $\pi/2$ if b is the partner of a, and $2\pi/3$ otherwise. On a register of $n \geq 2$ channels at zero holding, every $u \to v$ is a certain change of some word with drive*

$$\sum_k \varpi(u_k, v_k) = \frac{\pi}{2} \cdot \#\{k \,:\, u_k \neq v_k\} + \frac{\pi}{6} \cdot \#\{k \,:\, v_k \notin \{u_k, u_k'\}\}.$$

On two channels, a flip of one low bit costs exactly $\pi/2$.

Proof. Change one channel at a time. A partner costs two entangling tokens, $\pi/2$. Any two labels lie in a common triad, and on a triad the cycle $e^{2\pi B/3}$ and its inverse together move each label to each other one; that costs $2\pi/3$. Certain transfers compose: if $W_1 e_u = c_1 e_{u'}$ and $W_2 e_{u'} = c_2 e_{u''}$, then $W_2 W_1 e_u = c_1 c_2 e_{u''}$. The last claim combines the construction with Theorem 30.25(a). □

Whether these prices are also lower bounds, that is whether every higher-label change pays the extra sixth of a turn, is open (Problem 10).

Entangling tokens can also reach one channel through a second one that starts and ends at label 0, a returned helper. On two channels, a word *implements the parity phase γ* when for some unit ζ its ideal matrix sends $e_{(a,0)}$ to $\zeta e^{\mathrm{i}\gamma s(a)} e_{(a,0)}$ for every label a, where $S e_a = s(a) e_a$, $s(a) = \pm 1$. Which phases can the helper deliver when every entangling token has the helper as its target?

Theorem 30.31 (One crossing direction gives exactly the eighth turns)**.** *In classical mathematics, on two channels, let γ be real. Some word made of braid tokens on channel 2 and entangling tokens on the channels $(1, 2)$, with target 2, implements the parity phase γ if and only if γ is an integer multiple of $\pi/4$.*

Proof. Only multiples of $\pi/4$. Let J sum over the helper: $J(x \otimes y) = (\sum_m y_m)\, x$, from the register to channel 1. A braid matrix $e^{\theta B}$ is real orthogonal and fixes $\mathbb{1} = \sum_m e_m$, because $B\mathbb{1} = 0$; so its transpose fixes $\mathbb{1}$ too, and J is unchanged by a braid token on channel 2. The entangling token has ideal matrix $X = (1 + \mathrm{i} S_1 F_2)/\sqrt{2}$, since $(S_1 F_2)^2 = 1$, and F permutes labels, so $JX = \kappa J$ with $\kappa = (1 + \mathrm{i}S)/\sqrt{2}$, which multiplies e_a by $e^{\mathrm{i}\pi s(a)/4}$. Token by token, $JW = \kappa^N J$ for a word with N entangling tokens. Applied to $e_{(a,0)}$ for $a = 0$ and $a = 1$, of signs $+1$ and -1, an implementation gives $\zeta e^{\pm \mathrm{i}\gamma} = e^{\pm \mathrm{i}\pi N/4}$. Dividing, $e^{2\mathrm{i}\gamma} = e^{\mathrm{i}\pi N/2}$, so $\gamma - N\pi/4$ is a multiple of π.

Every multiple of $\pi/4$. A braid token on a triad $(p, q, 7)$ carries a real vector y to a real vector z whenever z agrees with y off the triad and has the same triad sum and the same triad sum of squares: the rotation fixes the component along the axis, which the sum determines, and turns the orthogonal component, of length fixed by the two sums, to any direction in its plane. Put $c_\pm = \frac{1 \pm \sqrt{3}}{4}$, $\tau = \sqrt{(1 - c_+)(1 + 3c_+)}$ and

$$p_1 = c_+ e_0 + \tfrac{1}{2}(1 - c_+ + \tau)\, e_1 + \tfrac{1}{2}(1 - c_+ - \tau)\, e_7, \qquad v = c_+(e_0 + e_1) + c_-(e_6 + e_7).$$

All three of e_0, p_1, v have sum 1 and sum of squares 1, because $2c_+ + 2c_- = 1$ and $2c_+^2 + 2c_-^2 = 1$. So a braid token on the triad $(0,1,7)$ carries e_0 to p_1, and one on $(1,6,7)$ carries p_1 to v; call their product R. The vector v is fixed by F, so $X(e_a \otimes v) = e^{i\pi s(a)/4} e_a \otimes v$, and the word R, the entangling token, R^{-1} (the two braid tokens reversed, with opposite angles) sends $e_{(a,0)}$ to $e^{i\pi s(a)/4} e_{(a,0)}$: it implements $\pi/4$ with $\zeta = 1$. Repeating it m times implements $m\pi/4$, and $m \in \{0, \dots, 7\}$ covers every multiple of $\pi/4$ modulo 2π. □

The obstruction is the helper's all-ones component, which no braid rotates and which counts only the crossings. With both directions it no longer applies; whether every phase is then reachable is Problem 23.

Holding

A register holds its configuration under a holding H while tokens run, and the holding acts as a free generator of size $\varepsilon\|H\|/\hbar$ per token. The quarter-turn floor already charges it in full. The finer floors are the following; we use units with $\hbar = \varepsilon = 1$, which by Definition 30.19 loses nothing.

Theorem 30.32 (The separated action bound)**.** *In classical mathematics: on every register, under every self-adjoint holding H, every word w that makes a certain change satisfies*

$$\frac{\pi}{2} \le \sqrt{2/3}\, B(w) + \frac{\pi}{4} N(w) + D, \qquad D = |w|\,\|H\|,$$

while its drive is $g(w) = B(w) + \frac{\pi}{4} N(w)$. Consequently a word with no entangling token and drift $D < \pi/4$ drives at least $7\pi/24$, and a word that drives less than $\pi/2$ contains at most one entangling token.

Proof. Apply Lemma 30.20 with R the projection onto e_v and $X_t = G_t - iH$. The off-diagonal row of X_t has norm at most that of G_t plus $\|H\|$, and the row norms of braid and entangling generators are $\sqrt{2/3}\,|\theta|$ (or 0) and $\frac{\pi}{4}$. With no entangling token and $D < \pi/4$, $\sqrt{2/3}\,B > \pi/4$, so $B > \frac{\pi}{4}\sqrt{3/2} > \frac{\pi}{4} \cdot \frac{7}{6} = \frac{7\pi}{24}$. Two entangling tokens alone drive $\pi/2$. □

At drift $\pi/4$ one entangling token flips a low bit for drive $\pi/4$ (Problem 21). Just below that drift a flip for little more than $\pi/4$ needs one entangling token and a little braid drive; with two tokens the word is forced.

Proposition 30.33 (Two-token flips below a quarter turn of holding)**.** *In classical mathematics, on two channels in units $\hbar = \varepsilon = 1$:*

(a) let a word of two tokens, exactly one of them entangling and every braid angle below $\pi/4$ in absolute value, under a self-adjoint holding H with drift $2\|H\| = \pi/4 - \delta$, $\delta > 0$, move the configuration $(0,0)$ to $(1,0)$ with certainty. Then the word is one entangling token on the channels $(2,1)$, with target 1, and one braid token on channel 1 on the triad $(0,1,7)$, in either order;

(b) consequently the following are equivalent: (i) there are $K > 0$ and $\delta_0 \in (0, \pi/4)$ such that for every $\delta \in (0, \delta_0)$ some two-token word with exactly one entangling token, under some self-adjoint holding of drift $\pi/4 - \delta$, moves $(0,0)$ to $(1,0)$ with certainty with braid drive at most $K\sqrt{\delta}$; (ii) the same with the word restricted to the two words of (a).

Proof. (a) Let R project onto the configurations whose channel 1 shows 1, and R' onto those whose channel 1 does not show 0. The change starts outside the range of each and ends inside it. Lemma 30.20 with $X_t = G_t - \mathrm{i}H$ and $\|RX_t(1-R)\| \le \|RG_t(1-R)\| + \|H\|$ gives $\sum_t \|RG_t(1-R)\| \ge \frac{\pi}{2} - 2\|H\| = \frac{\pi}{4} + \delta$, and likewise for R'. A token that commutes with the projection contributes 0, the entangling token at most $\pi/4$ and a braid token at most $|\theta| < \pi/4$; so the entangling token does not commute with R, which excludes the token with target 2, whose action on channel 1 is a diagonal sign. Then the braid token contributes at least $\delta > 0$ for R and for R'. A braid token on channel 2 commutes with both; on channel 1, one whose triad misses label 1 commutes with R, and one whose triad misses label 0 commutes with R'. The only triad containing 0 and 1 is $(0, 1, 7)$.

(b) The two words of (a) have one entangling token, braid drive $|\theta|$ and drift $2\|H\|$, so (ii) gives (i). Given (i), shrink δ_0 until $K\sqrt{\delta_0} < \pi/4$; then every braid angle of a witness is below $\pi/4$, and (a) applies. □

Theorem 30.34 (Diagonal holding)**.** *In classical mathematics: on one channel, under any holding diagonal in the labels and for any $\hbar, \varepsilon > 0$, every word that makes a certain change drives at least $\frac{\pi}{2}\sqrt{3/2} \approx 1.924$.*

Proof. A diagonal holding adds nothing to the off-diagonal row of X_t, so Lemma 30.20 with R onto e_v gives $\sqrt{2/3}\, g(w) \ge \pi/2$. □

A scalar holding $H = s \cdot 1$ multiplies every paid token by a phase, so it leaves the one-channel price at exactly $2\pi/3$ (Theorem 30.24). Between the scalar case and the floor of Theorem 30.34 lies the simplest nonscalar holding: $H = P_7$, the projection onto the label 7 of one channel. One braid pulse on the triad $(i, j, 7)$ under it, for a duration $\varepsilon > 0$, is up to a phase $\exp(\theta B - \mathrm{i}\varepsilon P_7)$, which is the identity off the triad and on the triad, in the coordinates $(i, j, 7)$, the matrix $A = \theta B - \mathrm{i}\varepsilon E$, E the projection onto the third coordinate. For a real root x of

$$p(x) = x^3 - \varepsilon x^2 - \theta^2 x + \tfrac{1}{3}\theta^2 \varepsilon,$$

a direct computation shows that $\omega(x) = (\theta^2 - \mathrm{i}\sqrt{3}\,\theta x,\ \theta^2 + \mathrm{i}\sqrt{3}\,\theta x,\ \theta^2 - 3x^2)$ satisfies $A\omega(x) = -\mathrm{i}x\,\omega(x)$. Write $\gamma(x) = x - 2\arctan(\sqrt{3}x/\theta)$ for $\theta \ne 0$.

Theorem 30.35 (One pulse under label-seven holding)**.** *In classical mathematics, on one channel with $H = P_7$, $\hbar = 1$ and $\varepsilon > 0$:*

(a) if one braid pulse of angle θ makes a certain change, then $|\theta| \ge 2\pi/3$, and the change exchanges the two labels of its triad other than 7;

(b) a certain single pulse exists: there are θ and $\varepsilon > 0$ for which the pulse makes a certain change;

(c) a word of two pulses, one of which has angle 0, drives at least $2\pi/3$ if it makes a certain change.

Proof. (a) A certain change must stay in the triad. Since $U = e^A$ commutes with A, $Ue_a = ce_b$ gives $\langle e_a, Ae_a\rangle = \langle Ue_a, AUe_a\rangle = \langle e_b, Ae_b\rangle$; the diagonal of A is $(0, 0, -\mathrm{i}\varepsilon)$, so $\{a, b\} = \{i, j\}$, say

$a = i$ and $b = j$. For every real root x of p, $\langle e_j, U\omega(x)\rangle = e^{-\mathrm{i}x}\omega_j(x)$ and also $= \langle U^* e_j, \omega(x)\rangle = c\,\omega_i(x)$. With $y = \sqrt{3}x/\theta$ this says $c(1 - \mathrm{i}y) = e^{-\mathrm{i}x}(1 + \mathrm{i}y)$, that is $c = e^{-\mathrm{i}\gamma(x)}$. Two roots therefore have γ-values differing by a multiple of 2π. Let $t = |\theta| < 2\pi/3$ and $s = t/\sqrt{3}$. From $p(-t) = -\frac{2}{3}\varepsilon t^2 < 0$, $p(-s) = 2s^3 > 0$, $p(0) > 0$ and $p(s) = -2s^3 < 0$, there are roots $-\alpha \in (-t, -s)$ and $\beta \in (0, s)$. If $\theta > 0$: arctan is concave on $[0, \sqrt{3}]$, so $\arctan y \geq \frac{\pi}{3} \cdot \frac{y}{\sqrt{3}}$ there, and $2\arctan(\sqrt{3}\alpha/t) - \alpha \geq (\frac{2\pi}{3t} - 1)\alpha > 0$; also it is below $2\pi/3$. So $\gamma(-\alpha) \in (0, \frac{2\pi}{3})$, likewise $\gamma(\beta) \in (-\frac{2\pi}{3}, 0)$, and their difference lies in $(0, \frac{4\pi}{3})$. If $\theta < 0$: $\gamma(\beta) > 0 > \gamma(-\alpha)$, and the difference is $\alpha + \beta + 2\arctan(\sqrt{3}\alpha/t) + 2\arctan(\sqrt{3}\beta/t)$. The roots of p sum to ε and multiply to $-\frac{1}{3}\theta^2\varepsilon$, so $\alpha\beta(\varepsilon + \alpha - \beta) = \frac{1}{3}t^2\varepsilon$ and, as $\alpha > \beta$, $\alpha\beta < t^2/3$. Then $\frac{3\alpha\beta}{t^2} < 1$ puts the two arctangents below $\frac{\pi}{2}$ in sum, and $3\alpha(\alpha + \beta) < 3\alpha^2 + t^2 < 4t\alpha$ (as $\frac{t}{3} < \alpha < t$) gives $\alpha + \beta < \frac{4t}{3} < \frac{8\pi}{9}$. The difference lies in $(0, 2\pi)$. In both cases no multiple of 2π fits; and $\theta = 0$ gives a diagonal U.

(b) Conversely, if p has three distinct roots whose γ-values agree modulo 2π, put $c = e^{-\mathrm{i}\gamma}$. Then $U^* e_j - \bar{c} e_i$ is orthogonal to the three eigenvectors $\omega(x)$, which form a basis, so $|\langle e_j, U e_i\rangle| = 1$. Take $\theta = -T$ with $T > 0$, so that $\gamma(x) = x + 2\arctan(\sqrt{3}x/T)$, and scale $x = Ty$, $\varepsilon = \mu T$: the roots solve $y^3 - \mu y^2 - y + \frac{\mu}{3} = 0$. For $s \in [\frac{1}{4}, \frac{1}{2}]$ let $\mu(s) = 3s(s^2 - 1)/(3s^2 - 1) > 0$, so that s is a root, and let the other roots be $y_\pm = \frac{1}{2}\big(\mu - s \pm \sqrt{\Delta}\big)$ with $\Delta = \mu^2 + 2\mu s - 3s^2 + 4$; then $y_- < s < y_+$. With $\mathscr{A}(y) = \arctan(\sqrt{3}y)$, the choice $T(s) = \big(2\pi - 2(\mathscr{A}(s) - \mathscr{A}(y_-))\big)/(s - y_-) > 0$ makes the γ-values of Ty_- and Ts differ by exactly 2π, and $F(s) = T(s)(y_+ - s) + 2(\mathscr{A}(y_+) - \mathscr{A}(s)) - 2\pi$ measures the other gap. At $s = \frac{1}{4}$: $\mu = \frac{45}{52}$, $\Delta = \frac{844}{169}$, $\frac{24}{13} < \sqrt{\Delta} < 3$, and $\mathscr{A}(\frac{1}{4}) \geq \frac{\pi}{8}$, $\mathscr{A}(y_-) \leq -\frac{\pi}{4}$, $\mathscr{A}(y_+) \leq \frac{5\pi}{12}$ (using $\tan\frac{5\pi}{12} = 2 + \sqrt{3}$); the first two bound the numerator of T by $\frac{5\pi}{4}$, and $\sqrt{\Delta} > \frac{24}{13}$ bounds $\frac{5}{4}(y_+ - \frac{1}{4})/(\frac{1}{4} - y_-)$ by $\frac{17}{12}$, so $F(\frac{1}{4}) < \frac{17\pi}{12} + \frac{7\pi}{12} - 2\pi = 0$. At $s = \frac{1}{2}$: $\mu = \frac{9}{2}$, $\Delta = 28$, and $\mathscr{A}(\frac{1}{2}) \leq \frac{\pi}{4}$, $\mathscr{A}(y_-) \geq -\frac{\pi}{3}$, $\mathscr{A}(y_+) \geq \frac{\pi}{3}$ bound the numerator of T below by $\frac{5\pi}{6}$, and $\sqrt{\Delta} < 8$ gives $\frac{5}{6}(y_+ - \frac{1}{2})/(\frac{1}{2} - y_-) > \frac{11}{6}$, so $F(\frac{1}{2}) > \frac{11\pi}{6} + \frac{\pi}{6} - 2\pi = 0$. By the intermediate value theorem $F(s_*) = 0$ for some s_*; the three roots at s_* then have γ-values agreeing modulo 2π. Numerically $s_* \approx 0.3672$, $T \approx 3.0216$ and $\varepsilon \approx 4.8358$; only the existence is proved.

(c) The idle pulse is a diagonal unitary, which leaves the moduli of all entries unchanged; so the word has the entry moduli of the other pulse alone, and (a) applies. □

So the conjecture that a positive label-seven holding forbids every exact single pulse is false; but every exact pulse costs at least the cycle, and the holding turns the cycle's continuous family of exact transfers into isolated resonances. Whether the cheapest resonance is attained, and where, is open (Problem 25); numerical searches find one at $|\theta| \approx 2.7071$, $\varepsilon \approx 8.4221$, below the large-detuning limit $\pi\sqrt{3}/2 \approx 2.7207$.

Below that limit an exact pulse cannot drift off to long durations.

Theorem 30.36 (Cheap exact pulses have bounded duration)**.** *In classical mathematics, on one channel with $H = P_7$ and $\hbar = 1$: if one braid pulse of angle θ on the triad $(i, j, 7)$, for a duration $\varepsilon > 0$, moves e_i to e_j with certainty, then*

$$\varepsilon^2\Big(\frac{\pi\sqrt{3}}{2} - |\theta|\Big) \leq 100.$$

Proof. Write $t = |\theta|$ and $L = \pi\sqrt{3}/2$. If $t \geq L$ there is nothing to show. By Theorem 30.35(a), $t \geq 2\pi/3$, so $L - t < 1$, and for $\varepsilon < 4$ the left side is below 16. Let $\varepsilon \geq 4$ and $t < L$, so $t^2 < 3\pi^2/4 < 15/2$.

Roots. The cubic p of Theorem 30.35 has $p(0) = t^2\varepsilon/3 > 0$ and $p(\varepsilon) = -2t^2\varepsilon/3 < 0$, and with $h = 4t^2/(3\varepsilon)$ a direct expansion gives $p(\varepsilon + h) = \frac{2}{3}t^2\varepsilon + \frac{5}{3}t^2h + h^3 > 0$. So p has roots $b < 0 < a < \varepsilon < c$ with $\varepsilon(c - \varepsilon) < 4t^2/3$, and $a + b + c = \varepsilon$, $abc = -t^2\varepsilon/3$.

The phase condition. By the proof of Theorem 30.35(a), the two roots have $\gamma(a) - \gamma(b) \in 2\pi\mathbb{Z}$, that is $g \mp 2(\arctan A - \arctan B) \in 2\pi\mathbb{Z}$ with $g = a - b > 0$, $A = \sqrt{3}a/t > 0 > B = \sqrt{3}b/t$, and the sign $\mp$ that of θ. Here $-AB = -3ab/t^2 = \varepsilon/c < 1$, and for positive u, v with $uv < 1$, $\arctan u + \arctan v = \frac{\pi}{2} - \arctan\frac{1-uv}{u+v}$; so $\arctan A - \arctan B = \frac{\pi}{2} - \arctan\rho$ with

$$\rho = \frac{1 + AB}{A - B} = \frac{t(c - \varepsilon)}{\sqrt{3}\, g\, c} > 0,$$

using $t^2 + 3ab = t^2(c - \varepsilon)/c$. Hence $g + \sigma(\pi - 2\arctan\rho) = 2\pi n$ for a sign σ and an integer n. As $0 < \pi - 2\arctan\rho < \pi$ and $g > 0$: if $\sigma = +1$ then $n \geq 1$ and $g > \pi$; if $\sigma = -1$ then $n \geq 0$ and $g \geq \pi - 2\arctan\rho$. Either way $g \geq \pi - 2\rho$.

The bounds. From $a + b = \varepsilon - c$ and $ab = -t^2\varepsilon/(3c)$, $g^2 = (c - \varepsilon)^2 + 4t^2\varepsilon/(3c)$. Since $\varepsilon(c - \varepsilon) < 10 \leq \frac{5}{8}\varepsilon^2$, $c < \frac{13}{8}\varepsilon$ and $g^2 > \frac{32}{39}t^2 > \frac{3}{4}t^2$, so $\sqrt{3}\, g > \frac{3}{2}t$, and then

$$\varepsilon^2\rho = \frac{\varepsilon t \cdot \varepsilon(c - \varepsilon)}{\sqrt{3}\, g\, c} < \frac{\varepsilon t \cdot 4t^2/3}{\frac{3}{2}t \cdot \varepsilon} = \frac{8}{9}t^2, \qquad \varepsilon^2 g^2 < \frac{16t^4}{9} + \frac{4t^2\varepsilon^2}{3},$$

the second because $\varepsilon(c - \varepsilon) < 4t^2/3$ and $\varepsilon/c < 1$. In particular $\rho < \frac{8}{9} \cdot \frac{15}{2} \cdot \frac{1}{16} < \frac{1}{2}$, so $g^2 \geq (\pi - 2\rho)^2 \geq \pi^2 - 4\pi\rho$, and

$$\varepsilon^2\Big(\frac{3\pi^2}{4} - t^2\Big) < \frac{3}{4}\Big(\frac{16t^4}{9} + 4\pi \cdot \frac{8}{9}t^2\Big) = \frac{4t^4 + 8\pi t^2}{3} < \frac{4 \cdot \frac{225}{4} + 8 \cdot 3.15 \cdot \frac{15}{2}}{3} = 138.$$

Finally $L^2 - t^2 = (L - t)(L + t)$ with $L + t > L > \frac{5}{2}$, so $\varepsilon^2(L - t) < 138/\frac{5}{2} < 56$. □

So an exact pulse cheaper than $\pi\sqrt{3}/2$ by η has $\varepsilon \leq 10/\sqrt{\eta}$; exchanging i and j turns B into $-B$, so this covers every certain single pulse. Numerically the best constant is about 0.97.

Proposition 30.37 (Repeated pulses are one longer pulse)**.** *In classical mathematics, under the holding P_7 with $\hbar = 1$ and $\varepsilon > 0$: m identical braid pulses of angle θ have exactly the entry moduli of one pulse of angle $m\theta$ and duration $m\varepsilon$, and drive $|m\theta|$. Hence every word of identical braid pulses that makes a certain change drives at least $2\pi/3$.*

Proof. Each paid pulse is a phase times e^X with $X = \theta B - \mathrm{i}\varepsilon P_7$, and $(e^X)^m = e^{mX}$ is the pulse with angle $m\theta$ and duration $m\varepsilon$. Apply Theorem 30.35(a). □

Repeating a pulse lengthens it. Cutting every pulse of a word into equal pieces does the converse, exactly, and it ties the price at one duration to the price at its multiples.

Proposition 30.38 (Subdividing pulses)**.** *In classical mathematics, on one channel with $\hbar = 1$: for a word w and $m \geq 1$ let $w^{(m)}$ replace each braid token of angle θ by m braid tokens of angle θ/m on the same triad. Call a word* certain *at duration ε when its paid word under the holding P_7 at that duration makes a certain change.*

(a) For every holding H, the entries of the paid word of $w^{(m)}$ at duration ε have the moduli of those of w at duration $m\varepsilon$; $w^{(m)}$ has the drive of w and $m|w|$ tokens.

(b) Under the holding P_7: if every certain word at duration ε drives at least $2\pi/3 + \eta$, then so does every certain word at duration $m\varepsilon$.

(c) Under P_7, the statement for every $\eta > 0$ there are $\varepsilon > 0$ and a certain word of drive below $2\pi/3 + \eta$ *is equivalent to* for all $\eta, \delta > 0$ and every count N there are $\varepsilon \in (0, \delta)$ and a certain word of at least N tokens with drive in $[\frac{\pi}{2}\sqrt{3/2}, \frac{2\pi}{3} + \eta)$.

(d) Under P_7, the statement for every $\varepsilon > 0$ and $\eta > 0$ some certain word at duration ε drives below $2\pi/3 + \eta$ *is equivalent to the same statement for $\varepsilon \geq 1$, with drive at least $\frac{\pi}{2}\sqrt{3/2}$; and it fails exactly when some $\varepsilon, \eta > 0$ make every certain word at every duration $m\varepsilon$, $m \geq 1$, drive at least $2\pi/3 + \eta$.*

Proof. (a) The m pieces of a paid token give $\left(e^{-\mathrm{i}\|G\|/m} e^{G/m - \mathrm{i}\varepsilon H}\right)^m = e^{-\mathrm{i}\|G\|} e^{G - \mathrm{i}m\varepsilon H}$, the paid token at duration $m\varepsilon$; drive and length are clear. (b) A certain word at duration $m\varepsilon$ that drove less would give, by (a), a certain word at duration ε with the same drive. (c) One direction is immediate. Conversely, given a word at duration ε, which has at least one token, take $m > \max(N, \varepsilon/\delta)$ and $w^{(m)}$ at duration ε/m; its drive is at least $\frac{\pi}{2}\sqrt{3/2}$ by Theorem 30.34. (d) For $\varepsilon < 1$ apply the statement at $m\varepsilon \geq 1$ and subdivide. If the statement fails at ε with some η, (b) carries the gap to every $m\varepsilon$; the converse is the case $m = 1$. □

The total holding exposure is the same in both words; only its grain changes. So a surcharge for exact change under the holding cannot be confined to faint holdings (Problem 20).

Holding can also do work. The next constructions show that the floors above are not empty statements about drive alone.

Proposition 30.39 (Holding that works)**.** *In classical mathematics, in units with $\hbar = \varepsilon = 1$:*

(a) on every register, for all configurations u, v there are a holding and a word of drive 0 with drift exactly $\pi/2$ that moves u to v with certainty;

(b) on every register, for every channel k, label $b \neq u_k$ and $0 \leq d \leq 2\pi/3$, there are a holding and a word with drift exactly d and drive exactly $2\pi/3 - d$ that sets channel k to b with certainty;

(c) for every flip of one low bit and every $D > \pi/6$ there are a holding and a word with drift at most D and drive below $\pi/2$ that make it.

Proof. (a) Let Σ exchange e_u and e_v and fix the other basis vectors; it is a self-adjoint involution of norm 1. With $H = -\frac{\pi}{2}\Sigma$ and one idle token, the paid word is $e^{\mathrm{i}\pi\Sigma/2} = \mathrm{i}\Sigma$.

(b) Let $G = \pm B$ be the generator, on channel k, of a triad containing u_k and b, with the sign chosen so that $e^{2\pi G/3}$ moves u_k to b. Take $H = \mathrm{i}\, dG$, which is self-adjoint of norm d, and one braid token $(\frac{2\pi}{3} - d)G$. The paid token is a phase times $\exp((\frac{2\pi}{3} - d)G + dG) = e^{2\pi G/3}$.

(c) Use (b) with b the partner and $\frac{\pi}{6} < d \le \min(D, \frac{2\pi}{3})$: the drive $\frac{2\pi}{3} - d$ is below $\frac{\pi}{2}$. □

How the price falls as holding is allowed is the main open question of this section; two curves frame it. The line $2\pi/3 - \frac{4}{3}D$ joins the zero-holding price to the pure-holding point $(\pi/2, 0)$ of (a), and every certain one-channel word is conjectured to lie on or above it (Problem 11). The aligned construction (b) runs along $2\pi/3 - D$, above the line. A geometric argument on the cone that a driven triad state sweeps suggests the exact price

$$\mathsf{P}(D) = \frac{2\pi}{3} - \arccos\frac{3\cos D - 1}{2}, \qquad 0 \le D \le \frac{\pi}{2}$$

(Problem 13); it runs from $\mathsf{P}(0) = 2\pi/3$ to $\mathsf{P}(\pi/2) = 0$ and reaches $\pi/2$ at $D_1 = \arccos\frac{1+\sqrt{3}}{3} \approx 0.4259$. If it is the price, the line is never approached at intermediate drift (Problem 12):

Proposition 30.40 (The conjectured curve lies above the line)**.** *In classical mathematics,* $\mathsf{P}(D) \ge \frac{2\pi}{3} - \frac{4}{3}D$ *for* $0 \le D \le \frac{\pi}{2}$*, with strict inequality for* $0 < D < \frac{\pi}{2}$.

Proof. The claim is $\arccos\frac{3\cos D-1}{2} \le \frac{4}{3}D$, that is $\frac{3\cos D-1}{2} \ge \cos\frac{4D}{3}$, since $\frac{4D}{3} \le \frac{2\pi}{3}$. With $c = \cos\frac{D}{3}$, $\cos D = 4c^3 - 3c$ and $\cos\frac{4D}{3} = 8c^4 - 8c^2 + 1$, and the difference is

$$\frac{3\cos D - 1}{2} - \cos\frac{4D}{3} = 8(1-c)\Big(c^2 - \frac{3}{4}\Big)\Big(c + \frac{1}{4}\Big),$$

which is nonnegative for $c \in [\frac{\sqrt{3}}{2}, 1]$ and positive inside. □

30.5 Open problems

Each problem is stated exactly, in the notation of the section that introduces it: in classical mathematics where it concerns classical objects, and on $\mathbb{R}_\delta$ where it asks for a price. Problems 1–25 belong to this chapter. Problems 26 onwards collect, by subject, what the rest of the book leaves open, each with a pointer to the result that proves what is known around it. Where a problem is equivalent to a sharper statement, both forms are given with the reason; numerical exploration and finite searches are reported as evidence, not proof.

Self-similar ledgers

1. *Rates with a hold letter.* For the three-letter joins $\sigma(a) = \mathsf{d}^{A(a)}\mathsf{c}^{B(a)}\mathsf{h}^{H(a)}$ given by nine counts: (i) if such a join is primitive and has rate r, then $-1 < r < 1$ and r is a root of a nonzero integer polynomial of degree at most 3; (ii) every irrational $r \in (-1, 1)$, other than $\sqrt{2}/2$, that is a root of a nonzero integer polynomial of degree at most 3 is the rate of a primitive such join. By Theorem 30.17 this is equivalent to: the rates of primitive three-letter joins are exactly the numbers in $(-1, 1)$ algebraic of degree at most 3.

2. *How fast a neutral ledger wanders.* For every primitive two-letter join with rate r and $|\mu| = 1$, the limit v of Theorem 30.15 satisfies $12v \le (\lambda^2 - 1)(1 + |r|)^2$. Proposition 30.16 gives equality at every integer $\lambda \ge 2$ with $r = 0$. Evidence: exact checks of many thousands of primitive neutral joins with $\lambda \le 6$ found no violation, and equality only at $r = 0$.

3. *The neutral variance law with a hold letter.* Let σ be a primitive three-letter join with rate r, and let $|\mu| = 1$ with $\delta_r(\sigma(a)) = \mu\,\delta_r(a)$ for each letter a. If σ does not keep balanced books, then V_k/k converges to a positive limit. This is equivalent to the three-letter form of Theorem 30.15.

4. *Extremal orderings of a Jordan census.* Let a primitive three-letter join with rate 0 have images of one length L, and let $u \ne 0$ be a letter weight and $c \ne 0$ a constant with $u(\sigma(a)) = u(a)$ and $\mathrm{bal}(\sigma(a)) = \mathrm{bal}(a) + c\,u(a)$ for every letter a. Then (i) among the joins with the same nine counts, one whose images are each sorted by u (nondecreasing or nonincreasing) has a limit of V_k/k^3 at least as large as every limit attained by another; and (ii) some join with the same counts has $V_k/k^3 \to 0$ if and only if all nonzero values of u have the same absolute value. Part (i) holds for the census of Theorem 30.18; what remains is equivalent to (i) for every other such census together with (ii).

Simultaneous approximation

5. *A pair from two quadratic fields.* There is $c > 0$ with $\sqrt{|k|}\max\bigl(|k(\sqrt{5}-2)-m|,\ |k\tfrac{3\sqrt{2}}{2}-n|\bigr) \ge c$ for all integers $k \ne 0$, m, n. By Theorem 30.5 this is equivalent to the exponent $\tfrac{3}{2}$ for three probes at φ, φ^2 and $\sqrt{2}$. Every exponent above $\tfrac{3}{2}$ follows from Schmidt's theorem.

6. *Cusick's bound.* $S(\alpha, \beta) \le \tfrac{2}{7}$ for all real α, β [79]. By Theorem 30.7 this is equivalent to: no two probes in $(1, \varphi^2)$ have separation score above $2\pi^2/7$.

7. *Unimodular bases of $\mathbb{Z}[\vartheta]$.* With $\vartheta = 2\cos(2\pi/7)$: for every $\eta > 0$ there are integers p, q, r, s with $ps - qr = 1$ and $\tfrac{2}{7} - \eta < S(p\vartheta + q\vartheta^2,\ r\vartheta + s\vartheta^2) < \tfrac{2}{7}$. By Proposition 30.8 this is equivalent to integral cubic probe designs approaching the score $2\pi^2/7$ from below.

8. *Index-two coordinates.* (i) $S(p\vartheta + q\vartheta^2,\ r\vartheta + s\vartheta^2) \le \tfrac{1}{14}$ whenever $ps - qr = 2$; (ii) for every $\eta > 0$ some such pair has $\tfrac{1}{14} - \eta < S < \tfrac{1}{14}$. Evidence: evaluating a formula of Cusick's type for these modules over all determinant-two integer matrices with entries up to 14 puts the supremum near 0.0649, below $\tfrac{1}{14} \approx 0.0714$; if that is right, (i) holds with a gap and (ii) fails.

Certain change

In the problems on holding, "certain word" means a word whose paid word makes a certain change, and units are $\hbar = \varepsilon = 1$ unless the holding is P_7, where $\hbar = 1$ and $\varepsilon > 0$ is free.

9. *Free phase gates.* On one channel, allow besides braid tokens free *phase gates*, diagonal unitaries $e_a \mapsto e^{\mathrm{i}\phi(a)}e_a$ of drive 0. Is there such a word that makes a certain change with

drive below $2\pi/3$? Equivalently: one with at least two nonidle braid tokens and some phase gate that is not a multiple of the identity. Proved: with at most one nonidle braid token the price stays $2\pi/3$, because the phase gates only multiply rows and columns by units, and with scalar phase gates only it stays $2\pi/3$, because they factor out. Conjectured answer: no.

10. *The higher-label surcharge.* On every register of $n \geq 2$ channels at zero holding, every certain change $u \to v$ with at least one channel moving outside its low-bit pair drives at least $\frac{\pi}{2} \cdot \#\{k : u_k \neq v_k\} + \frac{\pi}{6} \cdot \#\{k : v_k \notin \{u_k, u'_k\}\}$. This is equivalent to the exactness of the prices of Proposition 30.30. Proved: Theorem 30.26 and Proposition 30.27.

11. *The four-thirds floor.* On one channel, under every self-adjoint holding, every certain word satisfies $g(w) + \frac{4}{3}D \geq \frac{2\pi}{3}$. Proved: at drift 0, for scalar holding, and when $D \geq \pi/2$ or $g(w) \geq 2\pi/3$; what remains is equivalent to the case of nonempty words with $0 < D < \pi/2$, $0 < g(w) < 2\pi/3$ and $g(w) + D \geq \pi/2$, under holdings that are not multiples of the identity and cannot be shortened in norm by adding one. Indeed, adding $s \cdot 1$ to the holding multiplies the paid word by a unit phase, so the word stays certain, and a real s that minimises the norm can only lower D; and a certain word below the floor has all the listed properties, by Theorem 30.24, the scalar case and Theorem 30.32.

12. *Approaching the line.* For every $0 < d < \pi/2$ and $\eta > 0$ there are a holding and a certain one-channel word with drift at most d and drive at most $\frac{2\pi}{3} - \frac{4}{3}d + \eta$. Conjectured answer: no; the aligned construction of Proposition 30.39(b) stays $\frac{d}{3}$ above the line, and if the curve of Problem 13 is the price, this fails.

13. *The one-channel price under holding.* (i) Every certain one-channel word with drift $D \leq \pi/2$ drives at least $\mathsf{P}(D)$; (ii) for all labels $u \neq v$, every $D \in [0, \pi/2]$ and $\eta > 0$, some certain word $u \to v$ has drift at most D and drive at most $\mathsf{P}(D) + \eta$. Both endpoints are proved (Theorem 30.24 and Proposition 30.39(a)). Evidence: numerical optimisation matches P to six digits at several drifts, and searches over all eight labels never beat it.

14. *One holding for several channels.* On $n \geq 2$ channels, for changes $u \to v$ in which k channels change and none moves to its partner, the price at drift $D \in [0, \pi/2]$ is $k\big(\frac{2\pi}{3} - \arccos\big((3(\cos D)^{1/k} - 1)/2\big)\big)$: a lower bound for every certain word, approached within every $\eta > 0$. At $D = 0$ this contains the surcharge of Problem 10 for such changes. Conjectured answer: no.

15. *A threshold for flips.* On $n \geq 2$ channels, for a flip of one low bit: (i) every certain word with drift at most $D_1 = \arccos\frac{1+\sqrt{3}}{3}$ drives at least $\pi/2$; (ii) for every $D > D_1$ some certain word has drift at most D and drive below $\pi/2$. Proved: (ii) for $D > \pi/6$, and any word below $\pi/2$ has at most one entangling token (Theorem 30.32).

16. *The braid cost of leaving a pair.* On $n \geq 2$ channels at zero holding, a certain change that moves channel k off its low-bit pair spends at least $2\pi/3$ of braid drive on channel k. Proved: $\frac{\pi}{2}\sqrt{3/2}$ (Proposition 30.27), the angle budget $\sum \beta(\theta) \geq \pi/2$, and $2\pi/3$ whenever only one braid token on channel k is not idle (Theorem 30.28). Evidence that it is false: if exact rotations inside the low-bit pairs were free, searches reach 1.9512; see Problem 23.

17. *Is the pair floor sharp?* For every $\eta > 0$ some register of $n \geq 2$ channels has a certain change at zero holding that moves channel k off its low-bit pair with braid drive on channel k below $\frac{\pi}{2}\sqrt{3/2} + \eta$. By Corollary 30.29 the words may be taken with braid drive below $2\pi/3$ and at least two non-idle braid tokens on channel k, whose angles satisfy $\frac{\pi}{2} \leq \sum \beta(\theta) \leq \sqrt{2/3} \sum |\theta|$. A positive answer would settle Problem 16 in the negative. Evidence: with exact rotations inside the low-bit pairs free, searches level off near 1.945.

18. *Two pulses under label-seven holding.* Every certain word of two braid pulses under P_7 drives at least $2\pi/3$; more strongly, at least the least drive of a certain single pulse. Proved: when one angle is 0, and when the two pulses are identical. Evidence: scans find every certain two-pulse word at drive at least 2.7071.

19. *Many pulses under label-seven holding.* Every certain word of braid pulses under P_7 drives at least $2\pi/3$. Proved for words of identical pulses (Proposition 30.37) and with the floor $\frac{\pi}{2}\sqrt{3/2}$ for all words (Theorem 30.34); what remains is equivalent to words of at least two pulses, not all identical, with drive in $[\frac{\pi}{2}\sqrt{3/2}, \frac{2\pi}{3})$.

20. *Approaching the cycle under label-seven holding.* (i) For every $\eta > 0$ there are a duration $\varepsilon > 0$ and a certain one-channel word under P_7 with drive below $2\pi/3 + \eta$. (ii) For every $\varepsilon > 0$ and every $\eta > 0$ there is such a word at duration ε. By Proposition 30.38, (i) is equivalent to its form with durations below any $\delta > 0$ and words of at least any length N, (ii) to its form for $\varepsilon \geq 1$, and (ii) fails exactly when one surcharge η persists at all durations $\varepsilon, 2\varepsilon, 3\varepsilon, \ldots$. Together with Problem 19, (i) would make $2\pi/3$ the exact price under this holding, approached but perhaps not attained. Evidence: optimisation finds certain words at falling drive as pulses are added, about 2.15 with twenty pulses, and none below $2\pi/3$.

21. *A jump at a quarter turn of holding.* Is there $c > 0$ such that on every register of $n \geq 2$ channels, under every holding with drift below $\pi/4$, every certain flip of one low bit drives at least $\frac{\pi}{4} + c$? (At drift exactly $\pi/4$, one entangling token under the holding $-\frac{\pi}{4} S_l F_k$ makes the flip for drive $\pi/4$.) Equivalently: one-entangler words with braid drive $0 < B < \pi/24$ and drift in $[\frac{\pi}{4} - \sqrt{2/3}\, B, \frac{\pi}{4})$ that make the flip need $B \geq c$. For by Theorem 30.32, $\frac{\pi}{2} \leq \sqrt{2/3}\, B + N\frac{\pi}{4} + D$: below drift $\pi/4$ a word with no entangling token drives more than $\frac{\pi}{4}\sqrt{3/2} > \frac{\pi}{4} + \frac{\pi}{24}$, a word with two or more drives at least $\pi/2$, and a word with exactly one drives $\frac{\pi}{4} + B$, where $D \geq \frac{\pi}{4} - \sqrt{2/3}\, B$ forces $B > 0$; those with $B \geq \pi/24$ already pay $\frac{\pi}{4} + \frac{\pi}{24}$. Conjectured answer: no.

22. *A two-token approach to the quarter turn.* There are $K > 0$ and $\delta_0 \in (0, \pi/4)$ such that for every $\delta \in (0, \delta_0)$ some self-adjoint holding H on two channels with $2\|H\| = \pi/4 - \delta$ and some angle $|\theta| \leq K\sqrt{\delta}$ make one of the two words of Proposition 30.33(a), with braid angle θ, move $(0, 0)$ to $(1, 0)$ with certainty. By Proposition 30.33(b) this is equivalent to the same statement for all two-token words with one entangling token and braid drive at most $K\sqrt{\delta}$; a positive answer would answer Problem 21 in the negative. Evidence that it is false: optimisation over all holdings at $\delta = 0.1$ finds the least defect $1 - |\langle e_{(1,0)}, W e_{(0,0)} \rangle|^2$, namely $\sin^2 \delta$, at $\theta = 0$.

23. *Returned-helper rotations.* On two channels, for every real γ there are a word with no braid drive on channel 1 and a unit ζ such that, on every configuration whose channel 2 shows 0, the word agrees with $\zeta(\cos\gamma + \mathrm{i}\sin\gamma\, S)$ on channel 1 and the identity on channel 2, where S is the parity sign. By the identity $k(\cos\gamma + \mathrm{i}\sin\gamma\, S)k^{-1} = \cos\gamma + \sin\gamma\, SF$ for $k = (1 + \mathrm{i}F)/\sqrt{2}$, which uses only $F^2 = 1$ and $SF = -FS$, this is equivalent to the same statement for the pair rotations $e^{\gamma SF}$: on configurations whose channel 2 shows 0, one entangling token on the channels $(2, 1)$ acts as k on channel 1 (the sign of label 0 is $+1$), and $k^{-1} = k^7$, so conjugating by k maps admissible words to admissible words. With entangling tokens only on the channels $(1, 2)$, with target 2, exactly the multiples of $\pi/4$ are reachable (Theorem 30.31), so the problem needs entangling tokens on the channels $(2, 1)$ as well. Evidence: the Lie algebra generated by the relevant generators has dimension 136 and contains the needed ones; explicit words of up to 18 tokens were not found.

24. *Is the diagonal floor sharp?* For every $\eta > 0$ there are a diagonal holding and a certain one-channel word with drive below $\frac{\pi}{2}\sqrt{3/2} + \eta$. Evidence that it is false: searches with arbitrary diagonal holdings never beat the zero-holding cycle.

25. *The cheapest exact pulse.* There are θ and $\varepsilon > 0$ such that one pulse under P_7 makes a certain change, $2\pi/3 < |\theta| < \pi\sqrt{3}/2$, and no certain single pulse has smaller $|\theta|$. Proved: certain pulses exist and all have $|\theta| \geq 2\pi/3$ (Theorem 30.35), and one with $|\theta| = \pi\sqrt{3}/2 - \eta$ has $\varepsilon \leq 10/\sqrt{\eta}$ (Theorem 30.36), so a minimising sequence below $\pi\sqrt{3}/2$ stays at bounded duration. Evidence: $|\theta| \approx 2.70706$ at $\varepsilon \approx 8.42212$.

Logic and the prices of principles

In Problems 26–30, θ is the halting matrix of Lemma E.11, and a derivation *with no added rule* is a derivation in the calculus of Chapter 6 whose EM, LPO and MP entries are all false, at any tier of induction (Section 7.14).

26. *Decidability of the halting matrix.* Does $\forall e\, \forall x\, (\theta(e, x) \vee \neg\theta(e, x))$ have a derivation with no added rule? Its truth is decided by a program (Lemma E.11); the question is whether the calculus derives that decision.

27. Does $\mathrm{LPO}_\theta \to \mathrm{MP}_\theta$ have a derivation with no added rule? A positive answer to Problem 26 gives one, and then LPO_θ is exactly $\mathrm{WLPO}_\theta \wedge \mathrm{MP}_\theta$.

28. Does $\mathrm{WLPO}_\theta \to \mathrm{MP}_\theta^\vee$ have a derivation with no added rule? A positive answer to Problem 26 gives one.

29. Does $\mathrm{MP}_\theta \to \mathrm{MP}_\theta^\vee$ have a derivation with no added rule? With Problem 28, a positive answer puts $\mathrm{MP}_\theta^\vee$, which is not free (Theorem 7.59), in the meet of WLPO_θ and MP_θ. The calculus derives $\mathrm{MP}_\theta^\vee$ from the Σ_1 Markov letter at $\theta(e_1, x) \vee \theta(e_2, x)$; the question asks for it from the letter at θ itself.

30. With λ as in Section 7.14, do $\mathrm{WLPO}_\theta \to \mathrm{LLPO}_\lambda$ and $\mathrm{LLPO}_\lambda \to \mathrm{MP}_\theta^\vee$ have derivations with no added rule? As principles on binary sequences both steps hold (WLPO gives LLPO, Chapter 12, and LLPO gives $\mathrm{MP}^\vee$, Theorem 14.19); at the instances only the non-derivations of Table 7.5 are proved.

31. *The questions of Parts III and IV.* The questions about the computable calculus of δ's objects stated in Section 16.11, and those about linear algebra without unique choice stated in Section 14.9.

The continuum and analysis

32. *Zone averages and completeness.* Does the existence of zone averages for every F with a modulus of uniform continuity on the zone imply completeness for sequences of points with the modulus $k \mapsto k$? This is the converse of Theorem 12.101(1).

33. *Taylor's step with no choice.* Does Proposition 12.99 hold for pointwise derivatives without countable choice of natural numbers? A positive answer gives, with no choice, the six expansions of lattice gauge quantities of Remark 26.64; with uniform derivatives each already holds with no choice.

34. *Completeness from choice, inside HAS.* Does HAS derive every instance of Cauchy completeness, written as in Theorem 12.100(2), from the sentence of countable choice of natural numbers? On $\mathbb{R}_\delta$ the implication holds (Theorem 12.100(1)).

35. *Exact fixed points from LLPO alone.* Does LLPO, without countable choice, give an exact fixed point of every uniformly continuous map of δ's triangle to itself? With countable choice it does (Section 29.4); with LLPO and countable choice only at quantifier-free and existential formulas, the corresponding instance of exact roots is neither provable nor refutable (Theorem 12.70).

36. *The open rows of the price table.* The 22 rows of Appendix Q priced "open", each between the bounds it states after its price (Section 28.12): among them the Heine–Cantor theorem on $[0, 1]$ (row 1441), at least the decidable fan theorem and at most WLPO; Brouwer's theorem on the square for maps continuous at each point (row 1058), at least LLPO and at most WLPO; and a nonzero null vector of every singular matrix (row 684), at most LLPO with countable choice and with no lower bound proved. Five of the open rows have an exact price, in the sense of the table's key, in a principle that is not one of the table's letters: the derivative convention (row 10), Dini's theorem and the bounded convergence theorem for integrals (rows 552 and 553), the Heine–Cantor theorem on $[0, 1]$ (row 1441) and the continuous image of a compact set (row 1447). The problem is to price the open rows in the table's letters, or to show that no letter prices them.

The cost and the verdict

37. *The six exponentials on the line.* Let $e^{\ell_p} = p$ for $p = 2, 3, 5$, and let each $e^{d\ell_p}$ be a ratio. Is "d is an integer" provable on $\mathbb{R}_\delta$ with no omniscience principle and no choice principle? The rest of the verdict's transcendence step is reduced to this core with no choice (Section 22.12); what is missing is a free zero estimate for exponential polynomials in several frequencies.

38. *From a bound to a curvature.* Let G satisfy the composition law of Chapter 21 with $G(0) = 0$ and $|G| \le c$ near 0, where $\frac{3}{2} \le c < 2$. Does G have a curvature, with no omniscience principle? It does for $c < \frac{3}{2}$ and fails at $c = 2$ (Theorem 21.45).

39. *The sign of a cost.* Does LLPO imply that every solution G of the composition law with $G(0) = 0$ is ≥ 0 everywhere or ≤ 0 everywhere, with no curvature premise? The converse is proved (Theorem 21.47).

40. *Strong extensionality and Markov's principle.* Does strong extensionality of integer-valued maps on the line, $\mathrm{SE}_{\mathbb{Z}}$, imply MP? MP implies it, and with the WLPO function it is MP (Theorem 22.24, Proposition 22.25).

41. *The last bracket of the chain.* The exact price of step T2.6 of Table 22.1, dropping continuity and the order premises: it lies below the WLPO function (Theorem 22.14), and no lower bound is proved.

42. *Deciding a cost law.* The questions (O1) and (O2) stated with Definition 21.28.

The world's records

43. *The eight-tick comparison.* Derive that one comparison against the fundamental clock mode lasts eight ticks. By Theorem 24.67 this is exactly $A_0 = 2\pi\hbar$, and the comparison laws alone leave A_0 free.

44. *Locality from the light cone.* Derive chronological coverage, which is equivalent to locality (the time window): the law that fixes which retained records a comparison addresses. The light cone of Sections 24.1 and 24.5 orders postings and does not supply it.

45. *The act-rate law.* Derive whether the Hubble patch acts at most once per tick of the cosmic clock or in proportion to its volume; Theorem 24.63 proves what each law gives.

46. *One crossing per tick.* Derive that the patch's units form one unit system read on the patch's accounts (Proposition 24.74), or the identification it reduces to, that moves are posts.

47. *The price of access from locality.* Access with locality makes the capacity exceed every bound in classical mathematics. What does that step cost on $\mathbb{R}_\delta$?

48. *Certain flips.* Does some native word flip a register configuration with certainty, at zero boundary flux and unit action scale; and does the integer parameter m of the register drop out of that question? With the entangler reach problem of Section 24.9, this asks what the native register can do exactly.

49. *Count-only stores.* Does a count-only store that serves, beyond every bound, an infinite set of counts whose differences are all served keep the natural numbers real (Section 24.12)?

Fields and the second law

50. *The tick's gap on the line.* A contraction of χ^2 at every tick that is uniform in the ball, and a form of the gap on $\mathbb{R}_\delta$ (Section 26.5, after Theorem 26.32).

51. The statements (B) and (M) of Proposition 26.44, and the laws in three or more quantities left open with Theorem 26.53.

Gravity and the record cone

52. *The quartic cell form as a derivative.* On $\mathbb{R}_\delta$, at $\kappa = 1$, is the quartic cell form the fourth derivative of δ's own Regge action? This is equivalent to the two named premises on third-order rates used by Theorem 25.71 (Proposition 27.63).

53. *Beyond order K^4.* What is the exact branch of the graviton on the $\kappa = 1$ lattice, to all orders in the momentum, and what is the order-K^4 correction to the tree scattering of four gravitons? At order K^4 the dispersion law of the whole family is proved (Theorem 27.75).

54. *The body diagonal.* Does the zero mode of the body diagonal at $\kappa = 0$ change gauge-invariant amplitudes with two derivatives (Proposition 27.72)?

55. *A changed coupling.* Hinge shape costs cannot complete a changed coupling at quartic order (Theorem 27.71; Proposition 27.70 gives the direction of the change through cubic order). Can a cost of another kind, for example one nonlinear in the deficits, complete it?

56. *The premises of the graviton's action.* Derive from the foundation each named statement of Proposition 27.76: that the ledger writes a step by its cheapest record (tick-free writing); double entry together with the reading that the charge reads a strain only through what comparison events post (the coherent reading); double entry for the derived scalar record (locality); the purity of each post's lattice record, which is $\kappa = 1$; and the reading of the face directions through the records of the double steps (independent axes).

57. *The long-wave profile of a mass.* Is an actual lattice solution of a mass a native C^3 family solving the weak equations of first and second order (Theorem 25.70)? The question is open in classical mathematics too. And do the third-order weak equations give the third-order jet equation (Theorem 25.71)?

58. *Charts of the posting order.* (i) The grid $2 \times 3 \times 3 \times 3$ has a record-cone chart read on its stage events. (ii) So does $3 \times 3 \times 3 \times 3$. (iii) Every truncation $2 \times 3 \times [n]$ has one; every $n > 20$ is open, and this is the sphere-order form of the third question of Felsner, Fishburn and Trotter [201] for these orders. (iv) In every real normed space, a ladder of length n with all angular margins at least $\theta > 0$ has $\theta^2(n-1) \leq 1$; the bound with 2 in place of 1 is proved, and 1 is proved and sharp in one dimension (Theorem 24.85). Searches found no chart in (i)–(iii) and no counterexample to (iv) in small cases.

Learning

59. *Own rows at every level.* Theorem 29.69 at the levels $n \geq 4$, and a reduction from a learner to U_n in Theorem 29.68(4) that is uniform in the learner. And for a bare oracle O, not joined with the row table: if some learner computable from O learns $C_{n+2}(\mathbf{t}_n)$ with $n+2$ changes, must O compute U_n? The proof gives only that U_n is computable from O joined with the row table (Remark 29.70).

60. *Harrington's method.* Is Σ_2-LPO strictly stronger than LPO under the reading of maps of Chapter 21 (Section 29.9), and do Harrington's own bounds, in the standard order of codes, carry a totality decision (Section 29.10)?

Questions to work through

1. For $k = 1, \dots, 5$ find the integer m nearest to $k(\sqrt{5}-2)$ and check $6k\,|k(\sqrt{5}-2) - m| \geq 1$ directly. Which k comes closest to equality? (Lemma 30.3.)

2. Take the recognizers $\sim_{2,4}$ and $\sim_{5,6}$. Compute their join, and find counts a, b related by the join for which no count n has $n \sim_{2,4} a$ and $n \sim_{5,6} b$. (Theorem 30.9.)

3. Decide whether $\sqrt{3} - 1$ and $\sqrt{3}/3$ are rates of primitive two-letter joins, and for each one that is, write down a join that has it. (Theorem 30.13.)

4. For $\mathsf{d} \mapsto \mathsf{dcd}$, $\mathsf{c} \mapsto \mathsf{cdc}$, write out $\sigma^2(\mathsf{d})$, $\sigma^2(\mathsf{c})$ and check the coboundary $\pi(\mathsf{d}) = 0$, $\pi(\mathsf{c}) = 1$; show that $\mathsf{d} \mapsto \mathsf{ddc}$, $\mathsf{c} \mapsto \mathsf{ccd}$ admits none. (Theorem 30.14.)

5. On one channel, write words of drive exactly $2\pi/3$ that move label 0 to label 7 and 7 to 0 with certainty. Show that one braid token of angle $\pi/2$ on the triad $(0, 1, 7)$ moves no label with certainty, by computing its diagonal. (Theorems 30.22 and 30.24.)

6. Verify that $\omega(x) = (\theta^2 - \mathrm{i}\sqrt{3}\theta x,\ \theta^2 + \mathrm{i}\sqrt{3}\theta x,\ \theta^2 - 3x^2)$ satisfies $A\omega(x) = -\mathrm{i}x\,\omega(x)$ when $p(x) = 0$, row by row. (Theorem 30.35.)

Appendix A

Finite controls and approximation

Chapter 18 uses two approximation results. A finite alphabet of gates supplies descriptions that can be searched. A family of native rotations and couplings supplies the operations used in the positive-energy construction. This appendix proves both results. The common step is to isolate elementary matrix directions and then connect them across a register.

The argument has four stages. We first identify a sufficient collection of matrix directions. We explain why directions obtained by sums and commutators correspond to limits of available operations. We then obtain those directions from the native controls. Finally, an explicit eight-gate sequence starts the corresponding argument for the finite alphabet.

All spaces in this appendix have positive finite dimension. Matrix norms are operator norms unless stated otherwise. An approximation assertion means that for every positive error there is a finite sequence attaining it; it does not assert a bound on the length of that sequence.

A.1 The matrix directions that suffice

Write E_{uv} for the matrix with a single 1 in row u, column v. For distinct indices define

$$R_{uv} = E_{uv} - E_{vu}, \qquad I_{uv} = i(E_{uv} + E_{vu}), \qquad D_{uv} = i(E_{uu} - E_{vv}).$$

Here I_{uv} names an imaginary symmetric edge; the unsubscripted I continues to mean the identity. All three matrices are skew-Hermitian and have trace zero. The real vector space of all such matrices is denoted $\mathfrak{su}(d)$.

For matrices A, B, their commutator is $[A, B] = AB - BA$. The multiplication rule $E_{uv}E_{xy} = \delta_{vx}E_{uy}$ gives, for pairwise distinct u, v, w,

$$\begin{aligned} [R_{uv}, R_{vw}] &= R_{uw}, & [R_{uv}, I_{vw}] &= I_{uw}, \\ [I_{uv}, I_{vw}] &= -R_{uw}, & [R_{uv}, I_{uv}] &= 2D_{uv}. \end{aligned} \tag{A.1}$$

For example, in the first identity the only surviving products in $R_{uv}R_{vw}$ and $R_{vw}R_{uv}$ are E_{uw} and E_{wu}. Matrices supported on disjoint pairs commute.

Lemma A.1 (Connected real edges and one imaginary edge)**.** *Let $d \geq 3$. Suppose a real vector space L of skew-Hermitian matrices is closed under commutators. If L contains R_{uv} along the edges of a connected graph on the d indices, and contains one I_{pq} with $p \neq q$, then $\mathfrak{su}(d) \subseteq L$.*

Proof. Take a simple path between any two vertices. Repeatedly applying the first identity in (A.1) shortens that path, so every R_{uv} belongs to L. The given imaginary edge and the last identity give D_{pq}.

Commuting D_{pq} with R_{pr} gives I_{pr} when r is distinct from p, q, and gives $2I_{pq}$ when $r = q$. Thus every imaginary edge incident to p is present. If $u, v \neq p$ and $u \neq v$, then $[R_{up}, I_{pv}] = I_{uv}$. All

imaginary edges are therefore present, and their brackets with the real edges give all diagonal differences.

To see that these matrices span the claimed space, let $A^\dagger = -A$. For $u < v$, write $A_{uv} = x_{uv} + iy_{uv}$. Its two off-diagonal entries are exactly those of $x_{uv}R_{uv} + y_{uv}I_{uv}$. The remaining diagonal has the form $i\,\mathrm{diag}(t_1, \dots, t_d)$ with real t_j. Trace zero says $\sum_j t_j = 0$, so, after fixing an index p, this diagonal equals $\sum_{u \neq p} t_u D_{up}$. Every part lies in L. □

This lemma reduces a large spanning problem to a small graph. The rest of the proof constructs its edges from the operations actually supplied.

A.2 From finite operations to continuous directions

Let G be the norm closure of a group of available unitary operations. Define

$$L_G = \{X \,:\, X^\dagger = -X \text{ and } \exp(tX) \in G \text{ for every real } t\}.$$

The definition refers only to limits of finite operation sequences. We need to justify the algebraic manipulations used on L_G.

Lemma A.2 (A product-limit estimate)**.** *If A and B are contractions, then $\|A^n - B^n\| \le n\|A - B\|$. If two contraction-valued curves f, g satisfy $f(0) = g(0) = I$ and $f'(0) = g'(0)$, then, for fixed real t,*

$$\left\| f(t/n)^n - g(t/n)^n \right\| \longrightarrow 0.$$

Proof. The telescoping identity

$$A^n - B^n = \sum_{j=0}^{n-1} A^{n-1-j}(A - B)B^j$$

gives the first estimate. Equality of the derivatives says $\|f(h) - g(h)\| = o(|h|)$. Substituting $h = t/n$ into the estimate proves the limit. For $t = 0$ the difference is already zero. Applied to $f(h) = \exp(hX)\exp(hY)$ and $g(h) = \exp(h(X + Y))$, as in Proposition A.3, the limit is the finite-dimensional case of Trotter's product formula [25]. □

Proposition A.3 (The available directions form a closed Lie algebra)**.** *The set L_G is a closed real vector space, is closed under commutators, and is preserved by $X \mapsto UXU^\dagger$ for every $U \in G$.*

Proof. Zero and real scalar multiples follow directly from the definition. For $X, Y \in L_G$, the curves

$$f(h) = \exp(hX)\exp(hY), \qquad g(h) = \exp(h(X + Y))$$

are unitary, have value I at zero, and have derivative $X + Y$ there. The product-limit lemma shows that $\exp(t(X + Y))$ is a limit of elements of G, and hence is in G. This proves closure under sums. If $X_j \to X$ with $X_j \in L_G$, continuity of the adjoint and exponential gives $X^\dagger = -X$ and $\exp(tX) \in G$ for each t. Thus the vector space is closed.

Conjugation preserves it because $\exp(tUXU^\dagger) = U\exp(tX)U^\dagger$. Finally, for $X, Y \in L_G$, each difference quotient

$$\frac{\exp(hX)Y\exp(-hX) - Y}{h} \qquad (h \neq 0)$$

belongs to this real vector space. Its limit is $XY - YX$. Closedness therefore supplies the commutator. □

There is a useful variant. If $f(h) \in G$ is a differentiable unitary curve with $f(0) = I$ and $f'(0) = X$, compare it with $\exp(hX)$ in Lemma A.2. The powers $f(t/n)^n$ show that $X \in L_G$. A direction can thus enter through a small available curve even before its exponential has been identified.

Lemma A.4 (A trace-zero logarithm)**.** *Every unitary matrix U with determinant one is* $\exp X$ *for some $X \in \mathfrak{su}(d)$.*

Proof. The finite-dimensional spectral theorem gives an orthonormal eigenbasis and real phases θ_j in which U is diagonal with entries $e^{i\theta_j}$. Since $\det U = 1$, the sum $\sum_j \theta_j$ is an integer multiple of 2π. Subtract that sum from one of the phases. The exponentials are unchanged, and the adjusted phases sum to zero. The diagonal matrix of their multiples by i is skew-Hermitian with trace zero. Conjugating back gives X. □

Consequently, $\mathfrak{su}(d) \subseteq L_G$ implies that every determinant-one unitary is a norm limit of available finite sequences. No assertion about the rate of approximation has entered the argument.

A.3 Two native cells supply all matrix directions

One cell has basis labels $0, \dots, 7$. Put

$$\sigma = (1, -1, -1, 1, -1, 1, 1, -1), \qquad S = \operatorname{diag}(\sigma),$$

and let F interchange 0 with 1, 2 with 3, 4 with 5, and 6 with 7. Write $f(c)$ for the partner of c under this interchange. Thus $S^2 = F^2 = I$, both matrices are Hermitian, and $\sigma_{f(c)} = -\sigma_c$.

The local rotations have directions

$$C_{abc} = R_{ab} + R_{bc} + R_{ca}$$

for distinct indices. Their normalized generators are $C_{abc}/\sqrt{3}$. It suffices to supply the 21 choices C_{ab7} with $0 \le a < b < 7$, since

$$C_{abc} = C_{ab7} + C_{bc7} + C_{ca7}$$

when a, b, c avoid 7, with the signs adjusted by reversing a pair when necessary. The nonzero normalization does not change the real span of the directions.

For two cells, supply the couplings

$$E_{12} = \frac{I + iS \otimes F}{\sqrt{2}}, \qquad E_{21} = \frac{I + iF \otimes S}{\sqrt{2}}.$$

They are unitary because each tensor factor product is a Hermitian involution. Inverses are also available as finite sequences: if $E = (I + iX)/\sqrt{2}$ with $X^2 = I$, then $E^8 = I$.

Let L be a real Lie algebra containing the local cyclic directions on each cell and preserved by conjugation with both couplings. We will prove $\mathfrak{su}(64) \subseteq L$. This applies to L_G by Proposition A.3.

The couplings separate local real edges

For $X^2 = I$, direct multiplication gives

$$\mathrm{Ad}_E(Y) = \frac{Y + XYX}{2} + \frac{i}{2}(XY - YX), \qquad \mathrm{Ad}_E^2(Y) = XYX. \tag{A.2}$$

It follows that L contains both XYX and $i(XY - YX)$ whenever it contains Y. Indeed the second expression is $2\,\mathrm{Ad}_E(Y) - Y - \mathrm{Ad}_E^2(Y)$.

For an operator on the first cell, conjugation by $S \otimes F$ gives conjugation by S on that cell. The corresponding assertion holds for the second cell using $F \otimes S$. Since $SR_{ab}S = \sigma_a\sigma_b R_{ab}$, cyclic sums can be separated. If a, b have the same sign, choose k of the opposite sign. Then

$$\frac{C_{abk} + SC_{abk}S}{2} = R_{ab}.$$

If a, b have opposite signs, choose $k \neq a$ of the same sign as a. We already have R_{ak}. Subtracting it from C_{akb} leaves $Z = R_{kb} + R_{ba}$. The edge identities give

$$[R_{ak}, Z] = R_{ab} + R_{kb}, \qquad \frac{[R_{ak}, Z] - Z}{2} = R_{ab}.$$

There are four indices of each sign, so the required choices exist. Thus $R_{ab} \otimes I$ and $I \otimes R_{cd}$ belong to L for every distinct pair.

The couplings produce imaginary tensor directions

Write $P_c = E_{cc}$ and $Q_{ab} = E_{ab} + E_{ba}$. The imaginary commutator supplied by (A.2) gives

$$i[S, R_{ab}] \otimes F = i(\sigma_a - \sigma_b)Q_{ab} \otimes F.$$

For opposite signs the coefficient is nonzero, so $iQ_{ab} \otimes F$ belongs to L. Commuting with $I \otimes R_{c,f(c)}$ uses

$$[F, R_{c,f(c)}] = 2(P_{f(c)} - P_c)$$

and gives $iQ_{ab} \otimes (P_c - P_{f(c)})$ after a real rescaling. For $d \neq c$, commute this with $I \otimes R_{cd}$. The result is $iQ_{ab} \otimes Q_{cd}$, multiplied by 1 if $d \neq f(c)$ and by 2 if $d = f(c)$. Another commutator with $I \otimes R_{cd}$ gives a nonzero real multiple of $iQ_{ab} \otimes (P_d - P_c)$.

Interchanging the cells gives the complementary family. We have proved that L contains

$$\begin{aligned} &iQ_{ab} \otimes Q_{cd}, \quad iQ_{ab} \otimes (P_c - P_d) \quad \text{if } \sigma_a \neq \sigma_b,\ c \neq d, \\ &\qquad\qquad\quad i(P_a - P_b) \otimes Q_{cd} \quad \text{if } a \neq b,\ \sigma_c \neq \sigma_d. \end{aligned} \tag{A.3}$$

The sign conditions have been retained; removing them prematurely would hide the work done by the next step.

A tensor direction can be confined to one column

Suppose a, b have opposite signs and $d \neq e$. Choose c distinct from d, e with sign opposite to that of e. The following two matrices are available from (A.3):

$$A = iQ_{ab} \otimes Q_{cd}, \qquad B = i(P_a - P_b) \otimes Q_{ce}.$$

Their commutator is $R_{ab} \otimes Q_{de}$. To check this, use $Q_{ab}(P_a - P_b) = -R_{ab}$, $(P_a - P_b)Q_{ab} = R_{ab}$, and $Q_{cd}Q_{ce} = E_{de}$ for the three distinct indices. Commuting with $I \otimes R_{de}$ then gives $2R_{ab} \otimes (P_e - P_d)$.

Sum the latter differences over e. Since

$$\sum_e (P_e - P_d) = I - 8P_d,$$

and $R_{ab} \otimes I$ is already available, we obtain $R_{ab} \otimes P_d$. If a, b have the same sign, choose k of the opposite sign and use

$$[R_{ak} \otimes P_d, R_{kb} \otimes P_d] = R_{ab} \otimes P_d.$$

Thus every real edge within a fixed second coordinate is present:

$$R_{ab} \otimes P_c = R_{(a,c),(b,c)} \in L. \tag{A.4}$$

Every imaginary edge in such a column is present too. Given distinct b, e, choose a different from both with sign opposite to that of b. The available matrix

$$iQ_{ab} \otimes (P_c - P_{f(c)}) = I_{(a,c),(b,c)} - I_{(a,f(c)),(b,f(c))}$$

has two disjoint supports. Its commutator with $R_{(a,c),(b,c)}$ isolates $2D_{(a,c),(b,c)}$. Commuting that diagonal difference with $R_{(b,c),(e,c)}$ supplies $I_{(b,c),(e,c)}$, up to a nonzero real factor.

The columns connect

For $a \neq e$ and $c \neq d$, choose b different from a, e and of sign opposite to a. Equation (A.3) contains

$$iQ_{ab} \otimes Q_{cd} = I_{(a,c),(b,d)} + I_{(a,d),(b,c)}.$$

Commute this with the already available $I_{(b,d),(e,d)}$. The second summand has disjoint support and contributes zero; the first gives $-R_{(a,c),(e,d)}$.

We now have real edges between different first coordinates, whether or not their second coordinates agree. Two vertices with the same first coordinate can be joined through any vertex with a different first coordinate. These edges form a connected graph on the 64 states. We also have imaginary column edges. Applying Lemma A.1 proves

$$\mathfrak{su}(64) \subseteq L. \tag{A.5}$$

Every step used the stated cyclic rotations, the two ordered couplings, real linear combinations and commutators.

A.4 Overlapping pairs control a whole register

Two-cell control must still be connected across a longer register. The following argument makes that step explicit.

Lemma A.5 (Overlapping factors)**.** *Let A, B, C be nonempty finite index sets, with $|A|, |B| \geq 3$. A real Lie algebra containing*

$$\mathfrak{su}(A \times B) \otimes I_C \quad \text{and} \quad I_A \otimes \mathfrak{su}(B \times C)$$

contains $\mathfrak{su}(A \times B \times C)$.

Proof. For a real function δ on the indices, write $D(\delta) = i\,\mathrm{diag}(\delta)$. Direct multiplication gives

$$[D(\delta), R_{xy}] = (\delta_x - \delta_y) I_{xy}, \qquad [D(\delta), I_{xy}] = -(\delta_x - \delta_y) R_{xy}.$$

These equations allow a diagonal difference to select one part of an edge sum.

First fix a_0 and distinct b, b'. From the second supplied algebra we have

$$T = \sum_a R_{(a,b,c),(a,b',c')}.$$

Choose a_1, a_2 distinct from each other and from a_0. A diagonal difference from the first algebra, equal to 1 on the slice (a_0, b), -1 on (a_1, b) and zero elsewhere, turns T into the difference of its two corresponding imaginary edges. Commute once more with the diagonal difference on slices (a_0, b) and (a_2, b). Only the a_0 edge survives, with a minus sign. Hence $R_{(a_0,b,c),(a_0,b',c')}$ is available.

Next start from the first algebra's sum

$$T' = \sum_{c'} R_{(a,b,c'),(a',b',c')}, \qquad b \neq b'.$$

Choose b_2 outside $\{b, b'\}$. A diagonal difference from the second algebra, equal to 1 on the fibre (b, c) and -1 on (b_2, c), selects $I_{(a,b,c),(a',b',c)}$ from T'. A second commutator with that same diagonal gives the corresponding real edge.

We therefore have real edges whenever $b \neq b'$ and either the first or the third coordinate agrees. This graph is connected. In fact, choose b_2 different from both b, b' to join any two vertices by

$$(a, b, c) \longrightarrow (a, b_2, c') \longrightarrow (a', b', c').$$

The fibre construction also supplies an imaginary edge. Lemma A.1 finishes the proof. □

Theorem A.6 (Approximation by native controls)**.** *On a register of $n \geq 2$ eight-state cells, supply the local cyclic rotations and both ordered couplings $E_{j,j+1}$ and $E_{j+1,j}$ on every adjacent pair. Finite sequences of these operations approximate every determinant-one unitary on the register to arbitrary positive operator-norm accuracy.*

Proof. Take the closed group G generated by those operations. Its algebra L_G contains the local cyclic directions, and conjugation by the couplings preserves it. The two-cell argument gives all of $\mathfrak{su}(64)$ on each adjacent pair.

Induct on the number of cells. The case of two cells was proved above. For the induction step, separate the first cell, second cell and the remaining tail. Full control on the first pair and, by induction, on the second cell together with the tail satisfies Lemma A.5. Thus L_G contains all traceless skew-Hermitian matrices on the complete register. The logarithm lemma places every determinant-one unitary in G, which is precisely the asserted approximation property. □

A.5 A finite alphabet has an infinite-order seed

The search construction in Chapter 18 needs a finite alphabet with effectively accessible entries. Use binary coordinates for each eight-state cell. On one binary coordinate let

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}.$$

The third gate flips a target binary coordinate when both selected control coordinates are 1. The target differs from each control; the two controls are allowed to coincide. With coincident controls this is the controlled-NOT gate. This convention matters in the register argument below.

These gates are unitary and have finite-order inverses: $H^2 = I$, $P^4 = I$, and the controlled flip squares to I. Their finite words therefore form a group. Entries of every word lie in $\mathbb{Q}(\sqrt{2}, i)$.

On three binary coordinates, let T interchange states 6 and 7, and let $H_* = H_2 H_1 H_0$. Consider the eight-gate word

$$U = (H_* T)^2.$$

This is an exact choice of gates, with no angle selected from a numerical fit. Its infinite order will supply a continuous direction in the closure of the finite words.

Let e_j be the coordinate vectors and put

$$v = e_6 - e_7, \qquad w = e_1 - e_3 - e_5 + e_7.$$

The unnormalized Walsh matrix $W = 2\sqrt{2}\, H_*$ has entries $W_{jk} = (-1)^{j \cdot k}$, where the dot product is computed modulo two on the three binary coordinates. Expanding the two columns indexed by $6, 7$, and then the four columns indexed by $1, 3, 5, 7$, gives

$$Wv = 2w, \quad Ww = 4v, \quad Tv = -v, \quad Tw = w + v.$$

Consequently

$$Uv = -v - \tfrac{1}{2}w, \qquad Uw = v - \tfrac{1}{2}w. \tag{A.6}$$

The restriction to the plane spanned by v, w has characteristic polynomial $x^2 + \frac{3}{2}x + 1$ and eigenvalue

$$\lambda = \frac{-3 + i\sqrt{7}}{4}.$$

It has modulus one. It is not a root of unity. If $\lambda^k = 1$ for some positive integer k, both λ and its conjugate would be algebraic integers, so their sum $-3/2$ would be an algebraic integer. A rational algebraic integer must be an integer: substituting a reduced fraction into a monic integer polynomial shows that its denominator divides its numerator's highest power. The denominator must therefore be 1. This contradicts the sum $-3/2$.

The closure of the integer powers of any unit complex number of infinite order is the whole circle. Here is the needed elementary argument. Partition the circle into N equal arcs and

place $N+1$ powers in them. Two lie in the same arc. Their quotient, or its inverse, has argument β with $0 < \beta < 2\pi/N$. Successive multiples of β before one full turn leave gaps at most β, including the final gap. They approximate every point of the circle within $2\pi/N$ in angle. They are still integer powers of the original number. Letting N increase proves density.

To identify the resulting matrix direction, calculate $v^T v = 2$, $w^T w = 4$ and $v^T w = -1$. The orthogonal projection onto their plane is

$$Q = \frac{4vv^T + vw^T + wv^T + 2ww^T}{7}.$$

Set $B = vw^T - wv^T$ and $A = B/\sqrt{7}$. The same inner products give

$$Q^2 = Q, \qquad A^2 = -Q, \qquad QA = AQ = A.$$

Since $T = I - vv^T$ and $W^2 = 8I$, expansion of $(H_* T)^2$ gives

$$U = I - vv^T - \tfrac{1}{2}ww^T - \tfrac{1}{2}wv^T = I - \tfrac{7}{4}Q + \tfrac{1}{4}B.$$

For $|z| = 1$ define

$$M(z) = I - Q + \mathrm{Re}(z)Q + \mathrm{Im}(z)A.$$

The displayed relations show $M(z_1 z_2) = M(z_1)M(z_2)$, $M(\lambda) = U$, and $M(e^{it}) = \exp(tA)$. The last equality follows directly by separating even and odd terms of the exponential series, using $A^2 = -Q$. Density of the powers of λ, together with continuity of M, puts $\exp(tA)$ in the closure of finite words for every real t. Hence A, and therefore B, belongs to L_G for that closure.

A.6 Sign masks isolate the seed's edges

An infinite-order rotation alone need not control a whole space. We now use the other gates to separate its matrix entries.

The square P^2 is the binary sign gate $Z = \mathrm{diag}(1, -1)$. Products of such gates give every diagonal sign mask

$$D_s e_j = (-1)^{s \cdot j} e_j, \qquad s, j \in \{0, 1\}^3.$$

Conjugating by a mask multiplies R_{uv} by $(-1)^{s \cdot (u \oplus v)}$, where $\oplus$ is coordinatewise addition modulo two. The real linear combination

$$\Pi_d(Y) = \frac{1}{8} \sum_s (-1)^{s \cdot d} D_s Y D_s$$

therefore retains exactly the edges with $u \oplus v = d$. Indeed, the coefficient is the average of the character $(-1)^{s \cdot (d \oplus u \oplus v)}$. It is 1 if its index is zero; otherwise pairing s with the result of flipping a coordinate where that index is 1 makes the sum zero.

Expansion of $B = vw^T - wv^T$ yields

$$B = R_{61} - R_{63} - R_{65} + R_{67} - R_{71} + R_{73} + R_{75}.$$

Thus

$$\Pi_1(B) = R_{67}, \quad \Pi_2(B) = R_{75}, \quad \Pi_4(B) = R_{73}, \quad \Pi_6(B) = -R_{71}.$$

Each lies in L_G, because the masks are available words and L_G is a real vector space preserved by their conjugations.

The bit flip is also available: $X = HZH$. Conjugating R_{67} by flips of the higher coordinates gives R_{23}, R_{45} and R_{01}. The edges

$$01,\ 23,\ 45,\ 67,\ 71,\ 73,\ 75$$

form a connected graph on the eight labels. Finally, conjugating R_{67} by the phase gate on the lowest binary coordinate gives $-I_{67}$. Lemma A.1 now gives

$$\mathfrak{su}(8) \subseteq L_G. \tag{A.7}$$

The finite alphabet has acquired every local continuous direction through closure. The source of that continuity was the explicit infinite-order seed, not a continuously adjustable gate in the alphabet.

A.7 The finite alphabet connects the cells

It remains to obtain the native coupling used earlier. On one cell, the sign matrix S is the product of the three binary Z gates: it records the parity of the three bits. The matrix F is the flip X on the lowest bit.

For distinct binary coordinates a, b, a controlled-NOT with control a and target b conjugates Z_b to $Z_a Z_b$. This follows on a basis state: after the target flip, its sign is $(-1)^{x_b + x_a}$. Use three such controlled-NOT gates, controlled by the three bits of cell j and targeted at the lowest bit of cell k. Their product conjugates $Z_{k,0}$ to $S_j Z_{k,0}$. Conjugating next by H on that target bit changes $Z_{k,0}$ to F_k. Hence a finite word conjugates $iZ_{k,0}$ to iS_jF_k.

Equation (A.7) supplies $iZ_{k,0}$ as a local direction. Conjugation therefore supplies iS_jF_k, and its exponential at $\pi/4$ is

$$\exp\big(i\tfrac{\pi}{4} S_j F_k\big) = \frac{I + iS_jF_k}{\sqrt{2}} = E_{jk}.$$

This coupling belongs to the closed word group. The reversed ordered coupling is obtained by reversing the roles of the cells. All cyclic directions are already local members of $\mathfrak{su}(8)$. The native register argument thus applies inside this same closed group.

Theorem A.7 (Approximation by the finite alphabet)**.** *On any register of $n \geq 2$ eight-state cells, the gates H, P and the controlled flip with the stated repeated-control convention generate finite words dense in the determinant-one unitaries, in operator norm.*

Proof. The preceding construction puts the local cyclic rotations and adjacent ordered couplings in the closure of finite words. The proof of Theorem A.6 then places every determinant-one unitary in that closure. □

This is the existence statement needed by the rational search. The finite words can be enumerated, their entries can be enclosed by rational intervals, and their errors can be certified with slack. The search uses those facts to find an approximation; density ensures that a sufficient one will eventually be found. It supplies no practical running-time guarantee.

A.8 A complete record-and-field assembly

The pulse theorem needs a full register and a positive holding operator on it. We now give the record-and-field extension used for the native two-flux construction. The retained source is supplied on $14 + 2N$ eight-state cells with a positive interior Hamiltonian $H_{\mathrm{old}}(N, m)$. Here $N \geq 0$ fixes the size of its two reference banks, and $m \in \mathbb{Z}$ is a fixed source-flux parameter. Its material construction is outside this extension. The task here is to specify every added space and energy term, then check that the whole space has the form required by the pulse theorem.

We work in the dimensionless energy units of the following cell clock. The energy scale and $\hbar$ must be used consistently when applying Chapter 18.

One complete cell pair

Let $\mathscr{H} = \mathbb{C}^8$, with the S and F already specified in this appendix. Define

$$Q = iSF, \qquad P = \frac{I - Q}{2}, \qquad \bar{P} = I - P.$$

Because $SF = -FS$ and $S^2 = F^2 = I$, the matrix Q is a Hermitian involution and $P, \bar{P}$ are complementary orthogonal projections. They each have rank four. An explicit orthonormal basis for the range of $\bar{P}$ is given by the columns

$$B_j = \frac{e_{2j} - i\sigma_{2j}e_{2j+1}}{\sqrt{2}}, \qquad j = 0, 1, 2, 3.$$

Direct multiplication gives $QB = B$, $B^\dagger B = I_4$ and $BB^\dagger = \bar{P}$. The columns of FB form an orthonormal basis for the range of P, since $QF = -FQ$.

The eight-level clock is the positive matrix

$$C = T\,\mathrm{diag}(0, 1, \dots, 7)T^\dagger, \qquad T_{jk} = \frac{e^{-i\pi jk/4}}{\sqrt{8}}. \tag{A.8}$$

The finite geometric sum shows that the columns of T are orthonormal. Thus $0 \leq C \leq 7I$. Average its four conjugates to obtain

$$\bar{C} = \frac{C + SCS + FCF + QCQ}{4} \geq 0.$$

On the complete pair space $\mathscr{P} = \mathscr{H} \otimes \mathscr{H}$, introduce the isometry

$$J = \bar{P} \otimes B + P \otimes FB : \mathscr{H} \otimes \mathbb{C}^4 \longrightarrow \mathscr{P}.$$

Orthogonality of P and $\bar{P}$ gives

$$J^\dagger J = I_{32}, \qquad JJ^\dagger = \bar{P} \otimes \bar{P} + P \otimes P.$$

It puts the two cells in matching occupation sectors, while retaining all four reference coordinates in either sector. Define the full pair holding by

$$H_{\text{pair}} = J(C \otimes I_4)J^\dagger + (\bar{P}C\bar{P}) \otimes P + (PCP) \otimes \bar{P} + I_8 \otimes \bar{C} + 7I_{64}. \tag{A.9}$$

Every summand is positive. For example, $v^\dagger J(C \otimes I_4)J^\dagger v \geq 0$ by positivity of $C \otimes I_4$, and PCP is positive by the same argument. The two middle terms keep the clock on the unmatched occupation sectors. Equation (A.9) is therefore an operator on all 64 pair coordinates, with its seven-unit scalar baseline included.

For n pairs, write

$$\mathscr{B}_n = \mathscr{P}^{\otimes n}, \qquad H_n = \sum_{r=1}^{n} (H_{\text{pair}})_r.$$

The subscript r places the pair operator in slot r and identities in every other slot. The space has dimension $64^n = 8^{2n}$, and $H_n \geq 0$. For $n = 0$, the space is one-dimensional and the sum is zero. Each pair retains its complete clock and reference contribution.

Three fresh pairs and their connecting edge

The added record has three original cells and three compensating reference cells. Its full space is $\mathscr{R} = \mathscr{B}_3$, of dimension 8^6, and its internal holding is H_3. The source-minus-reference occupation is

$$Q_{\text{fresh}} = \sum_{r=1}^{3} (P \otimes I - I \otimes P)_r.$$

In the occupation basis of each cell, its eigenvalue is $\sum_r (u_r - v_r)$, where $u_r, v_r \in \{0, 1\}$. The four-dimensional coordinate inside each occupation sector is retained. Thus the fresh charge takes the integer values from -3 to 3 on the complete space, including sectors outside a neutral preparation.

The selected edge energy is

$$c(j) = 4\left[j \operatorname{arsinh}(j/2) - \sqrt{j^2 + 4} + 2\right], \qquad j \in \mathbb{Z}. \tag{A.10}$$

The coefficient four is part of this specified apparatus. To see positivity, put $t = \operatorname{arsinh}(j/2)$. The bracket becomes $2(t \sinh t - \cosh t + 1)$. This is even, vanishes at zero and has derivative $2t \cosh t \geq 0$ for $t \geq 0$. Hence $c(j) \geq 0$.

The connecting-edge operator E_{fresh} is diagonal in the occupation basis, with entry $c(\sum_r (u_r - v_r))$ on that coordinate. Equivalently, it is $c(Q_{\text{fresh}})$ by the finite spectral calculus. It is positive on the entire fresh record. The new edge contributes zero on the neutral preparation, but its operator remains in the full holding.

Fix also an exterior flux $\beta \in \mathbb{Z}$. With all operators in physical cell coordinates, the placed record holding is

$$H_{\text{placed}} = H_{\text{old}}(N, m) \otimes I_{\mathscr{R}} + I_{\text{old}} \otimes H_3 + c(\beta)I + I_{\text{old}} \otimes E_{\text{fresh}}. \tag{A.11}$$

The retained interior source holding excludes the exterior edge counted by $c(\beta)I$ here. This expression counts that edge once, adds the fresh connecting edge once, and preserves every old and fresh coordinate. Each term is positive. The old-source positivity is the supplied input; the added terms have just been constructed and proved positive.

Two finite field banks

Choose nonnegative integers a, b. The two banks have $2a$ and $2b$ complete pairs:

$$\mathscr{F}_{a,b} = \mathscr{B}_{2a} \otimes \mathscr{B}_{2b}, \qquad H_{\text{field}} = H_{2a} \otimes I + I \otimes H_{2b}.$$

Their full dimension is 8^{4a+4b}, and their holding is positive. A useful code in a bank of n pairs has a chain label $k \in \{0, \dots, n\}$ and two lists of n four-valued coordinates. The first k source cells are occupied and the others unoccupied; each reference cell has matching occupation. The two lists choose the coordinates inside the source and reference occupation sectors. Distinct code labels give orthogonal vectors, so this is an isometric encoding of dimension $(n+1)4^n4^n$ into the full 64^n-dimensional bank.

For the two banks, their chain labels k, l describe the integer pair $(k-a, l-b)$. These pairs cover the box $[-a, a] \times [-b, b]$. If W is the resulting field-code isometry, then

$$WW^\dagger + (I - WW^\dagger) = I.$$

The second projection keeps the unused coordinates visible. The holding operator remains H_{field} on the whole space; no invariance of the selected code under that holding is assumed.

These field labels do not themselves make the old-source parameter m or the exterior parameter β vary with the encoded state. They remain fixed in (A.11). A source coupling that reads those labels as variable physical flux would be an additional operator correspondence. It is not needed for the full-register pulse theorem proved here.

The whole register and the coordinate bijection

The complete carrier and holding are

$$\mathscr{K} = (\mathscr{H}^{\otimes(14+2N)} \otimes \mathscr{B}_3) \otimes \mathscr{F}_{a,b}, \tag{A.12}$$

$$H_{\text{reg}} = H_{\text{placed}} \otimes I_{\text{field}} + I_{\text{placed}} \otimes H_{\text{field}}. \tag{A.13}$$

Both terms are positive, hence $H_{\text{reg}} \geq 0$. This conclusion concerns the full holding, including the scalar edge energy and every unmatched or unused coordinate.

To identify its cell count, take a tuple of

$$n = 20 + 2N + 4a + 4b$$

labels from $\{0, \dots, 7\}$. Split it into successive blocks of lengths $14 + 2N$, 6, $2a$, $2a$, $2b$, $2b$. The first is the old source. Group the six fresh labels into three source–reference pairs. Pair the two $2a$-blocks coordinate by coordinate to form the first field bank, and do the same for the two $2b$-blocks. Reading the source and reference entries of each pair and concatenating the blocks reverses this map. This gives the required basis bijection with n eight-state cells. It proves the tensor-power form directly, rather than inferring it from finite dimension alone.

The permutation matrix of this bijection transports the input layout, holding, evolution and complete reader together. Conjugation preserves products, adjoints, exponentials and operator norms, so it preserves the pulse error bound. This is a change of mathematical coordinates. It does not supply a free physical permutation actuator. A separate unitary choice of occupation coordinates can likewise be applied to the full operators without compressing their space.

For a requested instrument, choose the register sizes large enough for the fixed layout of Chapter 18. Since $n \geq 20$, the requirement of at least two cells is already met. The positive-pulse theorem now applies to this specified H_{reg}, conditional on the retained positive source holding, preparation, complete readout, selectable controls and external switching. Shorter pulses still require the larger amplitudes in (18.14). The construction supplies the complete finite extension and its error accounting under precisely those interfaces.

Appendix B

Deriving the conserved native count

The recorder bound uses the sum of the projections P_c across the source and reference cells. This appendix derives its conservation from the two native conserved quantities, clock energy and occupation. The calculation is finite. It gives an expression for P_c using the two one-cell charges, identity, linear combinations and commutators. Those operations transport to sums of charges over arbitrarily many cells.

B.1 Separate the spectral gaps of the clock

Conjugate the one-cell operators by the Fourier matrix W of Chapter 19. Write

$$D = W^* K W = \operatorname{diag}(0, 1, \dots, 7), \qquad T = W^*(iSF)W.$$

Since $N = (I - iSF)/2$, we also have $T = I - 2W^*NW$. Thus D and T are available from the two charges and identity. The matrix T is Hermitian.

Let E_{ij} denote a matrix unit, with indices $0, \dots, 7$. For any matrix X,

$$[D, X]_{ij} = (i - j)X_{ij}.$$

The possible integer gaps are $-7, -6, \dots, 7$. For one of these gaps d, define the polynomial

$$p_d(t) = \prod_{\substack{-7 \le k \le 7 \\ k \ne d}} \frac{t - k}{d - k}.$$

Every denominator is nonzero. At an allowed integer gap, p_d is 1 for gap d and 0 for all the others. Therefore

$$\big(p_d(\operatorname{ad}_D)X\big)_{ij} = \begin{cases} X_{ij}, & i - j = d, \\ 0, & i - j \ne d, \end{cases} \qquad \operatorname{ad}_D(X) = [D, X]. \tag{B.1}$$

The left side is a finite linear combination of iterated commutators. This formula extracts a specified spectral gap without assuming that its matrix entries are independently controllable.

Put $A_d = p_d(\operatorname{ad}_D)T$. Because T is Hermitian, $A_{-d} = A_d^*$. Both sides are obtained by the same gap-filter construction; taking an adjoint introduces no extra generator.

B.2 The entries needed for the certificate

The Fourier entries can be calculated directly:

$$T_{ij} = \frac{1}{8} \sum_{k=0}^{7} \overline{\omega^{ki}}\, i s_k\, \omega^{f(k)j}. \tag{B.2}$$

Here s_k is the kth diagonal entry of S, and f interchanges 0 with 1, 2 with 3, 4 with 5, and 6 with 7. The powers of ω are

$$1,\ \frac{1-i}{\sqrt{2}},\ -i,\ -\frac{1+i}{\sqrt{2}},\ -1,\ \frac{-1+i}{\sqrt{2}},\ i,\ \frac{1+i}{\sqrt{2}}.$$

Only gaps 7, 5 and 3 are needed. Substituting these eight powers in (B.2) gives

$$\begin{aligned}
A_7 &= a_7 E_{70},\\
A_5 &= a_{50}E_{50} + a_{61}E_{61} + a_{72}E_{72},\\
A_3 &= a_{30}E_{30} + a_{30}E_{41} + a_{52}E_{52} + a_{63}E_{63} + a_{74}E_{74},
\end{aligned}$$

where

$$\begin{aligned}
a_7 &= \frac{-1+i-i\sqrt{2}}{4}, & a_{50} &= \frac{1+i+i\sqrt{2}}{4},\\
a_{61} &= \frac{-1+\sqrt{2}+i}{4}, & a_{72} &= \frac{1-\sqrt{2}-i}{4},\\
a_{30} &= \frac{-1+i+i\sqrt{2}}{4}, & a_{52} &= \frac{1-\sqrt{2}+i}{4},\\
a_{63} &= \frac{-1+\sqrt{2}-i}{4}, & a_{74} &= \frac{1-i-i\sqrt{2}}{4}.
\end{aligned}$$

All remaining entries in these three gap matrices vanish by (B.1).

The numbers needed below are their squared magnitudes. Set

$$\alpha = \frac{2+\sqrt{2}}{8}, \qquad \beta = \frac{2-\sqrt{2}}{8}, \qquad c_+ = 8+4\sqrt{2}, \quad c_- = 8-4\sqrt{2}.$$

The displayed coefficients give

$$\begin{aligned}
|a_7|^2 = |a_{61}|^2 = |a_{72}|^2 &= \beta,\\
|a_{50}|^2 = |a_{30}|^2 = |a_{74}|^2 &= \alpha,\\
c_+\beta = 1, \qquad c_-\alpha &= 1.
\end{aligned}$$

For example, $|a_{50}|^2 = (1+(1+\sqrt{2})^2)/16 = (2+\sqrt{2})/8$. These exact radical identities fix all coefficients in the count expression.

B.3 A finite expression for the odd-mode projection

First isolate the two end levels:

$$D_{70} = c_+[A_7, A_7^*] = E_{77} - E_{00}.$$

For a diagonal matrix D', the coefficient of E_{ij} in $[D', X]$ is multiplied by $D'_{ii} - D'_{jj}$. Hence

$$B_5 = [D_{70}, A_5] = a_{50}E_{50} + a_{72}E_{72}, \qquad B_3 = [D_{70}, A_3] = a_{30}E_{30} + a_{74}E_{74}.$$

The interior entries have disappeared. Their adjoints are likewise finite expressions: $B_5^* = [A_5^*, D_{70}]$ and $B_3^* = [A_3^*, D_{70}]$.

Introduce four diagonal matrices obtained by further commutators:

$$\begin{aligned} Q_{\mathrm{all}} &= [A_5, A_5^*], \\ Q_5 &= [B_5, B_5^*], \\ T_5 &= [[Q_5, B_5], B_5^*], \\ Q_3 &= [B_3, B_3^*]. \end{aligned}$$

The rule $E_{ij}E_{kl} = \delta_{jk}E_{il}$ makes their calculation short. The nonzero entries of A_5, B_5 and B_3 have different source and target indices, so cross terms vanish. We obtain

$$Q_{\mathrm{all}} = \alpha(E_{55} - E_{00}) + \beta(E_{66} - E_{11}) + \beta(E_{77} - E_{22}), \tag{B.3}$$

$$Q_5 = \alpha(E_{55} - E_{00}) + \beta(E_{77} - E_{22}), \tag{B.4}$$

$$T_5 = 2\alpha^2(E_{55} - E_{00}) + 2\beta^2(E_{77} - E_{22}), \tag{B.5}$$

$$Q_3 = \alpha(E_{33} - E_{00} + E_{77} - E_{44}). \tag{B.6}$$

For (B.5), the two entries of $[Q_5, B_5]$ acquire factors 2α and 2β. Commuting with B_5^* gives the stated squared coefficients.

Proposition B.1 (The count identity)**.** *The odd-mode projection is the following finite expression:*

$$\begin{aligned} P_{\mathrm{odd}} = \frac{1}{2}\big(I - D_{70} - c_+ Q_{\mathrm{all}} + (24 + 4\sqrt{2})Q_5 \\ - 16T_5 + c_- Q_3\big), \end{aligned} \tag{B.7}$$

where $P_{\mathrm{odd}} = \mathrm{diag}(0, 1, 0, 1, 0, 1, 0, 1)$.

Proof. All matrices on the right are diagonal. Set $d = 24 + 4\sqrt{2}$ and write $s = \sqrt{2}$. In addition to $c_+\beta = c_-\alpha = 1$, direct substitution using $s^2 = 2$ gives

$$(d - c_+)\alpha - 32\alpha^2 = 1, \qquad (d - c_+)\beta - 32\beta^2 = 1.$$

The diagonal entries inside the parentheses in (B.7), in index order, are therefore

$$\begin{gathered} 2 - (d - c_+)\alpha + 32\alpha^2 - c_-\alpha = 0, \\ 1 + c_+\beta = 2, \\ 1 - (d - c_+)\beta + 32\beta^2 = 0, \\ 1 + c_-\alpha = 2, \\ 1 - c_-\alpha = 0, \\ 1 + (d - c_+)\alpha - 32\alpha^2 = 2, \\ 1 - c_+\beta = 0, \\ (d - c_+)\beta - 32\beta^2 + c_-\alpha = 2. \end{gathered}$$

Thus the diagonal inside the parentheses is $(0, 2, 0, 2, 0, 2, 0, 2)$. Division by 2 proves the formula. □

Finally, the jth Fourier column is an eigenvector of the four-step shift with eigenvalue $\omega^{4j} = (-1)^j$. Consequently

$$W^* P_c W = P_{\mathrm{odd}}.$$

Conjugating (B.7) by W gives the required expression for P_c in I, K, N. Conjugation preserves linear combinations and commutators. The gap filters and all intermediate matrices above are explicit finite expressions in those operations.

B.4 Transport to a conserving recorder

For m cells, possibly accompanied by a neutral factor, write $\mathscr{L}_m(T) = \sum_{j=1}^{m} T^{(j)}$. This map preserves addition and scalar multiplication. Expanding the commutator of two such sums,

$$[\mathscr{L}_m(T), \mathscr{L}_m(T')] = \sum_{j,k} [T^{(j)}, T'^{(k)}] = \sum_{j} [T, T']^{(j)} = \mathscr{L}_m([T, T']).$$

Terms on different slots vanish. Also $\mathscr{L}_m(I) = mI$; this scalar multiple of identity commutes with every interaction.

If an operator U commutes with two matrices, it commutes with their products, sums, scalar multiples and commutator. Start with the two conserved collective charges. Apply these facts successively to the finite expression just proved. The result is

$$[U, \mathscr{L}_m(H_0)] = [U, \mathscr{L}_m(N)] = 0 \quad \Longrightarrow \quad [U, \mathscr{L}_m(P_c)] = 0,$$

using the nonzero factor in $H_0 = \varepsilon_0 K$. Taking $m = r + 2$ and grouping the first two slots as the source gives

$$\mathscr{L}_m(P_c) = A \otimes I + I \otimes R,$$

the count used in the recorder theorem. No assumption about the state of the r recorder cells occurs in this transport argument.

Appendix C

A toolkit for reading the proofs

A long proof becomes easier to follow when its small moves are familiar. This appendix explains the moves used throughout the book and points to places where they do actual work. It can be read before the main text or consulted when a particular step becomes difficult.

The notation guide in Appendix N gives shorter reminders. Here the aim is to explain how to use the notation while following an argument.

C.1 Read a theorem as a task

A theorem separates the data supplied from the conclusion to be obtained. Start by identifying both. In Theorem 2.3, the supplied data are a carrier, a start and a step. The conclusion asks for a map with two equations, and says that only one such map exists. There are therefore two tasks: construct the map and rule out a different one.

The construction comes first. Its recursive equations tell us what value to give the start and how to continue after a step. For uniqueness, the proof introduces an arbitrary competing map. It uses only the two equations required of that competitor, then proves agreement at every finite record. The competitor's arbitrariness is what lets the conclusion cover all permitted maps.

When the next theorem adds injectivity, do not carry that condition backward into the earlier proof. The unique reading exists even for a target that merges records. The conditions in Definition 2.4 enter the later proof that the reading is bijective. Separating the tasks prevents a uniqueness theorem from acquiring a stronger conclusion than it proves.

An efficient way to read a displayed calculation is to mark each equality by its source: a definition, the current hypothesis or an earlier result. If the source is unclear, inspect that one step before continuing. Long strings of equalities are usually several short substitutions in sequence.

C.2 Induction follows a finite construction

In Chapter 1, an orbit expression is built by the start constructor or by applying the step constructor to an earlier expression. A proof by structural induction follows these two cases. The induction hypothesis is the assertion already available for the earlier expression. The step case must turn that assertion into the assertion for the new one.

Consider the second round trip in Theorem 1.4:

$$e(r(a)) = a.$$

At the start, the defining equations of e and r give the result. In the step case, suppose the

displayed identity holds for a. Then

$$e(r(S(a))) = e(r(a) + 1) = S(e(r(a))) = S(a).$$

The first two equalities unfold definitions. The last uses the induction hypothesis inside S. That is the whole step. It does not assume the round-trip identity for the expression currently being proved.

The object of induction changes across the book. An induction on an orbit expression follows its start and step. An induction on a formula follows its logical constructors. An induction on a derivation follows its last rule and the derivations of its premises. These proofs have a common shape, but their cases come from different definitions.

Some induction hypotheses must quantify over an extra parameter. In the injectivity proof of Chapter 2, the statement for a says that *for every* c, equality of the two readings implies $a = c$. This lets the step proof use the hypothesis at the predecessor d of a newly examined expression $c = S(d)$. A hypothesis fixed at one earlier choice of c would not supply that step.

Recursion constructs an object by the same pattern. Induction proves a property of all objects so constructed. The distinction becomes especially important in Chapter 9, where the available construction rules are themselves being compared inside a specified language.

C.3 Maps, round trips and quotients

A map specifies one output for each input, so a relation with exactly one output for each input is a map; this reading is the principle of unique choice, and the paragraph before Definition 12.21 notes where it matters. A map is *injective* when equal outputs imply equal inputs, and *surjective* when every member of its stated target is an output. A bijection has both properties. The target matters: being onto the range of a map is different from being onto a larger ambient carrier.

The two round trips of Theorem 1.4 do different jobs. Reading and then expanding recovers the original tally. Expanding and then reading recovers the original number. Together they establish a reversible change of representation. Chapter 8 repeats this argument for lists of one mark and then uses both directions to transport quantified formulas. With only the first direction, as Theorem 8.8 shows, purely existential statements go up and purely universal statements come down, and Proposition 8.9 shows that no other quantifier prefix transfers.

A commuting diagram records an equality of routes. In Figure 2.1, the upper and right arrows mean "step, then translate." The left and lower arrows mean "translate, then step." The diagram commutes because both routes have the value required by (2.1). The drawing summarizes an equation; it does not prove the equation by itself.

A quotient has a different purpose. It deliberately identifies objects related by a specified equivalence relation. A representative is one object used to describe a class. In Chapter 4, a signed difference has representatives (a, b), and balance is the relation

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a + d = c + b.$$

An operation on representatives becomes an operation on classes only after we show that replacing the representatives by equivalent ones gives an equivalent output. This is the

well-definedness argument. Without it, the proposed answer could depend on a choice that the quotient has discarded.

For a reader r, the kernel relation identifies exactly those inputs with the same reading. Theorem 8.13 sends the class of x to $r(x)$. The definition of the kernel proves both that this is well-defined and that it is injective. Its target is the range, so an element of that target already comes with a preimage. This is why the theorem gives a bijection to the range without claiming that the reader reaches every value in a larger target space.

C.4 Quantifiers say what information must be supplied

The assertion $\forall x\, P(x)$ asks for the property at an arbitrary input. The assertion $\exists x\, P(x)$ is established constructively by giving an input together with its property. A negation $\neg P$ means that P would lead to an impossible case. These forms provide different information to a later proof.

For example, a proof of $\neg\neg\exists n\, P(n)$ rules out the denial of all witnesses. It does not, by its form alone, display a witness. Chapter 6 states the additional Markov principle used to make that passage for decidable predicates. On the continuum this passage is exactly the step from "not zero" to a sign (Theorem 12.11). The pairwise decisions and the global witness are separate inputs to the reasoning.

Quantifier order is just as important. The finite-prefix result of Chapter 10 gives a program for each supplied finite table. The program may depend on that table and its budget. A uniform name asks for one program that works at every precision. Moving the existence of the program ahead of the precision request changes what must be constructed. Theorem 11.11 proves that this change has real content in the specified classical comparison.

Freshness conditions enforce arbitrariness inside the proof language. When a universal variable is introduced, the old hypotheses are lifted past it. They cannot silently fix its value. When an existential witness is used, the final conclusion must not depend on which fresh witness was chosen. The lifting marks in Appendix D are the bookkeeping that enforces these conditions.

C.5 Equivalences and countermodels answer different questions

To prove $A \iff B$, prove both directions. In the catalogue results, one direction uses a logical principle to obtain a classification. The reverse direction turns a question about that principle into a particular congruence. Applying the proposed classification to that congruence supplies the desired answer. This reverse construction is what makes the logical price exact.

Merely finding an EM-labelled proof does not establish that its conclusion requires excluded middle. Chapter 6 gives two proofs of the same small disjunction with different ledgers. A ledger describes the route actually taken. A necessity result must exclude a cheaper route or recover the claimed resource from the conclusion.

A countermodel excludes a proposed rule by satisfying the retained rules while failing the

proposed conclusion. The two-chain model of Chapter 9 supports indexed recursion on both chains, but the designated start reaches only the first. The proof then exhibits an expressible family that distinguishes this failure.

Two checks are indispensable. The interpretation must satisfy every rule of the weaker language, including conversion and substitution. And the failed statement must really be expressible in that language. An informal property outside the language cannot serve as its separating judgment. Section 9.8 supplies the soundness argument that makes the concrete models usable for this purpose.

The same two checks govern the models that separate logical principles. In Sections 7.10 and 7.12 a formula takes its value among the open sets of a space, or in a chain, rather than among true and false. Lemma 7.21 is the soundness check: every derivable sentence takes the top value. The principles and the versions of the classification are sentences of the system, which is the expressibility check. A principle whose value lies below the top is then not derivable, and two principles with different values in one model are not equivalent.

C.6 Reading the finite matrices

Chapters 18 and 19 use finite complex matrices. A column vector lists the coordinates of a state vector. A matrix lists the coefficients of a linear map: each output coordinate is obtained by multiplying a row by the input column and adding the products. The order of a matrix product matters because the right-hand map acts first.

The adjoint, written A^* or $A^\dagger$, is the conjugate transpose. Both symbols mean the same operation here. A Hermitian matrix equals its adjoint. An orthogonal projection satisfies $P^* = P$ and $P^2 = P$; applying it twice has the same effect as applying it once. A unitary matrix satisfies $U^*U = UU^* = I$, so its adjoint reverses its action.

An isometry V may be rectangular. It satisfies $V^*V = I$ on its input space. Its columns have unit norm and are mutually orthogonal. The matrix VV^* acts on the larger output space and projects onto the embedded input. This is the distinction used in the dilation proof of Theorem 18.3: the complement $I - VV^*$ is a necessary part of the full operation.

Block notation groups coordinates without removing them. In the dilation, the two block rows specify outputs in the input-sized space and the larger space. To check the square of that block matrix, multiply block rows by block columns in order. The off-diagonal blocks vanish because the complement projection kills the image of V. The diagonal blocks become identities. The proof in the chapter displays all four blocks.

A state matrix ρ is positive with trace one. Positivity says that $x^*\rho x \geq 0$ for every column x; the trace is the sum of the diagonal entries. In a finite instrument, the output is divided into blocks indexed by the retained mark. The trace of a block is that mark's probability. The block itself also contains the returned quantum state. Keeping only its trace would discard part of the operation being compared.

The norm tells us which size of error is controlled. For an operator, its norm bounds its action on unit vectors. For a Hermitian state difference, the trace norm is the sum of the absolute eigenvalues. Chapter 18 uses the full trace norm, with no factor of one half. Chapter 19

instead bounds the absolute probability error of one sharp event, uniformly over all unit source vectors. The hypotheses and constants should be read with those different comparisons in mind.

C.7 Eigenvalues, positivity and evolution

An eigenvector of a matrix is a nonzero vector whose direction is unchanged by the matrix; the eigenvalue is its scalar multiplier. The finite spectral theorem says that a Hermitian matrix has an orthonormal basis of eigenvectors and real eigenvalues. In that basis, checking positivity means checking that the diagonal eigenvalues are nonnegative. Changing back to the original coordinates keeps the same operator.

This is the theorem used when Proposition 18.2 takes the square root of a positive Choi matrix: take the nonnegative square root of each eigenvalue and change coordinates back. The proof then reshapes its columns into branch matrices. The spectral theorem establishes the existence of this factorization. It does not by itself give the effective naming procedure established later by rational search.

The same method explains the exponential of a Hermitian energy operator. In an eigenbasis, evolution multiplies each coordinate by a complex phase determined by that coordinate's energy and the elapsed time. Each phase has absolute value one, so the evolution preserves lengths. A skew-Hermitian generator is the corresponding object with the factor of i already incorporated; its real-time exponential is unitary.

Products of evolutions require care. Two energy operators need not have the same eigenvectors, and their order can matter. This is why the pulse proof does not split the background holding energy from the control by pretending they commute. It bounds the difference of the full exponentials in (18.15). The scalar baseline can be separated because an identity operator commutes with every matrix. Appendix A then explains how products and commutators make further matrix directions available in the closure of finite operation sequences. A limiting existence proof and a practical bound on the required sequence length remain different questions.

C.8 Keeping both systems in the calculation

A tensor product keeps the coordinates of two systems together. A product vector $x \otimes p$ describes the particular separated preparation used in Chapter 19. An operator $A \otimes I$ acts on the first factor, and $I \otimes R$ acts on the second. The total count is their sum. Grouping several cells into the recorder factor does not require those cells to be mutually uncorrelated.

This distinction is visible in the variance calculation. The proof uses the product between the unknown source and the whole fixed recorder. It does not split the recorder into independent cell states. The projection-sum estimate therefore remains valid when its cells and neutral memory have internal correlations.

A retained reference plays another role. It is an additional system whose correlations with the input must survive the comparison of instruments. The construction in Chapter 18 leaves that reference untouched while comparing the complete output on the joint state. The error

bound is independent of the finite reference dimension. This is stronger information than a comparison of mark probabilities alone.

Whenever a tensor product appears, name the factors before manipulating the formula. Then identify which factors each operator acts on. That small step explains the commutation of operators on different slots, the role of the neutral pointer and the exact place where an initial product preparation is used.

Appendix D

The rules at a glance

This appendix collects the rules of Chapters 5 and 6 in one place. It is a reference for the specified arithmetic language. The typed language of Chapter 9 has its own formation, conversion and elimination rules; its three orbit constructions are recalled at the end.

D.1 Reading the notation

Terms are variables, 0, successors, sums and products. Formulas are equalities, $\bot$, conjunctions, disjunctions, implications and quantified formulas. Write $\neg\varphi$ for $\varphi \to \bot$. A context Γ is a finite list of hypotheses. In a displayed rule, the judgments above the line are premises and the judgment below is the conclusion. A judgment displayed without premises is an available leaf.

The expression $\varphi(t)$ means capture-avoiding substitution at variable position zero. The lift $\varphi\uparrow$ moves the free variables past a new position; $\Gamma\uparrow$ lifts every formula of the context. These operations are defined in Definition 5.2. In ordinary named-variable notation, they enforce the familiar requirement that the newly bound variable is fresh for the old hypotheses.

D.2 The arithmetic base

Hypothesis and equality

$$\frac{\varphi \in \Gamma}{\Gamma \vdash \varphi}, \qquad \Gamma \vdash t = t, \qquad \frac{\Gamma \vdash t = s \qquad \Gamma \vdash \varphi(t)}{\Gamma \vdash \varphi(s)}.$$

Start, step and operations

$$\Gamma \vdash \neg(S(t) = 0), \qquad \frac{\Gamma \vdash S(t) = S(u)}{\Gamma \vdash t = u}.$$

$$\Gamma \vdash t + 0 = t, \qquad \Gamma \vdash t + S(u) = S(t + u),$$
$$\Gamma \vdash t \cdot 0 = 0, \qquad \Gamma \vdash t \cdot S(u) = t \cdot u + t.$$

$$\frac{\Gamma \vdash \varphi(0) \qquad \Gamma \vdash \forall x\big(\varphi(x) \to \varphi(S(x))\big)}{\Gamma \vdash \forall x\, \varphi(x)}.$$

The induction variable in the last display is fresh for Γ. Parameters other than that variable may remain free. The formula φ may contain quantifiers; that fact affects its ledger entry.

D.3 The logical rules

Implication and falsehood

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi}, \qquad \frac{\Gamma \vdash \varphi \to \psi \qquad \Gamma \vdash \varphi}{\Gamma \vdash \psi}, \qquad \frac{\Gamma \vdash \bot}{\Gamma \vdash \chi}.$$

Conjunction

$$\frac{\Gamma \vdash \varphi \qquad \Gamma \vdash \psi}{\Gamma \vdash \varphi \wedge \psi}, \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \varphi}, \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \psi}.$$

Disjunction

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \varphi \vee \psi}, \qquad \frac{\Gamma \vdash \psi}{\Gamma \vdash \varphi \vee \psi},$$

$$\frac{\Gamma \vdash \varphi \vee \psi \qquad \Gamma, \varphi \vdash \chi \qquad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi}.$$

Quantifiers

$$\frac{\Gamma\uparrow \vdash \varphi}{\Gamma \vdash \forall \varphi}, \qquad \frac{\Gamma \vdash \forall \varphi}{\Gamma \vdash \varphi(t)}, \qquad \frac{\Gamma \vdash \varphi(t)}{\Gamma \vdash \exists \varphi},$$

$$\frac{\Gamma \vdash \exists \varphi \qquad \Gamma\uparrow, \varphi \vdash \psi\uparrow}{\Gamma \vdash \psi}.$$

The final conclusion of existential elimination is lifted in the second premise. It therefore cannot depend on the fresh witness position.

D.4 Additional schemas and their records

The three optional schemas are

$$\begin{aligned}
&\text{EM:} && \Gamma \vdash \varphi \vee \neg\varphi,\\
&\text{LPO:} && \Gamma \vdash \big[\forall x(\varphi(x) \vee \neg\varphi(x))\big]\\
&&& \qquad \to \big[(\exists x\, \varphi(x)) \vee (\forall x \neg\varphi(x))\big],\\
&\text{MP:} && \Gamma \vdash \neg\neg\exists x\, \varphi(x) \to \exists x\, \varphi(x).
\end{aligned}$$

The MP schema is restricted to quantifier-free φ. Its interpretation uses the decidability established in Lemma 6.2; it is not an unrestricted double-negation elimination rule. So

restricted, and with induction on every formula, it is derivable from the base rules: a quantifier-free formula is eventually constant in its searched variable (Lemma 6.8), and the existential is a finite search (Theorem 6.9). The LPO schema applies to every formula, and with induction on every formula it proves every instance of the EM schema (Theorem 6.10). At that level the three schemas therefore mark two strengths: MP with the base rules, and LPO with EM (Theorem 6.11). Without induction, or with induction on quantifier-free formulas only, they mark four (Propositions 6.13 and 6.14).

For a derivation, the ledger $(e, \ell, m; i)$ records, in order, occurrence of EM, LPO, MP and induction on a formula containing a quantifier. Combine premises by inclusive-or in each coordinate, then add the contribution of the current rule. Thus the ledger records occurrence, not multiplicity. An empty ledger has four false coordinates. A posit-free derivation has its first three coordinates false and may still have the fourth coordinate true.

These schemas are formulas of the arithmetic language. Principles about arbitrary propositions and predicates in the surrounding mathematics are stated separately in Definition 6.1. The soundness argument explains how the latter validate the former. A reader should not replace one level by the other merely because both use the same abbreviation. The two levels also differ in strength. In the metatheory the three principles keep their separate strengths; as schemas of this language, with induction on every formula, MP is derivable and LPO is equivalent to EM. Interpreting any derivation needs at most the metatheoretic LPO, applied to the predicates that formulas define, and interpreting all derivations with an EM entry, or all with an LPO entry, needs exactly that (Corollary 6.15).

Peano arithmetic, as presented in Chapter 8, uses the same formulas. It states successor injectivity as the axiom $S(t) = S(u) \to t = u$, makes excluded middle an ordinary rule and keeps no ledger. By Theorem 8.11 it proves exactly what these rules prove with EM allowed and LPO and MP excluded, from every finite context. Deleting excluded middle from it gives Heyting arithmetic, which proves exactly what the posit-free derivations prove (Corollary 8.12).

D.5 The three orbit constructions

Chapter 9 specifies a typed language with orbit type O, an empty type, dependent functions and identity types. An iterator has a fixed target type B, a starting value $b : B$ and an update $g : B \to B$:

$$I(B, b, g, 0) \equiv b, \qquad I(B, b, g, S(x)) \equiv g(I(B, b, g, x)).$$

An indexed recursor still has fixed target B, but its update receives the preceding orbit point as well as the preceding value:

$$R(B, b, g, 0) \equiv b, \qquad R(B, b, g, S(x)) \equiv g(x, R(B, b, g, x)).$$

Here $g : \Pi x : O(B \to B)$. A dependent eliminator has a family $P(x)$, a start $b : P(0)$ and a step $g : \Pi x : O(P(x) \to P(S(x)))$:

$$E(P, b, g, 0) \equiv b, \qquad E(P, b, g, S(x)) \equiv g(x, E(P, b, g, x)).$$

Its result at x has type $P(x)$.

The symbol $\equiv$ here denotes the conversion relation of that language. It includes the specified computation equations and their closure under constructors and capture-avoiding substitution. The language does not silently add equality reflection, function extensionality or a uniqueness rule for recursive definitions.

The optional discriminator has computation equations

$$D(0) \equiv \mathrm{Id}_O(0,0), \qquad D(S(x)) \equiv \mathbf{0}.$$

It is part of the language used for the relevant separation result. Full formation and identity-elimination rules remain in Chapter 9. These displays are a reminder of their inputs and computation equations, not permission to ignore the typing conditions in that chapter.

Appendix E

The prices of the three rules: supporting proofs

This appendix supplies the routine verifications behind four proofs of Chapter 6: the derivation of every restricted Markov instance (Theorem 6.9), the realizability argument (Theorem 6.11), and the two models with two stages (Propositions 6.13 and 6.14). The central steps are proved in the chapter; here we check the cases.

E.1 Arithmetic inside the rules

The derivations in this section are from no hypotheses, use no EM, LPO or MP rule, and leave their variables free; a derivation of an implication serves as a derived rule. Equality is symmetric and transitive, and is respected by S, $+$ and $\cdot$, by one use of equality substitution each together with reflexivity, as in the proof of Proposition 5.6. We use the abbreviations $u \le w$, $u < w$ and $\leftrightarrow$ of the proof of Theorem 6.9. The equivalence $\leftrightarrow$ is reflexive, symmetric and transitive, and is respected by $\wedge$, $\vee$ and $\to$; these are derivations by the propositional rules alone.

Lemma E.1 (Arithmetic facts)**.** *The following have derivations.*

1. *The laws* $0 + x = x$, $S(x) + y = S(x + y)$, $x + y = y + x$, $(x + y) + z = x + (y + z)$, $0 \cdot x = 0$, $S(x) \cdot y = x \cdot y + y$, $x \cdot y = y \cdot x$, $x \cdot (y + z) = x \cdot y + x \cdot z$, $(x + y) \cdot z = x \cdot z + y \cdot z$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$, $S(0) \cdot x = x$, *and cancellation* $a + c = b + c \to a = b$. *Each uses induction on a quantifier-free formula only.*
2. *For every quantifier-free* ψ, $\psi \vee \neg\psi$.
3. *Every number is* 0 *or a successor:* $x = 0 \vee \exists k\,(x = S(k))$.
4. *The order facts* $\neg(y < 0)$; $a + c = b \to (a < b \vee a = b)$; $y < S(n) \to (y < n \vee y = n)$; $c \le c + u$; $y \le b \to y \le a + b$; $0 \le u$; *and* $y \le a \to S(a) \le X \to y < X$.
5. *Comparison:* $x < y \vee y \le x$.
6. *Uniqueness of remainders:* $q \cdot d + r = p \cdot d + s \to r < d \to s < d \to r = s$.
7. *Cancellation of a nonzero factor:* $A \cdot S(m) = B \cdot S(m) \to A = B$.

Items 2, 3, 5, 6 and 7 use induction on formulas with quantifiers, and item 4 uses item 3.

Proof. Item 1. Follow the proofs of Lemma 3.2, Theorem 3.3, Proposition 3.4, Lemma 3.6, Proposition 3.7, Theorem 3.8 and Proposition 3.9, as Corollary 5.9 does for commutativity. Each induction there is on an equation, or for cancellation on the implication $a + c = b + c \to a = b$ in c, whose step uses injectivity; the other variables are parameters. The right distributive law follows from the left one and commutativity, and $S(0) \cdot x = 0 \cdot x + x = x$.

Item 2. Equations are decided by (6.9), which is derived in the proof of Theorem 6.10 by an induction on the formula $\forall y(x = y \vee \neg(x = y))$. The other cases of that proof, for $\bot$, $\wedge$, $\vee$ and $\to$, use no additional rule.

Item 3. Induct on x with the formula itself. At 0 take the left side; at $S(x)$ take the right side with witness x.

Item 4. From $y + S(j) = 0$ the recursive equation gives $S(y + j) = 0$, against separation. For the second fact split c by item 3: if $c = 0$ then $a = a + 0 = b$; if $c = S(k)$ then $a + S(k) = b$ is $a < b$. For the third, $y + S(j) = S(n)$ gives $S(y + j) = S(n)$, injectivity gives $y + j = n$, and the second fact applies. The fourth has witness u; for the fifth, $y + j = b$ gives $y + (a + j) = a + b$ by the laws of item 1; the sixth has witness u by $0 + u = u$. For the last, $y + j = a$ and $S(a) + j' = X$ give $y + S(j + j') = S(a + j') = S(a) + j' = X$.

Item 5. Induct on x, with y a parameter, on the formula $x < y \vee y \leq x$. At 0 split y by item 3: $y = 0$ gives $y \leq 0$ with witness 0, and $y = S(k)$ gives $0 + S(k) = y$. For the step, if $x + S(k) = y$ then $S(x) + k = y$, and the second order fact gives $S(x) < y$ or $S(x) = y$, the latter giving $y \leq S(x)$ with witness 0; if $y + k = x$ then $y + S(k) = S(x)$.

Item 6. Induct on q with the formula $\forall p\,(q \cdot d + r = p \cdot d + s \to r < d \to s < d \to r = s)$, keeping d, r, s as parameters, and split p by item 3. If $q = p = 0$ the equation is $r = s$. If $q = 0$ and $p = S(p')$, then $r = (p' \cdot d + s) + d$, and $r + S(k) = d$ gives $S((p' \cdot d + s) + k) + d = 0 + d$, so cancellation yields a successor equal to 0, which is impossible; the case $q = S(q')$, $p = 0$ is symmetric, using $s < d$. If $q = S(q')$ and $p = S(p')$, the equation reads $(q' \cdot d + r) + d = (p' \cdot d + s) + d$; cancel d and apply the induction hypothesis at p'.

Item 7. Induct on A with the formula $\forall B(A \cdot S(m) = B \cdot S(m) \to A = B)$, and split B by item 3. If $A = B = 0$ there is nothing to prove. If exactly one of them is 0, the other side is $k \cdot S(m) + S(m) = S(k \cdot S(m) + m)$, a successor equal to 0, which is impossible. If $A = S(a)$ and $B = S(k)$, the equation is $a \cdot S(m) + S(m) = k \cdot S(m) + S(m)$; cancellation gives $a \cdot S(m) = k \cdot S(m)$, the induction hypothesis at k gives $a = k$, and so $S(a) = S(k)$. □

Lemma E.2 (Polynomial forms)**.** *For every term s and every term X, there is a derivation of $s(X) = h_X(\vec{c})$, where $\vec{c}$ is the coefficient list of s defined in the proof of Theorem 6.9. It uses only the laws of Lemma E.1(1).*

Proof. Recursion on s in the metatheory. For x, $(0 \cdot X + S(0)) \cdot X + 0$ equals X because $0 \cdot X + S(0) = S(0)$ and $S(0) \cdot X = X$. A parameter p equals $0 \cdot X + p$, and 0 is $h_X()$. For $S(s)$, the form of s is 0 or $h \cdot X + c$, and $S(0) = 0 \cdot X + S(0)$ and $S(h \cdot X + c) = h \cdot X + S(c)$. For sums, recursion on the two lists uses $(a \cdot X + c) + (b \cdot X + d) = (a + b) \cdot X + (c + d)$ and $0 + h = h = h + 0$. For products, a separate recursion on the list $\vec{d}$ gives

$$c \cdot h_X(\vec{d}) = h_X(c \cdot \vec{d}),$$

from $c \cdot (h \cdot X + d) = (c \cdot h) \cdot X + c \cdot d$. Then recursion on the first list uses $0 \cdot h = 0$ and

$$(h \cdot X + c) \cdot g = (h \cdot g) \cdot X + c \cdot g = h_X(0, \vec{e}) + h_X(c \cdot \vec{d}),$$

where $g = h_X(\vec{d})$, $h \cdot g = h_X(\vec{e})$ by the recursion, and $h_X(0, \vec{e}) = h_X(\vec{e}) \cdot X + 0$; the sum case then combines the two forms. Each displayed identity follows from the laws of Lemma E.1(1), and the derivations are combined by the congruence of equality. □

E.2 Structures with two stages

Two-stage structures, the passage h, the environments $h\rho$ and the forcing relation $\rho \Vdash \varphi$ are defined in the proof of Proposition 6.13. Throughout this section the metatheory is classical, and satisfaction at the later stage, $M_1, \sigma \vDash \varphi$, is that of Definition 5.7. A context is forced at ρ when each of its formulas is.

Lemma E.3 (Persistence and binding)**.** *Let (M_0, M_1, h) be a two-stage structure, ρ an environment in M_0, t a term, φ a formula and a an element of M_0.*

1. *The value of t in M_1 under $h\rho$ is h of its value in M_0 under ρ, and $h[a, \rho] = [h(a), h\rho]$.*
2. *If $\rho \Vdash \varphi$, then $M_1, h\rho \vDash \varphi$.*
3. *$[a, \rho] \Vdash \varphi\uparrow$ if and only if $\rho \Vdash \varphi$; and $\rho \Vdash \varphi(t)$ if and only if $[E_\rho(t), \rho] \Vdash \varphi$, where $E_\rho(t)$ is the value of t in M_0. Replacing the variable x_0 of φ by $S(x_0)$ replaces its assigned element a by $S(a)$: $[a, \rho] \Vdash \varphi^S$ if and only if $[S(a), \rho] \Vdash \varphi$.*

Proof. For the first assertion induct on t. A variable takes the value $h(\rho(k))$ on both sides by the definition of $h\rho$; zero goes to zero; and the successor, sum and product cases are the three equations that h preserves. The identity of environments holds position by position.

For persistence induct on φ, with ρ arbitrary. For an equation apply h to both values and use the first assertion. Falsehood is never forced. Conjunction and disjunction follow from the induction hypotheses for their components. For an implication or a universal formula the later clause of the forcing definition is exactly the required statement (for $\forall\varphi$, the clause at every b is satisfaction of $\forall\varphi$ in M_1 under $h\rho$). For $\exists\varphi$, a witness a with $[a, \rho] \Vdash \varphi$ gives $M_1, h[a, \rho] \vDash \varphi$ by the induction hypothesis, that is $M_1, [h(a), h\rho] \vDash \varphi$, so $h(a)$ is a witness in M_1.

For the third assertion, Lemma 5.4 holds in every structure, with the same proof (as noted in the proof of Theorem 5.8), and it holds for lifting and substitution at every cutoff and position, since its proof inducts with those parameters arbitrary. Repeat that proof for forcing. Equations use the term identities of that lemma in M_0. The connectives other than implication follow from the induction hypotheses. In the clauses for implication and the universal quantifier, the later conditions are statements of satisfaction in M_1 under $h\rho$ or $[b, h\rho]$; by the first assertion the modified environments correspond under h, because $h[a, \rho] = [h(a), h\rho]$ and h carries the value of a substituted term in M_0 to its value in M_1; so the lemma in M_1 transforms them as required. The statement about φ^S is the substitution identity at the term $S(x_0)$, whose value at $[a, \rho]$ is $S(a)$, as in the final assertion of Lemma 5.4. □

The later stage is an ordinary structure read classically. Its soundness needs no new idea.

Lemma E.4 (Soundness at the later stage)**.** *Let M be a structure satisfying the six arithmetic rules, and read satisfaction in M classically. Every derivation with no induction rule is sound in M, whatever EM, LPO and MP rules it contains: if σ satisfies its context, σ satisfies its conclusion. If, moreover, M satisfies every instance*

$$\big(\varphi(0) \wedge \forall x\,(\varphi(x) \to \varphi(S(x)))\big) \to \forall x\,\varphi(x)$$

with φ quantifier-free, under every environment, then the same holds for every derivation whose induction formulas are all quantifier-free.

Proof. The base rules are covered by the first case of Theorem 5.8, and the induction rule on a quantifier-free formula by the assumed instances. The three additional rules are true in M because the metatheory is classical. An EM instance $\varphi \vee \neg\varphi$ holds since $\sigma \vDash \varphi$ either holds or fails. For an LPO instance, either some b has $[b, \sigma] \vDash \varphi$, which gives $\exists x\,\varphi$, or none has, which gives $\forall x\,\neg\varphi$. For an MP instance, if $\exists x\,\varphi$ were false then $\neg\exists x\,\varphi$ would hold and contradict $\neg\neg\exists x\,\varphi$; so $\exists x\,\varphi$ holds. □

Theorem E.5 (Soundness at the earlier stage)**.** *Let (M_0, M_1, h) be a two-stage structure, and let d be a derivation of $\Gamma \vdash \varphi$ whose EM and MP entries are false, so that its only additional rules are LPO rules. Assume one of the following:*

(a) d contains no induction rule; or

(b) every induction formula of d is quantifier-free, M_1 satisfies the quantifier-free induction instances of Lemma E.4, and the earlier stage forces quantifier-free induction*: for every quantifier-free ψ and every environment ρ in M_0, if $\rho \Vdash \psi(0)$ and $\rho \Vdash \forall x(\psi(x) \to \psi(S(x)))$, then $\rho \Vdash \forall x\,\psi(x)$.*

Then every environment ρ in M_0 that forces Γ forces φ.

Proof. Induct on d, keeping the context, the conclusion and ρ arbitrary. Lemma E.3 is used throughout; its third assertion is called the binding lemma below. The subderivations of d inherit the hypotheses (a) or (b), so Lemma E.4 applies to each of them in M_1.

Hypotheses, reflexivity and equality substitution. A hypothesis is forced because the context is. Reflexivity compares a value with itself. For substitution, the premises give equal values $E_\rho(t) = E_\rho(s)$ in M_0 and $\rho \Vdash \varphi(t)$, that is $[E_\rho(t), \rho] \Vdash \varphi$ by the binding lemma; replacing the value by the equal one and applying the binding lemma again gives $\rho \Vdash \varphi(s)$.

The arithmetic rules. The rule $\neg(S(t) = 0)$ is an implication. Its earlier clause holds because no successor is zero in M_0, and its later clause because no successor is zero in M_1. Injectivity and the four recursive equations are statements about values in M_0, where they hold.

Implication. For introduction, let d' derive $\Gamma, \varphi \vdash \psi$. If $\rho \Vdash \varphi$, then ρ forces the context Γ, φ, and the induction hypothesis gives $\rho \Vdash \psi$. For the later clause, persistence gives $M_1, h\rho \vDash \Gamma$; if also $M_1, h\rho \vDash \varphi$, then Lemma E.4 applied to d' gives $M_1, h\rho \vDash \psi$. For elimination, apply the earlier clause of the forced implication to the forced antecedent.

Conjunction, disjunction and falsehood. Introductions and the eliminations of conjunction are immediate. For disjunction elimination, $\rho \Vdash \varphi \vee \psi$ means that one side is forced; add it to the forced context and apply the induction hypothesis for that branch. Falsehood is never forced, so the rule for an impossible case is vacuous.

Universal quantifier. For introduction, let d' derive $\Gamma\uparrow \vdash \varphi$. For an element a of M_0, the binding lemma shows that $[a, \rho]$ forces $\Gamma\uparrow$, and the induction hypothesis gives $[a, \rho] \Vdash \varphi$. For an element b of M_1, persistence gives $M_1, h\rho \vDash \Gamma$, hence $M_1, [b, h\rho] \vDash \Gamma\uparrow$ by (5.3) in M_1, and Lemma E.4 applied to d' gives $M_1, [b, h\rho] \vDash \varphi$. For elimination at a term t, use the earlier clause at $a = E_\rho(t)$ and the binding lemma.

Existential quantifier. For introduction at t, the element $E_\rho(t)$ is a witness by the binding lemma. For elimination, the first premise gives an element a with $[a,\rho] \Vdash \varphi$. The environment $[a,\rho]$ forces $\Gamma\uparrow$ by the binding lemma, so it forces the context $\Gamma\uparrow, \varphi$ of the second premise, whose induction hypothesis gives $[a,\rho] \Vdash \psi\uparrow$. The binding lemma turns this into $\rho \Vdash \psi$.

Induction. Under (a) there is no such rule. Under (b) the induction formula ψ is quantifier-free, the induction hypotheses give $\rho \Vdash \psi(0)$ and $\rho \Vdash \forall x(\psi(x) \to \psi(S(x)))$, and the assumption on the earlier stage gives $\rho \Vdash \forall x \psi$.

The LPO rule. Its conclusion is the implication $\forall x(\varphi \vee \neg\varphi) \to (\exists x \varphi \vee \forall x \neg\varphi)$. Its later clause holds by Lemma E.4. For the earlier clause, assume $\rho \Vdash \forall x(\varphi \vee \neg\varphi)$; we use only its earlier part, that $[a,\rho] \Vdash \varphi$ or $[a,\rho] \Vdash \neg\varphi$ for every element a of M_0. Either some a has $[a,\rho] \Vdash \varphi$, and then $\rho \Vdash \exists x \varphi$, the left alternative. Or none has; then every a has $[a,\rho] \Vdash \neg\varphi$, whose later clause says $M_1, h[a,\rho] \nVdash \varphi$, that is $M_1, [h(a), h\rho] \vDash \neg\varphi$. The earlier part of $\rho \Vdash \forall x \neg\varphi$ is the statement that every a forces $\neg\varphi$. For the later part let b be an element of M_1. Since h is onto, $b = h(a)$ for some a, and then $M_1, [b, h\rho] \vDash \neg\varphi$. So $\rho \Vdash \forall x \neg\varphi$, the right alternative.

There are no EM or MP rules. Every rule has been covered. □

A derivation from no hypotheses is therefore forced at every environment, because the empty context is forced. Contrapositively, a formula that is not forced at some environment has no derivation from no hypotheses that meets the conditions of the theorem.

E.3 Realizers

Programs, the notation $\{r\}(a)\downarrow v$, realizability, the environments ρ_l and the list codes $\ulcorner l \urcorner$ are those of the proof of Theorem 6.11. The metatheory is classical. We say that a program *computes* a function when it returns the function's value on every input; the functions computed in this section are assembled from the constructors of Definition 10.5, the universal program and the fixing of part of an input, as described in that proof.

Lemma E.6 (Realizability respects binding)**.** *For every formula φ, term t, environment ρ, and numbers r and n: r realizes $\varphi\uparrow$ at $[n,\rho]$ exactly when it realizes φ at ρ; and r realizes $\varphi(t)$ at ρ exactly when it realizes φ at $[E_\rho(t),\rho]$. Replacing the induction variable by its successor replaces its value n by $n+1$. The realizers of a formula at ρ depend only on the values ρ gives to the free variables of the formula.*

Proof. The clauses of realizability use the environment only through the values of terms and the extension $[n,\rho]$, exactly as the clauses of satisfaction do. The proof of Lemma 5.4 therefore applies word for word, with "r realizes" in place of "satisfies", the induction keeping r arbitrary. The last assertion is proved by the same induction. □

Lemma E.7 (Quantifier-free formulas)**.** *Let φ be quantifier-free. If some number realizes φ at ρ, then $\rho \vDash \varphi$. There is a program q_φ that returns a value on every input and, for every list l with $\rho_l \vDash \varphi$, returns on $\ulcorner l \urcorner$ a number that realizes φ at ρ_l. Whether $\rho_l \vDash \varphi$ is computed from $\ulcorner l \urcorner$.*

Proof. Truth is computed by evaluating the terms of φ under ρ_l and following Lemma 6.2. The other two claims are proved together by induction on φ. For an equation, realizability is truth

and q_φ returns 0. Nothing realizes $\bot$, which is false, and $q_\bot$ returns 0. For $\varphi \wedge \psi$, a realizer is a pair of realizers, so both components are true, and q runs q_φ and q_ψ on $\ulcorner l \urcorner$ and pairs their outputs. For $\varphi \vee \psi$, a realizer realizes the side its tag names, which is therefore true; q decides φ at ρ_l and returns $\langle 0, v\rangle$ if it holds, with v the output of q_φ on $\ulcorner l \urcorner$, and $\langle 1, w\rangle$ otherwise, with w the output of q_ψ there. For $\varphi \to \psi$, let r realize the implication at ρ and suppose $\rho \vDash \varphi$. Let l list $\rho(0), \dots, \rho(m-1)$, where m exceeds the index of every variable of $\varphi \to \psi$. Then ρ_l agrees with ρ on the variables of φ and ψ, so q_φ supplies a realizer of φ at ρ_l, which realizes φ at ρ (Lemma E.6); $\{r\}$ turns it into a realizer of ψ, so ψ is true. For the program, $q_{\varphi\to\psi}$ returns the code of the constant program whose value is the output of q_ψ on $\ulcorner l \urcorner$. If $\rho_l \vDash \varphi \to \psi$ and a realizes φ, then φ is true, hence ψ is, and the constant value realizes ψ. □

Lemma E.8 (Formulas without implication)**.** *Let ψ be built from equations and $\bot$ by $\wedge$, $\vee$, $\forall$ and $\exists$. If some number realizes ψ at ρ, then $\rho \vDash \psi$. If $\rho \nvDash \psi$, every number realizes $\neg\psi$ at ρ.*

Proof. Induct on ψ with ρ arbitrary. A realized equation is true, and $\bot$ is never realized. The components of a realized conjunction, the named side of a realized disjunction, the witness of a realized existential statement and every instance of a realized universal statement are realized, hence true by the induction hypothesis, and so the whole formula is true. If ψ is false, nothing realizes it, and every number realizes $\psi \to \bot$ vacuously. □

Theorem E.9 (Derivations from MP rules are realized)**.** *For every derivation d of $\gamma_1, \dots, \gamma_k \vdash \varphi$ whose additional rules are all MP rules, with any induction rules, there is a program c_d such that for every list l and all numbers $a_1, \dots, a_k$ with a_i realizing γ_i at ρ_l, c_d returns on $\langle \ulcorner l \urcorner, \ulcorner (a_1, \dots, a_k) \urcorner\rangle$ a number that realizes φ at ρ_l.*

Proof. Induct on d. Write $\vec{a}$ for $(a_1, \dots, a_k)$ and m for the input $\langle \ulcorner l \urcorner, \ulcorner \vec{a} \urcorner\rangle$, and assume throughout that each a_i realizes γ_i at ρ_l. When a premise has a larger context, its program receives the extended list of realizers.

Hypothesis γ_i: return a_i. *Reflexivity, the four recursive equations and $\neg(S(t) = 0)$:* return 0. The equations are true, and $S(t) = 0$ has no realizer, so every number realizes its negation. *Injectivity:* the premise $S(t) = S(u)$ is realized, hence true, so $t = u$ is true and 0 realizes it. *Equality substitution:* the first premise is realized, so t and s have equal values E; a realizer of $\varphi(t)$ realizes φ at $[E, \rho_l]$, hence $\varphi(s)$ (Lemma E.6); return the output of the second premise's program.

Implication. For introduction from $\varphi, \gamma_1, \dots, \gamma_k \vdash \psi$ with program c', return the code of the program that, on b, runs c' on $\langle \ulcorner l \urcorner, \ulcorner (b, \vec{a}) \urcorner\rangle$; it is obtained from m by fixing part of an input. On every b realizing φ it returns a realizer of ψ. For elimination, run the two premise programs on m, obtaining r and a, and return the value of U on $\langle r, a\rangle$, which realizes the conclusion because r realizes the implication and a its antecedent.

Conjunction and disjunction. Pair the outputs of the two premises, or take the first or second component. For the introductions of a disjunction return $\langle 0, a\rangle$ or $\langle 1, a\rangle$. For disjunction elimination, run the program of the disjunction to obtain $\langle i, a\rangle$; if $i = 0$ run the program of the left branch on $\langle \ulcorner l \urcorner, \ulcorner (a, \vec{a}) \urcorner\rangle$, and otherwise the program of the right branch on the same input.

The rule for an impossible case. Return the output of the premise's program. That output would realize $\bot$, which is impossible, so no list l and realizers $\vec{a}$ meet the assumption and there is nothing to check.

Universal quantifier. Introduction is treated in the proof of Theorem 6.11. For elimination at a term t, run the premise's program to obtain r, compute the value n of t under ρ_l, and return the value of U on $\langle r, n\rangle$. It realizes the body at $[n, \rho_l]$, which is $\varphi(t)$ at ρ_l.

Existential quantifier. For introduction at t, return the pair of the value of t and the premise's output. For elimination, the first premise's program returns $\langle n, b\rangle$ with b realizing the body at $[n, \rho_l] = \rho_{(n,l)}$. The a_i realize $\gamma_i{\uparrow}$ at $\rho_{(n,l)}$. Run the second premise's program on $\langle \ulcorner (n, l)\urcorner, \ulcorner (b, \vec{a})\urcorner\rangle$; its output realizes $\psi{\uparrow}$ at $[n, \rho_l]$, hence ψ at ρ_l.

Induction and the MP rule are treated in the proof of Theorem 6.11. There are no EM or LPO rules.

In each case the program is assembled from the programs of the premises by the constructors of Definition 10.5, the universal program and the fixing of part of an input, so it is a program. □

E.4 The halting matrix

We read programs within an allowance, as the rows of Table 20.1 do. Say that program c returns v on input n *within allowance* K, written $c(n){\downarrow_K}v$, according to the following recursion on K and, for a fixed K, on the code c. Nothing is returned within allowance 0. With allowance $k+1$ nothing is returned on an input $n > k$; for $n \le k$:

- the zero, successor, left and right programs return 0, $n+1$ and the two components of n;
- pair(f, g) returns $\langle y, z\rangle$ when $f(n){\downarrow_{k+1}}y$ and $g(n){\downarrow_{k+1}}z$;
- $f \circ g$ returns v when $g(n){\downarrow_{k+1}}y$ and $f(y){\downarrow_{k+1}}v$;
- for $n = \langle a, w\rangle$, rec(f, g) returns v when $w = 0$ and $f(a){\downarrow_{k+1}}v$, or when $w = w' + 1$, rec$(f, g)(\langle a, w'\rangle){\downarrow_k}i$ and $g(\langle a, \langle w', i\rangle\rangle){\downarrow_{k+1}}v$;
- for $n = \langle a, w\rangle$, search(f) returns v when $f(n){\downarrow_{k+1}}x$ and either $x = 0$ and $v = w$, or $x > 0$ and search$(f)(\langle a, w+1\rangle){\downarrow_k}v$.

The parts f, g of a program have smaller codes than the program, by (10.4), and the repeated calls of rec and search have the smaller allowance k; so the recursion is well founded. It computes, from K, c and n, whether anything is returned and what: every call is to a smaller allowance or a smaller code, and the only search is the one built into search(f), here cut off by the allowance.

Lemma E.10 (Allowances)**.** *Program c returns v on input n exactly when $c(n){\downarrow_K}v$ for some allowance K; and if $c(n){\downarrow_K}v$, then $c(n){\downarrow_{K'}}v$ for every $K' \ge K$.*

Proof. For the second assertion induct on K, then on c: each clause at K is a clause at $K+1$, with the premises at allowance K or $K-1$ raised by the induction hypothesis. If $c(n){\downarrow_K}v$, an induction on K and c shows that c returns v on n: each clause is the matching evaluation clause

of Definition 10.5. Conversely, induct on the finite evaluation of c on n. Each evaluation clause has finitely many sub-evaluations, which return within allowances given by the induction hypothesis. Take K larger than all of them, larger than one plus the allowance of the repeated call of a recursion or search, and larger than n; by the second assertion every sub-evaluation returns within the required allowance K or $K-1$, and the clause gives $c(n)\downarrow_K v$. □

A table is read through Gödel's remainder function β of Section 20.3.3: the *cell* (k,c,n) of the table named by A,B is the number w with $\beta(A,B,\langle k,\langle c,n\rangle\rangle,w)$, which exists and is unique by Lemma K.1. A cell holding 0 makes no claim; a cell holding $v+1$ claims $c(n)\downarrow_k v$. The cell (K,c,n), holding $v+1$, is *justified* when $K=k+1$, $n\le k$, and the claim follows from one clause of the recursion above whose premises are read from cells of the same table, with the inputs of those cells bounded explicitly:

- c is 0, 1, 2 or 3 and v is the value that clause prescribes;
- $c=\mathrm{pair}(f,g)$, the cells (K,f,n) and (K,g,n) hold $y+1$ and $z+1$, and $v=\langle y,z\rangle$;
- $c=f\circ g$, the cell (K,g,n) holds $y+1$ with $y<K$, and the cell (K,f,y) holds $v+1$;
- $c=\mathrm{rec}(f,g)$ and $n=\langle a,w\rangle$; either $w=0$, $a<K$ and the cell (K,f,a) holds $v+1$, or $w=w'+1$, $\langle a,w'\rangle<k$, the cell $(k,c,\langle a,w'\rangle)$ holds $i+1$, $\langle a,\langle w',i\rangle\rangle<K$ and the cell $(K,g,\langle a,\langle w',i\rangle\rangle)$ holds $v+1$;
- $c=\mathrm{search}(f)$, $n=\langle a,w\rangle$, the cell (K,f,n) holds $x+1$, and either $x=0$ and $v=w$, or $x>0$, $\langle a,w+1\rangle<k$ and the cell $(k,c,\langle a,w+1\rangle)$ holds $v+1$.

Let $\theta(e,x)$ say: there are A,B such that every cell (k,c,n) with $k\le x$, $c\le e$ and $n\le x$ holds 0 or is justified, and the cell $(x,e,0)$ holds $v+1$ for some v. Written in the arithmetic language, $u\le w$ is $\exists j(u+j=w)$ and $u<w$ is written $S(u)\le w$, which is equivalent to the form used in Section E.1; pairs, including the code of a compound program, are read through the pairing graph G; the cell with index i is the w with $\exists q\,(A=q\cdot m+w\wedge w<m)$ for $m=B\cdot S(i)+S(0)$, which is $\beta(A,B,i,w)$ in another notation; and each bounded universal statement "for all $k\le x$" is $\forall k\,(S(x)\le k\vee\cdots)$. The alternatives above are joined by $\vee$, and the remaining letters are existentially quantified.

Lemma E.11 (The halting matrix)**.** *The formula $\theta(e,x)$ has the properties (i)–(iii) stated in the proof of Theorem 6.11.*

Proof. Property (i) holds by construction: the formula uses equations, $\wedge$, $\vee$, $\exists$ and $\forall$, and no implication.

For (ii), suppose first that A,B witness $\theta(e,x)$. We show that every claim of a cell (k,c,n) in the rectangle $k\le x$, $c\le e$, $n\le x$ is correct, by induction on k and then on c. The cells a justification reads lie in the rectangle: their allowances are K or k, their codes are c or the smaller codes of its parts, and their inputs are at most $n\le k$ or bounded explicitly by K or k. Each is read at a smaller allowance, or at the same allowance and a smaller code. By the induction hypothesis their claims are correct, and the clause then gives the claim of (k,c,n). Applied to the cell $(x,e,0)$, this gives $e(0)\downarrow_x v$. Conversely, suppose $e(0)\downarrow_x v$. Fill each cell of the rectangle with

$u+1$ if $c(n)\downarrow_k u$ and with 0 otherwise. Finitely many indices are involved, and a finite list of numbers is named by some A, B through β: by Lemma K.3, read in $\mathbb{N}$ by Theorem 5.5, one entry can be appended at a time. Each nonzero cell is justified by the clause that produced its value; the explicit bounds hold because a call within allowance K requires its input to be below K. The cell $(x, e, 0)$ holds $v+1$. So $\theta(e, x)$ is true exactly when program e returns a value on input 0 within allowance x, and the rest of (ii) is Lemma E.10.

For (iii), whether $e(0)\downarrow_x v$ for some v is computed by the recursion above, so by (ii) truth of $\theta(e, x)$ is decided. When it is true, search the pairs $\langle A, B\rangle = 0, 1, 2, \ldots$ for the first one whose table is justified on the rectangle and claims a value at $(x, e, 0)$. That check is a finite computation, since every cell is a remainder and every quantifier in it is bounded, and by (ii) the search stops. From A, B, e and x a realizer of $\theta(e, x)$ is then assembled clause by clause. Each existential quantifier in θ below A, B has a witness computed from the values at hand: a quotient and a remainder in β, a difference in $\leq$ or $<$, a component of a pair in G, the predecessor k of K, the code of a part, or the content of a cell. An equation is realized by 0, a conjunction by the pair of realizers of its parts, and a disjunction by the tag of a true alternative, found by checking the alternatives in turn, with its realizer. A bounded universal statement $\forall k\,(S(x) \leq k \vee \chi)$ is realized by the program that, on k, returns $\langle 0, \langle k - x - 1, 0\rangle\rangle$ when $S(x) \leq k$ and otherwise $\langle 1, \cdot\rangle$ with the realizer of χ at k, which is true there. Each of these steps is computed from the values of the free variables, so the whole realizer is. $\square$

Appendix F

Proofs for the continuum chapter

This appendix carries the routine verifications behind Theorems 12.22, 12.38 and 12.63. Notation is that of the chapter and of the proofs that refer here.

F.1 Realizers for the rules

We complete the proof that every theorem of the theory of Theorem 12.22 has a modified realizer, and every theorem of that theory with $\mathrm{MP}_{\mathrm{bin}}$ added a Kleene realizer. For a derivation of A from hypotheses $B_1, \dots, B_m$ we give one program c. Its input lists the values of the free number variables, the total programs assigned to the free sequence variables, and numbers $s_1, \dots, s_m$, one for each hypothesis. For the modified reading we prove two properties: if $s_j \in \mathrm{HRO}_{\tau(B_j)}$ for every j, then c halts with a value in $\mathrm{HRO}_{\tau(A)}$ (the *typing property*); if moreover $s_j \Vdash_{\mathrm{m}} B_j$ for every j, the value realizes A (the *realizing property*). For Kleene's reading we prove only that c halts with a realizer of A when every s_j realizes B_j. The program is built by induction on the derivation; in each case below c_1, c_2, c_3 are the programs of the premises, and "run c_1" means run it on the same input. A premise may have free variables that do not occur in the conclusion, as in implication elimination; its program is run with those number variables assigned 0 and those sequence variables assigned the program 0, which is total. Any assignment would do, because the premise is derived for every valuation. Two facts are used throughout. First, the realizers of a formula depend only on the values that the valuation gives to its free variables, so substituting a term for a variable amounts to assigning the term's value. Second, $\tau(A)$ does not change under substitution or renaming of variables.

Hypothesis B_j. Return s_j. Both properties are immediate.

An axiom. Each choice schema is realized, under every valuation, by the fixed program E of the proof of Theorem 12.22, which lies in the type of the schema; return E. In Kleene's reading the same holds for the program M realizing $\mathrm{MP}_{\mathrm{bin}}$.

From $\bot$ infer A. Return 0, which lies in $\mathrm{HRO}_{\tau(A)}$. If every s_j realized B_j, then c_1 would return a realizer of $\bot$, and there is none; so the realizing property holds vacuously. In Kleene's reading the same argument applies.

Conjunction. For $A \wedge B$ return $\pi(v_1, v_2)$, with v_1, v_2 the values of c_1, c_2; to eliminate, return $(v)_0$ or $(v)_1$. Both properties follow from the clauses for $\times$ and $\wedge$.

Disjunction. To introduce $A \vee B$ from A return $\pi(0, v)$, and from B return $\pi(1, v)$. To infer C from $A \vee B$, from C under the extra hypothesis A and from C under the extra hypothesis B: run c_1 to obtain r; if $(r)_0 = 0$, run c_2 with $(r)_1$ added as the realizer of the extra hypothesis, and otherwise run c_3 in the same way. By the typing property of c_1, the number r lies in $\mathrm{HRO}_{\tau(A)+\tau(B)}$, so $(r)_1$ lies in the type of the branch taken, and that branch halts in $\mathrm{HRO}_{\tau(C)}$. If the hypotheses are realized, r realizes $A \vee B$, so $(r)_1$ realizes the side named by the tag, and the branch returns a realizer of C.

Implication. Introduction is treated in the proof of the theorem. For elimination, run c_1 to obtain r and c_2 to obtain s, and return $\{r\}(s)$. By the typing property, $r \in \mathrm{HRO}_{\tau(A)\to\tau(B)}$ and $s \in \mathrm{HRO}_{\tau(A)}$, so $\{r\}(s)$ halts in $\mathrm{HRO}_{\tau(B)}$; this is where the typing property is used to show that a run halts. If the hypotheses are realized, s realizes A and r realizes $A \to B$, so $\{r\}(s)$ realizes B. In Kleene's reading the last sentence is the whole argument.

Number quantifiers. To introduce $\forall x\, A(x)$, where x is not free in any hypothesis, return a program computing $n \mapsto c_1(n, \ldots)$, the program of the premise with x given the value n. For every n it halts in $\mathrm{HRO}_{\tau(A)}$, since the types of the hypotheses do not depend on x; and if the hypotheses are realized, they remain realized when x is given the value n, so the value realizes $A(n)$. To instantiate at a term t, compute the value v of t, which needs finitely many runs of the total programs assigned to the sequence variables of t, run c_1 to obtain r, and return $\{r\}(v)$. To introduce $\exists x\, A(x)$ from $A(t)$, return $\pi(v, w)$ with v the value of t and w the value of c_1. To infer C from $\exists x\, A(x)$ and from C under the extra hypothesis $A(x)$, where x is not free in C or in the other hypotheses, run c_1 to obtain r and run c_2 with x given the value $(r)_0$ and with $(r)_1$ as the realizer of $A(x)$.

Sequence quantifiers. To introduce $\forall \alpha\, A(\alpha)$, where α is not free in any hypothesis, return a program computing $i \mapsto c_1(\ldots, i, \ldots)$, the program of the premise with α assigned i. On every total program i it halts, with a value of the right type and, if the hypotheses are realized, realizing $A(i)$. To instantiate at a sequence variable β, apply the value of c_1 to the program assigned to β. To introduce $\exists \alpha\, A(\alpha)$ from $A(\beta)$, return $\pi(i_\beta, w)$, where i_β is the program assigned to β, total by the definition of a valuation. To infer C from $\exists \alpha\, A(\alpha)$ and from C under the extra hypothesis $A(\alpha)$, with α not free in C or in the other hypotheses, run c_1 to obtain r; its component $(r)_0$ is total, by the typing property or by the realizing clause, and c_2 is run with α assigned $(r)_0$ and with $(r)_1$ as the realizer of $A(\alpha)$.

Equality. Reflexivity is realized by 0, a realizer of a true equation. To infer $A(s)$ from $t = s$ and $A(t)$, return the value of the program for $A(t)$. Its type is $\tau(A(s)) = \tau(A(t))$. If the hypotheses are realized, $t = s$ is realized and therefore true, the two terms have the same value, and a realizer of $A(t)$ realizes $A(s)$.

Arithmetic axioms. Each is realized by 0. The equations for $+$ and $\cdot$ are true. For $\neg(St = 0)$, the number 0 lies in $\mathrm{HRO}_{\mathbb{N}\to\mathbb{N}}$ and the false equation $St = 0$ has no realizer. For $St = Ss \to t = s$, the program 0 returns 0 on every input, and when its input realizes the equation $St = Ss$, that equation is true, so $t = s$ is true and realized by 0.

Induction. This case is treated in the proof of the theorem: with the values r_0 and r_1 of the two premise programs, the program returns a program for the primitive recursion $R(0) = r_0$, $R(n+1) = \{\{r_1\}(n)\}(R(n))$.

In every case the program is computed from the programs of the premises by an explicit procedure, so it exists by the standard facts recalled in the proof of the theorem.

F.2 Located cuts and Archimedean ordered fields

Throughout, $C = (\mathscr{L}_C, \mathscr{U}_C)$ and D, E are located cuts, and ratios are compared, added and multiplied as in Chapter 4; every comparison of two ratios is decided by an integer computation.

Lemma F.1 (Crossing)**.** *Every lower member of a located cut lies below every upper member. If no lower member of C is an upper member of D and no lower member of D is an upper member of C, then $C = D$. In particular a located cut is determined by its lower set, and also by its upper set.*

Proof. The first statement is Lemma 12.28(1). For the second statement let $q \in \mathscr{L}_C$, and take $q' > q$ in $\mathscr{L}_C$. Locatedness of D at $q < q'$ gives $q \in \mathscr{L}_D$ or $q' \in \mathscr{U}_D$, and the second is excluded because q' is a lower member of C. So $\mathscr{L}_C \subseteq \mathscr{L}_D$, and symmetrically $\mathscr{L}_D \subseteq \mathscr{L}_C$. The same argument with upper sets, using a smaller upper member and locatedness at the pair it forms, gives $\mathscr{U}_C = \mathscr{U}_D$. If C and D have the same lower set, a lower member of either is a lower member of the other, hence not an upper member of it, and the hypothesis holds; likewise for upper sets. □

Lemma F.2 (Approximation between given members)**.** *Let C be a located cut, $a_0 \in \mathscr{L}_C$, $a_1 \in \mathscr{U}_C$ and n a natural number. There are $q \in \mathscr{L}_C$ and $r \in \mathscr{U}_C$ with $a_0 \le q, r \le a_1$ and $r - q < 1/(n+1)$. No choice principle is used.*

Proof. Lemma 12.32 gives $q \in \mathscr{L}_C$ and $r \in \mathscr{U}_C$ with $r - q < 1/(n+1)$. Replace q by a_0 if $q < a_0$ and r by a_1 if $a_1 < r$; both comparisons are decided, membership is kept, and the gap only shrinks. □

Lemma F.3 (The ratios in an Archimedean ordered field)**.** *Let F be an Archimedean ordered field and $u \mapsto u_F$ the map from $\mathbb{Z}_\delta$ preserving 1, sums and products. Then:*

(a) if $0 < a + b$, then $0 < a$ or $0 < b$;

(b) $u_F < v_F$ exactly when $u < v$;

(c) multiplying by a positive element preserves $<$, the inverse of a positive element is positive, and multiplying by a positive element reflects $<$;

(d) for a ratio $q = \langle u; d\rangle$ the element $q_F = u_F d_F^{-1}$ depends only on q; $q \mapsto q_F$ preserves sums and products; and $q < r$ exactly when $q_F < r_F$.

Proof. (a) Cotransitivity at a applied to $0 < a + b$ gives $0 < a$ or $a < a + b$, and adding $-a$ to the second gives $0 < b$.

(b) By induction, $0 < n_F$ for every natural number $n \ge 1$: $0 < 1$, and adding n_F to $0 < 1$ gives $n_F < n_F + 1_F = (n+1)_F$. If $u < v$, then $0 < (v - u)_F$, and adding u_F gives $u_F < v_F$. Conversely, suppose $u_F < v_F$. The integers are compared by decision. If $v = u$, the order would not be irreflexive; if $v < u$, then $v_F < u_F$ and transitivity gives $u_F < u_F$. Hence $u < v$.

(c) If $x < y$ and $0 < c$, then $0 < y - x$, so $0 < (y - x)c$, and adding xc gives $xc < yc$. Let c be positive with inverse e. Take a natural number K with $c < K_F$; then $K + 1 \ge 1$, and $(K+1)_F$ is positive, so it has an inverse w. Since $(K+1)_F w = 1$ is positive, a summand of the sum of $K + 1$ copies of w is positive by (a), so $0 < w$. Cotransitivity applied to $0 < w$ gives $0 < e$ or $e < w$. In the second case, multiplying by c gives $1 < wc$, and multiplying $c < (K+1)_F$ by w gives $wc < 1$; so $1 < 1$, which is impossible. Hence $0 < e$. If $xc < yc$ with c positive, multiplying by the positive inverse of c gives $x < y$.

(d) For $d \geq 1$, d_F is positive by (b), so it has an inverse, unique because inverses in a commutative ring are. If $ue = vd$, then $u_F e_F = v_F d_F$, and multiplying by $d_F^{-1} e_F^{-1}$ gives $u_F d_F^{-1} = v_F e_F^{-1}$. Sums and products are preserved because $(de)_F^{-1} = d_F^{-1} e_F^{-1}$: $(ue + vd)_F (de)_F^{-1} = u_F d_F^{-1} + v_F e_F^{-1}$ and $(uv)_F (de)_F^{-1} = u_F d_F^{-1}\, v_F e_F^{-1}$. Finally $\langle u; d\rangle < \langle v; e\rangle$ means $ue < vd$, which by (b) means $(ue)_F < (vd)_F$; multiplying by the positive element $d_F^{-1} e_F^{-1}$, which (c) allows in both directions, this is $q_F < r_F$. □

For x in F, the cut $\kappa_F(x)$ has lower set $\{q : q_F < x\}$ and upper set $\{r : x < r_F\}$. The ratios form an Archimedean ordered field, with their order decided; their cut map sends q to the cut of the ratio q, $\kappa(q) = (\{p : p < q\}, \{p : q < p\})$. The sum, negation, product and order of located cuts are defined in the proof of Theorem 12.38.

Lemma F.4 (The box law)**.** *Let F be an Archimedean ordered field. If $A < x < A'$, $B < y < B'$, and the four corners AB, AB', $A'B$, $A'B'$ lie below Q, then $Q < xy$ is impossible; if the four corners lie above Q, then $xy < Q$ is impossible. For ratios with $a < a'$ and $b < b'$: if $a \leq x \leq a'$, $b \leq y \leq b'$ and q lies below the four corners, then $q < xy$; if r lies above them, then $xy < r$.*

Proof. In every commutative ring,

$$(A'-x)\big[(B'-y)(Q-AB)+(y-B)(Q-AB')\big]+(x-A)\big[(B'-y)(Q-A'B)+(y-B)(Q-A'B')\big] + \big((x-A)+(A'-x)\big)\big((y-B)+(B'-y)\big)(xy-Q) = 0.$$

If the corners lie below Q and $Q < xy$, every product on the left is a product of positive elements, so the left side is positive, and $0 < 0$. The second statement uses the same identity with every bracket negated. For ratios, the identity

$$(a'-a)(b'-b)(xy-q) = (a'-x)(b'-y)(ab-q) + (a'-x)(y-b)(ab'-q) + (x-a)(b'-y)(a'b-q) + (x-a)(y-b)(a'b'-q)$$

expresses $xy - q$, after the positive factor $(a'-a)(b'-b)$, as a sum of positive multiples of the corner differences with nonnegative weights whose sum is $(a'-a)(b'-b) > 0$; so it is positive. The upper statement is the same with $r - xy$. □

Lemma F.5 (The cut of an element)**.** *Let F be an Archimedean ordered field. For every x in F, $\kappa_F(x)$ is a located cut. For every ratio q, $\kappa_F(q_F) = \kappa(q)$. Moreover $x < y$ exactly when $\kappa_F(x) < \kappa_F(y)$, and $\kappa_F(x) = \kappa_F(y)$ implies $x = y$.*

Proof. By the Archimedean condition $-x < N_F$ and $x < M_F$ for natural numbers N, M, so $-N$ is a lower and M an upper member. If $q' < q$ and $q_F < x$, then $q'_F < q_F < x$; the upper set is closed upward in the same way. The lower set has no greatest element, as shown in the proof of Theorem 12.38, using Lemma F.3(c) to multiply by the positive element $x - q_F$; the upper set has no least element by the same argument applied to $-x$. A ratio in both sets would give $q_F < x < q_F$. Locatedness is cotransitivity, as shown in the proof of the theorem. For a ratio q, the lower set of $\kappa_F(q_F)$ is $\{p : p_F < q_F\} = \{p : p < q\}$ by Lemma F.3(d), and so $\kappa_F(q_F) = \kappa(q)$ by Lemma F.1.

Let $x < y$. The positive element $y - x$ has an inverse e, and some natural number N has $e < N_F$; then $N \geq 1$ and $1 < N_F(y - x)$, so $\langle 1; N\rangle_F < y - x$. Lemma 12.32 gives q in the lower and r in

the upper set of $\kappa_F(x)$ with $r - q < \langle 1; N\rangle$. Then $r_F < q_F + \langle 1; N\rangle_F < x + (y - x) = y$, so r lies in the upper set of $\kappa_F(x)$ and in the lower set of $\kappa_F(y)$: $\kappa_F(x) < \kappa_F(y)$. Conversely, a ratio r in both sets gives $x < r_F < y$. Finally, if $\kappa_F(x) = \kappa_F(y)$, then $x < y$ would give $\kappa_F(x) < \kappa_F(x)$, that is a ratio in both the lower and the upper set of one located cut; so $\neg(x < y)$, likewise $\neg(y < x)$, and the field axiom on such pairs gives $x = y$. □

Lemma F.6 (The field of located cuts)**.** *With the sum, negation, product and order defined in the proof of Theorem 12.38, and with zero and one the cuts of the ratios* 0 *and* 1*:*

(a) $C + D$, $-C$ and CD are located cuts;

(b) for every Archimedean ordered field F, $\kappa_F(x+y) = \kappa_F(x) + \kappa_F(y)$ and $\kappa_F(xy) = \kappa_F(x)\,\kappa_F(y)$; in particular cuts of ratios add and multiply as the ratios do;

(c) the located cuts form a commutative ring;

(d) they form an Archimedean ordered field.

Proof. (a) *Sum.* For $a \in \mathscr{L}_C$, $b \in \mathscr{L}_D$, $a' \in \mathscr{U}_C$ and $b' \in \mathscr{U}_D$, the ratio $a + b - 1$ is a lower and $a' + b' + 1$ an upper member. The lower set is closed downward by its definition, and if $q < a + b$ then $(q + a + b)/2$ is a larger member; the upper set is treated in the same way. A common member q would satisfy $a' + b' < q < a + b$ for lower members a, b and upper members a', b', while $a < a'$ and $b < b'$. For locatedness let $q < r$ and $\varepsilon = r - q$. By Lemma 12.32 choose $a \in \mathscr{L}_C$, $a' \in \mathscr{U}_C$ with $a' - a < \varepsilon/2$, and $b \in \mathscr{L}_D$, $b' \in \mathscr{U}_D$ likewise. Compare q with $a + b$. If $q < a + b$, then q is a lower member. Otherwise $a + b \le q$, and $a' + b' < a + b + \varepsilon \le r$, so r is an upper member.

Negation. Negation of ratios reverses their order, so every law of $-C$ is a law of C read in reverse; locatedness at $q < r$ is locatedness of C at $-r < -q$, with the two cases exchanged.

Product. Choose $a \in \mathscr{L}_C$, $a' \in \mathscr{U}_C$, $b \in \mathscr{L}_D$, $b' \in \mathscr{U}_D$. A ratio below the smallest of the four corners is a lower member, and a ratio above the largest an upper member. The lower set is closed downward by definition; if q lies below the four corners of some brackets, the mean of q and the smallest corner is a larger lower member. The upper set is treated in the same way. Suppose q is a lower member through brackets (a, a', b, b') and an upper member through brackets (c, c', d, d'). Let x be the larger of a and c, and y the larger of b and d; then x is a lower member of C, so $a \le x < a'$ and $c \le x < c'$, and likewise for y. The box law (Lemma F.4) for ratios gives $q < xy$ from the first brackets and $xy < q$ from the second, a contradiction.

For locatedness let $q < r$ and $\varepsilon = r - q$. Fix $a_0 \in \mathscr{L}_C$, $a_1 \in \mathscr{U}_C$, $b_0 \in \mathscr{L}_D$, $b_1 \in \mathscr{U}_D$, and a natural number $B \ge 1$ with $|a_0|, |a_1|, |b_0|, |b_1| \le B$. By Lemma F.2 choose $a \in \mathscr{L}_C$ and $a' \in \mathscr{U}_C$ with $a_0 \le a < a' \le a_1$ and $a' - a < \varepsilon/(8B)$, and b, b' for D likewise. All four ratios have absolute value at most B, so each corner differs from ab by at most $B\varepsilon/(8B) + B\varepsilon/(8B) = \varepsilon/4$; for instance $|a'b' - ab| \le |a'|\,|b' - b| + |b|\,|a' - a|$. Compare $ab - q$ with $\varepsilon/2$. If $ab - q > \varepsilon/2$, every corner exceeds $q + \varepsilon/4$, so q is a lower member. Otherwise $ab \le q + \varepsilon/2$, every corner is at most $q + 3\varepsilon/4 < r$, and r is an upper member.

(b) For the sum, suppose q is a lower member of $\kappa_F(x+y)$ and an upper member of $\kappa_F(x)+\kappa_F(y)$, through $x < a'_F$, $y < b'_F$ and $a' + b' < q$. Then $q_F < x + y < a'_F + b'_F < q_F$, which is impossible.

If q is a lower member of the sum through $q < a + b$ with $a_F < x$ and $b_F < y$, and an upper member of $\kappa_F(x + y)$, then $q_F < a_F + b_F < x + y < q_F$. By Lemma F.1 the two cuts are equal. For the product, suppose $q_F < xy$ and q lies above the four corners of ratios with $a_F < x < a'_F$ and $b_F < y < b'_F$. Since $q \mapsto q_F$ preserves products and order, the box law in F excludes $q_F < xy$. The other crossing is excluded in the same way, and Lemma F.1 gives equality. For the field of ratios, (b) and Lemma F.5 give $\kappa(q) + \kappa(r) = \kappa(q + r)$ and $\kappa(q)\kappa(r) = \kappa(qr)$.

(c) Every law is proved by Lemma F.1, or by comparing lower sets. Finitely many brackets $[a_i, a'_i]$ of one cut, with lower ends in its lower set and upper ends in its upper set, have a common ratio: the largest a_i, a lower member, which lies below every a'_j. Call it a *common point* of the brackets.

Addition. The lower set of $(C + D) + E$ is the set of q with $q < a + b + e$ for lower members a, b, e of C, D, E: if $q < s + e$ with $s < a + b$ this holds, and if $q < a + b + e$ then $s = a + b - (a + b + e - q)/2$ is a lower member of $C + D$ with $q < s + e$. The same description holds for $C + (D + E)$. Commutativity is visible in the definition. The lower set of $C + 0$ is $\mathcal{L}_C$: if $q < a + z$ with $z < 0$, then $q < a$; if $q < a$ with $a \in \mathcal{L}_C$, then $z = (q - a)/2$ is negative and $q < a + z$. The lower set of $C + (-C)$ is the set of negative ratios: if $q < a + b$ with $a \in \mathcal{L}_C$ and $-b \in \mathcal{U}_C$, then $a < -b$ and $q < 0$; conversely, if $q < 0$, Lemma 12.32 gives $a \in \mathcal{L}_C$, $a' \in \mathcal{U}_C$ with $a' - a < -q$, and $b = -a'$ satisfies $q < a + b$.

Multiplication. Commutativity is visible in the definition. For $C \cdot 1 = C$: if q is a lower member of $C \cdot 1$ through brackets (a, a', b, b'), with $b < 1 < b'$, the box law at the point $(a, 1)$ gives $q < a$, so q is a lower member of C and not an upper member. If q is an upper member of $C \cdot 1$, the box law at $(a', 1)$ gives $a' < q$, so q is an upper member of C and not a lower member. Lemma F.1 gives $C \cdot 1 = C$.

For associativity, suppose q is a lower member of $(CD)E$ and an upper member of $C(DE)$. The first means that q lies below the four corners of brackets $p \in \mathcal{L}_{CD}$, $p' \in \mathcal{U}_{CD}$, $e \in \mathcal{L}_E$, $e' \in \mathcal{U}_E$, where p and p' come with brackets of C and D; the second means that q lies above the four corners of brackets of C and of DE, which come with brackets of D and E. Let x, y, z be common points of all the brackets of C, of D and of E that occur. By the box law, $p < xy < p'$; by the box law again, $q < (xy)z$. In the same way $x(yz) < q$. Since $(xy)z = x(yz)$, this is impossible. The other crossing is excluded in the same way, and Lemma F.1 gives $(CD)E = C(DE)$.

For distributivity, suppose q is a lower member of $C(D + E)$ and an upper member of $CD + CE$. The first means that q lies below the four corners of brackets a, a' of C and s, s' of $D + E$, with $s < b_1 + e_1$ and $b'_1 + e'_1 < s'$ for lower members b_1, e_1 and upper members b'_1, e'_1 of D and E. The second means $t + w < q$ with t an upper member of CD and w an upper member of CE, each above the corners of its brackets. Let x, y, z be common points of the brackets of C, D and E that occur. Then $s < b_1 + e_1 \le y + z < b'_1 + e'_1 < s'$, so the box law gives $q < x(y + z)$; and it gives $xy < t$ and $xz < w$, so $xy + xz < q$. Since $x(y + z) = xy + xz$, this is impossible. The other crossing is the same with lower and upper exchanged.

These are the laws of a commutative ring.

(d) *Order.* No ratio lies in both sets of one cut, so $<$ is irreflexive. If $q \in \mathcal{U}_C \cap \mathcal{L}_D$ and $p \in \mathcal{U}_D \cap \mathcal{L}_E$, then $q < p$, so $p \in \mathcal{U}_C$ and $C < E$. For cotransitivity let $q \in \mathcal{U}_C \cap \mathcal{L}_D$, take a smaller $q' \in \mathcal{U}_C$, and apply the locatedness of E at $q' < q$: either $q' \in \mathcal{L}_E$ and $C < E$, or $q \in \mathcal{U}_E$ and $E < D$. If $\neg(C < D)$, every $q \in \mathcal{L}_D$ lies in $\mathcal{L}_C$: take a larger $q' \in \mathcal{L}_D$, and

locatedness of C at $q < q'$ gives $q \in \mathscr{L}_C$ or $q' \in \mathscr{U}_C$, the latter meaning $C < D$. So if neither $C < D$ nor $D < C$, the lower sets agree and $C = D$.

If $C < D$ through $q \in \mathscr{U}_C \cap \mathscr{L}_D$, take $p < q$ in $\mathscr{U}_C$ and $p' > q$ in $\mathscr{L}_D$, put $\delta = p' - p$, and choose $e \in \mathscr{L}_E$, $e' \in \mathscr{U}_E$ with $e' - e < \delta$. The ratio $p + e'$ lies above a sum of upper members of C and E, since C has an upper member below p; and $p + e' < p' + e$, a sum of lower members of D and E. So $p + e'$ lies in the upper set of $C + E$ and in the lower set of $D + E$, and $C + E < D + E$.

If $0 < C$ and $0 < D$, there are positive $a \in \mathscr{L}_C$ and $b \in \mathscr{L}_D$. With any upper members $a' > a$ and $b' > b$, the four corners are at least $ab > 0$, so 0 and then a positive ratio are lower members of CD, and $0 < CD$. The ratio $1/2$ shows $0 < 1$.

Inverses. Let $0 < C$, so that C has a positive lower member. Let C' have the lower and upper sets

$$\{q : \ qa' < 1 \text{ for some } a' \in \mathscr{U}_C\}, \qquad \{r : \ 1 < ra \text{ for some positive } a \in \mathscr{L}_C\}.$$

Every upper member of C is positive, so 0 is a lower member of C', and $2/a$ is an upper member for a positive lower member a. If $qa' < 1$ and $p < q$, then $pa' < 1$; and $q + (1 - qa')/(2a')$ is a larger lower member. If $1 < ra$, every $r'' > r$ has $1 < r''a$, and $r - (ra - 1)/(2a)$ is a smaller upper member. A common member q would satisfy $qa' < 1 < qa$ with $q > 0$ and $a < a'$, which is impossible. For locatedness let $q < r$. If $q \le 0$, then q is a lower member. Otherwise $0 < 1/r < 1/q$, and the locatedness of C gives $1/r \in \mathscr{L}_C$ or $1/q \in \mathscr{U}_C$. In the first case a larger lower member $a > 1/r$ is positive with $1 < ra$; in the second a smaller upper member $a' < 1/q$ has $qa' < 1$.

Now $CC' = 1$. If q is a lower member of CC', through brackets (a, a', b, b') with b a lower member of C' through $bc' < 1$, let x be a common point of the brackets of C that occur, together with a positive lower member of C; then $x > 0$, $x < c'$, and $xb < 1$ whether b is positive or not, so the box law gives $q < xb < 1$. If q is an upper member of CC', through brackets (a, a', b, b') with b' an upper member of C' through $1 < b'c$ for a positive $c \in \mathscr{L}_C$, let x be a common point of the brackets of C that occur, together with c; then $x \ge c$ and $1 < b'c \le b'x < q$. So no lower member of CC' is above 1, and no upper member is below 1, and Lemma F.1, with the cut of the ratio 1, gives $CC' = 1$.

Archimedean bound. For an upper member r of C take a natural number $n > r$. By (b) the sum of n copies of the cut of 1 is the cut of the ratio n, which has r as a lower member; so $C < n$. □

F.3 Limits in the recursive model

Lemma F.7 (Regular sequences read classically)**.** *Reasoning classically, every regular sequence s has a limit $\lim s$, a real number with $|s_n - \lim s| \le 1/(n+1)$ for every n. Two regular sequences name the same point exactly when their limits agree. A point x is positive exactly when $\lim x > 0$; hence $x < y$ exactly when $\lim x < \lim y$, and $x \le y$ exactly when $\lim x \le \lim y$. Two points are within $1/(k+1)$ exactly when their limits differ by at most $1/(k+1)$, and a ratio q lies in the lower set of the cut of x exactly when $q < \lim x$.*

Proof. Regularity makes (s_n) a Cauchy sequence of rational numbers, and letting m grow in $|s_n - s_m| \le 1/(n+1) + 1/(m+1)$ gives the bound. If $s \approx t$, the differences $s_n - t_n$ eventually

stay within every $1/(k+1)$, so the limits agree; conversely, equal limits give $|s_n - t_n| \le 1/(n+1)+1/(n+1)$, which is eventually within every tolerance. If $s_N > 2/(N+1)$, Lemma 12.16 gives $s_m > 1/(2N+2)$ from $m = 2N+1$ on, so $\lim s \ge 1/(2N+2) > 0$. If $\lim s = \delta > 0$, then $s_N \ge \delta - 1/(N+1) > 2/(N+1)$ once $N+1 > 3/\delta$. The stages of $y - x$ are $y_{2n+1} - x_{2n+1}$, whose limit is $\lim y - \lim x$; this gives the statement on $<$, and $\le$ is its negation. Being within $1/(k+1)$ is a pair of order statements, and the lower set of the cut of x consists of the ratios q with $\eta(q) < x$, which by the statement on $<$ holds exactly when $q = \lim \eta(q) < \lim x$. □

F.4 Stage estimates on the line

These lemmas complete the proof of Proposition 12.19. Throughout, s, t, u are regular sequences of ratios and $K_s = \lfloor |s_0| \rfloor + 3$, so that $|s_n| \le K_s$ for every n.

Lemma F.8 (Naming the same point)**.** *Let s and t be regular sequences.*

(a) If $|s_n - t_n| \le C/(n+1)$ for every n, for a fixed C, then $s \approx t$.

(b) $s \approx t$ exactly when $|s_n - t_n| \le 2/(n+1)$ for every n.

(c) A point is $\eta(0)$ exactly when it is neither positive nor negative.

Proof. (a) For the tolerance $1/(k+1)$, every $n \ge C(k+1)$ has $C/(n+1) < 1/(k+1)$. (b) One direction is (a). Conversely let $s \approx t$ and fix n. For any k take a stage $p \ge k$ beyond the one from which s and t agree within $1/(k+1)$. Then $|s_n - t_n| \le |s_n - s_p| + |s_p - t_p| + |t_p - t_n| \le 2/(n+1) + 2/(p+1) + 1/(k+1) \le 2/(n+1) + 3/(k+1)$, and the Archimedean step leaves $2/(n+1)$, as in the proof of Theorem 12.17. (c) The constant sequence 0 has no stage above $2/(n+1)$ or below $-2/(n+1)$, and positivity and negativity belong to the point. Conversely, if s names a point that is neither positive nor negative, no stage of s lies above $2/(n+1)$ or below $-2/(n+1)$; each of these comparisons is decided, so $|s_n| \le 2/(n+1)$ at every stage, and Lemma 12.15 gives $s \approx \eta(0)$. □

Lemma F.9 (Following ratios)**.** *Say that s follows ratios $f(M)$ with constant C when $|s_n - f(M)| \le C\big(1/(n+1) + 1/(M+1)\big)$ for all n and M.*

(a) Every s follows its own stages with constant 1, and the constant sequence of q follows the constant ratios q with constant 0.

(b) If s follows f with constant C and t follows g with constant D, then $s + t$ follows $f + g$ with constant $C + D$, $-s$ follows $-f$ with constant C, and, when $|f(M)| \le B$ for every M, st follows fg with constant $K_t C + BD$.

(c) If s follows f with constant C, t follows g with constant D, and $|f(M) - g(M)| \le F/(M+1)$ for every M, then $|s_n - t_n| \le (C+D)/(n+1)$ for every n, so $s \approx t$.

Proof. (a) is regularity. (b) The stage n of $s + t$ reads both sequences at $2n+1 \ge n$, and $1/(2n+2) \le 1/(n+1)$; negation changes no distance. The stage n of st is $s_j t_j$

with $j \geq n$, and $|s_j t_j - f(M)g(M)| \leq |t_j|\,|s_j - f(M)| + |f(M)|\,|t_j - g(M)|$, which is at most $(K_t C + BD)\big(1/(j+1) + 1/(M+1)\big)$. (c) For every M,

$$|s_n - t_n| \leq |s_n - f(M)| + |f(M) - g(M)| + |g(M) - t_n| \leq \frac{C+D}{n+1} + \frac{C+D+F}{M+1},$$

and the Archimedean step removes the last term. Then (a) of Lemma F.8 applies. □

Lemma F.10 (Positive points). *If $a + b > 0$, then $a > 0$ or $b > 0$. If $a > 0$ and $b > 0$, then $a + b > 0$ and $ab > 0$.*

Proof. Let s name a and t name b. Positivity belongs to the point, so the sum sequence has a positive stage N: $s_{2N+1} + t_{2N+1} > 2/(N+1)$. One of the two ratios, by a decided comparison, exceeds $1/(N+1) = 2/((2N+1)+1)$, a positive stage of a or of b. Now let $s_N > 2/(N+1)$ and $t_M > 2/(M+1)$, and let L be the larger of N and M. By Lemma 12.16, every $m \geq 2L+1$ has $s_m > 1/(2L+2)$ and $t_m > 1/(2L+2)$. The sum at stage $n = 2L+2$ reads stage $2n+1 \geq 2L+1$, so it exceeds $1/(L+1) > 2/(2L+3) = 2/(n+1)$. The product at stage $n = 2(2L+2)^2$ reads a stage $j \geq n \geq 2L+1$, so it exceeds $1/(2L+2)^2 > 2/(n+1)$. □

Lemma F.11 (The inverse of a point apart from zero). *Let s name x and suppose $|s_N| > 2/(N+1)$. Then every $m \geq 2N+1$ has $|s_m| > 1/(2N+2)$, the ratios $r_n = 1/s_{k(n)}$ with $k(n) = 4(N+1)^2(n+1) - 1$ form a regular sequence, and its point y has $xy = \eta(1)$.*

Proof. The first statement is Lemma 12.16 applied to s or to $-s$, according to the decided sign of s_N. Since $4(N+1)^2 \geq 2N+2$, every $k(n)$ is at least $2N+1$. For stages a, b of modulus above $1/(2N+2)$, $|1/a - 1/b| = |a-b|/(|a|\,|b|) \leq 4(N+1)^2|a-b|$, so

$$|r_m - r_n| \leq 4(N+1)^2\Big(\frac{1}{k(m)+1} + \frac{1}{k(n)+1}\Big) = \frac{1}{m+1} + \frac{1}{n+1}.$$

For the product, s follows the ratios $s_{k(M)}$ with constant 1, since $k(M) \geq M$, and the sequence r follows its own stages $1/s_{k(M)}$ with constant 1. The products of these ratios are all 1, and $|s_{k(M)}| \leq K_s$, so by Lemma F.9 the product sequence follows the constant ratios 1, as does $\eta(1)$; hence $xy = \eta(1)$. □

Lemma F.12 (Square roots by trisection). *Every point $x \geq 0$ has a square root $y \geq 0$, computed from any one regular sequence naming x, with no choice.*

Proof. We use four facts. (i) For every stage m of a sequence s naming x, $x \leq \eta(s_m + 1/(m+1))$ and $\eta(s_m - 1/(m+1)) \leq x$: the stages of $x - \eta(s_m + 1/(m+1))$ are $s_{2n+1} - s_m - 1/(m+1) \leq 1/(2n+2)$, never above $2/(n+1)$, and similarly for the other side. (ii) If $u \leq v$, then $u + w \leq v + w$, by item 1 of the proposition. (iii) If $u \geq 0$ and $v \geq 0$, then $u + v \geq 0$ and $uv \geq 0$. For the sum, $u + v < 0$ would make $-u$ or $-v$ positive by Lemma F.10. For the product, if $uv < 0$, then uv is invertible by item 3, so u and v are invertible, hence apart from 0, hence positive since they are not negative; then $uv > 0$, which contradicts $uv < 0$. In particular $0 \leq a \leq y$ gives $y\,y - a\,a = (y-a)(y+a) \geq 0$. (iv) Two points in an interval $[\eta(a), \eta(b)]$ are within $b - a$ of each other, by (ii).

Let s name x, take n_0 with $x < \eta(n_0)$ from item 4, and put $K = n_0 + 1$, so that $K \geq 1$ and $x \leq \eta(K^2)$. Define ratios $a_n \leq b_n$ by $a_0 = 0$, $b_0 = K$, and the following step. Let $w = b_n - a_n$, $c_1 = a_n + w/3$, $c_2 = a_n + 2w/3$, and $m = 6 \cdot 9^n$. Since $a_n \geq 0$,

$$c_2^2 - c_1^2 = \frac{w}{3}(c_1 + c_2) \geq \frac{w^2}{3},$$

and we will have $w = K(2/3)^n$, so the gap is at least $1/(3 \cdot 9^n)$, more than $2/(m+1)$. If $s_m > c_1^2 + 1/(m+1)$, put $a_{n+1} = c_1$ and $b_{n+1} = b_n$; then $\eta(c_1^2) < \eta(s_m - 1/(m+1)) \leq x$ by (i). Otherwise put $a_{n+1} = a_n$ and $b_{n+1} = c_2$; then $x \leq \eta(s_m + 1/(m+1)) \leq \eta(c_1^2 + 2/(m+1)) < \eta(c_2^2)$. Each step is a comparison of two ratios. By induction, $0 \leq a_n < b_n \leq K$, $b_n - a_n = K(2/3)^n$, (a_n) does not decrease, (b_n) does not increase, and $\eta(a_n^2) \leq x \leq \eta(b_n^2)$.

Put $\varphi(k) = 2K(k+1)$. Since $(3/2)^n \geq 1 + n/2$, $K(2/3)^{\varphi(k)} \leq K/(1 + K(k+1)) < 1/(k+1)$. The ratios $r_k = a_{\varphi(k)}$ satisfy $a_{\varphi(k)} \leq r_l < b_{\varphi(k)}$ for $l \geq k$, so $|r_k - r_l| < 1/(k+1)$, and r is regular; let y be its point. No stage of r is negative, so $y \geq 0$. Every stage r_l lies below $b_{\varphi(k)}$, and $a_{\varphi(k)} - r_l$ is at most 0 when $l \geq k$ and less than $1/(l+1)$ when $l < k$. So neither $y - \eta(b_{\varphi(k)})$ nor $\eta(a_{\varphi(k)}) - y$ has a stage above $2/(n+1)$, and $\eta(a_{\varphi(k)}) \leq y \leq \eta(b_{\varphi(k)})$. By (iii), $\eta(a_{\varphi(k)}^2) \leq y\,y \leq \eta(b_{\varphi(k)}^2)$, and x lies in the same interval, whose width is $(b-a)(b+a) \leq 2K(b-a) < 2K/(k+1)$ at $\varphi(k)$. By (iv), $y\,y$ and x are within every $1/(k+1)$ of each other, hence equal. $\square$

Appendix G

Proofs for the linear-algebra chapter

This appendix proves the rows of Table 14.1 on matrices, unique roots and monotone maps, and the results on unique choice used in Section 14.9. Notation and conventions, including the meaning of *no choice*, are those of Chapter 14.

G.1 Matrices

Proposition G.1 (Counting independent vectors)**.** *Independent vectors and injective matrices are counted with no principle.*

1. *If $v_0, \dots, v_{k-1}$ is an independent family of vectors of $\mathbb{R}^n_\delta$, then $k \le n$.*
2. *If an $n \times m$ matrix A has zero kernel, that is $Ax = 0$ gives $x = 0$, then $m \le n$.*

No omniscience principle and no choice is used.

Proof. For the first statement, $k \le n$ is a decided comparison of natural numbers, so it suffices to refute $n < k$. We show by induction on n that no family of more than n vectors of $\mathbb{R}^n_\delta$ is independent. For $n = 0$ every vector is the empty tuple, and the combination with every coefficient 1 vanishes, so independence of a nonempty family would give $1 = 0$.

Let $v_0, \dots, v_{k-1}$ be an independent family of vectors of $\mathbb{R}^{n+1}_\delta$ with $k \ge n + 2$; we derive a contradiction. Write a_i for the first coordinate of v_i and $\bar{v}_i$ for the others. Two cases each lead to a contradiction.

Every a_i is 0. A vanishing combination of the $\bar{v}_i$ is then a vanishing combination of the v_i, so the $\bar{v}_i$ are $k > n$ independent vectors of $\mathbb{R}^n_\delta$, against the induction hypothesis.

Some a_{i_0} is apart from 0. Let α be its inverse, and put $w_r = \bar{v}_r - a_r \alpha \bar{v}_{i_0}$ for $r \neq i_0$. These are $k - 1 > n$ vectors of $\mathbb{R}^n_\delta$, and they are independent. Indeed, let $\sum_{r \neq i_0} c_r w_r = 0$, put $S = \sum_{r \neq i_0} c_r a_r$, and give v_{i_0} the coefficient $c_{i_0} = -\alpha S$. The first coordinate of $\sum_i c_i v_i$ is $S - \alpha S a_{i_0} = 0$, and the others are

$$\sum_{r \neq i_0} c_r \big(w_r + a_r \alpha \bar{v}_{i_0} \big) - \alpha S \bar{v}_{i_0} = \sum_{r \neq i_0} c_r w_r = 0.$$

Independence of the v_i gives $c_r = 0$ for every r. This again contradicts the induction hypothesis.

Since the goal is a contradiction, neither case has to be decided. The second case shows that no a_i is apart from 0, so every a_i is 0 by Lemma F.8(c), and the first case applies.

For the second statement, $\sum_j c_j \operatorname{col}_j = Ac$, so a matrix with zero kernel has m independent columns in $\mathbb{R}^n_\delta$, and the first statement gives $m \le n$. □

Definition G.2 (Determinants, minors and rank)**.** The determinant of an $n \times n$ matrix M is defined by expansion along the first column: the 0×0 matrix has determinant 1, and for $n \geq 1$

$$\det M = \sum_{i<n} (-1)^i M_{i0} \det M^{(i,0)},$$

where $M^{(i,j)}$ is M with row i and column j deleted. For an $m \times n$ matrix A, an $r \times r$ *minor* is given by injective maps ρ from $\{0, \dots, r-1\}$ to $\{0, \dots, m-1\}$ and σ from $\{0, \dots, r-1\}$ to $\{0, \dots, n-1\}$; its entries are $A_{\rho(a)\sigma(b)}$. The matrix A has *rank* r when some $r \times r$ minor has a determinant $\neq 0$ and every $(r+1) \times (r+1)$ minor has determinant 0.

Proposition G.3 (Two identities of determinants)**.** *For every $(n+1) \times (n+1)$ matrix M over $\mathbb{R}_\delta$,*

$$\det M = \sum_{j \leq n} (-1)^j M_{0j} \det M^{(0,j)} \qquad \textit{and} \qquad \det M^{\mathsf{T}} = \det M.$$

No omniscience principle and no choice is used.

Proof. Both are identities of the ring operations, whose laws hold with no principle by Proposition 12.19. For the first, argue by induction on n. For $n = 0$ both sides are M_{00}. Let $n \geq 1$, and write $M^{\{0,i\},\{0,j\}}$ for M with rows $0, i$ and columns $0, j$ deleted. In the definition of $\det M$, the term $i = 0$ is $M_{00} \det M^{(0,0)}$. For $i \geq 1$, the first row of $M^{(i,0)}$ consists of the entries M_{0j} with $j \geq 1$, at position $j - 1$, and the induction hypothesis expands $\det M^{(i,0)}$ along that row. So

$$\det M = M_{00} \det M^{(0,0)} + \sum_{i \geq 1} \sum_{j \geq 1} (-1)^{i+j-1} M_{i0} M_{0j} \det M^{\{0,i\},\{0,j\}}.$$

In the right side of the claim, the term $j = 0$ is again $M_{00} \det M^{(0,0)}$. For $j \geq 1$, the first column of $M^{(0,j)}$ consists of the entries M_{i0} with $i \geq 1$, at position $i - 1$, and the definition expands $\det M^{(0,j)}$ along it. This gives the same double sum.

For the second identity, again by induction on n, the case $n = 0$ being clear: the definition, the induction hypothesis for the minors, and the first identity give

$$\det M^{\mathsf{T}} = \sum_i (-1)^i M_{0i} \det\big(M^{(0,i)}\big)^{\mathsf{T}} = \sum_i (-1)^i M_{0i} \det M^{(0,i)} = \det M. \qquad \square$$

Theorem G.4 (The rank costs WLPO)**.** *The assertion that every matrix over $\mathbb{R}_\delta$ has a rank which is also the rank of its transpose is equivalent to WLPO. Neither direction uses a choice principle.*

Proof. Assume WLPO, and let A be an $m \times n$ matrix. For each minor, WLPO decides whether its determinant is 0 or $\neq 0$. The minors of a given size are finitely many, since maps between finite sets can be listed and their injectivity decided; so "some $r \times r$ minor has a determinant $\neq 0$" is decided for each r. It holds for $r = 0$, since the empty minor has determinant 1. It fails for $r = m + 1$, since no map from $\{0, \dots, m\}$ into $\{0, \dots, m-1\}$ is injective. Let $r \leq m$ be the largest number for which it holds, found by a bounded search. Every $(r+1) \times (r+1)$ minor then has $\neg(\det \neq 0)$, hence determinant 0, since $x = 0$ says that x is neither positive nor negative and is therefore stable under double negation (Lemma F.8(c)). So A has rank r.

The minors of A^{T} are the transposes of those of A, obtained by exchanging ρ and σ, and by Proposition G.3 they have the same determinants; so r is also the rank of A^{T}.

Conversely, let x be a point, and let r be a rank of the 1×1 matrix (x). No map from $\{0,1\}$ into $\{0\}$ is injective, so $r \leq 1$. If $r = 1$, the only 1×1 minor has determinant $x \neq 0$; if $r = 0$, every 1×1 minor has determinant 0, so $x = 0$. Hence $x = 0 \vee x \neq 0$ for every point, which is WLPO by Corollary 12.18. □

For 2×2 matrices the definition gives $\det A = A_{00}A_{11} - A_{10}A_{01}$. Write $\operatorname{adj} A$ for the matrix with rows $(A_{11}, -A_{01})$ and $(-A_{10}, A_{00})$; then $A \operatorname{adj} A = \operatorname{adj} A \, A = (\det A)\, I$, and $\det(AB) = \det A \det B$, by computation in the ring. Call A *injective* when $Av = 0$ gives $v = 0$, *surjective* when every w is Av for some v, and *invertible* when $AB = BA = I$ for some B.

Theorem G.5 (Inverses of 2×2 matrices)**.** *For 2×2 matrices over $\mathbb{R}_\delta$:*

1. *every matrix with $\det A \neq 0$ is invertible if and only if $\mathrm{MP}_{\mathrm{bin}}$ holds;*
2. *every injective matrix is surjective if and only if $\mathrm{MP}_{\mathrm{bin}}$ holds;*
3. *every surjective matrix is injective.*

No choice is used.

Proof. The first statement. Under $\mathrm{MP}_{\mathrm{bin}}$ a determinant $\neq 0$ is apart from 0 (Corollary 12.18), with an inverse e, and $e \operatorname{adj} A$ inverts A. Conversely, let $x \neq 0$. The diagonal matrix D with entries x and 1 has determinant x; an inverse B gives $xB_{00} = 1$ from the entry $(0,0)$ of $DB = I$. So every $x \neq 0$ is invertible, which is $\mathrm{MP}_{\mathrm{bin}}$ by the eighth row of Table 12.2.

The second. Let A be injective. Then $\det A \neq 0$: if $\det A = 0$, the vectors $(A_{01}, -A_{00})$ and $(A_{11}, -A_{10})$ lie in the kernel, since

$$A\begin{pmatrix} A_{01} \\ -A_{00} \end{pmatrix} = \begin{pmatrix} 0 \\ -\det A \end{pmatrix}, \qquad A\begin{pmatrix} A_{11} \\ -A_{10} \end{pmatrix} = \begin{pmatrix} \det A \\ 0 \end{pmatrix},$$

so all four entries vanish, $A = 0$, and injectivity applied to $Ae_0 = 0$ would give $1 = 0$. Under $\mathrm{MP}_{\mathrm{bin}}$ the first statement makes A invertible, with inverse B, and $w = A(Bw)$. Conversely, let $x \neq 0$. The matrix D above is injective: $(xv_0, v_1) = 0$ gives $v_1 = 0$ and, by Lemma 14.4, $v_0 = 0$. A vector v with $Dv = e_0$ has $xv_0 = 1$, so x is invertible, and the eighth row of Table 12.2 gives $\mathrm{MP}_{\mathrm{bin}}$.

The third. Let A be surjective, and take u, w with $Au = e_0$ and $Aw = e_1$. The matrix B with columns u and w has $AB = I$, so $\det A \det B = 1$ and $\det A$ is invertible. If $Av = 0$, then $(\det A)\, v = \operatorname{adj} A\,(Av) = 0$, and multiplying by the inverse of $\det A$ gives $v = 0$. □

G.2 Unique roots and monotone maps

Lemma G.6 (Quotients exist up to double negation)**.** *Let a and b be points with $b \neq 0$.*

1. *It is not the case that no c has $bc = a$.*

2. *If* $0 \le a \le b$, *then some* c *has* $bc = a$ *exactly when exactly one point* c *of* $[0,1]$ *has* $bc = a$.

No omniscience principle and no choice is used.

Proof. If no c had $bc = a$, then b would have no inverse, since an inverse u would give $b(ua) = a$. So b would be neither positive nor negative, and $b = 0$ by Lemma F.8(c), contrary to $b \neq 0$. The second statement is Lemma 14.10. □

So for $0 \le a \le b$ with $b \neq 0$ the quotient is unique and cannot fail to exist. A stronger, quantitative uniqueness has a price. Say that a function f has a *quantitatively unique root* on $[0,1]$ when for every k there is a j such that any two points x, y of $[0,1]$ with $f(x)$ and $f(y)$ within $1/(j+1)$ of 0 are within $1/(k+1)$ of each other.

Theorem G.7 (Quantitative uniqueness costs $\mathrm{MP}_{\mathrm{bin}}$)**.** *The assertion that for all points* $0 \le a \le b$ *with* $b \neq 0$ *the function* $c \mapsto bc - a$ *has a quantitatively unique root on* $[0,1]$ *is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$. *Together with WLPO it gives LPO. No choice is used.*

Proof. Assume $\mathrm{MP}_{\mathrm{bin}}$, and let $0 \le a \le b$ with $b \neq 0$. Then b is apart from 0 and not negative, so $b > 0$, and its inverse d is positive, as in the proof of Lemma 14.5. Take a natural number $K \ge 1$ with $d \le K$. Given k, put $j + 1 = 2K(k+1)$. If $bx - a$ and $by - a$ are within $1/(j+1)$ of 0, then $|bx - by| \le 2/(j+1)$, and $|x - y| = d\,|bx - by| \le 2K/(j+1) = 1/(k+1)$.

Conversely, let $x \neq 0$, and apply the assertion to $a = 0$ and $b = |x|$, which satisfy $0 \le 0 \le |x|$ and $|x| \neq 0$. Take j for $k = 1$. The points 0 and 1 of $[0,1]$ are not within $1/2$ of each other. Cotransitivity of $1/(2j+2) < 1/(j+1)$ at $|x|$ gives $|x| > 1/(2j+2)$ or $|x| < 1/(j+1)$. In the second case $|x| \cdot 1$ and $|x| \cdot 0$ are both within $1/(j+1)$ of 0, so 1 and 0 would be within $1/2$; this is impossible. So $|x| > 0$, and x is apart from 0. Every nonzero point is therefore apart from 0, which is $\mathrm{MP}_{\mathrm{bin}}$ by Corollary 12.18. With WLPO, Proposition 12.9 gives LPO. □

The last row priced $\mathrm{MP}_{\mathrm{bin}}$ concerns maps of the whole line. A map f of $\mathbb{R}_\delta$ to itself is *continuous* when it is continuous at every point: for every x and k some j has $f(y)$ within $1/(k+1)$ of $f(x)$ whenever y is within $1/(j+1)$ of x. It is *injective* when $f(x) = f(y)$ gives $x = y$, and *strictly increasing*, respectively *strictly decreasing*, when $x < y$ gives $f(x) < f(y)$, respectively $f(y) < f(x)$.

Lemma G.8 (A barrier for nowhere-zero functions)**.** *Assume* $\mathrm{MP}_{\mathrm{bin}}$. *No function* g *continuous at every point of* $[0,1]$ *has* $g(0) \le 0 \le g(1)$ *and* $g(t) \neq 0$ *for every* t *in* $[0,1]$. *No choice is used.*

Proof. Suppose g is such a function. For each pair (n, j) with $j \le 2^n$ the point $y = g(j/2^n)$ is not zero, so $\mathrm{MP}_{\mathrm{bin}}$ makes it apart from 0 (Corollary 12.18). Exactly one of $y > 0$ and $y < 0$ then holds, and which one is computed from the point itself, as in the proof of Theorem 14.30: search any regular sequence naming y for its first stage above $2/(N+1)$ or below $-2/(N+1)$. The search is proved to terminate, because y is apart from 0, and two sequences naming y give the same answer, since a point is not both positive and negative. Put $B(n,j) = 1$ when $y < 0$ and $B(n,j) = 0$ when $y > 0$, and $B(n,j) = 0$ for the pairs with $j > 2^n$. This table is computed from g with no selection: $B(n,j) = 1$ gives $g(j/2^n) \le 0$, and $B(n,j) = 0$ gives $g(j/2^n) \ge 0$. The second part of the proof of Theorem 12.54, which uses continuity only at the limit point it constructs, turns this table into a root t of g in $[0,1]$, contrary to $g(t) \neq 0$. □

Theorem G.9 (Continuous injective maps are monotone at the price $\mathrm{MP}_{\mathrm{bin}}$)**.** *Consider the assertion that every continuous injective map of $\mathbb{R}_\delta$ to itself is strictly increasing or strictly decreasing.*

1. *The assertion implies $\mathrm{MP}_{\mathrm{bin}}$. No choice is used.*
2. *$\mathrm{MP}_{\mathrm{bin}}$ implies the assertion. No choice is used.*

Proof. For the first statement, let $a \neq 0$. The map $x \mapsto ax$ is continuous, since $|ax - ay| \leq K|x - y|$ for a natural number K with $|a| \leq K$, and it is injective by Lemma 14.4. If it is strictly increasing, then $a \cdot 0 < a \cdot 1$ and $a > 0$; if strictly decreasing, $a < 0$. So every nonzero point is apart from 0, which is $\mathrm{MP}_{\mathrm{bin}}$ by Corollary 12.18.

For the second, assume $\mathrm{MP}_{\mathrm{bin}}$, and let f be continuous and injective. Then $f(0) \neq f(1)$, and the seventh row of Table 12.2 gives $f(0) < f(1)$ or $f(1) < f(0)$. In the second case $-f$ is continuous and injective with $-f(0) < -f(1)$, and f is strictly decreasing if $-f$ is strictly increasing. So we may assume $f(0) < f(1)$ and show that f is strictly increasing.

Let $x < y$. Then $f(x) \neq f(y)$, so $f(x) < f(y)$ or $f(y) < f(x)$; we refute the second. Suppose $f(y) < f(x)$, and put

$$g(t) = f(xt) - f\big(1 + (y-1)t\big).$$

Then $g(0) = f(0) - f(1) \leq 0$ and $g(1) = f(x) - f(y) \geq 0$. The two arguments of f stay apart on $[0,1]$. By Lemma 12.52, item 1, there is a ratio ε with $0 < \varepsilon \leq 1$ and $\varepsilon < y - x$. For t in $[0,1]$,

$$1 + (y-1)t - xt = (1-t) + t(y-x) \geq (1-t)\varepsilon + t\varepsilon = \varepsilon > 0,$$

since $1 - t \geq 0$ and $t \geq 0$. So $f(xt) \neq f(1 + (y-1)t)$ by injectivity, that is $g(t) \neq 0$. The function g is continuous at every point of $[0,1]$: the maps $t \mapsto xt$ and $t \mapsto 1 + (y-1)t$ multiply distances by at most a natural number K bounding $|x|$ and $|y-1|$, so continuity of f at xt and at $1 + (y-1)t$, each for the tolerance $1/(2k+2)$, and the triangle inequality give a modulus for g at t. Lemma G.8 now gives a contradiction. Hence $f(x) < f(y)$. □

G.3 Without unique choice

Theorem G.10 ($\mathrm{MP}^\vee_{\mathrm{seq}}$ is $\mathrm{MP}^\vee$ with unique choice for Π^0_1 relations)**.** *The sequence form of $\mathrm{MP}^\vee$ is $\mathrm{MP}^\vee$ together with a choice principle.*

1. *$\mathrm{MP}^\vee_{\mathrm{seq}}$ is equivalent to $\mathrm{MP}^\vee$ together with unique choice of bits for Π^0_1 relations.*
2. *$\mathrm{MP}_{\mathrm{bin}}$ implies $\mathrm{MP}^\vee_{\mathrm{seq}}$, and hence unique choice of bits for Π^0_1 relations.*

No choice is used.

Proof. From $\mathrm{MP}^\vee_{\mathrm{seq}}$ to $\mathrm{MP}^\vee$. Apply $\mathrm{MP}^\vee_{\mathrm{seq}}$ to the matrix all of whose rows are the given sequence.

From $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ *to unique choice.* Let $r(i,b,n)$ be such that exactly one bit passes each row, in the sense of Definition 14.20. Call n a *failure* of row i when $r(i,0,n) = 0$ or $r(i,1,n) = 0$. Build a first-failure row F_i: $F_i(2n) = 1$ when n is the least failure of row i and $r(i,0,n) = 0$, and $F_i(2n+1) = 1$ when n is the least failure, $r(i,0,n) = 1$ and $r(i,1,n) = 0$; all other entries are 0, and each is a bounded search. A row has at most one least failure, so F_i has at most one entry 1. It is not all zero: otherwise, by induction on n, no n would be a failure, both bits would pass row i, and uniqueness would fail. So $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ gives a binary sequence f.

Suppose $f(i) = 1$, so the even entries of F_i vanish. We show that the bit 0 passes row i. Fix n; the value $r(i,0,n)$ is decided, so suppose it is 0. Then n is a failure, and a bounded search finds the least failure $j \le n$. Since $F_i(2j) = 0$, $r(i,0,j) = 1$, and so $r(i,1,j) = 0$: the bit 1 does not pass. The bit that passes is therefore 0, and $r(i,0,n) = 1$, which is a contradiction. So $r(i,0,n) = 1$ for every n. If $f(i) = 0$, the odd entries of F_i vanish, and in the same way the least failure j below any n with $r(i,1,n) = 0$ has $r(i,0,j) = 0$, so the bit 1 passes row i. The sequence $1 - f$ chooses the passing bit of every row. *The converse* is Lemma 14.21.

From $\mathrm{MP}_{\mathrm{bin}}$. Let $(\alpha_i(n))$ be a matrix as in Definition 14.18. Each row is not all zero, so $\mathrm{MP}_{\mathrm{bin}}$ gives an index with $\alpha_i(n) = 1$, and the least such index n_i is found by a search that is proved to terminate. It is the only entry 1 of the row. Put $f(i) = 1$ when n_i is odd and $f(i) = 0$ when it is even; then the entries of the other parity vanish. Unique choice of bits for Π^0_1 relations follows from the first statement. □

A failure index would name the surviving bit directly. Searching for one costs exactly $\mathrm{MP}_{\mathrm{bin}}$, Markov's principle for binary sequences.

Theorem G.11 (Searching for a failed test costs $\mathrm{MP}_{\mathrm{bin}}$)**.** *Let $r(i,b,n)$ be such that exactly one bit passes each row. Call n a failure of row i when $r(i,0,n) = 0$ or $r(i,1,n) = 0$.*

1. *For every i it is not the case that row i has no failure.*
2. *The assertion that, for every such r, every row has a failure is equivalent to* $\mathrm{MP}_{\mathrm{bin}}$.
3. *That assertion implies unique choice of bits for Π^0_1 relations.*
4. *WLPO implies that assertion if and only if WLPO implies* $\mathrm{MP}_{\mathrm{bin}}$.

No choice is used.

Proof. If row i had no failure, both bits would pass it, and uniqueness would fail. This is the first statement.

For the second, assume $\mathrm{MP}_{\mathrm{bin}}$. For row i let $e(n) = 1$ when n is a failure. By the first statement e is not all zero, and $\mathrm{MP}_{\mathrm{bin}}$ gives a failure. Conversely, assume the assertion, and let α be a binary sequence that is not all zero. Put $r(i,0,n) = 1$ and $r(i,1,n) = 1 - \alpha(n)$ for all i and n. The bit 0 passes every row, and the bit 1 passes a row only if α is all zero, which is excluded; so exactly one bit passes each row. The assertion gives a failure n of row 0, which cannot be one of the bit 0; so $r(0,1,n) = 0$, that is $\alpha(n) = 1$.

For the third, let every row have a failure. The least failure n_i of row i is found by a search that is proved to terminate. Put $f(i) = 1$ when $r(i,0,n_i) = 0$, and $f(i) = 0$ otherwise, in which

case $r(i, 1, n_i) = 0$. In either case the bit $1 - f(i)$ fails row i at n_i, so the passing bit is $f(i)$. The fourth statement follows from the second. □

WLPO together with $\mathrm{MP}_{\mathrm{bin}}$ is LPO (Proposition 12.9), which intuitionistic second-order arithmetic does not derive from WLPO (Theorem 7.38); so a search for failed tests is not a route from WLPO to unique choice.

Theorem G.12 (What remains without unique choice)**.** *The following hold with no choice.*

1. *Scalar pair division is equivalent to WLPO together with affine trisection control. The same holds for basis selection for every finite family, and for rank–nullity, in place of scalar pair division.*
2. *WLPO implies scalar pair division if and only if WLPO implies affine trisection control.*
3. *Scalar pair division implies unique choice of bits for Π^0_1 relations if and only if it implies absolute-quotient collection for every sequence of nonzero points given by regular sequences. These implications hold exactly when scalar pair division, basis selection for every finite family and rank–nullity are each equivalent to WLPO together with unique choice of bits for Π^0_1 relations, with no choice in either direction.*
4. *WLPO implies unique choice of bits for Π^0_1 relations if and only if both the implication of the second statement and those of the third hold. In that case scalar pair division, basis selection for every finite family and rank–nullity are each equivalent to WLPO, with no choice in either direction.*

Proof. The first statement. Theorem 14.15 makes scalar pair division WLPO together with dominated division by nonzero points, and Theorem 14.27 makes the latter affine trisection control. Theorems 14.13 and 14.14 give the other two assertions.

The second. Under WLPO, the first statement makes scalar pair division equivalent to affine trisection control.

The third. Under scalar pair division WLPO holds, so LLPO and $\mathrm{MP}^{\vee}$ hold (Theorem 14.19). Given $\mathrm{MP}^{\vee}$, unique choice of bits for Π^0_1 relations is equivalent to $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ (Lemma 14.21 and Theorem G.10), and $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ is equivalent to absolute-quotient collection (Theorem 14.30). This gives the first sentence. For the second, Corollary 14.29 proves each of the three assertions from WLPO and that principle, and each assertion gives scalar pair division (Theorems 14.13 and 14.14), hence WLPO. So each assertion is equivalent to WLPO with that principle exactly when it implies the principle, and by the same two theorems this is the case for one of them exactly when it is for scalar pair division.

The fourth. If WLPO implies unique choice of bits for Π^0_1 relations, then WLPO implies scalar pair division by Corollary 14.29, and scalar pair division implies that choice, since it gives WLPO. Conversely, assume the implication of the second statement and those of the third, and assume WLPO. Then scalar pair division holds, and so does unique choice of bits for Π^0_1 relations. In that case WLPO implies each of the three assertions, and each of them implies WLPO by Theorem 14.16, with no choice in either direction. □

Appendix H

Proofs for certified computation

This appendix proves Proposition 15.26, Theorem 15.29 and Corollary 15.30. Notation is that of Chapter 15. Section H.1 is a bookkeeping of integers. Sections H.2 and H.3 are classical real analysis: they use the real numbers, their completeness and right derivatives, as the statements they prove do.

H.1 The size of every operand

A ratio is stored as a pair (u, e) of an integer and a positive integer, without reduction. Say that it has *width* b if $|u| \le 2^b$ and $e \le 2^b$. The ratio then has at most $2b + 3$ binary digits in all, counting a sign: $\mathrm{bl}(|u|) \le b + 1$ and $\mathrm{bl}(e) \le b + 1$. Every stored ratio $q = (u, e)$ has the width $\mathrm{cap}(q) = \mathrm{bl}(|u|) + \mathrm{bl}(e)$.

Lemma H.1 (Width rules)**.** *Let q and r have widths a and b.*

(a) A width b is also a width b' for every $b' \ge b$.

(b) The product qr has width $a + b$, and the sum $q + r$, with numerator $u_q e_r + u_r e_q$ over $e_q e_r$, has width $a + b + 1$.

(c) For a natural number $k \ge 1$, q/k has width $a + \mathrm{bl}(k)$, and 2^{-e} has width e.

(d) If $|q| \le V \le 2^h$ and the denominator of q is at most 2^w, then q has width $h + w$.

(e) A sum of L ratios of width b, formed term by term from 0, has width $(b + 1)L$, and so has each of its partial sums.

Proof. (a) is immediate. (b) $|u_q u_r| \le 2^{a+b}$ and $e_q e_r \le 2^{a+b}$; and $|u_q e_r + u_r e_q| \le 2 \cdot 2^{a+b}$. (c) The numerator is unchanged and the denominator is multiplied by $k < 2^{\mathrm{bl}(k)}$; and 2^{-e} is $(1, 2^e)$. (d) $|u| = |q|\, e \le 2^h 2^w$. (e) Induction on L with (b): the sum of a ratio of width b and a partial sum of width $(b + 1)(L - 1)$ has width $(b + 1)L$; the empty sum 0 has width 0. □

Let the problem have a coefficient field. Put

$$s = \max_i \Big(\mathrm{cap}(c_i) + \sum_{(a,j)} \mathrm{cap}(a) + \sum_{(b,j,k)} \mathrm{cap}(b) \Big), \qquad L = \max_i \big(\#A\text{-terms} + \#B\text{-terms} \big),$$

the sums and counts running over the terms of coordinate i, and $s_0 = \max_i \mathrm{cap}(y_{0,i})$. Then every c_i and every coefficient of the field has width s, every coordinate has at most L terms of each kind, and every $y_{0,i}$ has width s_0. Let $r = \mathrm{bl}(M) + \mathrm{bl}(R) + 1$.

Lemma H.2 (Stored coefficients)**.** *Let $\|z\| \le M$, and let every coordinate of z have denominator exactly 2^w. In the table at z with rounding $[\cdot]_w$, every coefficient a_k has denominator 2^w and width $\mathrm{bl}(M) + \mathrm{bl}(R)\,k + w$. For $w = N$ and $k \le N$ this is at most $\rho(N) = r(N+1)$. If the region certificate holds at level n_0 and $n \ge n_0$, every row stored during the S steps at level n has width $\rho(N(n))$.*

Proof. The rounding $[\cdot]_w$ produces the denominator 2^w, and $a_0 = z$ has it by hypothesis. By Lemma 15.14, $\|a_k\| \le MR^k \le 2^{\mathrm{bl}(M)}2^{\mathrm{bl}(R)k}$, and Lemma H.1(d) gives the width. Expanding $r(N+1)$ shows that it is at least $\mathrm{bl}(M) + \mathrm{bl}(R)k + N$ when $k \le N$. At level n, each step starts from $y_k^{(n)}$, which has denominator $2^{N(n)}$ and, by Theorem 15.20, norm at most M. □

Lemma H.3 (New coefficients before rounding)**.** *If $f_0, \dots, f_k$ have coordinates of width b, then* $\mathrm{next}(f, k)$*, for the coefficient field, has coordinates of width*

$$v(b,k) = s + (s+b+1)L + 1 + (k+1)\big((s+2b+1)L+1\big) + 1 + \mathrm{bl}(k+1),$$

and so do all partial sums formed while computing it.

Proof. By Lemma H.1: c_i has width s; each term $a\,u_j$ of $(Af_k)_i$ has width $s+b$, and the at most L terms sum to width $(s+b+1)L$; adding c_i gives $s + (s+b+1)L + 1$. Each term $b\,u_j v_k$ of a coordinate of $B(f_i, f_{k-i})$ has width $s+2b$, so the coordinate has width $(s+2b+1)L$; the Cauchy sum of $k+1$ such vectors has width $(k+1)\big((s+2b+1)L+1\big)$. One more addition and the division by $k+1$ give $v(b,k)$. □

Lemma H.4 (Taylor sums before rounding)**.** *If $f_0, \dots, f_N$ have coordinates of width b, then $S_N(f)$ and all its partial sums have coordinates of width $b + (\sigma N + b + 1)N$.*

Proof. Let $\theta(0) = b$ and $\theta(j+1) = \theta(j) + \sigma(j+1) + b + 1$: the term $2^{-\sigma(j+1)} f_{j+1}$ has width $\sigma(j+1) + b$, and adding it costs one more. So $S_j(f)$ has width $\theta(j)$, and induction on $j \le N$ gives $\theta(j) \le b + (\sigma N + b + 1)j$. □

With $b = \rho(N)$ and $k \le N$, $\mathrm{bl}(k+1) \le N+1$, so $v(\rho(N), k) \le v^*(N)$, where $v^*(N)$ is $v(\rho(N), N)$ with its last summand replaced by $N+1$. Define the *capacity*

$$\mathscr{B}(N) = s_0 + s + \rho(N) + v^*(N) + \big(\rho(N) + (\sigma N + \rho(N) + 1)N\big) + N + 3. \tag{H.1}$$

Expanding, $\mathscr{B}(N) = q_P N^2 + l_P N + c_P$ with

$$q_P = 2rL + \sigma + r, \qquad l_P = 3r + (s + 5r + 1)L + 4, \qquad c_P = s_0 + 2s + 2r + (2s + 3r + 2)L + 7.$$

Proof of Proposition 15.26. The computation of a_m at level $n = \max(m, n_0)$, with $N = N(n)$, consists of the following operations, and in each case the operands and results have width at most one of the summands of (H.1), hence at most $\mathscr{B}(N)$ by Lemma H.1(a).

(i) The rounding of y_0: operands of width s_0, results of width $\rho(N)$ by Lemma H.2.

(ii) In each step, the products, sums and divisions that form next of the stored rows: field coefficients of width s, stored rows of width $\rho(N)$ (Lemma H.2, which is where the region certificate is used), and every intermediate value of width at most $v^*(N)$ by Lemma H.3.

(iii) The rounding of each new row: operand of width $v^*(N)$, result of width $\rho(N)$.

(iv) The scalings and additions of the Taylor sum: width at most $\rho(N) + (\sigma N + \rho(N) + 1)N$ by Lemma H.4; its rounding gives the next stored vector, of width $\rho(N)$.

(v) The final rounding to the integer a_m, from a stored value of width $\rho(N)$.

A rounding $[q]_w$ of $q = (u, e)$ of width b is the floor division of $2u2^w + e$ by $2e$, whose integers have absolute value at most 2^{b+w+2} and 2^{b+1}; here $w \le N$, and the last summand $N + 3$ of (H.1) covers them. Divisions are by $k + 1 \le N + 1$. The expansion of $\mathscr{B}$ follows from $\rho(N) = rN + r$ by collecting powers of N. Finally N^2, N and 1 are at most $(N + 1)^2$, and $N + 1 = \max(m, n_0) + \kappa + 1 \le m + n_0 + \kappa + 1$, so $\mathscr{B}(N) \le (q_P + l_P + c_P)(m + n_0 + \kappa + 1)^2$. The count of digits is the remark at the start of this section. $\square$

The quadratic growth comes from the sums formed without reduction: the Cauchy sums of next, which contribute $2rL$ to q_P, and the Taylor sum, which contributes $\sigma + r$. A sum formed without reduction multiplies the denominators of its terms. For the Taylor sum of stored rows these are $2^{\sigma j + N}$, $j \le N$, and their product is a power of two with exponent $(N + 1)(\sigma N/2 + N)$. So this bookkeeping gives a capacity quadratic, not linear, in N.

The answer itself needs far fewer digits than the capacity allows. The width $W_m = 4\mathscr{B}(N) + 4$ of the answer word in Definition 15.32 is therefore always enough.

Lemma H.5 (The answer fits its width)**.** *Let a problem have a coefficient field and satisfy the region certificate at level n_0. Then the answer a_m of Theorem 15.22 satisfies $|a_m| < 2^{W_m + 1}$.*

Proof. Let $z = y_S^{(\max(m, n_0))}$. By Theorem 15.20, $|z_i| \le M$, and $M \le 2^{\mathrm{bl}(M)} - 1$. So

$$|a_m| \le 2^m |z_i| + \tfrac{1}{2} \le 2^{m + \mathrm{bl}(M)} - 2^m + \tfrac{1}{2} < 2^{m + \mathrm{bl}(M)}.$$

Since $N = \max(m, n_0) + \kappa \ge m$, $m + \mathrm{bl}(M) \le \mathrm{bl}(M) + N + 1 \le \rho(N) \le \mathscr{B}(N) \le W_m + 1$. $\square$

H.2 The classical solution value

Throughout this section $\mathbb{R}^d$ carries the sup norm, A and B are the real linear and bilinear maps of Theorem 15.29, and $f(x) = c + Ax + B(x, x)$.

Lemma H.6 (The field on the doubled ball)**.** *On the ball of radius $2M$, f is Lipschitz with constant $L = \alpha + 4\beta M$.*

Proof. By linearity and bilinearity, $f(x) - f(x') = A(x - x') + B(x - x', x) + B(x', x - x')$, so $\|f(x) - f(x')\| \le \big(\alpha + \beta\|x\| + \beta\|x'\|\big)\|x - x'\| \le (\alpha + 4\beta M)\|x - x'\|$. $\square$

Lemma H.7 (The Taylor polynomial nearly solves the equation). *Let z be a rational vector with $\|z\| \le M$, let a_j be the exact coefficients at z, let $N \ge 1$, and put $\Pi(\tau) = \sum_{j=0}^{N} \tau^j a_j$. Then*

$$\Pi'(\tau) - f(\Pi(\tau)) = -\tau^N A a_N - \sum_{\substack{i,l \le N \\ i+l \ge N}} \tau^{i+l} B(a_i, a_l).$$

If $0 \le \tau$ and $R\tau \le 1/8$, then $\|\Pi'(\tau) - f(\Pi(\tau))\| \le \big(\alpha M + \beta M^2 (N+1)^2\big) 8^{-N}$ and $\|\Pi(\tau)\| \le 2M$.

Proof. The real maps agree with A and B on rational vectors, so the exact coefficients satisfy $(j+1)a_{j+1} = c\,[j=0] + A a_j + \sum_{i \le j} B(a_i, a_{j-i})$ in $\mathbb{R}^d$. Hence

$$\Pi'(\tau) = \sum_{j<N} (j+1)\tau^j a_{j+1} = c + \sum_{j<N} \tau^j A a_j + \sum_{i+l<N} \tau^{i+l} B(a_i, a_l),$$

the last sum collecting the Cauchy sums by antidiagonals, while $f(\Pi(\tau)) = c + \sum_{j \le N} \tau^j A a_j + \sum_{i,l \le N} \tau^{i+l} B(a_i, a_l)$ by linearity and bilinearity. Subtracting gives the identity. By Lemma 15.14, $\|a_j\| \le M R^j$. So $\|\tau^N A a_N\| \le \alpha M (R\tau)^N \le \alpha M 8^{-N}$, and each of the at most $(N+1)^2$ terms of the double sum is at most $\beta M^2 (R\tau)^{i+l} \le \beta M^2\, 8^{-N}$, because $i + l \ge N$ and $R\tau \le 1/8$. Finally $\|\Pi(\tau)\| \le M \sum_j (R\tau)^j \le M \sum_j 8^{-j} \le 2M$. □

The next two lemmas are the classical Gronwall inequality, first for a single function with a one-sided growth bound, then for two approximate solutions with right derivatives in a region where the field is Lipschitz.

Lemma H.8 (A one-sided comparison). *Let $\delta : [a, a+h] \to \mathbb{R}$ be continuous, with $h > 0$, and let $L, \varepsilon \ge 0$. Suppose that for every $t \in [a, a+h)$ and every $\lambda > L\delta(t) + \varepsilon$ there is $\zeta > 0$ with $\delta(s) - \delta(t) \le \lambda (s - t)$ for all $s \in [t, t+\zeta)$. Then $\delta(t) \le e^{L(t-a)}\big(\delta(a) + \varepsilon(t-a)\big)$ for every $t \in [a, a+h]$.*

Proof. Fix $\eta > 0$ and put

$$\psi(t) = e^{L(t-a)}\big(\delta(a) + \eta + (\varepsilon + \eta)(t-a)\big).$$

Then

$$\psi'(t) = L\psi(t) + e^{L(t-a)}(\varepsilon + \eta) \ge L\psi(t) + \varepsilon + \eta.$$

We show $\delta \le \psi$ on $[a, a+h]$. Otherwise the set U of points where $\delta > \psi$ is nonempty and open in $[a, a+h]$; let t^* be its infimum. The point a is not in U, since $\delta(a) < \psi(a)$; as U is open, $t^* \notin U$, so $\delta(t^*) \le \psi(t^*)$. Points of U approach t^*, necessarily from the right, so $\delta(t^*) \ge \psi(t^*)$ by continuity. Hence $\delta(t^*) = \psi(t^*)$ and $t^* < a + h$. Take $\lambda = L\delta(t^*) + \varepsilon + \eta/2$, which is at most $\psi'(t^*) - \eta/2$. By hypothesis $\delta(s) - \delta(t^*) \le \lambda (s - t^*)$ for s slightly larger than t^*, while $\psi(s) - \psi(t^*) > \lambda (s - t^*)$ for such s, since $\psi'(t^*) > \lambda$. So $\delta(s) < \psi(s)$ on an interval $(t^*, t^* + \zeta')$, which contains points of U: a contradiction. So $\delta \le \psi$ for every $\eta > 0$, and letting η tend to 0 gives the claim. □

The infimum of a set of real numbers and the passage to the limit are where the completeness of the classical real numbers enters.

Lemma H.9 (Gronwall for approximate solutions)**.** *Let f be Lipschitz with constant $L \geq 0$ on a set $\Omega \subseteq \mathbb{R}^d$, and let $h > 0$. Let $u, v : [a, a+h] \to \mathbb{R}^d$ be continuous, with $u(t), v(t) \in \Omega$ and right derivatives $u'(t)$, $v'(t)$ at every $t \in [a, a+h)$ satisfying $\|u'(t) - f(u(t))\| \leq \varepsilon_u$ and $\|v'(t) - f(v(t))\| \leq \varepsilon_v$. Then, with $\varepsilon = \varepsilon_u + \varepsilon_v$,*

$$\|u(a+h) - v(a+h)\| \leq e^{Lh}\big(\|u(a) - v(a)\| + \varepsilon h\big).$$

Proof. Let $\delta(t) = \|u(t) - v(t)\|$, a continuous function. At $t \in [a, a+h)$ the right derivatives give $\delta(s) - \delta(t) \leq \|(u(s) - v(s)) - (u(t) - v(t))\| = (s-t)\big(\|u'(t) - v'(t)\| + o(1)\big)$ as $s \downarrow t$, and $\|u'(t) - v'(t)\| \leq \|f(u(t)) - f(v(t))\| + \varepsilon \leq L\delta(t) + \varepsilon$. So the hypothesis of Lemma H.8 holds, and its conclusion at $t = a + h$ is the claim. □

Lemma H.10 (One step against the true solution)**.** *Under the hypotheses of Theorem 15.29, let $n \geq n_0$, $h = 2^{-\sigma}$, $k < S$ and $e_k = \|y(kh) - y_k^{(n)}\|$. Then*

$$e_{k+1} \leq e^{Lh}\big(e_k + (\alpha M + \beta M^2)2^{-n}h\big) + (M+2)2^{-n}.$$

Proof. Let $z = y_k^{(n)}$ and $N = N(n)$. By Theorem 15.20, $\|z\| \leq M$. Since $8R \leq 2^{\sigma}$, $Rh \leq 1/8$. On $[kh, (k+1)h]$ compare $u = y$, which solves the equation exactly and stays in the ball of radius M, with $v(t) = \Pi(t - kh)$, which by Lemma H.7 misses the equation by at most $\varepsilon_N = (\alpha M + \beta M^2(N+1)^2)8^{-N}$ and stays in the ball of radius $2M$. Lemmas H.6 and H.9 give $\|y((k+1)h) - \Pi(h)\| \leq e^{Lh}(e_k + \varepsilon_N h)$, because $\Pi(0) = z$. Since $(N+1)^2 \leq 4^N$ and $N \geq n$, $\varepsilon_N \leq (\alpha M + \beta M^2)2^{-N} \leq (\alpha M + \beta M^2)2^{-n}$. Next, $\Pi(h) = S_N(a(z)) = \Phi_N(z)$, and by Lemma 15.17(b) with $N_0 = N$ and the estimate in the proof of Lemma 15.19, $\|\Phi_N(z) - y_{k+1}^{(n)}\| \leq (M+2)2^{-N} \leq (M+2)2^{-n}$. Add the two bounds. □

Proof of Theorem 15.29. Put $\delta' = (\alpha M + \beta M^2)2^{-n}h + (M+2)2^{-n}$. We show $e_k \leq e^{Lkh}(2^{-n} + k\delta')$ for $k \leq S$. At $k = 0$, $e_0 = \|y_0 - [y_0]_{N(n)}\| \leq 2^{-(N(n)+1)} \leq 2^{-n}$. If the bound holds at k, Lemma H.10 and $1 \leq e^{Lh}$ give $e_{k+1} \leq e^{Lh}(e_k + \delta') \leq e^{L(k+1)h}(2^{-n} + (k+1)\delta')$. At $k = S$ we have $Sh = T$, so

$$e_S \leq e^{LT}\Big(1 + S\big((\alpha M + \beta M^2)2^{-\sigma} + M + 2\big)\Big)2^{-n} = C_P\, 2^{-n}.$$

For the answers, fix m and i. For every $M' \geq m$, Theorem 15.22 and the bound just proved, at the level $\max(M', n_0) \geq n_0$, give

$$\big|a_m/2^m - y_i(T)\big| \leq 2^{-m} + C_P\, 2^{-M'}.$$

Letting M' grow gives $|a_m/2^m - y_i(T)| \leq 2^{-m}$. □

For the oscillator, the realization is $A(x_0, x_1) = (x_1, -x_0)$, with $\|Ax\| \leq \|x\|$, and $B = 0$; the solution $(\sin t, \cos t)$ has the derivative $(\cos t, -\sin t) = Ay(t)$ and norm at most $1 \leq M = 2$. With $\alpha = 1$, $\beta = 0$, $S = 32$ and $\sigma = 5$ the constant is $C_P = e\big(1 + 32(2 \cdot 2^{-5} + 4)\big) = 131e$, so the trajectory at level n ends within $131e \cdot 2^{-n} < 2^{9-n}$ of $(\sin 1, \cos 1)$. The constant does not enter the answers: the last step of the proof removes it, because the levels agree among themselves to within $2^{-(m+2)}$ by Theorem 15.21.

H.3 Polynomial fields

Proof of Corollary 15.30. The quadratized field is a coefficient field, and its sums define a real linear map (here 0) and a real bilinear map B on $\mathbb{R}^{d_Q}$ that agree with the rational maps on rational vectors; the triangle inequality over the terms, as in the proof of Proposition 15.25, gives $\|B(w,w')\| \le \beta\|w\|\,\|w'\|$. For $x \in \mathbb{R}^d$ let $W(x) = (x^\alpha)_{|\alpha|\le D}$. Collecting the terms of the variable w_α,

$$B\big(W(x),W(x)\big)_\alpha = \sum_{i\colon \alpha_i>0} \sum_{(q,\xi)\in p_i} \alpha_i q\, x^{\alpha-e_i}x^{\xi} = \sum_i \alpha_i x^{\alpha-e_i}p_i(x),$$

because $\alpha - e_i$ and ξ have degree at most D and so are among the variables (Theorem 15.28(a)). By the product rule for right derivatives, $t \mapsto x(t)^\alpha$ has the right derivative $\sum_i \alpha_i x(t)^{\alpha-e_i}x_i'(t) = \sum_i \alpha_i x(t)^{\alpha-e_i}p_i(x(t))$. Hence $w(t) = W(x(t))$ is continuous on $[0,T]$, starts at the monomials of x_0, has the right derivative $B(w(t),w(t))$ on $[0,T)$, and stays in the ball of radius M by hypothesis. Theorem 15.29 for the quadratized problem gives $|a_m/2^m - w_{e_i}(T)| \le 2^{-m}$, and $w_{e_i}(T) = x_i(T)$. For the last assertion, if $\|x(t)\| \le r$ with $r \ge 1$, then $|x(t)^\alpha| \le r^{|\alpha|} \le r^D \le M$ for $|\alpha| \le D$. □

The quadratized system has no other solutions with these initial values. The proof is again classical; it uses that a continuous function on a closed interval is bounded.

Proposition H.11 (Quadratized solutions are monomial vectors)**.** *Let $w : [0,T] \to \mathbb{R}^{d_Q}$ be continuous, with $w(0) = W(x_0)$ and right derivative $B(w(t),w(t))$ at every $t \in [0,T)$, for the quadratized field of p. Let $x(t) = (w_{e_i}(t))_{i<d}$. Then $w(t) = W(x(t))$ for every $t \in [0,T]$, and x is continuous, $x(0) = x_0$, and x has the right derivative $p(x(t))$ at every $t \in [0,T)$.*

Proof. Let $\pi(t) = W(x(t))$ and $E = w - \pi$. Both w and π are continuous on $[0,T]$, so they are bounded there, say $\|w\| \le C_1$ and $\|\pi\| \le C_2$. Write $f(w) = B(w,w)$. By the product rule, π_α has the right derivative $\sum_i \alpha_i x^{\alpha-e_i} f(w)_{e_i}$. At a monomial vector the identity in the proof of Corollary 15.30 reads $f(\pi)_\alpha = \sum_i \alpha_i x^{\alpha-e_i}p_i(x) = \sum_i \alpha_i x^{\alpha-e_i} f(\pi)_{e_i}$, because $f(\pi)_{e_i} = p_i(x)$. So E has the right derivative

$$E'_\alpha = \big(f(w) - f(\pi)\big)_\alpha - \sum_i \alpha_i x^{\alpha-e_i}\big(f(w) - f(\pi)\big)_{e_i}.$$

Each $x^{\alpha-e_i}$ is a coordinate of π, so $|x^{\alpha-e_i}| \le C_2$, and $\sum_i \alpha_i \le D$. Bilinearity gives $f(w) - f(\pi) = B(w-\pi,w) + B(\pi,w-\pi)$, so $\|f(w) - f(\pi)\| \le \beta(C_1+C_2)\|E\|$. Hence $\|E'\| \le K\|E\|$ with $K = (1 + DC_2)\beta(C_1+C_2)$. Since $E(0) = W(x_0) - W(x_0) = 0$, Lemma H.8 (for $T > 0$; for $T = 0$ there is nothing to prove) with $\delta = \|E\|$, $L = K$ and $\varepsilon = 0$ gives $E = 0$ on $[0,T]$. So $w = \pi$, and x, made of coordinates of w, is continuous with $x(0) = x_0$ and right derivative $f(w)_{e_i} = f(\pi)_{e_i} = p_i(x)$. □

With Corollary 15.30 this shows that the quadratization neither loses nor adds solutions: on $[0,T]$, solutions of $x' = p(x)$ from x_0 and solutions of the quadratized system from the monomials of x_0 correspond through $x \mapsto W(x)$ and $w \mapsto (w_{e_i})_i$.

Appendix I

The machine of Theorem 15.33

How many steps does it take a one-tape binary machine to add one to a number it keeps on its tape? For the machine of Theorem 15.33 the answer is exact. Suppose a counter holding 3 is kept on track 5 of the scratch area, and the head rests on the cell just left of the input. The program inc_5 of Lemma I.14 walks six cells left and then along track 5 until it meets the counter's first blank cell (62 units), writes a mark there, steps one row further and writes a blank (19 units), and walks back along the origin track to where it started (81 units): $162 = 36 \cdot 3 + 2 \cdot 5 + 44$ units in all, and the counter now holds 4. Every count in this appendix is of this kind: a program, the cells it reads and writes, and the exact or bounded number of units it takes.

Theorem 15.33 says that one control table H and one exponent λ serve every certified problem. Section 15.9 proves it from three lemmas and the stages of Table 15.1. This appendix constructs every stage and proves its bound, so that the proof of the theorem is complete.

The first six sections lay the foundation. They fix the layout of the tape (Section I.1), the language of programs and the rules that bound sequences and loops (Section I.2), and the small programs that every stage is built from: moving along the tape (Section I.3), counters and columns (Section I.4) and binary arithmetic (Section I.5). Section I.6 bounds every integer the machine stores.

The stages follow in the order in which they run; each section gives the programs, their contracts and their step bounds with fixed exponents. Section I.7 constructs the two stages that read the request, Parse and Decode. Sections I.8 and I.9 construct Schedule, which computes the problem's constants, the schedule and the width. Section I.10 constructs Initialize, which writes the first stored vector. Section I.11 constructs the coefficient table of a step, one rounded entry at a time, and Section I.12 its Taylor sum, which together make one stored step. Section I.13 repeats the step S times and selects the requested coordinate, Section I.14 rounds it to the precision m, and Section I.15 writes the answer word and erases the scratch area. Section I.16 composes the stages into the program of Theorem 15.33, computes its exponent λ and its constant K, and proves the theorem (Theorem I.209).

I.1 The tape: rows, tracks and marks

The machine has one tape of bits and no registers. Everything it remembers between instructions is written on the tape, and the input lies in the cells $0, 1, 2, \ldots$ until the machine has read it. The scratch area is the negative half, cut into rows of sixteen cells.

Definition I.1 (Cells, rows and tracks)**.** For a row $r \in \mathbb{N}$ and a track $k \in \{0, 1, \dots, 15\}$ put

$$\mathrm{cell}(r, k) = -2 - 16r - k.$$

Row r consists of the sixteen cells $\mathrm{cell}(r, 0), \dots, \mathrm{cell}(r, 15)$, that is, of $-2-16r$ down to $-17-16r$; track k consists of the cells $\mathrm{cell}(0, k), \mathrm{cell}(1, k), \dots$, sixteen apart. Cell -1 is the *root*. When no confusion is possible we write (r, k) for $\mathrm{cell}(r, k)$ and $\tau(r, k)$ for the bit a tape τ holds there.

Every cell $x \le -2$ is $\mathrm{cell}(r, k)$ for exactly one pair: write $-2 - x = 16r + k$ with $0 \le k < 16$. One step left from (r, k) reaches $(r, k+1)$ when $k < 15$; sixteen steps left reach $(r+1, k)$. A track is therefore a copy of the half-line $0, 1, 2, \dots$ read downwards with stride sixteen, and Lemma I.8 below runs ordinary one-tape programs on it.

Definition I.2 (Marks on tracks)**.** Let τ be a tape, k a track and $a, r, r_0, n \in \mathbb{N}$.

(a) Track k *carries the counter* a if $\tau(j, k) = 1$ for $j < a$ and $\tau(a, k) = 0$. The cell (a, k) is the counter's *hole*; nothing is said about the cells beyond it.

(b) Track k *holds the value* a if it carries the counter a and $\tau(j, k) = 0$ for every $j > a$.

(c) τ has *origin marks up to* r if $\tau(0, 1) = 0$ and $\tau(j, 1) = 1$ for $1 \le j \le r$.

(d) τ is a *workspace of* r *rows* if $\tau(-1) = 1$, $\tau(x) = 0$ for every $x \ge 0$, $\tau(j, 0) = 1$ for $j < r$, $\tau(r, 0) = 0$, and every cell of every row $j > r$ holds 0. Track 0 holds the row *headers* and row r is the *sentinel row*.

(e) Track 14 carries a *retained word of length* n *from row* r_0 if $\tau(r_0 + j, 14) = 1$ for $j < n$ and $\tau(r_0 + n, 14) = 0$.

A counter is a unary numeral that can be advanced by writing two cells. The same track may carry other marks beyond the hole; a value has none.

Definition I.3 (Streams, buffers and signed words)**.** Let $w = w_0 w_1 \cdots w_{\ell-1}$ be a word of bits.

(a) w lies *on the presence track* p *and the value track* v *from row* r_0 if $\tau(r_0 + j, p) = 1$ and $\tau(r_0 + j, v) = w_j$ for $j < \ell$ and $\tau(r_0 + \ell, p) = 0$. The *binary stream* is the word on tracks 6 and 7 from row 0; the *saved request* is the word on tracks 3 and 4 from row 0.

(b) The *output buffer holds* w *with capacity* R if track 11 carries the counter ℓ, $\tau(j, 12) = 1$ for $j < \ell$ and $\tau(j, 12) = 0$ for $\ell \le j \le R$, $\tau(j, 13) = w_j$ for $j < \ell$, τ has origin marks up to R, and $\ell \le R$.

(c) For an integer z and $W \in \mathbb{N}$, the *signed word* $\mathrm{sw}_W(z)$ is the bit $[z < 0]$ followed by the $W + 1$ lowest binary digits of $|z|$, least significant first: $W + 2$ bits. The integer z *fits* W if $|z| < 2^{W+1}$; then these digits are all the binary digits of $|z|$, and $\mathrm{sw}_W(z)$ determines z. For a list of integers, $\mathrm{sw}_W(z_1, \dots, z_d)$ is the concatenation of their signed words.

(d) For bits $b_0 \cdots b_{n-1}$ put $\mathrm{val}(b) = \sum_j b_j 2^j$ and $\mathrm{sval}_n(b) = \mathrm{val}(b) - 2^n b_{n-1}$ (the reading in two's complement); $\bar{b}$ is the word of complemented bits.

The answer word of Definition 15.32 is the frame of $\mathrm{sw}_{W_m}(a_m)$. Table I.1 lists what each track holds between stages; a stage that uses a track for scratch says so in its contract.

Tracks	Use
0	row headers of the workspace
1	origin marks: every walk back to the root follows this track
2	the step counter of the trajectory loop; scratch elsewhere
3, 4	the saved request: presence and bits of every frame's payload
5, 8, 9, 10	unary registers: after the schedule, σ, N, W_m and $\mathscr{B}(N)$ on 5, 8, 9 and 10 (with S on 6 and κ on 7); between steps, m, N, W_m and σ on 5, 8, 9 and 10
6, 7	the binary stream: the stored vector $\mathrm{sw}_{W_m}(Z_k)$, and at the end the answer word
11, 12, 13	the output buffer: the coefficient table of a step, and the arithmetic area of the coefficient stage
14	retained unary words, such as the step count S
15	a mark at the first row of every frame of the saved request after the first

Table I.1. The tracks of the scratch area and what they hold between stages.

Lemma I.4 (Where the marks are)**.** *Let τ be a workspace of r rows. Every cell holding* 1 *is the root or a cell of a row $j \le r$, and every counter carried by a track is at most $r + 1$.*

Proof. The first claim restates the definition: cells $x \ge 0$ and the cells of rows $j > r$ hold 0. If track k carried a counter $a \ge r + 2$, the cell $(r + 1, k)$ would hold 1. □

So a single walk over the rows marked on track 0 finds every cell the machine has written. The last stage does exactly that when it erases the scratch area.

I.2 Programs, counted runs and loops

Programs are the structured programs of Section 15.9: the primitives skip, $\mathsf{write}\,0$, $\mathsf{write}\,1$, left and right; sequences $P;Q$; branches $\mathsf{branch}(P_0, P_1)$, which run P_0 on a scanned 0 and P_1 on a scanned 1; and loops $\mathsf{loop}(P)$, which run P while the scanned bit is 1. Every primitive and every test costs one unit, including the final test of a loop. By Lemma 15.34, units are steps of the compiled control table.

A configuration is a pair (h, τ) of a head position and a tape. We write

$$(h, \tau) \xrightarrow[P]{t} (h', \tau')$$

when P has a counted execution in t units from (h, τ) to (h', τ'), and say that P runs from (h, τ) to (h', τ') *within b* units if it does so in some $t \le b$. Program names are set in sans serif, with tracks as subscripts.

Lemma I.5 (Moves and lists)**.** *Let left^n be n moves left followed by skip, and right^n likewise. For programs $P_1, \dots, P_k$ let $[P_1, \dots, P_k]$ be $P_1; (P_2; (\cdots; (P_k; \mathsf{skip}) \cdots))$.*

(a) *left^n runs from (h, τ) to $(h - n, \tau)$, and right^n from (h, τ) to $(h + n, \tau)$, in $n + 1$ units.*

(b) *If P_i runs in t_i units from the end of P_{i-1}, then $[P_1, \dots, P_k]$ runs in $t_1 + \cdots + t_k + 1$ units and $P_1; \cdots; P_k$ in $t_1 + \cdots + t_k$ units.*

Proof. Each move and the final skip cost one unit, and a sequence adds the units of its parts. □

Loops are bounded by an invariant that says, before each pass, where the head is and what the tape holds.

Lemma I.6 (Loops)**.** *Let P be a program, $n, b \in \mathbb{N}$, and $I_0, \dots, I_n$ sets of configurations. Suppose that for every $j < n$ each configuration in I_j scans a 1 and P runs from it within b units to a configuration in I_{j+1}, and that each configuration in I_n scans a 0. Then for every $j \le n$, $\mathsf{loop}(P)$ runs from each configuration in I_j within $(n-j)(b+1)+1$ units to a configuration in I_n.*

Proof. Induction on $n - j$. For $j = n$ the loop tests a 0 and stops: one unit. For $j < n$ the loop tests a 1 (one unit), runs P within b units to a configuration in I_{j+1}, and by the induction hypothesis finishes within $(n-j-1)(b+1)+1$ more units; in all $(n-j)(b+1)+1$. □

A stage of the machine is a program that starts and ends with the head on the root, together with a contract.

Definition I.7 (Stage contracts)**.** Fix a problem and write $X = m + n_0 + \kappa + 1$. A *contract* for a program P consists of two families of properties of tapes, Pre_r and $\mathrm{Post}_{r'}$ indexed by row counts, and a bound $K(X + r + 1)^g$. The program *meets* the contract if for every i, m, r and every tape τ satisfying Pre_r there are t, r' and τ' with

$$(-1, \tau) \xrightarrow[P]{t} (-1, \tau'), \qquad \tau' \text{ satisfies } \mathrm{Post}_{r'}, \qquad t + r' + 1 \le K(X + r + 1)^g.$$

The constant K may depend on the problem (the field, y_0, M, T and n_0) but is fixed before i, m, r and τ; the exponent g is a fixed natural number. When the rows do not change, a contract may bound t alone.

Lemma 15.36(c) composes two such contracts in sequence. The next two lemmas let one program written for a plain half-line run on a track, or on every other cell of a block, without being rewritten by hand.

Lemma I.8 (Running a program on a track)**.** *For a program P and $s \ge 1$ let $P^{\uparrow s}$ replace every left by left^s and every right by right^s. Fix $o \in \mathbb{Z}$, and say that (h, τ) carries (g, σ) if $h = o + sg$ and $\tau(o + si) = \sigma(i)$ for every $i \in \mathbb{Z}$. If $(g, \sigma) \xrightarrow[P]{t} (g', \sigma')$ and (h, τ) carries (g, σ), then $P^{\uparrow s}$ runs from (h, τ) within $(s+1)t$ units to a configuration (h', τ') that carries (g', σ'), with $\tau'(x) = \tau(x)$ for every x that is not of the form $o + si$.*

Proof. Induction on the counted execution of P. A write at virtual position g writes the cell $o + sg$, which is on the track, and costs one unit; a test reads $\tau(o + sg) = \sigma(g)$, so it takes the same branch or makes the same loop decision. A move becomes s moves and a skip, $s + 1$ units, and leaves the carried configuration moved by one. Sequences, branches and loops add the units of their parts, so every unit of P becomes at most $s + 1$ units of $P^{\uparrow s}$, and no cell off the track is written. □

Lemma I.9 (Mirror image). *Let $\overline{P}$ exchange* left *and* right *in P, and let the reflection of (g,σ) be $(-g,\, i \mapsto \sigma(-i))$. If $(g,\sigma) \xrightarrow[P]{t} (g',\sigma')$, then $\overline{P}$ runs in t units from the reflection of (g,σ) to the reflection of (g',σ'). Consequently, for $P^{\downarrow s} = (\overline{P})^{\uparrow s}$ and (h,τ) with $h = o - sg$ and $\tau(o - si) = \sigma(i)$ for all i, $P^{\downarrow s}$ runs within $(s+1)t$ units to (h',τ') with $h' = o - sg'$, $\tau'(o - si) = \sigma'(i)$, and $\tau'(x) = \tau(x)$ off the cells $o - si$.*

Proof. Induction on the execution: reflection exchanges a move left with a move right, preserves the scanned bit and turns a write at g into a write at $-g$. The second claim is Lemma I.8 applied to the reflected run. □

Track k is the descending sequence $\mathrm{cell}(j,k) = \mathrm{cell}(0,k) - 16j$. So a program P written for the half-line $0,1,2,\ldots$ runs on track k as $P^{\downarrow 16}$ with $o = \mathrm{cell}(0,k)$, within 17 times its units, and touches no other track. The arithmetic programs of Section I.5 are written for blocks of consecutive cells and used in this way.

I.3 Moving along the tape

The head finds its way by marks, never by addresses: control states are positions in the program, and a row number exists only as a counter on a track. Three scans do all the walking.

Lemma I.10 (Scans). *Let* $\mathsf{scanL} = \mathsf{loop}(\mathsf{left}^{16})$, $\mathsf{scanR} = \mathsf{loop}(\mathsf{right}^{16})$ *and* $\mathsf{scan}_1 = \mathsf{loop}(\mathsf{left})$.

(a) If $\tau(h-16j) = 1$ for $j < n$ and $\tau(h-16n) = 0$, then $(h,\tau) \xrightarrow[\mathsf{scanL}]{18n+1} (h-16n,\tau)$.

(b) If $\tau(h+16j) = 1$ for $j < n$ and $\tau(h+16n) = 0$, then $(h,\tau) \xrightarrow[\mathsf{scanR}]{18n+1} (h+16n,\tau)$.

(c) If $\tau(h-j) = 1$ for $j < n$ and $\tau(h-n) = 0$, then $(h,\tau) \xrightarrow[\mathsf{scan}_1]{2n+1} (h-n,\tau)$.

Proof. Lemma I.6 with $I_j = \{(h \mp 16j, \tau)\}$ (or $h - j$) and a body of 17 units (or 1): $n \cdot 18 + 1$ and $n \cdot 2 + 1$ units. □

Lemma I.11 (Going to a counter and back). *For tracks k and l put*

$$\mathsf{seek}_k = \mathsf{left}^{k+1};\mathsf{scanL}, \qquad \mathsf{home}_k = \mathsf{right}^{k-1};\mathsf{scanR};\mathsf{right}^2 \quad (k \geq 1),$$
$$\mathsf{pres}_k = [\mathsf{seek}_k, \mathsf{left}], \qquad \mathsf{val}_k = [\mathsf{seek}_k, \mathsf{left}^2],$$

and let $\mathsf{shift}_{k\to l}$ *be* left^{l-k} *if $k \leq l$ and* right^{k-l} *otherwise.*

(a) If track k carries the counter a, then seek_k *runs from $(-1,\tau)$ to $(\mathrm{cell}(a,k),\tau)$ in $18a+k+3$ units,* pres_k *to $(\mathrm{cell}(a,k+1),\tau)$ in $18a+k+5$ units and* val_k *to $(\mathrm{cell}(a,k+2),\tau)$ in $18a+k+7$ units.*

(b) If τ has origin marks up to a and $k \geq 1$, then home_k *runs from $(\mathrm{cell}(a,k),\tau)$ to $(-1,\tau)$ in $18a+k+4$ units.*

(c) $\mathsf{shift}_{k\to l}$ *runs from* $(\mathrm{cell}(r,k),\tau)$ *to* $(\mathrm{cell}(r,l),\tau)$ *in* $|k-l|+1$ *units.*

None of them writes.

Proof. (a) left^{k+1} takes $k+2$ units from -1 to $-2-k=\mathrm{cell}(0,k)$. The cells $\mathrm{cell}(0,k)-16j=\mathrm{cell}(j,k)$ hold 1 for $j<a$ and 0 for $j=a$, so scanL takes $18a+1$ units (Lemma I.10). One or two further moves and the final skip of the list give the other two counts. (b) right^{k-1} takes k units from $\mathrm{cell}(a,k)$ to $\mathrm{cell}(a,1)$. Going right along track 1 from row a the cells of rows $a,a-1,\dots,1$ hold 1 and that of row 0 holds 0, so scanR takes $18a+1$ units and stops at $\mathrm{cell}(0,1)=-3$; right^{2} takes 3 units to -1. (c) The two cells are $|k-l|$ apart in the same row. □

Lemma I.12 (Finding a mark, and writing a block)**.** *Let* $\mathsf{toggle}=\mathsf{branch}(\mathsf{write}\,1,\mathsf{write}\,0)$, $\mathsf{findL}=\mathsf{toggle};\mathsf{loop}(\mathsf{toggle};\mathsf{left};\mathsf{toggle});\mathsf{toggle}$*, and for bits* $b_0\cdots b_{\ell-1}$ *let* $\mathsf{put}(b_0\cdots b_{\ell-1})$ *be* $\mathsf{write}\,b_0;\mathsf{left};\mathsf{write}\,b_1;\mathsf{left};\cdots;\mathsf{write}\,b_{\ell-1};\mathsf{left};\mathsf{skip}$.

(a) toggle *complements the scanned cell in* 2 *units.*

(b) If $\tau(h-j)=0$ *for* $j<n$ *and* $\tau(h-n)=1$*, then* $(h,\tau)\xrightarrow[\mathsf{findL}]{6n+5}(h-n,\tau)$.

(c) $\mathsf{put}(b)$ *runs from* (h,τ) *in* $2\ell+1$ *units to* $(h-\ell,\tau')$*, where* $\tau'(h-j)=b_j$ *for* $j<\ell$ *and* $\tau'=\tau$ *elsewhere.*

Proof. (a) A branch and a write. (b) The first toggle complements the scanned cell, so the loop's test sees 1 exactly when the original bit was 0. A pass restores the current cell, moves left and complements the next cell: $1+2+1+2=6$ units with its test. After n passes the scanned cell is the complemented 1, the test fails (1 unit) and the last toggle restores it: $2+6n+1+2$ units, and the tape is unchanged. (c) Two units per bit and the skip. □

I.4 Counters, columns and rows

A counter is advanced by writing its hole and the cell below it. This section collects the programs that do this, copy and add counters, clear tracks and add rows, and it proves the rule by which a body is repeated once for every mark of a word.

Lemma I.13 (Advancing a counter)**.** *Let* $\mathsf{next}=[\mathsf{write}\,1,\mathsf{left}^{16},\mathsf{write}\,0]$. *It runs from* $(\mathrm{cell}(r,k),\tau)$ *in* 20 *units to* $(\mathrm{cell}(r+1,k),\tau')$*, where* τ' *is* τ *with* (r,k) *set to* 1 *and* $(r+1,k)$ *set to* 0.

Proof. $1+17+1+1$ units, by Lemma I.5. □

Lemma I.14 (Incrementing, resetting and marking)**.** *Let $2 \le k \le 15$ and let track k carry the counter a. Put*

$$\begin{aligned}
\mathsf{inc}_k &= \mathsf{seek}_k; \mathsf{write}\,1; \mathsf{left}^{16}; \mathsf{write}\,0; \mathsf{home}_k,\\
\mathsf{reset}_k &= [\mathsf{seek}_k, \mathsf{write}\,1, \mathsf{home}_k, \mathsf{left}^{k+1}, \mathsf{write}\,0, \mathsf{home}_k],\\
\mathsf{zero}_k &= \mathsf{left}^{k+1}; \mathsf{write}\,0; \mathsf{home}_k,\\
\mathsf{mark}_k &= [\mathsf{seek}_k, \mathsf{left}, \mathsf{write}\,1, \mathsf{right}, \mathsf{next}, \mathsf{home}_k] \quad (k \le 14),
\end{aligned}$$

and let inc_k^c be c copies of inc_k followed by skip. Each runs from $(-1,\tau)$ to $(-1,\tau')$:

(a) if τ has origin marks up to $a+1$, inc_k takes $36a+2k+44$ units and τ' is τ with (a,k) set to 1 and $(a+1,k)$ to 0, so track k carries $a+1$;

(b) if τ has origin marks up to $a+c$, inc_k^c takes $18c^2+(36a+2k+26)c+1$ units and leaves the counter $a+c$;

(c) if τ has origin marks up to a, reset_k takes $36a+4k+16$ units and τ' is τ with (a,k) set to 1 and $(0,k)$ to 0, so track k carries 0;

(d) if $\tau(0,1)=0$, zero_k takes $2k+7$ units and sets $(0,k)$ to 0, whatever track k held;

(e) if τ has origin marks up to $a+1$, mark_k takes $36a+2k+49$ units, sets $(a,k+1)$ to 1 and advances the counter on track k to $a+1$.

Track 1 is not written, so the origin marks are kept.

Proof. All counts come from Lemmas I.5, I.11 and I.13. (a) seek_k ($18a+k+3$) reaches the hole; $\mathsf{write}\,1$, left^{16} and $\mathsf{write}\,0$ (19) move the hole to row $a+1$; home_k from row $a+1$ takes $18(a+1)+k+4$. (b) The i-th increment starts at the counter $a+i$ and takes $36(a+i)+2k+44$ units; the sum over $i<c$ plus the final skip is $18c^2+(36a+2k+26)c+1$. (c) $(18a+k+3)+1+(18a+k+4)+(k+2)+1+(k+4)+1$: home_k from row 0 costs $k+4$. (d) $(k+2)+1+(k+4)$. (e) $(18a+k+3)+1+1+1+20+(18(a+1)+k+4)+1$. □

Lemma I.15 (Copying and adding counters)**.** *Let k, l and c be distinct tracks in $\{2,\dots,15\}$, let τ have origin marks up to R, and put*

$$\begin{aligned}
\mathsf{copy}_{k\to l} &= \mathsf{zero}_l; \mathsf{left}^{k+1}; \mathsf{loop}(\mathsf{shift}_{k\to l}; \mathsf{next}; \mathsf{shift}_{l\to k}); \mathsf{home}_k,\\
\mathsf{acc}^c_{k\to l} &= \mathsf{zero}_c; \mathsf{seek}_c; \mathsf{shift}_{c\to k}; \mathsf{loop}(\Gamma); \mathsf{home}_k,\\
\Gamma &= \mathsf{shift}_{k\to c}; \mathsf{next}; \mathsf{home}_c; \mathsf{inc}_l; \mathsf{seek}_c; \mathsf{shift}_{c\to k}.
\end{aligned}$$

(a) If track k carries $n \le R$, then $\mathsf{copy}_{k\to l}$ runs from $(-1,\tau)$ to $(-1,\tau')$ in $n(2|k-l|+41)+2k+2l+14 \le 100(n+1)$ units, where τ' is τ with (j,l) set to 1 for $j<n$ and (n,l) to 0: track l carries n.

(b) If track k carries b, track l carries a and $a+b \le R$, then $\mathsf{acc}^{c}_{k\to l}$ runs from $(-1,\tau)$ to $(-1,\tau')$ in

$$36b^2 + \big(36a + 2c + 2l + 2|k-c| + 92\big)b + 3c + k + |c-k| + 16 \le 500(a+b+1)^2$$

units; in τ' track l carries $a+b$, track c carries b, and only the cells (j,l), $a \le j \le a+b$, and (j,c), $j \le b$, have changed.

Proof. (a) zero_l $(2l+7)$ makes $(0,l)$ a hole, and left^{k+1} $(k+2)$ reaches cell$(0,k)$. Let I_j be the configuration at cell(j,k) in which $(i,l) = 1$ for $i < j$ and $(j,l) = 0$. From I_j with $j < n$ the body moves to cell(j,l), advances the hole there and returns to cell$(j+1,k)$: $2(|k-l|+1)+20$ units, so I_{j+1}. At I_n the scanned cell is the hole of track k. Lemma I.6 gives $n(2|k-l|+23)+1$ units, and home_k from row n takes $18n+k+4$. The sum is the stated count, and $2|k-l|+41 \le 67$ and $2k+2l+14 \le 74$ give the bound.

(b) zero_c, seek_c (counter 0) and $\mathsf{shift}_{c\to k}$ take $(2c+7)+(c+3)+(|c-k|+1)$ units and reach cell$(0,k)$. Let J_j be the configuration at cell(j,k) with j on track c and $a+j$ on track l. For $j < b$ the pass from J_j takes 1 unit for the test and

$$(|k-c|+1) + 20 + (18(j+1)+c+4) + (36(a+j)+2l+44) + (18(j+1)+c+3) + (|c-k|+1)$$

units, by Lemmas I.11, I.13 and I.14: $36(a+2j+1)+2c+2l+2|k-c|+74$ in all, and it ends in J_{j+1}. The final test at (b,k) costs 1 and home_k costs $18b+k+4$. Summing over $j < b$ gives the displayed count. With all tracks at most 15 and $|k-c| \le 13$, the count is at most $36(a+b)^2 + 178(a+b) + 89 \le 500(a+b+1)^2$. □

Lemma I.16 (Clearing tracks)**.** *Let $2 \le k \le 15$ and*

$$\mathsf{clear}_k = \mathsf{left}; \mathsf{loop}(\mathsf{left}^k; \mathsf{write}\,0; \mathsf{right}^k; \mathsf{left}^{16}); \mathsf{left}^k; \mathsf{write}\,0; \mathsf{home}_k,$$

and for tracks $k_1, \dots, k_s$ let $\mathsf{clear}_{k_1,\dots,k_s}$ be $\mathsf{clear}_{k_1}; \cdots; \mathsf{clear}_{k_s}; \mathsf{skip}$. If $\tau(j,0) = 1$ for $j < r$, $\tau(r,0) = 0$ and τ has origin marks up to r, then clear_k runs from $(-1,\tau)$ to $(-1,\tau')$ in $(2k+39)r + 2k + 8$ units, where τ' is τ with (j,k) set to 0 for every $j \le r$. So $\mathsf{clear}_{k_1,\dots,k_s}$ takes $\sum_i \big((2k_i+39)r + 2k_i + 8\big) + 1$ units.

Proof. left reaches the header $(0,0)$. A pass from the header $(j,0)$, $j < r$, clears (j,k), returns to the header and moves to $(j+1,0)$: $1 + (k+1) + 1 + (k+1) + 17 = 2k+21$ units with its test. The test at the sentinel $(r,0)$ fails (1 unit); left^k and $\mathsf{write}\,0$ clear (r,k) ($k+2$ units); and home_k from row r takes $18r+k+4$. □

Lemma I.17 (A new row)**.** *Let*

$$\mathsf{grow} = \mathsf{left}; \mathsf{scanL}; \mathsf{write}\,1; \mathsf{left}; \mathsf{scanR}; \mathsf{right}^2; \\ \mathsf{left}; \mathsf{scanL}; \mathsf{put}(\beta); \mathsf{right}^{15}; \mathsf{scanR}; \mathsf{right}^2,$$

where $\beta = \beta_0 \cdots \beta_{15}$ has $\beta_k = 1$ exactly for $k \in \{1,2,5,8,11,14\}$. If $\tau(j,0) = 1$ for $j < r$, $\tau(r,0) = 0$, $\tau(r+1,0) = 0$ and τ has origin marks up to r, then grow runs from $(-1,\tau)$ to $(-1,\tau')$ in $72r+99$ units, where τ' is τ with $(r,0)$ set to 1 and row $r+1$ overwritten by β (track k receiving β_k). In particular, if τ is a workspace of r rows with origin marks up to r, then τ' is a workspace of $r+1$ rows with origin marks up to $r+1$, and rows $j \le r$ are unchanged except for their header.

Proof. The first six parts promote the sentinel. left (1) reaches $(0,0)$, scanL $(18r+1)$ the sentinel header $(r,0)$, which is set to 1; left (1) reaches $(r,1)$, and scanR along the origin marks $(18r+1)$ and right^2 (3) return to the root: $36r+8$ units. The rest appends the block. Now the headers of rows $0,\dots,r$ hold 1 and $(r+1,0)$ holds 0, so left and scanL take $18(r+1)+2$ units to $(r+1,0)$; $\mathsf{put}(\beta)$ (33 units, Lemma I.12) writes β_k at $\mathrm{cell}(r+1,0)-k=(r+1,k)$ and stops at $(r+2,0)$; right^{15} (16) reaches $(r+1,1)$, which now holds $\beta_1=1$; and scanR $(18(r+1)+1)$ with right^2 (3) returns to the root. In all $36r+8+36r+91=72r+99$ units. The new row $r+1$ has header $\beta_0=0$ and origin mark $\beta_1=1$, every row below it is untouched and therefore blank, and row r now has header 1. □

The pattern β also fills tracks $2,5,8,11$ and 14 of the new row with 1. A counter on one of these tracks whose hole lies in an older row is unaffected, since only the cells below the hole matter.

The basic loop of the machine repeats a body once for every mark of a word. The step loop of the trajectory is of this kind: its body is one Taylor step, repeated once for each of the S marks of a retained word.

Lemma I.18 (Repeating a body along a word)**.** *Let $c\in\{2,12\}$ and put*

$$\mathsf{repeat}_c(P)=\big[\mathsf{seek}_c;\mathsf{left}^{14-c},\ \mathsf{loop}\big([\mathsf{home}_{14},P,\mathsf{seek}_c,\mathsf{next},\mathsf{left}^{14-c}]\big),\ \mathsf{home}_{14}\big].$$

Let $r_0,n,b\in\mathbb{N}$ and let $R_0,\dots,R_n$ be properties of tapes such that every tape with R_j has origin marks up to r_0+n. Suppose that for every $j<n$ and every tape τ with R_j on which track c carries r_0+j and track 14 a retained word of length n from row r_0, P runs from $(-1,\tau)$ within b units to some $(-1,\tau_1)$ such that τ_1 agrees with τ on the cells (i,c), $i\le r_0+n$, and $(r_0+i,14)$, $i\le n$, and the tape τ_2 obtained from τ_1 by setting (r_0+j,c) to 1 and (r_0+j+1,c) to 0 satisfies R_{j+1}. Then from every tape with R_0 on which track c carries r_0 and track 14 the retained word, $\mathsf{repeat}_c(P)$ runs to a tape with R_n, on which track c carries r_0+n and track 14 still carries the word, within

$$n\big(36(r_0+n)+b+58\big)+36r_0+18n+38$$

units.

Proof. The first part, seek_c followed by left^{14-c}, takes $(18r_0+c+3)+(15-c)=18r_0+18$ units and ends at $(r_0,14)$, the first mark of the word. Let I_j be the set of configurations at $(r_0+j,14)$ whose tape has R_j, carries r_0+j on track c and carries the word. A pass from I_j with $j<n$ scans the mark $(r_0+j,14)=1$, returns to the root $(18(r_0+j)+18$ units), runs P (at most b), goes to the hole of track c, which by hypothesis is still at row r_0+j $(18(r_0+j)+c+3)$, moves it by next (20), steps to $(r_0+j+1,14)$ $(15-c)$ and ends the list (1). The tape is then τ_2, which has R_{j+1}, carries r_0+j+1 on track c and still carries the word, because $c\neq 14$. So the pass ends in I_{j+1} within $36(r_0+j)+b+57$ units, at most $36(r_0+n)+b+57$. At I_n the scanned cell $(r_0+n,14)$ is the word's 0. Lemma I.6 bounds the loop by $n(36(r_0+n)+b+58)+1$, and home_{14} from row r_0+n and the final skip add $18(r_0+n)+19$. □

When the body has a polynomial bound and leaves the number of rows alone, repetition raises the exponent by one and moves the repetition count into the constant.

Lemma I.19 (Iterating a stage on fixed rows)**.** *In Lemma I.18 let $r_0 = 0$, let every tape with some R_j be a workspace of r rows with $n \le r$, and let $b = K_b Y^g$ with $Y \ge r + 1$. Then the number t of units of* $\mathsf{repeat}_c(P)$ *satisfies*

$$t + r + 1 \le \big(n(36n + K_b + 58) + 18n + 40\big)\, Y^{g+1}.$$

Proof. Since $Y \ge 1$, $Y^g \le Y^{g+1}$, $1 \le Y^{g+1}$ and $r + 1 \le Y \le Y^{g+1}$. So $t \le n(36n + 58) + nK_b Y^g + 18n + 38 \le (n(36n + K_b + 58) + 18n + 38)Y^{g+1}$, and adding $r + 1 \le Y^{g+1}$ gives the claim with room to spare. □

With $Y = X + r + 1$ this is the form of Definition I.7: the trajectory loop, whose body is one Taylor step on a fixed allocation of rows and whose count is the number S of steps, adds one to the exponent of the step and puts S into the constant only, whatever the precision.

I.5 Binary arithmetic in rows

The arithmetic programs work on a block of consecutive cells, cut into rows of six or twelve cells, with one presence cell per row. They are written for an ordinary tape and run where they are needed through Lemmas I.8 and I.9: on one track of the scratch area (stride sixteen, descending), or on the even cells of a block of twelve-cell rows (stride two), in which case the odd cells keep their contents. All the words are written least significant bit first.

Definition I.20 (Rows of a block)**.** Let σ be a tape, $q \ge 2$ and $h \in \mathbb{Z}$. The *q-cell rows at h* are the blocks $h + qj, \dots, h + qj + q - 1$, $j \in \mathbb{Z}$; cell $h + qj + c$ is *column c* of row j, and column 0 holds the *presence* bits. The block has *n marked rows* if $\sigma(h + qj) = 1$ for $0 \le j < n$ and $\sigma(h + qn) = 0$; it is *closed on the left* if also $\sigma(h - q) = 0$. Column c *holds* the word w if $\sigma(h + qj + c) = w_j$ for $j < |w|$. Row n, the first unmarked row, is the *end row*; some programs write in its columns.

Lemma I.21 (Ripple addition)**.** *In six-cell rows, let column* 1 *and column* 2 *hold the operands, column* 3 *the carries, column* 4 *the sum; column* 5 *is unused. Put*

$$\begin{aligned}
\mathsf{add} &= \mathsf{loop}(F), \qquad F = \mathsf{right};\mathsf{branch}(F_0, F_1),\\
F_a &= \mathsf{right};\mathsf{branch}(\mathsf{right};C_{a0},\ \mathsf{right};C_{a1}),\\
C_{ab} &= \mathsf{branch}\big(E(a \oplus b, \mathrm{maj}(a,b,0)),\ E(a \oplus b \oplus 1, \mathrm{maj}(a,b,1))\big),\\
E(u,v) &= \mathsf{right};\mathsf{write}\,u;\mathsf{right}^5;\mathsf{write}\,v;\mathsf{left}^3,
\end{aligned}$$

where maj *is the majority of three bits. Let the block at h have n marked rows, columns* 1 *and* 2 *hold words x and y of length n, and $c_0 = \sigma(h+3)$. Define $c_{j+1} = \mathrm{maj}(x_j, y_j, c_j)$ and $z_j = x_j \oplus y_j \oplus c_j$. Then* add *runs from (h, σ) to $(h + 6n, \sigma')$ in exactly $20n + 1$ units, where σ' is σ with column* 4 *of row j set to z_j for $j < n$ and column* 3 *of row j set to c_j for $1 \le j \le n$; and*

$$\mathrm{val}(z) + 2^n c_n = \mathrm{val}(x) + \mathrm{val}(y) + c_0.$$

Proof. Let I_j be the configuration at $h+6j$ in which the first j sums and carries have been written. From I_j, $j < n$, the presence bit is 1; F moves to column 1 and branches on x_j, moves to column 2 and branches on y_j, moves to column 3 and branches on c_j (this cell was written in the previous

pass, or is c_0); E writes z_j in column 4, moves five cells to column 3 of row $j+1$, writes c_{j+1} and returns to $h+6(j+1)$. With the test this is $1+(1+1)+(1+1)+(1+1)+(1+1+6+1+4) = 20$ units. At I_n the presence bit is 0. Lemma I.6 gives $20n+1$ units. The identity follows from $(a \oplus b \oplus c) + 2\,\mathrm{maj}(a,b,c) = a+b+c$ by induction on n. □

Corollary I.22 (Subtraction, comparison and signed sums)**.** *In Lemma I.21:*

(a) if column 2 holds $\bar{y}$ and $c_0 = 1$, then $\mathrm{val}(z) + 2^n c_n = \mathrm{val}(x) - \mathrm{val}(y) + 2^n$, so $c_n = 1$ exactly when $\mathrm{val}(y) \le \mathrm{val}(x)$, and then $\mathrm{val}(z) = \mathrm{val}(x) - \mathrm{val}(y)$;

(b) if $n \ge 1$, $u = \mathrm{sval}_n(x)$, $v = \mathrm{sval}_n(y)$ and $c_0 = 0$, then $\mathrm{sval}_n(z) = u + v$ whenever $-2^{n-1} \le u + v < 2^{n-1}$; with $\bar{y}$ in column 2 and $c_0 = 1$, $\mathrm{sval}_n(z) = u - v$ whenever $-2^{n-1} \le u - v < 2^{n-1}$;

(c) if $n \ge 1$, column 1 holds a word with $\mathrm{sval}_n = q$, column 2 holds 0^n, and $c_0 = g$, then $\mathrm{sval}_n(z) = q + g$ whenever $-2^{n-1} \le q + g < 2^{n-1}$. If $Z = 2^{k+1}q + 2^k g + \rho$ with $0 \le \rho < 2^k$, this is $\mathrm{near}(Z/2^{k+1})$.

Proof. (a) $\mathrm{val}(\bar{y}) = 2^n - 1 - \mathrm{val}(y)$, and $0 \le \mathrm{val}(z) < 2^n$. (b) $\mathrm{sval}_n(w) \equiv \mathrm{val}(w)$ modulo 2^n for every word, so $\mathrm{sval}_n(z) \equiv u + v + c_0$, and $\mathrm{sval}_n(\bar{y}) = -1 - v$. Both $\mathrm{sval}_n(z)$ and the target lie in $[-2^{n-1}, 2^{n-1})$, which contains one integer of each residue class. (c) The first claim is (b). For the second, $\mathrm{near}(Z/2^{k+1}) = \lfloor (Z + 2^k)/2^{k+1} \rfloor = q + \lfloor (2^k g + \rho + 2^k)/2^{k+1} \rfloor$, and $2^k g + \rho + 2^k$ lies in $[2^k, 2^{k+1})$ if $g = 0$ and in $[2^{k+1}, 2^{k+1} + 2^k)$ if $g = 1$. □

Lemma I.23 (Adding and subtracting one)**.** *Let* $\mathsf{carry} = \mathsf{loop}(\mathsf{write}\,0; \mathsf{right}); \mathsf{write}\,1$, $\mathsf{borrow} = \mathsf{toggle}; \mathsf{loop}(\mathsf{write}\,1; \mathsf{right}; \mathsf{toggle})$ *and* $\mathsf{borrow}^\circ = \mathsf{borrow}; \mathsf{left}; \mathsf{scan}_1; \mathsf{right}$.

(a) If $\sigma(h+j) = 1$ for $j < q$ and $\sigma(h+q) = 0$, carry *runs in $3q+2$ units to $h+q$, setting the cells $h, \dots, h+q-1$ to 0 and $h+q$ to 1.*

(b) If $\sigma(h+j) = 0$ for $j < q$ and $\sigma(h+q) = 1$, borrow *runs in $5q+3$ units to $h+q$, setting $h, \dots, h+q-1$ to 1 and $h+q$ to 0; if also $\sigma(h-1) = 0$,* borrow$^\circ$ *makes the same change and returns to h, in $7q+6$ units.*

Read as a word from h, (a) adds one to a word $1^q 0u$ and (b) subtracts one from a word $0^q 1u$.

Proof. (a) Each of the q passes costs 3 units with its test, the last test 1, and the final write 1. (b) The first toggle (2 units) makes the loop test see 1 exactly at the original 0s; each pass writes 1, moves right and complements the next cell (4 units, 5 with the test); at $h+q$ the complemented 1 reads 0 and the loop stops (1 unit). For borrow$^\circ$, left, a scan over the q new 1s up to the 0 at $h-1$ ($2q+1$) and right add $2q+3$ units. The values follow from $\mathrm{val}(0^q 1u) = \mathrm{val}(1^q 0u) + 1$. □

Lemma I.24 (Negation). *Let the two-cell rows at h have $\ell \geq 1$ marked rows whose column 1 holds a word w with $\mathrm{val}(w) > 0$, and let $\sigma(h-1) = 0$. Let*

$$\mathsf{neg} = \mathsf{right}; (\mathsf{borrow}^{\circ})^{\uparrow 2}; \mathsf{left}; \mathsf{loop}\big(\mathsf{right}; \mathsf{branch}(\mathsf{write}\,1, \mathsf{write}\,0); \mathsf{right}\big).$$

Then neg *runs from h to $h + 2\ell$ within $30(\ell+1)$ units and leaves in column 1 a word w' with $\mathrm{val}(w') \equiv -\mathrm{val}(w)$ modulo 2^{ℓ}; no other cell changes.*

Proof. Let $q < \ell$ be the position of the first 1 of w. The value cells $h+1+2i$ form a tape with stride two whose cell -1 is $h-1$. By Lemmas I.23 and I.8, $(\mathsf{borrow}^{\circ})^{\uparrow 2}$ subtracts one from w within $3(7q+6)$ units and returns to $h+1$. The last loop complements every value cell, 5 units a row and 1 for the final test. In all at most $1 + 3(7q+6) + 1 + 5\ell + 1 \leq 26\ell + 21$ units. Complementing ℓ bits maps the value v to $2^{\ell} - 1 - v$, so the result is $2^{\ell} - 1 - (\mathrm{val}(w) - 1) = 2^{\ell} - \mathrm{val}(w)$. □

The remaining programs use twelve-cell rows. Their columns are: 0 presence; 1 and 3 the original operands; 2 and 4 working words; 5 a row mark; 6, 8 and 10 the carry, sum and spare cells of the six-cell adder run on the even columns by $\mathsf{add}^{\uparrow 2}$; 9 a result.

Lemma I.25 (Walking over twelve-cell rows). *Let the twelve-cell rows at h have n marked rows and be closed on the left, and let $s, t < 12$. Put*

$$\begin{aligned}
\mathsf{back} &= \mathsf{left}^{12}; \mathsf{loop}(\mathsf{left}^{12}); \mathsf{right}^{12}, \qquad \mathsf{fwd} = \mathsf{loop}(\mathsf{right}^{12}),\\
\mathsf{cp}_{s\to t} &= \mathsf{loop}\big(\mathsf{right}^{s}; \mathsf{branch}(G_0, G_1)\big), \qquad G_b = \mathsf{left}^{s}; \mathsf{right}^{t}; \mathsf{write}\,b; \mathsf{left}^{t}; \mathsf{right}^{12},\\
\mathsf{cl}_t &= \mathsf{loop}(\mathsf{right}^{t}; \mathsf{write}\,0; \mathsf{left}^{t}; \mathsf{right}^{12}), \qquad \mathsf{z}_t = \mathsf{right}^{t}; \mathsf{write}\,0; \mathsf{left}^{t},\\
\mathsf{ret} &= \mathsf{right}^{5}; \mathsf{toggle}; \mathsf{loop}(\mathsf{toggle}; \mathsf{right}^{12}; \mathsf{toggle}); \mathsf{right}^{7}.
\end{aligned}$$

(a) For $k \leq n$, back *runs from $h + 12k$ to h in $14k + 27$ units, and* fwd *from h to $h + 12n$ in $14n + 1$ units.*

(b) From h, $\mathsf{cp}_{s\to t}$ copies column s into column t in $(2s + 2t + 20)n + 1$ units, cl_t sets column t to 0 in $(2t + 17)n + 1$ units, and both end at $h + 12n$; z_t sets column t of the current row to 0 in $2t + 3$ units.

(c) If column 5 holds 0 in rows $j < k$ and 1 in row $k < n$, then ret *runs from h to $h + 12(k+1)$ in $18k + 17$ units and sets column 5 of row k to 0.*

No other cells are written.

Proof. (a) left^{12} reaches row $k-1$; each marked row costs 14 units in the loop and the closed left end 1; right^{12} returns to row 0: $13 + 14k + 1 + 13$. (b) A pass of $\mathsf{cp}_{s\to t}$ costs $1+(s+1)+1+(s+1)+(t+1)+1+(t+1)+13$ units and one of cl_t costs $1+(t+1)+1+(t+1)+13$; the final test adds 1. (c) The first toggle complements the mark cell of row 0, so the loop runs exactly over the rows before k, restoring each and complementing the next (18 units a pass with its test); at row k the complemented mark reads 0. In all $6 + 2 + 18k + 1 + 8$ units, ending seven cells past the mark of row k, that is at $h + 12(k+1)$. □

Lemma I.26 (Doubling). *Let*

$$\mathsf{dbl} = \mathsf{loop}\big(\mathsf{right}^2; \mathsf{branch}(D_0, D_1)\big), \qquad D_u = \mathsf{right}^{20}; \mathsf{write}\, u; \mathsf{left}^{12}; \mathsf{branch}(D_0', D_1'),$$
$$D_b' = \mathsf{left}^8; \mathsf{write}\, b; \mathsf{right}^{10}.$$

If the twelve-cell rows at h have n marked rows, column 2 holds a and column 10 of row 0 holds e, then dbl *runs from h to $h + 12n$ in $62n + 1$ units and leaves in column 2 the word $e\, a_0 \cdots a_{n-2}$, whose value is $2\,\mathrm{val}(a) + e$ modulo 2^n; it writes only columns 2 and 10 (including column 10 of the end row).*

Proof. A pass from row j reads a_j in column 2, stores it in column 10 of row $j + 1$, reads the bit stored in column 10 of row j (the incoming bit: e for $j = 0$, a_{j-1} otherwise), writes it in column 2 and moves to row $j + 1$: $1 + 3 + 1 + 21 + 1 + 13 + 1 + 9 + 1 + 11 = 62$ units with its test. □

Lemma I.27 (Multiplication). *Let the twelve-cell rows at h have n marked rows and be closed on the left, with x in column 1 and y in column 3. Let*

$$\mathsf{mul} = \mathsf{cp}_{1\to 2}; \mathsf{back}; \mathsf{cl}_4; \mathsf{back}; \mathsf{cl}_5; \mathsf{back}; \mathsf{loop}(\Pi),$$
$$\Pi = \mathsf{right}^5; \mathsf{write}\, 1; \mathsf{left}^2; \mathsf{branch}\big(\mathsf{left}^3; \mathsf{back},\ \mathsf{left}^3; \mathsf{back}; \Sigma\big); \mathsf{z}_{10}; \mathsf{dbl}; \mathsf{back}; \mathsf{ret},$$
$$\Sigma = \mathsf{z}_6; \mathsf{add}^{\uparrow 2}; \mathsf{back}; \mathsf{cp}_{8\to 4}; \mathsf{back}.$$

Then mul *runs from h to $h + 12n$ within $240n^2 + 304n + 85 \le 400(n+1)^2$ units and leaves in column 4 the n lowest bits of $\mathrm{val}(x)\,\mathrm{val}(y)$. Columns 0, 1 and 3 are unchanged. If $n \ge 1$, $\mathrm{sval}_n(x) = u$, $\mathrm{sval}_n(y) = v$ and $-2^{n-1} \le uv < 2^{n-1}$, then column 4 reads uv in two's complement.*

Proof. The preparation copies x into column 2 and clears columns 4 and 5: $(26n + 1) + (25n + 1) + (27n + 1)$ units for the passes and $3(14n + 27)$ for the returns, by Lemma I.25. Let I_k ($k \le n$) be the configuration at row k in which column 2 holds the n lowest bits of $2^k\,\mathrm{val}(x)$, column 4 those of $\mathrm{val}(x) \cdot \big(\mathrm{val}(y) \bmod 2^k\big)$, column 5 is 0, and columns 0, 1, 3 are as given. A pass Π from I_k, $k < n$, marks row k in column 5 and reads y_k (10 units, then a branch). If $y_k = 1$, it returns to row 0 and runs Σ: z_6 clears the carry-in (15 units), $\mathsf{add}^{\uparrow 2}$ adds column 2 to column 4 into column 8 on the even columns within $3(20n + 1)$ units (Lemmas I.21 and I.8), and the copy of column 8 into column 4 with its two returns costs $(44n+1)+2(14n+27)$: at most $132n+73$. With left^3 and the return from row k the branch costs at most $14k + 132n + 105$. Then z_{10}, dbl and back double column 2 ($23 + (62n+1) + (14n+27)$ units), and ret removes the mark and moves to row $k + 1$ ($18k + 17$). The pass ends in I_{k+1} within $32k + 208n + 183 \le 240n + 183$ units. At I_n the presence bit is 0. Lemma I.6 bounds the loop by $n(240n + 184) + 1$, and the sum is $240n^2 + 304n + 85$. The two's complement claim follows as in Corollary I.22(b), since sval_n agrees with val modulo 2^n. □

Division is long division with restoring, one quotient bit per row, from the most significant bit down. Its single step is a subtraction that is undone when it would go negative.

Lemma I.28 (A restoring step)**.** *In six-cell rows at h, closed on the left, let column 1 hold x, column 2 hold $\bar{y}$, and $c_0 = 1$. Put* $\mathsf{rstep} = \mathsf{add}; \mathsf{right}^3; \mathsf{branch}(\beta_1, \beta_4)$ *with* $\beta_s = \mathsf{left}^3; \mathsf{left}^6; \mathsf{loop}(\mathsf{right}^s; \mathsf{branch}(\gamma_{s0}, \gamma_{s1})); \mathsf{right}^6$ *and* $\gamma_{sb} = \mathsf{right}^{4-s}; \mathsf{write}\, b; \mathsf{left}^{10}$*. Then* rstep *runs from h to h in exactly $40n + 25$ units. Afterwards the carry cell of the end row holds $e = [\mathrm{val}(y) \le \mathrm{val}(x)]$ and column 4 holds a word of value $\mathrm{val}(x) - e\,\mathrm{val}(y)$. If $\mathrm{val}(x) < 2\,\mathrm{val}(y)$, then $e = \lfloor \mathrm{val}(x)/\mathrm{val}(y) \rfloor$ and column 4 holds $\mathrm{val}(x) \bmod \mathrm{val}(y)$.*

Proof. By Corollary I.22(a), add ($20n + 1$ units, ending at the end row) leaves e in the end row's carry cell and, if $e = 1$, the difference in column 4. right^3 reads e (4 units and the branch 1). β_s moves to the last marked row and walks back to the closed left end copying column s into column 4: column 1 (restoring x) if $e = 0$, column 4 onto itself if $e = 1$. It costs $4 + 7 + 20n + 1 + 7$ units, a pass being $1 + (s + 1) + 1 + (5 - s) + 1 + 11 = 20$. The total is $(20n + 1) + 5 + (20n + 19)$. If $\mathrm{val}(x) < 2\,\mathrm{val}(y)$ the quotient is 0 or 1, and it is 1 exactly when $\mathrm{val}(y) \le \mathrm{val}(x)$. □

Lemma I.29 (Division)**.** *Let the twelve-cell rows at h have $n \ge 1$ marked rows and be closed on the left, column 1 hold x and column 4 hold $\bar{y}$, with $0 < \mathrm{val}(y) \le 2^{n-1}$. Let*

$$\begin{aligned}
\mathsf{div} &= \mathsf{cl}_2; \mathsf{back}; \mathsf{cl}_5; \mathsf{back}; \mathsf{fwd}; \mathsf{left}^{12}; \mathsf{loop}(\Delta); \mathsf{right}^{12},\\
\Delta &= \mathsf{right}^5; \mathsf{write}\ 1; \mathsf{left}^4; \mathsf{branch}(\Delta_0, \Delta_1),\\
\Delta_u &= \mathsf{left}; \mathsf{back}; \mathsf{right}^{10}; \mathsf{write}\, u; \mathsf{left}^{10}; \mathsf{dbl}; \mathsf{back};\\
&\qquad \mathsf{right}^6; \mathsf{write}\ 1; \mathsf{left}^6; \mathsf{rstep}^{\uparrow 2}; \mathsf{cp}_{8\to 2}; \mathsf{back};\\
&\qquad \mathsf{fwd}; \mathsf{right}^6; \mathsf{branch}(\Theta_0, \Theta_1),\\
\Theta_b &= \mathsf{left}^6; \mathsf{back}; \mathsf{ret}; \mathsf{left}^3; \mathsf{write}\, b; \mathsf{left}^{21}.
\end{aligned}$$

Then div *runs from h to h within $310n^2 + 389n + 84 \le 500(n + 1)^2$ units and leaves the quotient $\lfloor \mathrm{val}(x)/\mathrm{val}(y) \rfloor$ in column 9 and the remainder $\mathrm{val}(x) \bmod \mathrm{val}(y)$ in column 2; columns 0, 1 and 4 are unchanged.*

Proof. Write $y = \mathrm{val}(y)$. The preparation clears columns 2 and 5, walks to the end row and steps back to row $n - 1$: $(21n + 1) + (27n + 1) + 2(14n + 27) + (14n + 1) + 13$ units. For $k = n - 1, \dots, -1$ let I_k be the configuration at row k in which column 2 holds the remainder R_k of $\lfloor \mathrm{val}(x)/2^{k+1} \rfloor$ by y, column 9 holds the corresponding quotient bits in rows $k + 1, \dots, n - 1$, and column 5 is 0; $R_{n-1} = 0$. A pass Δ from I_k, $k \ge 0$, marks row k, reads x_k and returns to row 0 ($13 + 2 + (14k + 27)$ units). It writes x_k into column 10 of row 0 (23 units) and doubles column 2 with incoming bit x_k (Lemma I.26, $62n + 1$, and back, $14n + 27$): column 2 now holds $2R_k + x_k < 2y \le 2^n$, without overflow. It sets the carry-in (15) and runs rstep on the even columns, within $3(40n + 25)$ units (Lemma I.28): the end row's carry cell receives $e = \lfloor (2R_k + x_k)/y \rfloor$, and column 8 the remainder, which $\mathsf{cp}_{8\to 2}$ and back ($(40n+1)+(14n+27)$) move to column 2. Then Θ_e walks to the end row, reads e, returns, finds the marked row k with ret, writes e in its column 9 and steps to row $k - 1$ ($28n + 18k + 87$ units). The pass ends in I_{k-1} within $278n + 32k + 298 \le 310n + 298$ units: since $\lfloor \mathrm{val}(x)/2^k \rfloor = 2\lfloor \mathrm{val}(x)/2^{k+1} \rfloor + x_k$, the new remainder is R_{k-1} and e is the next quotient bit. At I_{-1} the scanned presence bit is the closed left end. Lemma I.6 bounds the loop by $n(310n + 299) + 1$; the final right^{12} costs 13. □

Lemma I.30 (Signed quotients)**.** *In Lemma I.29 let* $\mathrm{val}(x) = |z|$ *for an integer* z *with* $|z| < 2^n$*, and let the cell* $h-1$ *hold* $[z < 0]$*. There is a program* $\mathsf{sdiv} = \mathsf{div}; \mathsf{fix}$*, fixed once for all, that runs from* h *to* $h + 12n$ *within* $500(n+1)^2 + 270n + 175 \le 945(n+1)^2$ *units and leaves in the cell* $h-1$ *the bit* $[\lfloor z/y \rfloor < 0]$ *and in column* 8 *the* n *binary digits of* $|\lfloor z/y \rfloor|$*,* $y = \mathrm{val}(y)$*: the cell* $h-1$ *followed by column* 8 *is* $\mathrm{sw}_{n-1}(\lfloor z/y \rfloor)$*.*

Proof. Let q and ρ be the quotient and remainder of $|z|$ by y, left by div in columns 9 and 2. Then $\lfloor z/y \rfloor = q$ if $z \ge 0$ and $-(q + [\rho \ne 0])$ if $z < 0$, and the sign bit is unchanged: $z < 0$ forces $\lfloor z/y \rfloor \le -1$. The program fix has three parts. First it copies column 2 into column 4, returns, clears column 8 of row 0 and runs on the columns 0, 4 and 8 (stride four) a loop that writes into column 8 of each row the OR of the remainder bits of the earlier rows: $(32n+1) + (14n+27) + 19 + 5(11n+1)$ units, so at most $101n + 52$, and the end row's column 8 then holds $[\rho \ne 0]$. Second it reads that bit, returns to row 0, reads the sign in $h-1$, and writes $[z < 0] \wedge [\rho \ne 0]$ into the carry-in of row 0: $14n + 64$ units. Third it copies the quotient into column 2, clears column 4 and adds on the even columns: column 8 receives $q + [z < 0][\rho \ne 0]$, which is less than 2^n because $q + 1 \le |z|$ when $\rho \ne 0$; this costs $(42n+1) + (25n+1) + 2(14n+27) + 3(20n+1) \le 155n + 59$ units. The sum of the three parts is $270n + 175$. □

The last routine turns a binary number into a counter, one mark per unit of its value, so its cost grows with the value of the number and not only with its length.

Lemma I.31 (From binary to unary)**.** *There is a fixed program* $\mathsf{toUnary}$ *with the following property. Let the binary stream hold* w*, of length* ℓ *and value* n*, let track* 8 *carry the counter* r *and track* 9 *hold the value* r*, and let* τ *have origin marks up to* R *with* $\ell \le R$ *and* $r + n + 1 \le R$*. Then* $\mathsf{toUnary}$ *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $n\big(152\ell + 36(r+n) + 203\big) + 49\ell + 67 \le 1000(\ell + r + n + 1)^2$ *units; in* τ' *track* 8 *carries* $r + n$*, track* 9 *holds the value* $r + n$*, the origin marks are kept, and only tracks* 7*,* 8*,* 9 *and* 10 *have changed.*

Proof. The program is $\mathsf{Z}; \mathsf{loop}(\mathsf{home}_{10}; \mathsf{dec}; \mathsf{mark}_8; \mathsf{Z}); \mathsf{home}_{10}$. Here Z writes into the cell $(0, 10)$ the OR of the bits of the stream: it walks down the stream, carrying a running OR in track 10 one row ahead (30 or 31 units a row), and copies the final OR back to row 0; it costs at most $49\ell + 52$ units and ends at $(0, 10)$. $\mathsf{dec} = \mathsf{left}^8; (\mathsf{borrow})^{\downarrow 16}; \mathsf{home}_7$ subtracts one from the stream's bits on track 7 (Lemmas I.23 and I.9; the first 1 lies at a row $q < \ell$): at most $9 + 17(5q+3) + (18q+11) \le 103\ell + 71$ units. mark_8 ($36(r+j) + 65$ units in pass j) appends a mark to the value on track 9 and advances the counter on track 8. So a pass costs at most $14 + (103\ell + 71) + (36(r+j) + 65) + (49\ell + 52)$ units, and the flag it leaves is 1 exactly when the stream is still nonzero. Lemma I.6 with n passes, the first Z and the final home_{10} (14 units) give the bound. Every routine writes only tracks 7 to 10. □

I.6 The machine's own integers

Proposition 15.26 bounds the ratios of the procedure of Section 15.6. The machine computes differently: it keeps integer numerators over the fixed denominator 2^N, forms one integer numerator per new coefficient over a common denominator, and sums the Taylor series by Horner's rule on these integers, with one rounding at the end. The next proposition bounds

the integers it actually holds. Throughout, the problem has a coefficient field, $n = \max(m, n_0)$, $N = N(n)$, $W = W_m = 4\mathscr{B}(N) + 4$, and s, L and $\rho(N)$ are as in Section H.1. For a coordinate j let D_j be the product of the denominators of its coefficients (c_j, its A-coefficients and its B-coefficients), so that $D_j q$ is an integer for each of them.

Proposition I.32 (The machine's integers)**.**

(a) Let Z be an integer vector with $\|Z/2^N\| \le M$, and let $r_0 = Z, r_1, \dots, r_N$ be the integer rows of the table built from it as in Lemma 15.35. Every entry of every r_l fits W. In particular, under the region certificate, every stored vector Z_k, $k \le S$, and every row built from it fit W.

(b) Let Z be as in (a), $l < N$, $s_N = 2^N$, and let U be the integer with

$$c_j[l = 0] + (Af_l)_j + \sum_{i \le l} B(f_i, f_{l-i})_j = \frac{U}{D_j s_N^2}, \qquad f_i = r_i / s_N.$$

Then $r_{l+1,j} = \big\lfloor (2U + D_j s_N(l+1)) / (2D_j s_N(l+1)) \big\rfloor$, and both $|2U + D_j s_N(l+1)|$ and $2D_j s_N(l+1)$ are less than 2^{3W-1}.

(c) Let $z_0, \dots, z_N$ fit W, let $\sigma \ge 0$, $E = \sigma N$ and $H = W + \sigma N + N + 4$, and let $T_0 = z_0$ and $T_{k+1} = 2^\sigma T_k + z_{k+1}$ for $k < N$, the partial sums of Horner's rule. Then

$$|T_k| < (k+1)\, 2^{W+1+\sigma k} \quad \text{and} \quad |2^q T_k| < 2^{H-2} \qquad \text{whenever } k \le N \text{ and } \sigma k + q \le E;$$

and the rounded sum $\lfloor (2T_N + 2^E) / (2 \cdot 2^E) \rfloor$ fits $H - 1$.

(d) On a workspace of r rows every counter is at most $r + 1$.

(e) $W_m \le \big(4(q_P + l_P + c_P) + 4\big) X^2$ with $X = m + n_0 + \kappa + 1$.

Proof. (a) By Lemma 15.35, $r_l / 2^N$ is the coefficient a_l of the table at $Z/2^N$ with rounding $[\cdot]_N$, and by Lemma H.2 it has width $\rho(N)$ for $l \le N$, so $|r_l| \le 2^{\rho(N)}$. By (H.1), $\rho(N) \le \mathscr{B}(N) < W + 1$. The stored vectors lie in the ball by Theorem 15.20.

(b) Since $\mathrm{next}(f, l)$ is the displayed sum divided by $l + 1$, $r_{l+1,j} = \mathrm{near}(s_N \mathrm{next}(f,l)_j) = \mathrm{near}(U / (D_j s_N (l+1)))$, and $\mathrm{near}(u/v) = \lfloor (2u + v)/(2v) \rfloor$ for $v > 0$. Let w_j be the sum of $|D_j q|$ over the coefficients q of coordinate j. Clearing denominators,

$$U = [l = 0]\, D_j c_j s_N^2 + \sum_{(a,t)} D_j a\, r_{l,t}\, s_N + \sum_{i \le l} \sum_{(b,t,u)} D_j b\, r_{i,t}\, r_{l-i,u},$$

the sums running over the A-terms and B-terms of coordinate j. By (a) every $|r_{i,t}| < 2^{W+1}$, and $s_N \le 2^W$ because $N \le W$. So $|U| \le (l+1) w_j 2^{2W+2}$ and

$$|2U + D_j s_N (l+1)| \le (l+1)(2w_j + D_j) 2^{2W+2} < (l+2)(Q_j + 1) 2^{2W+2}, \qquad Q_j = 2w_j + 2D_j + 1,$$

and likewise $2D_j s_N(l+1) < (l+2)(Q_j+1)2^{2W+2}$. Each factor x is less than $2^{\mathrm{bl}(x)}$, so both integers are less than 2^β with $\beta = \mathrm{bl}(Q_j + 1) + \mathrm{bl}(l+2) + 2W + 2$. Now D_j is a product of denominators e whose binary lengths sum to at most s, so $D_j < 2^s$; each $|D_j q| = |u| D_j / e \le 2^{2s}$ for $q = u/e$; and coordinate j has at most $L + 1$ coefficients. Hence $Q_j + 1 = 2w_j + 2D_j + 2 \le (2L + 6) 2^{2s}$

and $\mathrm{bl}(Q_j + 1) \le 2s + 2L + 6$. Also $\mathrm{bl}(l + 2) \le l + 2 \le N + 1$. By (H.1), $\mathscr{B}(N) \ge s + L + N + 3$, since $v^*(N) \ge L$. Therefore $\beta \le 2W + 2s + 2L + N + 9 \le 2W + 2\mathscr{B}(N) + 3 \le 3W - 1$.

(c) This is Lemma I.158(d), where the machine's use of these integers is described; the argument is short. $|T_0| = |z_0| < 2^{W+1}$, and

$$|T_{k+1}| \le 2^{\sigma}|T_k| + |z_{k+1}| < (k+1)2^{W+1+\sigma(k+1)} + 2^{W+1} \le (k+2)2^{W+1+\sigma(k+1)}.$$

If $\sigma k + q \le E$, then $|2^q T_k| < (k+1)2^{W+1+E} \le 2^{H-3}$, because $k+1 \le N+1 \le 2^N$. For an integer x, $|\lfloor x/2^E \rfloor| \le |x|$, so the rounded sum, which is $\lfloor T_N/2^E \rfloor$ or that number plus one (Lemma I.158(c)), has absolute value at most $2^{H-2} < 2^H$.

(d) is Lemma I.4. (e) By Proposition 15.26, $\mathscr{B}(N) \le (q_P + l_P + c_P)X^2$, and $4 \le 4X^2$. □

So the coefficient stage works with words of $3W_m$ bits and the Taylor sum with words of $W_m + \sigma N + N + 4$ bits, both at most a constant depending on the problem times X^2, since σ depends only on the problem and $N < X$; and every unary number the machine writes is bounded by the rows in use, which each stage bound controls through Definition I.7. The coefficients and the initial point are read in binary from the saved request, whose length is at most a constant depending on the problem times X.

I.7 Reading the request

The request for a problem of dimension $d = 2$ begins with the frame 1110010: three unary marks, a separator and the three binary digits 0, 1, 0 of 2, least significant first. The first stage of the machine spends exactly 965 units on these seven cells. It turns the three marks into three rows of the scratch area (475 units), copies the digits into rows 0, 1 and 2 of tracks 3 and 4 (490 units), and leaves a single mark on cell 6, the last cell of the frame, to show where reading resumes. Every later frame is read in the same way, and the row at which its bits begin is marked on track 15.

So the request does not stay where it was written. The stage Parse consumes the input cells one by one and moves the bits of every frame into tracks 3 and 4 of the rows, with the frame starts on track 15 (Theorem I.42). When it ends, the only 1 left in the cells $0, 1, 2, \ldots$ is that mark, on the last cell of the request. The stage Decode erases the mark, reads the saved frames back and writes every natural number of the request in unary on tracks 9 and 10 (Theorem I.58). The signed integers of the payload stay in binary on tracks 3 and 4, where the later stages read them.

I.7.1 The raw marker

The machine reads the request from left to right. One mark on the input cells separates what has been read from what has not: the cells before it have been erased, the cell under it is the last one consumed, and the cells after it are still the request.

Definition I.33 (Raw marker)**.** A tape τ has a *raw marker at* $c \in \mathbb{N}$ if $\tau(-1) = 1$, $\tau(c) = 1$ and $\tau(x) = 0$ for $0 \le x < c$. For a word v, the cells *after the marker hold* v if $\tau(c + 1 + j) = v_j$ for $j < |v|$ and $\tau(c + 1 + j) = 0$ for $j \ge |v|$.

Lemma I.34 (Following the marker)**.** *Let* $\overline{\mathsf{scan}}_1 = \mathsf{loop}(\mathsf{right})$ *and* $\mathsf{findR} = \overline{\mathsf{findL}}$ *be the mirror images (Lemma I.9) of* scan_1 *and* findL*, and put*

$$\mathsf{raw} = [\mathsf{right}, \mathsf{findR}], \qquad \mathsf{unraw} = [\mathsf{left}, \mathsf{findL}], \qquad \mathsf{home}_0 = \mathsf{left}; \mathsf{scanR}; \mathsf{right}^2,$$
$$\mathsf{adv} = \mathsf{write}\,0; \mathsf{right}; \mathsf{write}\,1; \mathsf{left}; \mathsf{findL},$$
$$\mathsf{rd}(P_0, P_1) = \mathsf{right}; \mathsf{branch}\big(\mathsf{left}; \mathsf{adv}; P_0,\ \mathsf{left}; \mathsf{adv}; P_1\big).$$

(a) If $\tau(h + j) = 1$ *for* $j < n$ *and* $\tau(h + n) = 0$*, then* $\overline{\mathsf{scan}}_1$ *runs from* (h, τ) *to* $(h + n, \tau)$ *in* $2n + 1$ *units. If* $\tau(h + j) = 0$ *for* $j < n$ *and* $\tau(h + n) = 1$*, then* findR *runs from* (h, τ) *to* $(h + n, \tau)$ *in* $6n + 5$ *units.*

(b) Let τ *have a raw marker at* c*. Then* raw *runs from* $(-1, \tau)$ *to* (c, τ)*, and* unraw *from* (c, τ) *to* $(-1, \tau)$*, each in* $6c + 7$ *units. The program* adv *runs from* (c, τ) *to* $(-1, \tau^+)$ *in* $6c + 15$ *units, where* τ^+ *is* τ *with the cell* c *set to* 0 *and the cell* $c + 1$ *set to* 1*; so* τ^+ *has a raw marker at* $c + 1$*. If* $b = \tau(c + 1)$ *and* P_b *runs from* $(-1, \tau^+)$ *in* t *units to a configuration* γ*, then* $\mathsf{rd}(P_0, P_1)$ *runs from* (c, τ) *to* γ *in* $6c + 18 + t$ *units.*

(c) If τ *has origin marks up to* k*, then* home_0 *runs from* $(\mathrm{cell}(k, 0), \tau)$ *to* $(-1, \tau)$ *in* $18k + 5$ *units.*

Apart from adv*, and* P_b *inside* rd*, none of these programs writes.*

Proof. (a) Reflection exchanges left and right, so by Lemma I.9 the two programs take the units of scan_1 and findL on the reflected tape: $2n + 1$ and $6n + 5$ (Lemmas I.10(c) and I.12(b)). (b) right takes one unit to cell 0; the cells $0, \dots, c - 1$ hold 0 and c holds 1, so findR takes $6c + 5$; with the final skip, $6c + 7$. For unraw, left reaches $c - 1$ and findL passes the c zeros $c - 1, \dots, 0$ to the root, which holds 1: again $1 + (6c + 5) + 1$. The first four primitives of adv write the two cells and return to c (4 units); now the cells $c, c - 1, \dots, 0$ hold 0 and the root holds 1, so findL takes $6(c + 1) + 5$. In rd, right, the test of $\tau(c + 1) = b$ and left take 3 units, adv takes $6c + 15$, and P_b takes t. (c) left moves from $\mathrm{cell}(k, 0)$ to $\mathrm{cell}(k, 1)$; along track 1 the rows $k, \dots, 1$ hold 1 and row 0 holds 0, so scanR takes $18k + 1$ units to $\mathrm{cell}(0, 1) = -3$, and right^2 three more. □

I.7.2 Rows for the first frame

The first frame of the request is $1^L 0 b$ with b the binary digits of d and $L = |b| \ge 1$. Before the scratch area has rows, the machine has no counters; it uses the unary length of the frame instead, and makes one row for each mark. Three sixteen-bit patterns are written as rows: β^{first} has 1 exactly on tracks 0 and 14, β^{next} on tracks $0, 1, 2, 5, 8, 11, 14$, and β^{sent} on tracks $1, 2, 5, 8, 11, 14$; the last is the pattern β of Lemma I.17. For a pattern β put

$$\mathsf{block}(\beta) = \mathsf{left}; \mathsf{scanL}; \mathsf{put}(\beta); \mathsf{right}^{15}; \mathsf{scanR}; \mathsf{right}^2,$$

the second half of grow, and

$$\begin{aligned}
&\mathsf{alloc} = \mathsf{pre}; \mathsf{loop}(\mathsf{hdr}); \mathsf{epi},\\
&\mathsf{pre} = \overline{\mathsf{scan}}_1; \mathsf{write}\,1; \mathsf{left}; \mathsf{write}\,0; \mathsf{left}; \mathsf{scan}_1; \mathsf{left}; \mathsf{put}(\beta^{\mathrm{first}}); \mathsf{right}^{17}; \mathsf{right},\\
&\mathsf{hdr} = \mathsf{cut}; \mathsf{block}(\beta^{\mathrm{next}}); \mathsf{right}, \qquad \mathsf{cut} = \overline{\mathsf{scan}}_1; \mathsf{left}; \mathsf{write}\,0; \mathsf{left}; \mathsf{scan}_1,\\
&\mathsf{epi} = \mathsf{left}; \mathsf{block}(\beta^{\mathrm{sent}}); \mathsf{write}\,1.
\end{aligned}$$

Lemma I.35 (Allocating the first rows)**.**

(a) Let $k \geq 1$, $\tau(j,0) = 1$ for $j < k$, $\tau(k,0) = 0$, $\tau(0,1) = 0$ and $\tau(j,1) = 1$ for $1 \leq j < k$, and let β be a pattern with $\beta_1 = 1$. Then block(β) *runs from $(-1,\tau)$ to $(-1,\tau')$ in $36k + 55$ units, where τ' is τ with row k overwritten by β (track t receiving β_t).*

(b) Let b be a word of length $L \geq 1$ and v a word. From the head on cell 0 of the input tape of $1^L 0 bv$, alloc *runs to $(-1,\tau_A)$ in $20L^2 + 80L + 55$ units, where $\tau_A(-1) = 1$, $\tau_A(x) = 0$ for $0 \leq x < L$, $\tau_A(L) = 1$, the cells after L hold bv, row 0 holds β^{first}, rows $1, \ldots, L-1$ hold β^{next}, row L holds β^{sent}, and every cell of a row $j > L$ holds 0.*

So τ_A has a raw marker at L, headers 1 in rows $j < L$, sentinel row L and origin marks up to L, and tracks 2, 5, 8 and 11 carry the counter 0.

Proof. (a) left (1 unit) reaches $(0,0)$ and scanL the first blank header $(k,0)$ $(18k+1)$; put(β) (33) writes β_t at cell$(k,0) - t = (k,t)$ and stops at $(k+1,0)$; right^{15} (16) reaches $(k,1)$, which now holds $\beta_1 = 1$. Along track 1 the rows $k, \ldots, 1$ hold 1 and row 0 holds 0, so scanR and right^2 return to the root in $18k+4$ units. In all $36k + 55$.

(b) While marks remain the root holds 0; it is set to 1 at the end. pre: $\overline{\mathsf{scan}}_1$ runs over the L marks to the separator at L $(2L+1)$, which write 1 turns into the marker; left; write 0; left (3) erase the mark $L-1$; scan$_1$ runs left over the remaining marks $L-2, \ldots, 0$ to the root $(2(L-1)+1)$; left, put(β^{first}) and right^{17} $(1+33+18)$ write row 0 and come back to the root, and right reaches cell 0: $4L + 57$ units. For $j \leq L-1$ let I_j be the configuration at cell 0 in which the cells $0, \ldots, L-2-j$ hold 1 (so $m = L-1-j$ marks remain), the cells $L-1-j, \ldots, L-1$ and the root hold 0, cell L holds 1, rows $0, \ldots, j$ hold $\beta^{\mathrm{first}}, \beta^{\mathrm{next}}, \ldots, \beta^{\mathrm{next}}$, and every cell of a row $> j$ holds 0. After pre the configuration lies in I_0. From I_j with $j < L-1$ the scanned cell holds 1. cut runs right over the m marks $(2m+1)$, erases the mark $m-1$ (3) and runs left over the other $m-1$ to the root $(2(m-1)+1)$: $4m+3$ units. Then block(β^{next}) writes row $j+1$ by (a), whose hypotheses hold with $k = j+1$ $(36(j+1)+55)$, and right returns to cell 0. With its test the pass takes $4(L-1-j) + 36j + 96$ units and ends in I_{j+1}. At I_{L-1} cell 0 holds 0, so the loop makes exactly $L-1$ passes (Lemma I.6), and

$$\sum_{j<L-1} \big(4(L-1-j) + 36j + 96\big) + 1 = 20(L-1)^2 + 80(L-1) + 1.$$

epi returns to the root, writes row L by (a) with $k = L$ $(1 + 36L + 55)$ and sets the root (1). The total is $(4L+57) + 20(L-1)^2 + 80(L-1) + 1 + (36L+57) = 20L^2 + 80L + 55$. The patterns give the last sentence: track 1 holds 0 in row 0 and 1 in rows $1, \ldots, L$, and tracks $2, 5, 8, 11$ hold 0 in row 0. □

The bits b are then copied into the rows, one per row, with track 2 as the counter of copied bits. Put

$$\begin{aligned}\mathsf{emit}_b &= \mathsf{left}^3;\mathsf{scanL};\mathsf{write}\,1;\mathsf{left};\mathsf{write}\,1;\mathsf{left};\mathsf{write}\,b;\mathsf{left}^{14};\mathsf{write}\,0;\mathsf{right}^2,\\ \mathsf{pk} &= \mathsf{home}_0;\mathsf{right};\mathsf{findR};\mathsf{rd}(\mathsf{emit}_0,\mathsf{emit}_1),\\ \mathsf{pack} &= \mathsf{left};\mathsf{loop}(\mathsf{pk});\mathsf{home}_0, \qquad \mathsf{boot} = \mathsf{alloc};\mathsf{pack}.\end{aligned}$$

Lemma I.36 (Copying the first frame)**.**

(a) *If track 2 carries the counter s, then emit_b runs from $(-1,\tau)$ to $(\mathrm{cell}(s+1,0),\tau')$ in $18s+29$ units, where τ' is τ with $(s,2)$ and $(s,3)$ set to 1, $(s,4)$ set to b and $(s+1,2)$ set to 0: the counter on track 2 advances to $s+1$.*

(b) *In Lemma I.35(b), pack runs from $(-1,\tau_A)$ to $(-1,\tau_B)$ in $36L^2+53L+7$ units, where τ_B is τ_A with the cells $L,\dots,2L-1$ set to 0, the cell $2L$ set to 1, the cells $(j,2)$ and $(j,3)$ set to 1 and $(j,4)$ set to b_j for $j<L$, and $(L,2)$ set to 0. So boot runs from cell 0 of the input tape of 1^L0bv to $(-1,\tau_B)$ in $56L^2+133L+62$ units.*

Proof. (a) left^3 (4) reaches $(0,2)$ and scanL the hole $(s,2)$ ($18s+1$); the next six primitives write $(s,2)$, $(s,3)$, $(s,4)$ (5 units), left^{14} reaches $(s+1,2)$ (15), and $\mathsf{write}\,0;\mathsf{right}^2$ (4) clears it and stops at $(s+1,0)$.

(b) For $k\le L$ let J_k be the configuration at $(k,0)$ whose tape is τ_A with the marker moved to $L+k$, the bits $b_0,\dots,b_{k-1}$ copied as stated and track 2 carrying k. The first left (1 unit) enters J_0. From J_k, $k<L$, the header $(k,0)$ holds 1; pk returns to the root ($18k+5$, Lemma I.34(c)), steps to cell 0 and finds the marker ($1+6(L+k)+5$), reads $b_k=\tau(L+k+1)$ and moves the marker by rd ($6(L+k)+18$ plus the continuation), and emit_{b_k} copies it ($18k+29$, by (a)): $36k+12(L+k)+58$ units, ending in J_{k+1}. At J_L the header $(L,0)$ holds 0. The loop takes $\sum_{k<L}(36k+12(L+k)+59)+1 = 36L^2+35L+1$ units and home_0 from row L takes $18L+5$; with the first left, $36L^2+53L+7$. Adding Lemma I.35 gives $56L^2+133L+62$. □

For $d=2$, $L=3$: $475+490=965$ units, as in the example that opens this section.

I.7.3 Reading one frame

Every later frame is read into the binary stream (tracks 6 and 7) and then appended to the saved word on tracks 3 and 4. The state of the tape between two frames is the following.

Definition I.37 (Reader state)**.** Let $c,R\in\mathbb{N}$ and let w be a word. A tape τ is a *reader state* (c,R,w) if

(a) τ has a raw marker at c;

(b) track 5 carries the counter 0 and track 2 the counter $|w|$, and $|w|\le R$;

(c) w lies on tracks 3 and 4 from row 0;

(d) τ has origin marks up to R, $\tau(j,0)=1$ for $j<R$ and $\tau(R,0)=0$;

(e) $\tau(x) = 0$ for every $x \le \mathrm{cell}(R+1, 0)$, that is, every cell of every row $j > R$ holds 0.

By Lemma I.36, τ_B is a reader state $(2L, L, b)$ whose cells after the marker hold v, and tracks 6, 7, 9, 10, 12, 13 and 15 hold 0 in every row. Reading a frame uses the programs

$$\begin{aligned}
\mathsf{hmark} &= [\mathsf{left}, \mathsf{adv}, \mathsf{mark}_5, \mathsf{raw}, \mathsf{right}],\\
\mathsf{ghmark} &= \mathsf{left}; \mathsf{unraw}; \mathsf{grow}; \mathsf{raw}; \mathsf{right}; \mathsf{hmark},\\
\mathsf{store}_b &= [\mathsf{seek}_5, \mathsf{left}^2, \mathsf{write}\, b, \mathsf{right}^2, \mathsf{next}, \mathsf{left}], \qquad \mathsf{cut}_6 = \mathsf{pres}_5; \mathsf{write}\, 0; \mathsf{home}_6,\\
\mathsf{pbit} &= [\mathsf{home}_6, \mathsf{raw}, \mathsf{rd}(\mathsf{store}_0, \mathsf{store}_1)],\\
\mathsf{xfer} &= [\mathsf{raw}, \mathsf{right}, \mathsf{loop}(\mathsf{ghmark}), \mathsf{left}, \mathsf{adv}, \mathsf{cut}_6, \mathsf{reset}_5,\\
&\qquad \mathsf{pres}_5, \mathsf{loop}(\mathsf{pbit}), \mathsf{home}_6],
\end{aligned}$$

which read the length of a frame (one new row per mark) and then its bits, and the programs

$$\begin{aligned}
\mathsf{fa}_b &= \mathsf{home}_7; \mathsf{emit}_b; \mathsf{home}_0; \mathsf{inc}_5; \mathsf{pres}_5, \qquad \mathsf{fa} = \mathsf{left}; \mathsf{branch}(\mathsf{fa}_0, \mathsf{fa}_1),\\
\mathsf{seal} &= \mathsf{seek}_2; \mathsf{left}; \mathsf{write}\, 0; \mathsf{home}_3, \qquad \mathsf{app} = \mathsf{pres}_5; \mathsf{loop}(\mathsf{fa}); \mathsf{home}_6; \mathsf{seal}; \mathsf{reset}_5,\\
\mathsf{start} &= \mathsf{seek}_2; \mathsf{left}^{13}; \mathsf{write}\, 1; \mathsf{home}_{15}, \qquad \mathsf{save} = \mathsf{xfer}; \mathsf{reset}_5; \mathsf{app},
\end{aligned}$$

which mark the first row of the frame on track 15 and append its bits to the saved word.

Lemma I.38 (Reading the length and the bits of a frame)**.** *Let τ have a raw marker at c whose cells after the marker hold $1^n 0 f v$ with $|f| = n \ge 1$; let track 5 carry 0, let τ have origin marks up to R, $\tau(j, 0) = 1$ for $j < R$ and $\tau(R, 0) = 0$, and let $\tau(x) = 0$ for $x \le \mathrm{cell}(R+1, 0)$. Then* xfer *runs from $(-1, \tau)$ to $(-1, \tau')$ in*

$$102n^2 + (36c + 72R + 298)n + 12c + 104$$

units. The tape τ' has a raw marker at $c + 2n + 1$ and the cells after it hold v; f lies on tracks 6 and 7 from row 0; track 5 carries n; τ' has origin marks up to $R + n$, $\tau'(j, 0) = 1$ for $j < R + n$, $\tau'(R + n, 0) = 0$, and every cell of a row $j > R + n$ holds 0. In rows $j \le R$ the tracks other than 0, 5, 6 and 7 are unchanged, and track 15 is unchanged in every row.

Proof. raw and right ($6c + 8$) reach the first mark $c + 1$. For $j \le n$ let H_j be the configuration at $c + 1 + j$ whose tape has the marker at $c + j$ followed by $n - j$ marks and a 0, carries j on track 5, has $R + j$ rows in the sense of the hypothesis, and has $(i, 6) = 1$ for $i < j$. From H_j, $j < n$, the scanned cell is a mark. ghmark returns to the root ($1 + 6(c + j) + 7$), adds a row ($72(R + j) + 99$, Lemma I.17), comes back past the marker ($6(c + j) + 7 + 1$), and hmark moves the marker onto the mark (Lemma I.34(b)), sets $(j, 6)$ to 1 and advances the counter on track 5 by mark_5 ($36j + 59$, Lemma I.14(e)), and returns to $c + j + 2$: $12(c + j) + 36j + 90$ units. The pass takes $24(c + j) + 36j + 72(R + j) + 205$ units and ends in H_{j+1}; the loop takes $\sum_{j<n}(24c + 72R + 206 + 132j) + 1 = 66n^2 + (24c + 72R + 140)n + 1$ units. At H_n the scanned cell is the separator. left and adv ($1 + 6(c + n) + 15$) consume it; cut_6 ($36n + 21$) writes the terminator $(n, 6) = 0$; reset_5 ($36n + 36$) and pres_5 (10) reach $(0, 6)$. For $k \le n$ let G_k be the configuration at $(k, 6)$ with the marker at $c + n + 1 + k$, track 5 carrying k and $(i, 7) = f_i$ for $i < k$. From G_k, $k < n$, the presence cell holds 1; pbit returns to the root ($18k + 10$), finds the marker ($6(c + n + 1 + k) + 7$), reads f_k and moves the marker ($6(c + n + 1 + k) + 18$), and store_{f_k} writes

$(k, 7)$, advances the counter and stops at $(k+1, 6)$ $(18k+37)$: $36k + 12(c+n+1+k) + 73$ units, ending in G_{k+1}. At G_n the terminator stops the loop, which took $24n^2 + (12(c+n+1)+50)n+1$ units, and home_6 and the final skip add $18n + 11$. The sum is the stated count. The writes are on tracks 0 (headers of the new rows), 5, 6 and 7, and in the new rows $R+1, \dots, R+n$, whose pattern β^{sent} has 0 on track 15. □

Lemma I.39 (Appending a frame to the saved word)**.** *Let τ have origin marks up to R, and let track 2 carry s.*

(a) If track 5 carries k and $s+1, k+1 \le R$, then fa_b runs from $(\mathrm{cell}(k,7), \tau)$ to $(\mathrm{cell}(k+1,6), \tau')$ in $72k + 36s + 145$ units, where τ' is τ changed as by emit_b in Lemma I.36(a) and with the counter on track 5 advanced to $k+1$.

(b) If a word f of length n lies on tracks 6 and 7 from row a, track 5 carries a, and $s+n, a+n \le R$, then app runs from $(-1, \tau)$ to $(-1, \tau')$ in

$$54n^2 + (72a + 36s + 184)n + 72a + 36s + 71$$

units, where in τ' the cells $(s+j, 3)$ hold 1 and $(s+j, 4)$ hold f_j for $j < n$, $(s+n, 3)$ holds 0, track 2 carries $s+n$, track 5 carries 0, and only tracks 2, 3, 4 and 5 have changed. So if w of length s lies on tracks 3 and 4 from row 0, afterwards wf does.

(c) If $s \le R$, then seal runs from -1 to -1 in $36s + 14$ units and sets $(s, 3)$ to 0, and start runs from -1 to -1 in $36s + 39$ units and sets $(s, 15)$ to 1.

Proof. (a) home_7 from row k $(18k+11)$, emit_b $(18s+29$, ending at $(s+1, 0))$, home_0 from row $s+1$ $(18s+23)$, inc_5 $(36k+54)$ and pres_5 from the new hole $(18k+28)$. (b) pres_5 $(18a+10)$ reaches $(a, 6)$. Let F_j, $j \le n$, be the configuration at $(a+j, 6)$ in which the first j bits of f have been appended and tracks 2 and 5 carry $s+j$ and $a+j$. A pass from F_j, $j < n$, tests the presence 1, moves to $(a+j, 7)$, branches on f_j and runs fa_{f_j}: $72(a+j) + 36(s+j) + 148$ units with the test, ending in F_{j+1}. At F_n the terminator $(a+n, 6) = 0$ stops the loop, which took $54n^2 + (72a + 36s + 94)n + 1$ units. home_6 $(18(a+n)+10)$, seal $(36(s+n)+14$, by (c)) and reset_5 $(36(a+n)+36$, Lemma I.14(c)) complete the count. (c) seek_2 $(18s+5)$, one move, one write and home_3 $(18s+7)$ for seal; for start, seek_2, left^{13} (14), the write and home_{15} $(18s+19)$. □

Lemma I.40 (Reading one frame)**.** *Let τ be a reader state (c, R, w) with $s = |w|$ whose cells after the marker hold $1^n 0 f v$, $|f| = n \ge 1$. Put $\mathsf{raw}^+ = \mathsf{raw}; \mathsf{right}$, $\mathsf{unraw}^+ = \mathsf{left}; \mathsf{unraw}$,*

$$\mathsf{rf} = \mathsf{unraw}^+; \mathsf{start}; \mathsf{save}; \mathsf{raw}^+,$$
$$X(c, R, s, n) = 156n^2 + (36c + 72R + 36s + 518)n + 12c + 36s + 211.$$

(a) save runs from $(-1, \tau)$ in $X(c, R, s, n)$ units, and rf runs from $(c+1, \tau)$ to $(c+2n+2, \tau')$ in

$$(6c + 8) + (36s + 39) + X(c, R, s, n) + \big(6(c + 2n + 1) + 8\big)$$

units.

(b) τ' *is a reader state* $(c + 2n + 1, R + n, wf)$ *whose cells after the marker hold* v.

(c) $\tau'(r, 15) = \tau(r, 15)$ *for every row* $r \neq s$, *and* $\tau'(s, 15) = 1$.

Proof. unraw^+ takes $1 + (6c + 7)$ units from $c + 1$ to the root, and start $(36s + 39)$ sets $(s, 15)$. Then xfer takes $102n^2 + (36c + 72R + 298)n + 12c + 104$ units (Lemma I.38; its hypotheses are parts (a), (b), (d) and (e) of the reader state). It leaves f on tracks 6 and 7 from row 0 and the counter n on track 5, and it does not touch tracks 2, 3, 4 in rows $j \leq R$, which hold w, its terminator and the counter $s \leq R$. reset_5 $(36n + 36)$ sets track 5 to 0, and app with $a = 0$ $(54n^2 + (36s + 184)n + 36s + 71$, Lemma I.39) appends f to w, which fits since $s + n \leq R + n$. The three counts add up to X. Finally raw^+ $(6(c + 2n + 1) + 8)$ finds the moved marker and steps past it. Part (b) collects the conclusions of Lemmas I.38 and I.39; part (c) holds because only start writes on track 15. □

I.7.4 The stage Parse

Fix a problem with a coefficient field of dimension d, a level n_0, a coordinate $i < d$ and a precision m, and let u be the request of Definition 15.32. It is a concatenation of frames,

$$u = 1^{|f_0|}0 f_0\, 1^{|f_1|}0 f_1 \cdots 1^{|f_F|}0 f_F,$$

where f_0 is the $\mathrm{bl}(d) + 1$ lowest binary digits of d, followed by one frame for every natural number of the shape, the frame of the number q of payload integers, one frame for each of these q signed integers, and the frames of M, T, n_0, i and m. Every f_k is nonempty. Put $l_1 = |f_0|$ and $n_k = |f_k|$, and

$$w_u = f_0 f_1 \cdots f_F, \qquad s_k = l_1 + n_1 + \cdots + n_{k-1} \quad (1 \leq k \leq F + 1),$$

so that s_k is the position in w_u of the first bit of f_k, and $s_{F+1} = |w_u|$.

Definition I.41 (The saved request)**.** A tape τ *carries the saved request* of u if w_u lies on tracks 3 and 4 from row 0 (Definition I.3(a)) and, for every row r, $\tau(r, 15) = 1$ if r is one of $s_1, \ldots, s_F$ and $\tau(r, 15) = 0$ otherwise.

The saved request keeps the whole request: every bit of every frame, in order, and the first row of every frame after the first. The first frame starts at row 0 and needs no mark. The program of the stage is

$$\mathsf{Parse} = \mathsf{boot}; \mathsf{frames}, \qquad \mathsf{frames} = \mathsf{raw}^+; \mathsf{loop}(\mathsf{rf}); \mathsf{unraw}^+.$$

Theorem I.42 (Parse)**.** *Let* $c_k = 2l_1 + \sum_{1 \leq j < k}(2n_j + 1)$ *for* $1 \leq k \leq F + 1$, *so that* $c_{F+1} = |u| - 1$, *and let*

$$B_k = (6c_k + 8) + (36s_k + 39) + X(c_k, s_k, s_k, n_k) + \big(6(c_k + 2n_k + 1) + 8\big).$$

Then Parse *runs from the head on cell* 0 *of the input tape of* u *to* $(-1, \tau_P)$ *in exactly*

$$C_P(u) = 56l_1^2 + 133l_1 + 62 + (12l_1 + 8) + \sum_{k=1}^{F}(1 + B_k) + 1 + (6|u| + 2)$$

units, and:

(a) τ_P *is a reader state* $(|u|-1, |w_u|, w_u)$ *and* $\tau_P(x) = 0$ *for every* $x \geq |u|$*; so the only cell* $x \geq 0$ *that holds* 1 *is* $|u|-1$*, the last cell of the request;*

(b) τ_P *carries the saved request of* u*;*

(c) $C_P(u) \leq 100000(|u|+1)^3 \leq 100000(c_{\mathrm{req}}+2)^3 X^3$*, where* $X = m + n_0 + \kappa + 1$ *and* c_{req} *is the length of the request without its last two frames (those of* i *and* m*), plus* $2d+7$*.*

The exponent of the stage is 3*; its constant* $100000(c_{\mathrm{req}}+2)^3$ *depends only on the problem and* n_0*.*

Proof. By Lemma I.36, boot takes $56l_1^2 + 133l_1 + 62$ units and leaves a reader state $(2l_1, l_1, f_0)$ whose cells after the marker hold $1^{n_1}0\,f_1 \cdots 1^{n_F}0\,f_F$, with track 15 blank. raw^+ $(12l_1+8)$ steps past the marker. For $1 \leq k \leq F+1$ let P_k be the set of configurations at $c_k + 1$ whose tape is a reader state $(c_k, s_k, f_0 \cdots f_{k-1})$, whose cells after the marker hold $1^{n_k}0\,f_k \cdots 1^{n_F}0\,f_F$, and whose track 15 holds 1 exactly in the rows $s_1, \ldots, s_{k-1}$. The configuration after raw^+ is in P_1, since $c_1 = 2l_1$ and $s_1 = l_1$. From P_k, $k \leq F$, the scanned cell is the first mark of the frame of f_k, and by Lemma I.40 with $R = s = s_k$ the body rf takes exactly B_k units and ends in P_{k+1}. At P_{F+1} the scanned cell is $c_{F+1} + 1 = |u|$, which holds 0. So the loop makes exactly F passes and takes $\sum_k (1 + B_k) + 1$ units, and unraw^+ from $|u|$ takes $6(|u|-1)+8$. The final tape is in P_{F+1}, which gives (a) and (b).

(c) Put $N = |u| + 1$ and $K = 4l_1 + 4(n_1 + \cdots + n_F) + 4F + 1$; since $|u| = 2l_1 + 1 + \sum_k (2n_k + 1)$, $K \leq 4N$. For every k, $c_k + 2s_k + n_k + 1 \leq K$, and $c_k, F \leq K$. With $x = c + 2s + n + 1$,

$$X(c,s,s,n) = 156n^2 + (36c + 108s + 518)n + 12c + 36s + 211 \leq 1000x^2,$$

since each of n, $c + 2s$ and 1 is at most x. Hence $B_k \leq 1000K^2 + 36K + 61 \leq 1100K^2$, the loop takes at most $F(1100K^2 + 1) + 1 \leq (1101F + 1)K^2$ units, and the part frames at most $(6K + 8) + (1101K + 1)K^2 + (6K + 8) \leq 1200K^3 \leq 76800N^3$. As $l_1 \leq N$, boot takes at most $251N^3$. So $C_P(u) \leq 100000N^3$. For the second inequality, the frame of a natural number n has $2\,\mathrm{bl}(n) + 3 \leq 2n + 3$ cells, $i \leq d - 1$ and $m \leq X$, so $|u| + 1 \leq c_{\mathrm{req}} + 2m \leq (c_{\mathrm{req}} + 2)X$. □

I.7.5 Reading a saved frame

The second stage reads the saved word back, one frame at a time. Track 5 counts the saved bits already read, and the marks on track 15 tell where a frame ends.

Definition I.43 (Frames on the saved tracks)**.** Let $a \in \mathbb{N}$ and let $g_1, \ldots, g_p$ be nonempty words; put $a_1 = a$ and $a_{j+1} = a_j + |g_j|$. A tape τ *lays out* $g_1, \ldots, g_p$ *from row* a if, for every $j \leq p$, $\tau(a_j + i, 3) = 1$ and $\tau(a_j + i, 4) = (g_j)_i$ for $i < |g_j|$, $\tau(a_j + i, 15) = 0$ for $0 < i < |g_j|$, and $\tau(a_j + |g_j|, 3) = 0$ or $\tau(a_j + |g_j|, 15) = 1$; and if $\tau(a_{p+1}, 3) = 0$.

The condition reads only tracks 3, 4 and 15, so a program that leaves these tracks unchanged keeps every layout. A tape that lays out $g_1, \ldots, g_p$ from row a lays out $g_2, \ldots, g_p$ from row $a + |g_1|$.

Lemma I.44 (The saved request as frames)**.** *A tape that carries the saved request of u lays out $f_0, f_1, \dots, f_F$ from row 0.*

Proof. The word $w_u = f_0 \cdots f_F$ lies on tracks 3 and 4 from row 0, which gives the presence and value cells and $\tau(|w_u|, 3) = 0$. Inside f_k, the rows $s_k + i$ with $0 < i < n_k$ lie strictly between s_k and s_{k+1}, so they carry no mark on track 15; the same holds for the rows $1, \dots, l_1 - 1$ inside f_0. The row after f_k is s_{k+1}, which is marked if $k < F$; for $k = F$ it is $|w_u|$, whose presence cell holds 0. □

Copying a frame needs a test that tells, at a row a, whether the current frame goes on: whether $(a, 3)$ holds 1 and $(a, 15)$ holds 0. Put

$$\begin{aligned}
\mathsf{src} &= \mathsf{seek}_5; \mathsf{right}^2, \qquad \mathsf{sealb} = \mathsf{seek}_8; \mathsf{right}^2; \mathsf{write}\,0; \mathsf{home}_6,\\
\mathsf{probe} &= \mathsf{branch}\big(\mathsf{left}^7; \mathsf{write}\,0; \mathsf{right}^5; \mathsf{left}^5,\\
&\qquad\qquad \mathsf{left}^{12}; \mathsf{branch}(\mathsf{right}^5; \mathsf{write}\,1,\ \mathsf{right}^5; \mathsf{write}\,0)\big),\\
\mathsf{cpy}_b &= \mathsf{left}^2; \mathsf{write}\,1; \mathsf{left}; \mathsf{write}\,b; \mathsf{left}; \mathsf{next}; \mathsf{right}^5,\\
\mathsf{bin}_b &= \mathsf{seek}_8; \mathsf{right}^4; \mathsf{cpy}_b; \mathsf{home}_3, \qquad \mathsf{cs}_b = \mathsf{home}_4; \mathsf{bin}_b; \mathsf{inc}_5; \mathsf{src},\\
\mathsf{cs} &= \mathsf{left}; \mathsf{branch}(\mathsf{cs}_0, \mathsf{cs}_1), \qquad \mathsf{fc} = \mathsf{right}^7; \mathsf{cs}; \mathsf{probe},\\
\mathsf{fetch} &= \mathsf{src}; \mathsf{cs}; \mathsf{probe}; \mathsf{loop}(\mathsf{fc}); \mathsf{home}_{10}; \mathsf{sealb}.
\end{aligned}$$

probe starts at $(a, 3)$, writes the answer into the flag cell $(a, 10)$ and stops there, in 22 units on either branch. The first bit of a frame is copied without a test: the first row of every frame after the first carries a mark on track 15.

Lemma I.45 (Copying a saved frame)**.** *Let τ have origin marks up to R, let track 5 carry a and track 8 carry o, and let τ lay out g (and possibly further frames) from row a, with $n = |g|$, $a + n \le R$ and $o + n \le R$. Then* fetch *runs from $(-1, \tau)$ to $(-1, \tau')$ in*

$$F(a, o, n) = 54(n-1)^2 + (72a + 36o + 306)(n-1) + 108a + 72o + 294 \le 1000(a + o + n + 1)^2$$

units. In τ', g lies on tracks 6 and 7 from row o, track 5 carries $a + n$ and track 8 carries $o + n$. Only tracks 5, 6, 7, 8 and 10 have changed, and only in rows $j \le R$.

Proof. cpy_b runs from $(o, 4)$ to $(o + 1, 3)$ in 33 units, writing $(o, 6) = 1$ and $(o, 7) = b$ and advancing the counter on track 8; so bin_b takes $(18o + 11) + 5 + 33 + (18o + 25) = 36o + 74$ units. From $(k, 3)$ with tracks 5 and 8 carrying k and o', cs reads $(k, 4)$ and copies it: $2 + (18k + 8) + (36o' + 74) + (36k + 54) + (18k + 29) = 72k + 36o' + 167$ units, ending at $(k + 1, 3)$ with both counters advanced. So src $(18a + 11)$, cs and probe (22) copy g_0 and set the flag of row $a + 1$. For $1 \le i \le n$ let C_i be the configuration at $(a + i, 10)$ in which $g_0, \dots, g_{i-1}$ lie on tracks 6, 7 from row o (without terminator), tracks 5 and 8 carry $a + i$ and $o + i$, and the flag holds 1 exactly when $i < n$; this is what the layout says about the rows $a + 1, \dots, a + n$. A pass from C_i, $i < n$, moves to $(a + i, 3)$ (8 units), copies g_i and sets the next flag: $72(a + i) + 36(o + i) + 198$ units with its test. So the loop takes $\sum_{1 \le i < n} (72(a + i) + 36(o + i) + 198) + 1 = 54(n-1)^2 + (72a + 36o + 252)(n-1) + 1$ units. home_{10} from row $a + n$ $(18(a + n) + 14)$ and sealb, which writes the terminator $(o + n, 6) = 0$ $(36(o + n) + 25)$, complete the count. Each of $n - 1$, a, o and 1 is at most $a + o + n + 1$, which gives the bound. □

A frame of the payload is skipped rather than copied:

$$\mathsf{sp} = \mathsf{next};\mathsf{right}^2;\mathsf{probe}, \qquad \mathsf{sb} = \mathsf{right}^5;\mathsf{sp},$$
$$\mathsf{skipf} = \mathsf{seek}_5;\mathsf{sp};\mathsf{loop}(\mathsf{sb});\mathsf{home}_{10}.$$

Lemma I.46 (Skipping a saved frame)**.** *If track* 5 *carries* a*,* τ *has origin marks up to* R *and lays out* g *from row* a*, and* $a+|g| \le R$*, then* skipf *runs from* -1 *to* -1 *in* $36a+70|g|+16 \le 200(a+|g|+1)^2$ *units; afterwards track* 5 *carries* $a+|g|$*, and only tracks* 5 *and* 10 *have changed, only in rows* $j \le R$*.*

Proof. seek_5 ($18a+8$) reaches the hole; sp advances it (20), steps to $(a+1,3)$ (3) and probes (22): 45 units. Each of the $|g|-1$ passes moves from the flag to the hole (6) and runs sp: 52 units with the test. The final test and home_{10} add $1+18(a+|g|)+14$. □

I.7.6 Appending a unary word

Every natural number read from the saved request is appended, in unary, to a temporary record on tracks 13 (presence) and 14 (values), with track 11 as its length. For a number n write $\langle n\rangle = 1^n0$, and for numbers $n_1,\dots,n_k$ write $\langle n_1,\dots,n_k\rangle = \langle n_1\rangle\cdots\langle n_k\rangle$.

Definition I.47 (Reading configuration)**.** Let w be a word and $a, R \in \mathbb{N}$. A tape τ is a *reading configuration* (w,a,R) if

(a) track 11 carries the counter $|w|$; $\tau(j,13)=1$ and $\tau(j,14)=w_j$ for $j<|w|$, and $\tau(j,13)=0$ for every $j \ge |w|$; and $|w| \le R$;

(b) τ has origin marks up to R, $\tau(j,0)=1$ for $j<R$, $\tau(R,0)=0$, and $\tau(x)=0$ for every $x \le \mathrm{cell}(R+1,0)$;

(c) track 5 carries the counter a, track 8 the counter 0, and track 9 holds the value 0.

By (b), a reading configuration whose root holds 1 and whose cells $x \ge 0$ hold 0 is a workspace of R rows; none of the programs below writes a cell $x \ge -1$. Appending uses the programs

$$\mathsf{rec}_b = \mathsf{seek}_{11};\mathsf{left}^2;\mathsf{write}\ 1;\mathsf{left};\mathsf{write}\ b;\mathsf{right}^3;\mathsf{next};\mathsf{home}_{11},$$
$$\mathsf{grec}_b = \mathsf{grow};\mathsf{rec}_b, \qquad \mathsf{at}_9 = \mathsf{seek}_8;\mathsf{left},$$
$$\mathsf{ub} = \mathsf{write}\ 0;\mathsf{home}_9;\mathsf{grec}_1;\mathsf{inc}_8;\mathsf{at}_9,$$
$$\mathsf{word} = \mathsf{reset}_8;\mathsf{at}_9;\mathsf{loop}(\mathsf{ub});\mathsf{home}_9;\mathsf{grec}_0;\mathsf{reset}_8.$$

Lemma I.48 (Appending to the record)**.**

(a) If track 11 *carries* r *and* τ *has origin marks up to* $r+1$*,* rec_b *runs from* -1 *to* -1 *in* $36r+77$ *units, setting* $(r,13)$ *to* 1 *and* $(r,14)$ *to* b *and advancing the counter on track* 11*.*

(b) If τ *satisfies (a) and (b) of Definition I.47 with* w *and* R*, then* grec_b *runs from* -1 *to* -1 *in* $72R+36|w|+176$ *units to a tape that satisfies (a) and (b) with* wb *and* $R+1$*; in rows* $j \le R$ *only tracks* 0, 11, 13 *and* 14 *change.*

(c) *If τ satisfies (a) and (b) with w and R, track* 8 *carries n and track* 9 *holds the value n, with $n \le R$, then* word *runs from* -1 *to* -1 *in*

$$W(|w|,R,n) = 90n^2 + n(72R + 36|w| + 389) + 72R + 36|w| + 298 \le 1000(R + |w| + n + 1)^2$$

units to a tape that satisfies (a) and (b) with $w\langle n\rangle$ and $R + n + 1$, on which track 9 *holds the value* 0 *and track* 8 *carries* 0*. Tracks* 3, 4 *and* 15 *are unchanged, and tracks* 2, 5 *and* 12 *are unchanged in rows $j \le R$.*

Proof. (a) seek_{11} ($18r + 14$), five primitives and right^3 that write the two cells and return to the hole ($3 + 3 + 4$), next (20) and home_{11} from row $r + 1$ ($18r + 33$). (b) grow ($72R + 99$, Lemma I.17) adds row $R + 1$; its pattern has 0 on track 13, so (a) persists, and then rec_b ($36|w| + 77$) appends b. (c) reset_8 ($36n + 48$) makes track 8 carry 0 and at_9 (12) reaches $(0, 9)$. For $j \le n$ let U_j be the configuration at $(j, 9)$ in which track 9 holds 1 exactly in the rows $j, \dots, n-1$, track 8 carries j, and (a), (b) hold with $w1^j$ and $R+j$. A pass from U_j, $j < n$, clears $(j, 9)$ (1), returns to the root ($18j + 13$), appends a 1 ($72(R + j) + 36(|w| + j) + 176$), advances track 8 ($36j+60$) and returns to $(j+1, 9)$ ($18j+30$): with the test, $72R+36|w|+180j+281$ units, ending in U_{j+1}. At U_n the cell $(n, 9)$ holds 0, and the loop has taken $90n^2 + n(72R + 36|w| + 191) + 1$ units. Then home_9 ($18n + 13$), grec_0 ($72(R + n) + 36(|w| + n) + 176$) and reset_8 ($36n + 48$). The writes lie on tracks 0, 8, 9, 11, 13, 14 and in new rows, whose pattern has 0 on tracks 3, 4, 15; the bound follows termwise as in Lemma I.45. □

A number arrives in binary, from a saved frame; it is turned into unary by the program toUnary of Lemma I.31, with one difference: the rows its counter needs do not exist yet, so every pass first adds one. With Z and dec the programs of the proof of Lemma I.31, put

$$\mathsf{toUnary}^{+} = \mathsf{Z}; \mathsf{loop}\big(\mathsf{home}_{10}; \mathsf{grow}; \mathsf{left}^{11}; \mathsf{home}_{10}; \mathsf{dec}; \mathsf{mark}_8; \mathsf{Z}\big); \mathsf{home}_{10}.$$

Lemma I.49 (From binary to unary with new rows)**.** *Let the binary stream hold a word b of length ℓ and value n, let track* 8 *carry r and track* 9 *hold the value r, and let τ satisfy (b) of Definition I.47 with R, where $\ell, r \le R$. Then* $\mathsf{toUnary}^{+}$ *runs from* -1 *to* -1 *within*

$$n\big(152\ell + 36(r + n) + 72(R + n) + 328\big) + 49\ell + 67 \le 1000(\ell + r + R + n + 1)^2$$

units to a tape on which the binary stream holds 0^ℓ, track 8 *carries $r + n$, track* 9 *holds the value $r + n$, and (b) holds with $R + n$. Tracks* 3, 4, 13 *and* 15 *are unchanged, and in rows $j \le R$ only tracks* 0, 7, 8, 9 *and* 10 *change.*

Proof. This is the proof of Lemma I.31 with three more parts in each pass. Z ($49\ell + 52$ at most) writes the OR of the stream into $(0, 10)$; the loop runs once for each unit of n. In pass j the program returns to the root (14), adds row $R + j + 1$ ($72(R + j) + 99$), steps to $(0, 10)$ (12) and runs the old body: home_{10} (14), dec (at most $103\ell + 71$), which subtracts one from the stream, mark_8 ($36(r + j) + 65$), which appends a mark to the value on track 9 and advances track 8, and Z. The new row lies beyond the holes of the counters on tracks 8 and 9, and its pattern has 0 on track 9. So a pass costs at most $152\ell + 36(r + j) + 72(R + j) + 327$ units, the n passes and the last test at most $n(152\ell + 36(r + n) + 72(R + n) + 328) + 1$, and the final home_{10} 14. Only the stream values (track 7), the counters and the OR cells (tracks 8, 9, 10) and the headers of the new rows are written in rows $j \le R$. □

One natural number of the request is read by

$$\mathsf{nat} = \mathsf{fetch}; \mathsf{reset}_8; \mathsf{toUnary}^+; \mathsf{word}.$$

Lemma I.50 (Reading one natural number)**.** *Let τ be a reading configuration (w, a, R) that lays out* $\mathrm{bits}(n)$ *from row a (possibly followed by further frames), where* $\mathrm{bits}(n)$ *is the word of the $\ell = \mathrm{bl}(n) + 1$ lowest binary digits of n, and let $a + \ell \le R$. Then* nat *runs from -1 to -1 within*

$$F(a, 0, \ell) + (36\ell + 48) + n\big(152\ell + 36n + 72(R + n) + 328\big) + 49\ell + 67 + W(|w|, R + n, n) \le 10000\,\sigma^2$$

units, $\sigma = a + \ell + R + |w| + n + 1$, to a reading configuration $(w\langle n\rangle, a + \ell, R + 2n + 1)$. Tracks 3, 4 *and* 15 *are unchanged, tracks* 2 *and* 12 *are unchanged in rows $j \le R$, the cells $(j, 13)$ and $(j, 14)$ with $j < |w|$ are unchanged, and so is every cell $x \ge -1$.*

Proof. fetch with $o = 0$ (Lemma I.45) copies $\mathrm{bits}(n)$ onto the binary stream, advances track 5 to $a + \ell$ and leaves ℓ on track 8, which reset_8 ($36\ell + 48$) sets back to 0. Track 9 holds 0, so Lemma I.49 applies with $r = 0$: the value n moves to tracks 8 and 9 and the rows become $R + n$. Neither program touches tracks 11, 13 and 14 in rows $j \le R$, nor track 5 beyond its own advance, so the record w and the counter $a + \ell$ survive. Then word (Lemma I.48(c) with $R + n$ rows) appends $\langle n\rangle$ and clears tracks 8 and 9. The bound adds $F(a, 0, \ell) \le 1000\sigma^2$, $36\ell + 48 \le 84\sigma^2$, the conversion $\le 1000\sigma^2$ and $W(|w|, R + n, n) \le 1000(2\sigma)^2$. □

Lemma I.51 (Reading several natural numbers)**.** *Let τ be a reading configuration (w, a, R) that lays out* $\mathrm{bits}(c_1), \dots, \mathrm{bits}(c_k), g_1, \dots, g_p$ *from row a, and let $a + \Lambda \le R$, where Λ is the total length of these frames. Let $\lambda = |\,\mathrm{bits}(c_1) \cdots \mathrm{bits}(c_k)|$ and*

$$\sigma = a + \Lambda + R + |w| + 4(c_1 + \cdots + c_k) + 3k + 1.$$

Then the list $[\mathsf{nat}, \dots, \mathsf{nat}]$ *of k copies runs from -1 to -1 within*

$$k \cdot 10000\,\sigma^2 + 1$$

units to a reading configuration $(w\langle c_1, \dots, c_k\rangle, a + \lambda, R + 2(c_1 + \cdots + c_k) + k)$ that lays out $g_1, \dots, g_p$ from row $a + \lambda$; tracks 3, 4, 15 *and every cell $x \ge -1$ are unchanged, and tracks* 2 *and* 12 *are unchanged in rows $j \le R$.*

Proof. Induction on k, by Lemma I.50; the final skip of the list is the $+1$. The layout of the remaining frames survives because tracks 3, 4 and 15 do. The scale does not grow: after the first number, a, R and $|w|$ grow by ℓ_1, $2c_1 + 1$ and $c_1 + 1$, while Λ, $4\sum c$ and $3k$ fall by ℓ_1, $4c_1$ and 3; and the scale $a + \ell_1 + R + |w| + c_1 + 1$ of the first number is at most σ. □

I.7.7 Rows, groups and the payload

The shape of the field is a group of d rows for the A-terms and a group of d rows for the B-terms. Row i of a group is the number n of terms of coordinate i followed by their $n\alpha$ indices, with arity $\alpha = 1$ for A-terms and $\alpha = 2$ for B-terms. The number n just read is the count of

the loop that reads the rest of the row, and the dimension, the first word of the record, is the count of the loop over rows. Put

$$\mathsf{row}_\alpha = [\mathsf{copy}_{11\to12}, \mathsf{nat}, \mathsf{repeat}_{12}([\mathsf{nat}, \dots, \mathsf{nat}])] \quad (\alpha \text{ copies of } \mathsf{nat}),$$
$$\mathsf{skips} = [\mathsf{copy}_{11\to12}, \mathsf{nat}, \mathsf{repeat}_{12}(\mathsf{skipf})].$$

$\mathsf{copy}_{11\to12}$ puts on track 12 the row where the next word of the record begins; after nat that word is $\langle n\rangle$, which is a retained word of length n on track 14, and repeat_{12} (Lemma I.18) runs its body once for each of its marks.

Lemma I.52 (One row of the shape). *Let $\alpha \in \{1, 2\}$ and let τ be a reading configuration (w, a, R) that lays out* $\mathrm{bits}(n)$*, then* $\mathrm{bits}(e_1), \dots, \mathrm{bits}(e_{n\alpha})$*, then $g_1, \dots, g_p$ from row a, with $a + \Lambda \le R$ as in Lemma I.51. Let $\ell_n = |\mathrm{bits}(n)|$ and let σ' be the scale of Lemma I.51 for the record $w\langle n\rangle$, the row $a + \ell_n$, $R + 2n + 1$ rows and the numbers $e_1, \dots, e_{n\alpha}$. Then row_α runs from -1 to -1 within*

$$100(|w| + 1) + 10000(a + \ell_n + R + |w| + n + 1)^2 + n\big(36(|w| + n) + 10000\alpha\sigma'^2 + 59\big) + 36|w| + 18n + 39$$

units to a reading configuration $(w\langle n, e_1, \dots, e_{n\alpha}\rangle, a', R+2(n+\sum e)+n\alpha+1)$ that lays out $g_1, \dots, g_p$ from the row a' after these frames. Tracks 3, 4, 15 *and the cells $x \ge -1$ are unchanged, and track* 2 *is unchanged in rows $j \le R$.*

Proof. $\mathsf{copy}_{11\to12}$ (Lemma I.15(a)) takes at most $100(|w| + 1)$ units and makes track 12 carry $|w|$. nat (Lemma I.50) appends $\langle n\rangle$, leaving track 12 alone in rows $j \le R$; now track 14 carries a retained word of length n from row $r_0 = |w|$. In Lemma I.18 with $c = 12$, let R_j say that j entries of the row have been read: the tape is a reading configuration with $w\langle n, e_1, \dots, e_{j\alpha}\rangle$ that lays out the remaining frames. The body, α copies of nat, runs within $\alpha \cdot 10000\sigma'^2 + 1$ units (Lemma I.51, whose scale for the remaining numbers is at most σ'); it does not change track 12 in rows $j \le R$, nor the cells of the record already written, which include the word $\langle n\rangle$. Lemma I.18 gives $n(36(r_0 + n) + b + 58) + 36r_0 + 18n + 38$ with $b = 10000\alpha\sigma'^2 + 1$, and the list adds one unit. □

Lemma I.53 (A group of rows). *Let τ be a reading configuration (w, a, R) with $d + 1 \le |w|$, on which track* 14 *carries a retained word of length d from row* 0 *and track* 2 *carries* 0*, and which lays out the frames of d rows of arity α followed by $g_1, \dots, g_p$ from row a, with $a + \Lambda \le R$. Let S be the scale of Lemma I.51 for all the numbers of the d rows. Then $\mathsf{repeat}_2(\mathsf{row}_\alpha)$ runs from -1 to -1 within*

$$d\big(36d + 10^6S^3 + 58\big) + 18d + 38$$

units to a reading configuration whose record is w followed by the unary words of all numbers of the d rows, which lays out $g_1, \dots, g_p$ from the row after these frames, and on which track 2 *carries d and track* 14 *still carries the retained word; tracks* 3, 4, 15 *and the cells $x \ge -1$ are unchanged.*

Proof. Lemma I.18 with $c = 2$, $r_0 = 0$, $n = d$ and the body row_α, whose cost is at most 10^6S^3: in Lemma I.52 each of $|w_j| + 1$, the first scale, σ' and the count n is at most S, and $\alpha \le 2$, so the count there is at most $100S + 10000S^2 + S(72S + 20000S^2 + 59) + 54S + 39 \le 10^6S^3$. The body keeps track 2 in rows $j \le R$ and the cells of the record already written, which contain the retained word. □

Lemma I.54 (Skipping the payload). *Let τ be a reading configuration (w, a, R) that lays out* bits(q)*, then q nonempty words $z_1, \dots, z_q$, then $g_1, \dots, g_p$ from row a, with $a + \Lambda \le R$; put $\ell_q = |\,\mathrm{bits}(q)|$ and $\zeta = |z_1 \cdots z_q|$. Then* skips *runs from -1 to -1 within*

$$100(|w|+1) + 10000(a + \ell_q + R + |w| + q + 1)^2 + q\big(36(|w|+q) + 200(a+\ell_q+\zeta+1)^2 + 58\big) + 36|w| + 18q + 39$$

units to a reading configuration $(w\langle q\rangle, a + \ell_q + \zeta, R + 2q + 1)$ that lays out $g_1, \dots, g_p$ from row $a + \ell_q + \zeta$; tracks 3, 4, 15 *and the cells $x \ge -1$ are unchanged.*

Proof. As for Lemma I.52, with the body skipf (Lemma I.46), which reads the next z_j from row $a_j \le a + \ell_q + \zeta$ within $200(a + \ell_q + \zeta + 1)^2$ units, changes only tracks 5 and 10 and adds no row. The signed integers are never converted: their bits stay on tracks 3 and 4. □

I.7.8 The numeric record

When every number has been read, the record on tracks 13 and 14 holds $\langle c_1, \dots, c_L\rangle$. The last part of the stage copies it to tracks 9 and 10, preceded by the unary word of its number L of words, which it counts on the way. With track 5 as the reading counter of the record and track 8 as the writing counter put

$$\begin{aligned}
\mathsf{tpres} &= \mathsf{seek}_5; \mathsf{left}^8, \qquad \mathsf{tadv} = \mathsf{right}^8; \mathsf{next}; \mathsf{left}^8,\\
\mathsf{tskip} &= \mathsf{right}^9; \mathsf{next}; \mathsf{left}^8,\\
\mathsf{femit}_b &= \mathsf{seek}_8; \mathsf{left}; \mathsf{write}\,1; \mathsf{left}; \mathsf{write}\,b; \mathsf{right}^2; \mathsf{next}; \mathsf{home}_8,\\
\mathsf{temit}_b &= \mathsf{home}_{14}; \mathsf{femit}_b; \mathsf{tpres}; \mathsf{tadv},\\
\mathsf{fseal} &= \mathsf{seek}_8; \mathsf{left}; \mathsf{write}\,0; \mathsf{home}_9; \mathsf{reset}_8,\\
\mathsf{scan}(V_0, V_1) &= \mathsf{tpres}; \mathsf{loop}\big(\mathsf{left}; \mathsf{branch}(V_0, V_1)\big); \mathsf{home}_{13},\\
\mathsf{fin} &= \mathsf{scan}(\mathsf{temit}_1, \mathsf{tskip}); \mathsf{femit}_0; \mathsf{reset}_5;\\
&\qquad \mathsf{scan}(\mathsf{temit}_0, \mathsf{temit}_1); \mathsf{fseal}.
\end{aligned}$$

Lemma I.55 (Writing the numeric record). *Let $c_1, \dots, c_L$ be numbers and $v = \langle c_1, \dots, c_L\rangle$. Let $\tau(j, 13) = 1$ and $\tau(j, 14) = v_j$ for $j < |v|$ and $\tau(|v|, 13) = 0$, let tracks* 5 *and* 8 *carry* 0*, let τ have origin marks up to R, and let $|v| + L + 1 \le R$. Then* fin *runs from -1 to -1 within $2000(R+1)^2$ units to a tape τ' with $\tau'(0, 8) = 0$ on which the word $\langle L, c_1, \dots, c_L\rangle$ lies on tracks* 9 *and* 10 *from row* 0*. Tracks* 3, 4, 13, 14 *and* 15 *are unchanged, and if τ is a workspace of R rows, so is τ'.*

Proof. femit_b appends b to tracks 9 and 10 at the counter o on track 8 and advances it, in $36o + 68$ units. In $\mathsf{scan}(V_0, V_1)$ the head stands at $(k, 13)$, where k is the counter on track 5; while the presence cell holds 1 the body reads the value v_k and runs V_{v_k}, which ends at $(k+1, 13)$ with track 5 advanced: tskip in 39 units, temit_b in $36k + 36o + 141$ units. So each pass costs at most $72R + 144$ units with its test. The first scan emits one 1 for each 0 of v, that is, 1^L; femit_0 completes $\langle L\rangle$; reset_5 $(36|v| + 36)$ rewinds the record; the second scan copies v after it; fseal $(72o + 74)$ writes the terminator at $o = L + 1 + |v|$ and resets track 8, so that $(0, 8) = 0$. With the two walks tpres and home_{13} of each scan and $|v|, L \le R$, the total is at most $2000(R+1)^2$. Only tracks 5, 8, 9 and 10 of rows $j \le R$ are written. □

I.7.9 The stage Decode

Let $c_1, \dots, c_L$ be the natural numbers of the request u in order: d; for each coordinate the number of its A-terms and their indices; for each coordinate the number of its B-terms and their index pairs; the number q of payload integers; and M, T, n_0, i, m. The *value size* of the request is

$$V(u) = |u| + c_1 + \cdots + c_L + 1.$$

It counts the unary length of every number, because Decode writes every number in unary. It is still linear in X: if $c_{\mathrm{num}} = c_{\mathrm{req}} + \Sigma_0 + d + 3$, where Σ_0 is the sum of the numbers c_j other than i and m, then $V(u) \le c_{\mathrm{num}} X$, since $|u| + 1 \le c_{\mathrm{req}} + 2m$, $i \le d$ and $m \le X - 1$.

Definition I.56 (The decoded request)**.** A tape τ is a *decoded request* of u with r rows if

(a) τ carries the saved request of u (Definition I.41);

(b) $\tau(0, 8) = 0$, and the word $\langle L, c_1, \dots, c_L \rangle$ lies on tracks 9 and 10 from row 0;

(c) τ is a workspace of r rows.

The program of the stage is

$$\begin{aligned} \mathsf{Decode} &= \big(\mathsf{erase}; (\mathsf{init}; \mathsf{walk})\big); (\mathsf{reset}_5; \mathsf{fin}), \qquad \mathsf{erase} = \mathsf{raw}; \mathsf{write}\,0; \mathsf{unraw}, \\ \mathsf{init} &= \mathsf{clear}_9; \mathsf{clear}_{13}; \mathsf{zero}_2; \mathsf{zero}_5; \mathsf{zero}_8; \mathsf{zero}_{11}; \mathsf{zero}_{12}; \mathsf{skip}, \\ \mathsf{walk} &= [\mathsf{zero}_2, \mathsf{zero}_{11}, \mathsf{nat}, \mathsf{repeat}_2(\mathsf{row}_1), \mathsf{zero}_2, \mathsf{repeat}_2(\mathsf{row}_2), \\ &\qquad \mathsf{skips}, [\mathsf{nat}, \mathsf{nat}, \mathsf{nat}, \mathsf{nat}, \mathsf{nat}]]. \end{aligned}$$

It reads d, then the group of A-rows and the group of B-rows, each counted by the first word of the record, then q and the q frames of the payload, which it skips, and finally M, T, n_0, i and m.

Lemma I.57 (Preparing the scratch area and reading d)**.** *Let $R = |w_u|$. From $(-1, \tau_P)$ the program* erase; init; [zero$_2$, zero$_{11}$, nat] *runs to $(-1, \tau_1)$ in t_1 units with $t_1 + (R + 2d + 1) + 1 \le 200000\, V(u)^2$, and τ_1 is a reading configuration $(\langle d \rangle, l_1, R + 2d + 1)$ that lays out $f_1, \dots, f_F$ from row l_1, carries the saved request of u, carries 0 on tracks 2 and 12, and is a workspace of $R + 2d + 1$ rows.*

Proof. By Theorem I.42(a), erase finds the marker at $c = |u| - 1$, erases it and returns ($12c + 15$ units, Lemma I.34); the tape is now a workspace of R rows. init clears tracks 9 and 13 in rows $j \le R$ ($57R + 26$ and $65R + 34$, Lemma I.16); they are 0 in the rows beyond. It then puts the hole of tracks $2, 5, 8, 11, 12$ in row 0 (112 units, Lemma I.14(d)). The result is a reading configuration $(\varnothing, 0, R)$, with $\varnothing$ the empty record, that carries 0 on tracks 2 and 12 and, by Lemma I.44, lays out $f_0, \dots, f_F$ from row 0. As $c \le |u|$ and $R \le |u|$, these two parts take at most $309V(u)$ units. zero$_2$ and zero$_{11}$ (11 and 29 units) change nothing. nat reads $f_0 = \mathrm{bits}(d)$ within $10000(l_1 + R + d + 1)^2 \le 160000V(u)^2$ units (Lemma I.50). It keeps tracks 3, 4 and 15, tracks 2 and 12 in rows $j \le R$, and the cells $x \ge -1$. Adding the final skip of the list and $R + 2d + 2 \le 3V(u)$ gives the bound. □

Theorem I.58 (Decode). *From $(-1,\tau_P)$, the program* Decode *runs to $(-1,\tau_D)$ in some number t of units, where τ_D is a decoded request of u with*

$$r' = |w_u| + 2(c_1 + \cdots + c_L) + L$$

rows, and

$$t + r' + 1 \le K_d\, V(u)^8, \qquad K_d = 10^{16} + 400082.$$

Consequently $t + r' + 1 \le K_d\, c_{\mathrm{num}}^8 X^8$. With Pre_r the tape τ_P left by Parse *on a request (with $r = |w_u|$ rows) and $\mathrm{Post}_{r'}$ the decoded requests with r' rows,* Decode *meets a contract in the sense of Definition I.7 with exponent 8 and constant $K_d\, c_{\mathrm{num}}^8$, which depends only on the problem and n_0.*

Proof. Write $V = V(u)$ and $R = |w_u|$; note $R \le |u|$ and $1 \le V$. The program $\mathsf{erase};(\mathsf{init};\mathsf{walk})$ is the program of Lemma I.57 followed by the list

$$[\mathsf{repeat}_2(\mathsf{row}_1), \mathsf{zero}_2, \mathsf{repeat}_2(\mathsf{row}_2), \mathsf{skips}, [\mathsf{nat}, \mathsf{nat}, \mathsf{nat}, \mathsf{nat}, \mathsf{nat}]],$$

up to the final skip of the first list, which the joint list does not repeat; so it takes at most $t_1 + t_2$ units, where t_2 is the count of this list from τ_1. Every scale that occurs below is at most $12V$, by the same count as in Lemma I.51: the reading row plus the remaining frames is $R \le V$, the rows are at most $2V$, the record is at most V, and the numbers still to be read with their count are at most V.

First group. On τ_1 the record is $\langle d\rangle$, so track 14 carries a retained word of length d from row 0, and track 2 carries 0. Lemma I.53 with $\alpha = 1$ reads the d rows of A-terms within $d(36d + 10^6(12V)^3 + 58) + 18d + 38 \le 10^{12}V^4$ units, as $d \le V$.

zero_2 (11 units) returns the hole of track 2 to row 0; the retained word on track 14 is part of the record and has not changed. Lemma I.53 with $\alpha = 2$ reads the d rows of B-terms within $10^{12}V^4$ units.

Payload. Lemma I.54 reads q and skips the q signed frames. There the record has length $|w| \le V$, $q \le V$, and the rows are at most $2V$, so the first scale of Lemma I.54 is at most $6V$ and $a+\ell_q+\zeta+1 \le R+1 \le 2V$; its count is at most $100(V+1)+10000(6V)^2+V(72V+800V^2+58)+54V+39 \le 10^{12}V^4$.

Last numbers. Lemma I.51 with $k = 5$ reads M, T, n_0, i and m within $5\cdot 10000(12V)^2+1 \le 10^{12}V^4$ units.

With the final skip, $t_2 \le 4 \cdot 10^{12}V^4 + 12 \le 10^{16}V^4$, and the tape is a reading configuration $(\langle c_1, \dots, c_L\rangle, R, r')$ that carries the saved request and is a workspace of r' rows: each number c added $2c + 1$ rows. Since $t_1 \le 200000V^2$,

$$t_1 + t_2 \le (10^{16} + 200000)V^6.$$

Then reset_5 ($36R + 36 \le 72V$ units) sets the reading counter to 0, and fin (Lemma I.55) writes $\langle L, c_1, \dots, c_L\rangle$ on tracks 9 and 10 within $2000(r'+1)^2$ units. Its hypothesis $|\langle c_1, \dots, c_L\rangle|+L+1 \le r'$ holds because $|\langle c_1, \dots, c_L\rangle| = c_1 + \cdots + c_L + L$ and $L + 1 \le R + c_1 + \cdots + c_L$. It keeps tracks 3, 4, 15 and the workspace. As $r' + 1 \le 2V \le 10V$,

$$t + r' + 1 \le (10^{16} + 200000)V^6 + 72V + 200000V^2 + 10V \le (10^{16} + 400082)V^8.$$

The last claim follows from $V(u) \le c_{\mathrm{num}}X$ and $X \le X + r + 1$. □

After Decode every cell $x \geq 0$ holds 0 again (part (c) of Definition I.56): the request now exists only in the rows, as the saved request on tracks 3, 4 and 15 and as the numeric record on tracks 9 and 10. The exponent 8 arises as $4 + 2 + 2$: the traversal of the request is bounded in the fourth power of the value size, and each of the two compositions around it, with the first number and with the final copy, raises the exponent by two while keeping the bound in one variable.

Corollary I.59 (Reading the request in sequence)**.** *The program* Parse; Decode *runs from the head on cell* 0 *of the input tape of* u *to* $(-1, \tau_D)$*, a decoded request of* u *with* r' *rows, in* $t_P + t_D$ *units, where* $t_P = C_P(u)$ *and* t_D *is the count of Theorem I.58, and*

$$t_P + t_D + r' + 1 \leq A\, X^{11}, \qquad A = 100000(c_{\mathrm{req}} + 2)^3 + K_d\, c_{\mathrm{num}}^8.$$

Proof. By Theorem I.42(c), $t_P \leq 100000(c_{\mathrm{req}} + 2)^3 X^3$, and by Theorem I.58, $t_D + r' + 1 \leq K_d\, c_{\mathrm{num}}^8 X^8$. Since $X \geq 1$, both powers are at most X^{11}. Add. □

The bound has the form of the hypothesis of Lemma 15.36(a) with $N = X$ and $e = 11$, so the stages after Decode compose with it directly. The constant A depends on the problem and n_0 only; the exponent $11 = 8 + 3$ is a fixed number.

I.8 The problem's constants: preparing the request and folding the field

Where does the machine find the number β of a field? Take $y' = -y^2$ with $y_0 = \frac{1}{2}$. Its request (Definition 15.32) has the shape $d = 1$, no A-term, one B-term with indices $(0, 0)$, then the payload count 6 and the six integers $1, 2$ (the numerator and denominator of y_0), $0, 1$ (of c_0) and $-1, 1$ (of the coefficient of the B-term), followed by M, T, n_0, i and m. After the request has been read (Section I.7), the bits of every frame lie on tracks 3 and 4 of the scratch area, with a mark on track 15 at the first bit of each frame after the first. By Definition 15.24, β is the largest, over the coordinates, of the sum of $\lceil |b| \rceil$ over the B-terms of the coordinate; here $\beta = \lceil |-1/1| \rceil = 1$. The machine finds it in four moves: it writes $|-1| = 1$ and the denominator 1 in unary, divides the first by the second with rounding up, adds the result over the B-terms of the coordinate, and keeps the largest row sum. The number 1 is then appended to a list kept on the tape, the *archive*. The same pattern, a number for each coordinate and then the largest of them, gives γ, α, the capacity s of the field, the row length L and the capacity s_0 of the initial point (Section H.1); the capacities count binary digits, which the machine reads off the lengths of the frames. For this request the archive ends as

$$(p, \gamma, \alpha, \beta, s, L, s_0) = (11, 0, 0, 1, 3, 1, 3),$$

where $p = 11$ is the row of the saved request at which the payload begins: the frames before it are those of $d = 1$ (2 bits), of the shape numbers 0, 1, 0, 0 ($1 + 2 + 1 + 1$ bits) and of $q = 6$ (4 bits).

This section constructs the first half of the stage "Schedule" of Table 15.1: from the decoded request to the archive of these seven numbers. Section I.9 computes σ, S, κ, N, the capacity polynomial and the width from the archive. The three programs of this section are

$$\begin{aligned} \mathsf{Prepare} &= \mathsf{norm};\mathsf{grow};\mathsf{norm};\mathsf{prec};\mathsf{clear}_{2,5,6,7,12,13,14},\\ \mathsf{MetaBank} &= \mathsf{Meta};\mathsf{Bank},\\ \mathsf{Summ} &= \mathsf{clear}_9;\mathsf{Gam};\mathsf{pass}(\mathsf{row}_A,2);\mathsf{pass}(\mathsf{row}_B,3);\\ &\qquad \mathsf{pass}(\mathsf{row}_W,4);\mathsf{pass}(\mathsf{row}_L,5);\mathsf{Ini}, \end{aligned}$$

run in this order. Prepare (Section I.8.1) reads the precision m into unary and blanks the scratch tracks. Meta archives the payload row p and Bank writes in unary the magnitude of every integer of the request (Section I.8.3). Summ archives γ (Gam), then the four row folds for α, β, s and L, then s_0 (Ini) (Section I.8.7). Every routine they use that the foundation sections do not prove is constructed and counted below, from the routines of Sections I.3 to I.5.

The objects on the tape

Throughout, ω is the saved request of Definition I.41: the bits of d followed by the bits of every later frame of the request, on track 3 (presence) and track 4 (bits) from row 0, with the frame-start marks on track 15. The payload integers are $z_1, \dots, z_q$, in the order of Definition 15.32: the numerators and denominators of y_0, of c, of the A-coefficients coordinate by coordinate, and of the B-coefficients. In this section a *frame of* ω means the bits of a frame of the request as they lie in ω: the unary length and the separator are not stored, and the first bit of every frame after the first carries the frame-start mark. The frame of an integer z has $\mathrm{bl}(|z|)+2$ bits: the bit $[z<0]$ and the $\mathrm{bl}(|z|)+1$ lowest binary digits of $|z|$. We write p for the row of ω at which the frame of z_1 begins, and p_j for the row at which the frame of z_j begins, so that $p_1 = p$ and $p_{j+1} = p_j + \mathrm{bl}(|z_j|) + 2$. For $n \in \mathbb{N}$ let $u(n) = 1^n 0$, and for numbers $n_1, \dots, n_k$ let $\mathrm{rec}(n_1, \dots, n_k) = u(k)\,u(n_1)\cdots u(n_k)$; the *numeric record* of Definition I.56 is this word on tracks 9 (presence) and 10 (values) from row 0, with $\tau(0,8) = 0$.

Definition I.60 (Bank, archive and gather state)**.** Fix a problem, i and m, and let τ be a tape and $R \in \mathbb{N}$.

(a) The *bank list* is $\mathbf{b} = (d,\ \text{the shape numbers},\ q,\ |z_1|, \dots, |z_q|,\ M,\ T,\ n_0,\ i,\ m)$: the control numbers of the request in their order, with the magnitude of every payload integer inserted after q. The tape carries the *bank* if the word $\mathrm{rec}(\mathbf{b})$ lies on tracks 9 and 10 from row 0 and $\tau(0,8) = 0$. It has a *bank view* within R rows if there are a tape τ_0 that carries the bank and is a workspace of at most R rows, and $\tau(j,10) = \tau_0(j,10)$ for every j: only the values of the bank are kept, and track 9 is free.

(b) For numbers $v_1, \dots, v_k$, τ holds the *archive* $(v_1, \dots, v_k)$ *within* R *rows* if ω lies on tracks 3 and 4 from row 0 (so $\tau(|\omega|,3) = 0$), the word $u(v_1)\cdots u(v_k)$ lies on tracks 3 and 4 from row $|\omega|+1$, and

$$|\omega| + 1 + \sum_{j \le k} (v_j + 1) \le R.$$

In particular every v_j is at most R.

(c) τ is in the *gather state* with archive $(v_1, \dots, v_k)$ and R rows if it holds that archive within R rows, has a bank view within R rows, has on track 15 the frame-start marks of the saved request, is a workspace of R rows with origin marks up to R, and $v_1 = p$.

The archive is an extension of the saved request: its words follow ω on the same two tracks, after the blank cell $(|\omega|, 3)$ that ends ω. So ω and the archive together are one finite stream, and the final erase of Section I.15 clears both with the rest of the rows.

The contracts

The stage results are stated here and proved at the ends of Sections I.8.1 and I.8.7. A *prepared request with r rows* is a decoded request with r rows (Definition I.56) that has origin marks up to r, holds the value m on track 11, and has 0 in every cell of tracks 2, 5, 6, 7, 12, 13 and 14.

Lemma I.61 (Preparing the request)**.** *From every decoded request with r rows,* Prepare *runs from the root to the root in some t units to a prepared request with $r+1$ rows, and $t \le 4000(r+2)^2$. The saved request and the numeric record are kept.*

Theorem I.62 (The archive of the seven constants)**.** *Fix a problem. There are constants K_1 and K_2, depending only on the problem, with the following properties for every i, m, r and tape τ.*

(a) If τ is a prepared request with r rows, then MetaBank *runs from the root to the root in some t units to a tape that holds the archive (p) within r' rows, carries the bank on tracks 9 and 10, keeps the saved request, is a workspace of r' rows with origin marks up to r', and satisfies*

$$t + (r' + m) + 2 \le K_1(r + m + 2)^{24}.$$

(b) If τ is in the gather state with archive (p) and r rows, then Summ *runs from the root to the root in some t units to a tape in the gather state with archive*

$$(p,\ \gamma,\ \alpha,\ \beta,\ s,\ L,\ s_0)$$

and $r + \rho_S$ rows, where ρ_S depends only on the problem, and

$$t + (r + \rho_S + m) + 2 \le K_2(r + m + 2)^3.$$

(c) If τ is a prepared request with r rows, then MetaBank; Summ *runs from the root to the root in some t units to a tape that holds the archive $(p, \gamma, \alpha, \beta, s, L, s_0)$ within r'' rows, keeps the saved request, is a workspace of r'' rows with origin marks up to r'', and*

$$t + (r'' + m) + 2 \le (K_1 + K_2K_1^3)(r + m + 2)^{72}.$$

Here γ, α and β are the bounds of Definition 15.24, and s, L and s_0 those of Section H.1.

I.8.1 Preparing the request

A decoded request leaves the scratch tracks in an unknown state, and it need not have origin marks. The first routine writes them from the row headers.

Lemma I.63 (The origin corridor)**.** *Let*

$$\mathsf{corr} = \mathsf{left}^2;\mathsf{write}\,0;\mathsf{right}^1;\mathsf{branch}(\mathsf{right}^1,\ \mathsf{corr}^+),$$
$$\mathsf{corr}^+ = \mathsf{left}^{16};\mathsf{loop}(\mathsf{left};\mathsf{write}\,1;\mathsf{left}^{15});\mathsf{left};\mathsf{write}\,1;\mathsf{home}_1,$$

and $\mathsf{norm} = \mathsf{corr};\mathsf{clear}_{2,5,6,7,11,12,13,14}$*. If τ is a workspace of r rows, then* corr *runs from $(-1,\tau)$ to $(-1,\tau')$ in 9 units if $r=0$ and in $37r+13$ units if $r\ge 1$, where τ' is τ with $(0,1)$ set to 0 and $(j,1)$ set to 1 for $1\le j\le r$. So τ' is a workspace of r rows with origin marks up to r. Then* norm *runs from $(-1,\tau)$ to the root within $489r+218$ units, and in addition sets every cell of tracks 2, 5, 6, 7, 11, 12, 13 and 14 to 0; no other track changes.*

Proof. left^2, $\mathsf{write}\,0$ and right^1 ($3+1+2$ units) clear $(0,1)=-3$ and stop on the header $(0,0)$. If $r=0$ the header is 0, and the branch and right^1 return to the root: 9 units. If $r\ge 1$, left^{16} (17 units) reaches $(1,0)$. Let I_j, for $1\le j\le r$, be the configuration at $(j,0)$ in which $(i,1)=1$ for $1\le i<j$. A pass from I_j with $j<r$ scans the header 1, writes the origin mark of row j and moves to $(j+1,0)$: 19 units with its test. At I_r the header is the sentinel 0. Lemma I.6 gives $19(r-1)+1$ units. Then left and $\mathsf{write}\,1$ mark row r, and home_1 from row r costs $18r+5$ (Lemma I.11). In all $7+17+19(r-1)+1+2+18r+5=37r+13$. Only track 1 is written, in rows $j\le r$, so τ' is again a workspace. For norm, Lemma I.16 applies to τ' and adds $\sum_k\big((2k+39)r+2k+8\big)+1=452r+205$ units over the eight tracks, whose numbers sum to 70; cells of rows $j>r$ are 0 already. The total is at most $489r+218$. □

The precision m is the last word of the numeric record. The next routine reads a unary word backwards, from its last mark to the blank cell before it, and counts its marks on track 11; it serves twice in this section.

Lemma I.64 (Reading a word backwards)**.** *Let*

$$\mathsf{tail} = \mathsf{loop}(\mathsf{home}_{10};\mathsf{inc}_{11};\mathsf{seek}_2;\mathsf{right}^{16};\mathsf{write}\,0;\mathsf{left}^8);\mathsf{home}_{10}.$$

Let $s,n,R\in\mathbb{N}$ with $s+n\le R$, and let τ have origin marks up to R, carry the counter $s+n$ on track 2 and the counter 0 on track 11, and satisfy $\tau(s,10)=0$ and $\tau(s+j,10)=1$ for $1\le j\le n$. Then tail *runs from $(\mathrm{cell}(s+n,10),\tau)$ to $(-1,\tau')$ in exactly*

$$n\big(36(s+n)+113\big)+1+18s+14$$

units. In τ' track 2 carries s and track 11 carries n; only the cells $(j,2)$, $s\le j<s+n$, and $(j,11)$, $j\le n$, differ from τ.

Proof. For $j\le n$ let I_j be the configuration at $\mathrm{cell}(s+n-j,10)$ in which track 2 carries $s+n-j$ and track 11 carries j. From I_j with $j<n$ the scanned cell is a mark. Put $h=s+n-j$. The pass returns to the root ($18h+14$ units), advances the counter on track 11 ($36j+66$, Lemma I.14; the origin marks reach $j+1\le R$), goes to the hole $(h,2)$ ($18h+5$), steps to $(h-1,2)$ and writes 0 there (18), so that track 2 now carries $h-1$, and moves to $(h-1,10)$ (9). With its test it costs $36h+36j+113=36(s+n)+113$ units and ends in I_{j+1}. At I_n the scanned cell is $(s,10)=0$. Lemma I.6 gives $n(36(s+n)+113)+1$ units, and home_{10} from row s costs $18s+14$. □

Lemma I.65 (The precision). *Let* $\mathsf{last} = \mathsf{acc}^{14}_{9\to 2}; \mathsf{seek}_2; \mathsf{right}^{32}; \mathsf{write}\,0; \mathsf{left}^{8}$ *and* $\mathsf{prec} = \mathsf{last}; \mathsf{tail}$. *Let τ be a workspace of R rows with origin marks up to R that carries the numeric record $w = \mathrm{rec}(n_1, \dots, n_k)$, $k \geq 1$, with $|w| \leq R$, and let track 2 carry the counter 0 and track 11 hold the value 0. Then* prec *runs from the root to the root within $1000(R+1)^2$ units to a tape on which track 11 holds the value n_k; only tracks 2, 11 and 14 change, so the workspace and the origin marks are kept.*

Proof. Put $\ell = |w|$. Track 9 carries the counter ℓ, so by Lemma I.15(b) $\mathsf{acc}^{14}_{9\to 2}$ makes track 2 carry ℓ in $36\ell^2 + 134\ell + 72$ units, writing only tracks 2 and 14. seek_2 reaches $(\ell, 2)$ $(18\ell + 5)$, right^{32} moves two rows back to $(\ell - 2, 2)$ (33), $\mathsf{write}\,0$ makes track 2 carry $\ell - 2$, and left^{8} reaches $(\ell - 2, 10)$ (9). The last word $u(n_k)$ of w occupies the rows $\ell - n_k - 1, \dots, \ell - 1$ of track 10, and the word before it, which exists because $k \geq 1$, ends with its blank cell at row $s = \ell - n_k - 2$. So Lemma I.64 applies with this s and $n = n_k$, and takes $n_k(36(\ell - 2) + 113) + 1 + 18s + 14$ units; track 11, blank before, now holds the value n_k. Since $n_k \leq \ell \leq R$ the total is at most $72\ell^2 + 211\ell + 135 \leq 1000(R+1)^2$. □

Proof of Lemma I.61. Let τ be a decoded request with r rows. Its control numbers end with m, so its numeric record is $\mathrm{rec}(n_1, \dots, n_k)$ with $n_k = m$, $k \geq 1$. The five parts of $\mathsf{Prepare}$ act as follows.

(i) norm writes the origin marks up to r and blanks the eight scratch tracks, within $489r + 218$ units (Lemma I.63).

(ii) grow makes a workspace of $r + 1$ rows with origin marks up to $r + 1$ in $72r + 99$ units (Lemma I.17). It writes the header $(r, 0)$ and the new row $r + 1$, which was blank and receives 0 on tracks 3, 4, 9, 10 and 15.

(iii) norm on $r + 1$ rows, within $489(r + 1) + 218$ units, blanks the marks that grow left on tracks 2, 5, 11 and 14 of row $r + 1$.

(iv) The numeric record has length at most $r + 1$, because its presence marks lie in rows $j \leq r$; track 2 carries 0 and track 11 holds 0. Lemma I.65 with $R = r + 1$ writes the value m on track 11 within $1000(r + 2)^2$ units.

(v) $\mathsf{clear}_{2,5,6,7,12,13,14}$ on $r + 1$ rows costs $391(r + 1) + 175$ units (Lemma I.16).

No part writes tracks 3, 4, 9, 10 or 15 outside the blank row $r + 1$, nor the cell $(0, 8)$, so the saved request and the numeric record are kept, and the result is a prepared request with $r + 1$ rows. The total is at most $1441r + 1590 + 1000(r + 2)^2 \leq 4000(r + 2)^2$. □

I.8.2 Counters, growth and the archive

Most routines of the schedule move a unary number from one track to another, or step a counter back by one. In this subsection and the rest of the section, "within R rows" includes that the tape is a workspace of R rows with origin marks up to R.

Lemma I.66 (Clean copies and steps back). *For distinct tracks $k, l \in \{2, \dots, 15\}$ put $\mathsf{cc}_{k\to l} = \mathsf{clear}_l; \mathsf{copy}_{k\to l}$ and $\mathsf{retr}_k = \mathsf{seek}_k; \mathsf{right}^{16}; \mathsf{write}\,0; \mathsf{home}_k$.*

(a) If τ is a workspace of R rows with origin marks up to R and track k carries a counter $n \le R$, then $\mathsf{cc}_{k\to l}$ runs from the root to the root within $(2l+39)R+2l+9+100(n+1) \le 300(R+1)$ units, and afterwards track l holds the value n; only track l changes.

(b) If track k carries the counter $j+1$ and τ has origin marks up to j, then retr_k runs from the root to the root in $36j + 2k + 43$ units and sets (j, k) to 0, so that track k carries j.

Proof. (a) clear_l followed by the final skip costs $(2l + 39)R + 2l + 9$ units and leaves track l blank in every row (Lemma I.16; rows beyond R are blank already). $\mathsf{copy}_{k\to l}$ costs at most $100(n + 1)$ and writes 1 in (j, l) for $j < n$ and 0 in (n, l) (Lemma I.15(a)), so track l holds the value n. With $l \le 15$ the count is at most $169R + 139$. (b) seek_k reaches the hole $(j + 1, k)$ in $18(j + 1) + k + 3$ units, right^{16} and $\mathsf{write}\,0$ (18) clear (j, k), and home_k from row j costs $18j + k + 4$ (Lemma I.11). □

New rows are added by the count held on track 12, one grow per unit, with track 2 as the loop counter.

Lemma I.67 (Growing by a count). *Let $\mathsf{rdc} = \mathsf{seek}_2; \mathsf{shift}_{2\to 12}$ and*

$$\mathsf{alloc} = \mathsf{grow}; \mathsf{rdc}; \mathsf{loop}(\mathsf{home}_{12}; \mathsf{grow}; \mathsf{inc}_2; \mathsf{rdc}); \mathsf{home}_{12}.$$

Let τ be a workspace of R rows with origin marks up to R on which track 2 carries the counter 0 and track 12 holds the value c. Then alloc runs from the root to the root within $2000(R + c + 2)^2$ units to a workspace of $R + 1 + c$ rows with origin marks up to $R + 1 + c$ on which track 12 still holds c. Besides the new rows and the headers of the rows $R, \dots, R + c$, only track 2 is written.

Proof. The marks of the value c lie in rows $j \le R$, so $c \le R + 1$. The first grow ($72R + 99$ units, Lemma I.17) gives a workspace of $R + 1$ rows; its new row has 0 on track 12 and track 2 still carries 0. rdc ($5 + 11$ units) reaches $(0, 12)$. For $j \le c$ let I_j be the configuration at $(j, 12)$ whose tape is a workspace of $R + 1 + j$ rows with origin marks up to $R + 1 + j$, carries j on track 2 and holds c on track 12. A pass from I_j, $j < c$, scans the mark $(j, 12)$, returns to the root $(18j + 16)$, adds a row $(72(R + 1 + j) + 99)$, advances track 2 $(36j + 48)$ and reads $(j + 1, 12)$ $(18j + 34)$: $72(R+1)+144j+198$ units with its test, ending in I_{j+1}. At I_c the scanned cell is the blank $(c, 12)$. With the final home_{12} $(18c+16)$ the total is $72R+132+72Rc+72c^2+216c \le 2000(R+c+2)^2$. □

The archive of Definition I.60 is read and written through a cursor on track 13. The routines are

$$\begin{aligned}
\mathsf{start} &= \mathsf{cc}_{3\to 13}; \mathsf{inc}_{13}, \qquad \mathsf{rv} = \mathsf{seek}_{13}; \mathsf{shift}_{13\to 4},\\
\mathsf{rd} &= \mathsf{clear}_{12}; \mathsf{rv}; \mathsf{loop}(\mathsf{home}_4; \mathsf{inc}_{12}; \mathsf{inc}_{13}; \mathsf{rv}); \mathsf{home}_4; \mathsf{inc}_{13},\\
\mathsf{term} &= \mathsf{seek}_{13}; \mathsf{shift}_{13\to 3}; \mathsf{write}\,0; \mathsf{home}_3, \qquad \mathsf{init} = \mathsf{grow}; \mathsf{grow}; \mathsf{start}; \mathsf{term},\\
\mathsf{load}_{k,l} &= \mathsf{start}; \mathsf{rd}^k; \mathsf{rd}; \mathsf{cc}_{12\to l}, \qquad \mathsf{prefix}_k = \mathsf{start}; \mathsf{rd}^k,\\
\mathsf{trunc}_k &= \mathsf{start}; \mathsf{rd}^k; \mathsf{term},
\end{aligned}$$

where rd^k is k copies of rd followed by skip; and, for appending,

$$\begin{aligned}\mathsf{sbuf} &= \mathsf{cc}_{13\to 2};\mathsf{cc}_{11\to 6};\mathsf{inc}_6;\mathsf{cc}_{11\to 7};\mathsf{zero}_5,\\ \mathsf{append} &= \mathsf{cc}_{11\to 12};\mathsf{zero}_2;\mathsf{alloc};\mathsf{sbuf};\mathsf{fapp};\mathsf{cc}_{2\to 13},\\ \mathsf{tc} &= \mathsf{right};\mathsf{write}\,0;\mathsf{right};\mathsf{loop}(\mathsf{right}^{16};\mathsf{left};\mathsf{write}\,1;\mathsf{right});\mathsf{right}^2,\\ \mathsf{end} &= \mathsf{seek}_{13};\mathsf{shift}_{13\to 3};\mathsf{scanL};\mathsf{tc};\mathsf{cc}_{2\to 13}, \qquad \mathsf{push} = \mathsf{start};\mathsf{end};\mathsf{append}.\end{aligned}$$

Here fapp is the frame-append routine of Lemma I.39(b), named app there and used here with $a = 0$: from a tape on which track 2 carries w, track 5 carries 0 and a word x lies on the binary stream (tracks 6 and 7 from row 0), with origin marks up to $w + |x|$, it writes x on tracks 3 and 4 from row w, writes 0 in $(w + |x|, 3)$, leaves track 2 carrying $w + |x|$ and track 5 carrying 0, changes no other track and no row before w of tracks 3 and 4, and takes $54|x|^2 + (36w + 184)|x| + 36w + 71$ units.

Lemma I.68 (The archive routines)**.** *Let τ be within R rows and let ω lie on tracks 3 and 4 from row 0. Each routine runs from the root to the root.*

(a) If $|\omega| + 1 \le R$, start makes track 13 hold the value $|\omega| + 1$, within $1000(R + 1)$ units; only track 13 changes.

(b) If track 13 carries a, $\tau(a + j, 4) = 1$ for $j < n$, $\tau(a + n, 4) = 0$ and $a + n + 1 \le R$, then rd makes track 12 hold the value n and track 13 carry $a + n + 1$, within $10000(R + 1)^2$ units; only tracks 12 and 13 change. So rd^k passes k consecutive words $u(n_1)\cdots u(n_k)$ of track 4 within $(k + 1)\cdot 10000(R + 1)^2$ units.

(c) If τ holds an archive within R rows whose entries, numbered from 0, are $v_0, v_1, \ldots$, if it has an entry v_k, and if $5 \le l \le 11$, then $\mathsf{load}_{k,l}$ makes track l hold the value v_k, within $(k + 3)\cdot 20000(R + 1)^2$ units; only tracks 12, 13 and l change.

(d) If track 13 carries $a \le R$, term sets $(a, 3)$ to 0 in $36a + 35$ units.

(e) init leaves the archive $()$ within $R + 2$ rows, with the value $|\omega| + 1$ on track 13; every track $k \ge 4$, $k \ne 13$, that carried a counter at most R still carries it; and its count t satisfies $t + (R + 2) + 2 \le 3000(R + 3)$.

(f) If track 13 carries $w \le R$ and track 11 carries $n \le R$, then append writes the word $u(n)$ on tracks 3 and 4 from row w followed by 0 in $(w + n + 1, 3)$, leaves the rows before w of tracks 3 and 4 and all of track 15 unchanged, makes track 13 hold $w + n + 1$, ends within $R + 1 + n$ rows, keeps every counter at most R on tracks 8 to 11, and takes at most $100000(R + w + n + 3)^2$ units.

(g) If track 13 carries a and a word of length ℓ lies on tracks 3 and 4 from row a, with $a + \ell \le R$, then end makes track 13 hold the value $a + \ell$ within $2000(R + 1)$ units; only tracks 2 and 13 change.

(h) If τ holds the archive $(v_1, \ldots, v_k)$ within R rows, track 11 carries $n \le R$ and track 9 carries $c \le R$, then push leaves the archive $(v_1, \ldots, v_k, n)$ within $R + 1 + n$ rows, track 9 still carrying c, track 10 and track 15 unchanged, within $10^6(R + n + 2)^2$ units.

(i) If τ holds the archive $(v_1, \dots, v_k, w_1, \dots, w_l)$ within R rows, then prefix_k makes track 13 carry $e = |\omega| + 1 + \sum_{j \le k}(v_j + 1)$ and changes only tracks 12 and 13; trunc_k moreover writes 0 in $(e, 3)$, so that the tape holds the archive $(v_1, \dots, v_k)$ within R rows, with the value e on track 13. Each takes at most $(k+3) \cdot 20000(R+1)^2$ units.

Proof. (a) Track 3 carries $|\omega|$. $\mathsf{cc}_{3\to 13}$ costs at most $65R + 35 + 100(|\omega| + 1)$ (Lemma I.66) and inc_{13} costs $36|\omega| + 70$; the sum is at most $201R + 135$. The new hole of track 13 is at row $|\omega| + 1$, and the cells beyond it were blank.

(b) clear_{12} costs $63R + 33$ and rv, from the counter a, $18a + 26$, ending at $(a, 4)$. For $j \le n$ let I_j be the configuration at $(a + j, 4)$ on which track 13 carries $a + j$ and track 12 holds the value j. A pass from I_j, $j < n$, scans a mark, returns to the root $(18(a + j) + 8)$, advances tracks 12 and 13 ($36j + 68$ and $36(a + j) + 70$) and reads the next cell ($18(a + j) + 44$): $72(a + j) + 36j + 191$ units with its test. At I_n the scanned cell is $(a + n, 4) = 0$; home_4 and inc_{13} add $54(a + n) + 78$. Since $a + n < R$ the total is at most $108R^2 + 326R + 138 \le 10000(R + 1)^2$. For rd^k, each copy starts at the first cell of the next word.

(c) start puts the cursor at the first word $u(v_0)$ of the archive, rd^k passes $u(v_0), \dots, u(v_{k-1})$, the next rd leaves v_k on track 12, and $\mathsf{cc}_{12\to l}$ copies it. The archive inequality gives the hypotheses of (a) and (b). The total is at most $1000(R + 1) + (k + 2) \cdot 10000(R + 1)^2 + 300(R + 1)$.

(d) $(18a + 16) + 11 + 1 + (18a + 7)$.

(e) Two rows are added ($144R + 270$ units). The marks of ω lie in rows $j \le R$, so $|\omega| + 1 \le R + 2$, and start ($\le 1000(R + 3)$) and term at row $|\omega| + 1$ ($\le 36(R + 2) + 35$) write the empty word at row $|\omega| + 1$. The sum plus $(R + 2) + 2$ is at most $1181R + 3381$. A counter at most R has its hole in an old row, which grow does not write, and start and term write only track 13 and $(|\omega| + 1, 3)$.

(f) $\mathsf{cc}_{11\to 12}$ and zero_2 prepare Lemma I.67 with $c = n$, which gives $R' = R + 1 + n$ rows; the new rows do not disturb counters whose holes lie in old rows, so track 13 still carries w and track 11 carries n. sbuf makes track 2 hold w, track 6 hold $n + 1$, track 7 hold n and track 5 carry 0: the binary stream is the word $u(n)$ of length $n + 1$. Then fapp writes $u(n)$ from row w with its terminator, and $\mathsf{cc}_{2\to 13}$ copies the new end $w + n + 1 \le R'$. The parts cost at most $300(R+1)+11$, $2000(R+n+2)^2$, $900(R'+1)+36n+73$, $54(n+1)^2+(36w+184)(n+1)+36w+71$ and $300(R' + 1)$, together less than $100000(R + w + n + 3)^2$. Tracks 8 to 11 and 15 are written only in the new rows, and tracks 3 and 4 only from row w.

(g) seek_{13} and $\mathsf{shift}_{13\to 3}$ ($18a + 27$) reach $(a, 3)$; scanL ($18\ell + 1$) reaches the blank $(a + \ell, 3)$. With $h = a + \ell$, tc moves to $(h, 2)$ and blanks it, moves to $(h, 1)$, and then, while the origin mark of the current row is 1, steps to the previous row, writes 1 on track 2 and returns to track 1 (21 units a pass); at the blank $(0, 1)$ it stops and right^2 reaches the root. It costs $21h + 7$ units and leaves track 2 carrying h. $\mathsf{cc}_{2\to 13}$ copies it. The total is at most $204R + 170$.

(h) By (a) and (g), $\mathsf{start}; \mathsf{end}$ puts on track 13 the end $e = |\omega| + 1 + \sum_j (v_j + 1) \le R$ of the archive within $3000(R + 1)$ units, and (f) with $w = e$ appends $u(n)$ there within $100000(2R + n + 3)^2 \le 400000(R + n + 2)^2$ units. The new end is $e + n + 1 \le R + 1 + n$, and the word ω and the old entries, which lie before row e, are kept. Track 9 is a counter of tracks 8 to 11, and no part writes track 10 or 15 outside the new blank rows.

(i) By (a) and (b), within $1000(R+1)+(k+1)\cdot 10000(R+1)^2$ units. term adds $36e+35$ and cuts the archive after its k-th entry; the span only shrinks. $\square$

I.8.3 The payload address and the magnitude bank

The decoding stage of Section I.7 reads the frames of the saved request in order and appends each control number, in unary, to a temporary record on tracks 13 and 14. The front of the schedule runs the same traversal twice more: once over the frames before the payload, to find the payload row p, and once over all frames, writing the magnitude of each payload integer instead of skipping it. We recall the routines and their contracts. A tape is in the *traversal state* $T(v,s,R)$ if it is a workspace of R rows with origin marks up to R, the word v lies on tracks 13 (presence) and 14 (values) from row 0, track 13 is blank from row $|v|$ on, $|v|\le R$, track 11 carries $|v|$, track 5 (the reading cursor) carries s, track 9 is blank and track 8 carries 0. We say that the frame of a natural number n *begins at row* s if the saved request carries the $\mathrm{bl}(n)+1$ bits of n on rows $s,\dots,s+\mathrm{bl}(n)$, with no frame-start mark on rows $s+1,\dots,s+\mathrm{bl}(n)$, followed by a frame-start mark or the end of ω.

Lemma I.69 (The decoding routines)**.** *Let*

$$\mathsf{ini}=\mathsf{clear}_9;\mathsf{clear}_{13};\mathsf{zero}_2;\mathsf{zero}_5;\mathsf{zero}_8;\mathsf{zero}_{11};\mathsf{zero}_{12};\mathsf{skip},$$

and let app, row_a *(*$a=1,2$*) and* fin *be the routines* nat, row_α *and* fin *of Section I.7. Let the tape keep the saved request.*

(a) On a workspace of R rows with origin marks up to R, ini *runs from the root to the root in $122R+172$ units to a tape in $T(\langle\rangle,0,R)$ on which tracks 2 and 12 carry 0.*

(b) If the tape is in $T(v,s,R)$, the frame of n begins at row s and $s+\mathrm{bl}(n)+1\le R$, then app *runs from the root to the root, within $10000\big(s+\mathrm{bl}(n)+1+R+|v|+n+1\big)^2$ units, to a tape in $T\big(vu(n),\ s+\mathrm{bl}(n)+1,\ R+2n+1\big)$ that keeps the saved request, the cells of tracks 2 and 12 in rows $j\le R$ and those of tracks 13 and 14 in rows $j<|v|$.*

(c) Let $a\in\{1,2\}$, and let the frames of ω from row s begin with, for each of d coordinates in turn, the frame of a count k followed by ak frames of natural numbers; let $c_1,\dots,c_g$ be the numbers of these frames and $s'=s+\sum_j(\mathrm{bl}(c_j)+1)$. If the tape is in $T(v,s,R)$, carries on track 14 a retained word of length d from row r_0 with $r_0+d+1\le|v|$, and carries r_0 on track 2, then $\mathsf{repeat}_2(\mathsf{row}_a)$ *runs from the root to the root to a tape in $T\big(vu(c_1)\cdots u(c_g),\ s',\ R+\sum_j(2c_j+1)\big)$ that keeps the saved request and the retained word, with r_0+d on track 2, within*

$$d\big(36(r_0+d)+10^6Z^3+58\big)+36r_0+18d+38$$

units, where $Z=|\omega|+R+|v|+4\sum_j c_j+3g+1$.

(d) If the word $u(n_1)\cdots u(n_k)$ lies on tracks 13 and 14 from row 0, tracks 5 and 8 carry 0, the tape is within R rows and $\sum_j(n_j+1)+k+1\le R$, then fin *runs from the root to the root within $2000(R+1)^2$ units to a tape within R rows that carries $\mathrm{rec}(n_1,\dots,n_k)$ on tracks 9 and 10, with $\tau(0,8)=0$, and keeps tracks 3, 4, 13, 14 and 15.*

Proof. (a) Lemmas I.16 and I.14(d): $(57R+26)+(65R+34)+(11+17+23+29+31)+1$. Track 9 and track 13 are then blank, and tracks 2, 5, 8, 11 and 12 carry 0, which is $T((),0,R)$.

The other parts are results of Section I.7 in the notation of this section. A tape in $T(v,s,R)$ is a reading configuration (v,s,R) (Definition I.47) that is a workspace, and a frame of n that begins at row s is the layout of bits(n), possibly followed by further frames, from row s (Definition I.43). Part (b) is Lemma I.50, whose scale is the one displayed in (b); the tape stays a workspace, since nat writes no cell $x \geq -1$. Part (c) is Lemma I.53 with the retained word and the counter of track 2 starting at row r_0 instead of 0: its proof, Lemma I.18 with $c=2$ from row r_0 and the body bound of Lemma I.52, gives the stated count, and the scale of the numbers read there is at most Z, because the frames read lie in ω. Part (d) is Lemma I.55 with $v = u(n_1)\cdots u(n_k)$ and $L=k$, since $\mathrm{rec}(n_1,\dots,n_k) = u(k)\,u(n_1)\cdots u(n_k)$. □

The magnitude of a payload integer is read by skipping its sign bit.

Lemma I.70 (A signed frame read as its magnitude)**.** *Let* $\mathsf{sapp} = \mathsf{inc}_5;\mathsf{app}$*. If the frame of the payload integer z begins at row s of ω, the tape is in $T(v,s,R)$ and keeps the saved request, and $s+\mathrm{bl}(|z|)+2 \leq R$, then* sapp *runs from the root to the root, within* $10100\,Y^2$ *units with* $Y = s+\mathrm{bl}(|z|)+2+R+|v|+|z|+1$*, to a tape in* $T\big(vu(|z|),\ s+\mathrm{bl}(|z|)+2,\ R+2|z|+1\big)$ *with the same preservation as in Lemma I.69(b).*

Proof. inc_5 ($36s+54$ units) moves the reading cursor to $s+1$ and writes only track 5. The frame of z consists of $[z<0]$ at row s and the $\mathrm{bl}(|z|)+1$ binary digits of $|z|$ on the rows after it, with no frame-start mark before the next frame. So the frame of the natural number $|z|$, in the sense above, begins at row $s+1$, and Lemma I.69(b) applies. Since $36s+54 \leq 100Y^2$, the total is at most $10100Y^2$. □

Lemma I.71 (The payload row)**.** *Let*

$$\mathsf{paddr} = \mathsf{ini};\mathsf{app};\mathsf{repeat}_2(\mathsf{row}_1);\mathsf{zero}_2;\mathsf{repeat}_2(\mathsf{row}_2);\mathsf{app};\mathsf{cc}_{14\to9};\mathsf{cc}_{9\to14},$$
$$\mathsf{Meta} = \mathsf{paddr};\mathsf{cc}_{5\to11};\mathsf{init};\mathsf{append}.$$

Let τ keep the saved request and be within R rows. Let D be the numeric size of the request ($|u|$ plus the sum of the control numbers plus 1*, as in Theorem I.58) and $N = R+D+1$.*

(a) paddr *runs from the root to the root in some t units to a tape within S rows, $R \leq S \leq 2N$, that keeps the saved request, carries the counter p on track* 5 *and holds the value d on track* 14*, and $t+S+1 \leq 4\cdot 10^{12}N^4$. No cell of tracks* 3*,* 4 *and* 15 *changes.*

(b) Meta *runs from the root to the root in some t units to a tape that holds the archive (p) within $R_1 = S+3+p$ rows and keeps the saved request, and $t+R_1+1 \leq 10^{32}N^8$.*

Proof. Every frame of ω is a frame of the request, so $|\omega| \leq |u| \leq D$; there are at most $|u|$ control numbers, and each is at most D. Let $c_1, c_2, \dots$ be the numbers of the frames before the payload: d, the shape numbers and q. Their sum Σ and their count satisfy $2\Sigma + (\text{count}) \leq 2D$.

(a) ini ($122R+172 \leq 300N$ units) gives $T((),0,R)$. The frame of d begins at row 0; app appends $u(d)$, within $10000(R+2D+1)^2 \leq 40000N^2$ units. The record now begins with the retained

word $u(d)$ of length d from row 0, and track 2 carries 0. The next frames are the linear shape group, for each of the d coordinates the count of its A-terms and their indices, so Lemma I.69(c) applies with $a = 1$ and $r_0 = 0$. Throughout the traversal of the prefix the capacity is at most $R + 2\Sigma + (\text{count}) \le R + 2D$, the record has length at most $2D$, and $|\omega| \le D$, so $Z \le 14N$ and the group costs at most $d(36d + 10^6(14N)^3 + 58) + 18d + 38 \le 3.1 \cdot 10^9 N^4$ units. zero_2 (11 units) resets track 2, and the bilinear group is read in the same way with $a = 2$. The last app reads q, within $10000(8N)^2$ units; the reading cursor now stands at the first payload frame, row p. In all at most $7 \cdot 10^9 N^4$ units, and the final capacity is $S = R + 2\Sigma + (\text{count}) \le 2N$, and $S \ge R$. The record begins with $u(d)$, so track 14 carries the counter d; $\mathsf{cc}_{14\to 9}$ and $\mathsf{cc}_{9\to 14}$ ($\le 600(S+1)$ units, Lemma I.66) make it a value. The bound follows. The parts ini, zero_2 and the two copies write only tracks 2, 5, 8, 9, 11, 12, 13 and 14, and app and $\mathsf{repeat}_2(\mathsf{row}_a)$ keep tracks 3, 4 and 15 (Lemmas I.50 and I.53), so no cell of these tracks changes.

(b) $p \le |\omega| \le S$, so $\mathsf{cc}_{5\to 11}$ ($\le 300(S+1)$) puts the value p on track 11. init leaves the archive $()$ within $S + 2$ rows with $|\omega| + 1$ on track 13 and keeps the counter p on track 11, and append with $w = |\omega| + 1$ and $n = p$ leaves the archive (p) within $S + 3 + p$ rows (Lemma I.68(e), (f)). These three parts and the $3 + p$ new rows cost at most $10^6(S+3)^2$. With (a), $t + R_1 + 1 \le 4 \cdot 10^{12} N^4 + 10^6(2N+3)^2 \le 10^{32} N^8$. □

Lemma I.72 (The magnitude bank)**.** *Let*

$$\begin{aligned}\mathsf{Bank} &= \mathsf{grow};\mathsf{ini};\mathsf{app};\big[\mathsf{repeat}_2(\mathsf{row}_1),\mathsf{zero}_2,\mathsf{repeat}_2(\mathsf{row}_2),\mathsf{pay},\\ &\qquad \mathsf{app},\mathsf{app},\mathsf{app},\mathsf{app},\mathsf{app}\big];\mathsf{reset}_5;\mathsf{fin},\\ \mathsf{pay} &= \big[\mathsf{copy}_{11\to 12},\ \mathsf{app},\ \mathsf{repeat}_{12}(\mathsf{sapp})\big].\end{aligned}$$

There is a constant c_B, depending only on the problem, with the following property. If τ keeps the saved request and is within R rows, then Bank *runs from the root to the root in some t units to a tape within $R' = R + 1 + 2\sum_j \mathbf{b}_j + |\mathbf{b}|$ rows that carries the bank, keeps tracks* 3, 4 *and* 15*, and*

$$t + R' + 1 \le c_B(R + m + 1)^3.$$

Proof. grow and ini give $T((), 0, R+1)$ and tracks 2 and 12 carrying 0. The first app and the two group traversals proceed as in Lemma I.71, appending d and the shape numbers. In pay, $\mathsf{copy}_{11\to 12}$ puts on track 12 the current length ℓ of the record (Lemma I.15(a)), and app appends $u(q)$, which is a retained word of length q on track 14 from row ℓ. Lemma I.18 with $c = 12$, $r_0 = \ell$, $n = q$, and R_j the traversal state after $u(|z_1|)\cdots u(|z_j|)$ with the reading cursor at row p_{j+1}, runs sapp once for each payload integer (Lemma I.70): sapp does not change track 12 in the rows of the capacity, nor the retained word, which lies in rows $j < |\nu|$, as the repeat rule requires. The five app read M, T, n_0, i and m. The record is now $u(\mathbf{b}_1)\cdots u(\mathbf{b}_K)$, $K = |\mathbf{b}|$, the capacity is R', and reset_5 ($36|\omega| + 36$ units) returns the reading cursor to 0. There is one frame per bank number, so $K \le |\omega| \le R + 1$, and Lemma I.69(d) applies: fin writes $\mathrm{rec}(\mathbf{b})$ on tracks 9 and 10. No part writes tracks 3, 4 or 15.

For the count let G be $|\mathbf{b}|$ plus the sum of the bank numbers other than m plus the number of bits of the frames of ω other than that of m; G depends only on the problem, and $|\omega| \le G + m + 1$. During the traversal every reading position is at most $|\omega|$, every record length at most $G+m+1$, every capacity at most $R + 1 + 3G + 2m$ and every appended number at most $G + m$. Hence

the argument Y of each of the at most G appends (Lemma I.69(b) and Lemma I.70) is at most $c(R+m+1)$ with $c = 7G+6$, and each costs at most $10100c^2(R+m+1)^2$; in each group traversal $Z \le c'(R+m+1)$ with $c' = 12G+8$, so each costs at most $(10^6c'^3d+72d^2+76d+38)(R+m+1)^3$; the repeat over the payload adds at most $q(36(2G+q)+58)+72G+18q+38$ besides its bodies; and grow, ini, zero_2, $\mathsf{copy}_{11\to12}$, reset_5 and fin, together with $R'+1$, cost at most $3000(R'+1)^2 \le 3000c^2(R+m+1)^2$, because $R'+1 \le R+2+3G+2m$. All these coefficients depend only on the problem, and $(R+m+1)^2 \le (R+m+1)^3$; their sum is the constant c_B. □

Two bounds of this shape compose with a product of exponents.

Lemma I.73 (Composing schedule bounds)**.** *Let $a, b \ge 1$ and $x, y, z, t, u, K, L \in \mathbb{N}$. If $t+y+2 \le K(x+2)^a$ and $u+z+2 \le L(y+2)^b$, then*

$$t+u+z+2 \le (K+LK^b)(x+2)^{ab}.$$

Proof. $y+2 \le K(x+2)^a$, so $(y+2)^b \le K^b(x+2)^{ab}$ and $u+z+2 \le LK^b(x+2)^{ab}$. Also $t \le K(x+2)^a \le K(x+2)^{ab}$, because $a \le ab$ and $x+2 \ge 1$. Add. □

The variable x will be the rows in use plus the precision m: the bound of a stage is then a polynomial in $r+m+2$, and m enters only there.

Proof of Theorem I.62(a). Let τ be a prepared request with r rows. It keeps the saved request and is within r rows. By Lemma I.71(b), Meta ends within r_1 rows with the archive (p) and $t_1 + r_1 + 1 \le 10^{32}N^8$, $N = r + D + 1$. By Theorem I.58, $D \le c_{\mathrm{num}}X$ with a constant c_{num} depending only on the problem, and $X = m + n_0 + \kappa + 1 \le (n_0+\kappa+2)(r+m+2)$; so $N \le c_M(r+m+2)$ with $c_M = c_{\mathrm{num}}(n_0+\kappa+2)+1$. Since $m+1 \le (r+m+2)^8$,

$$t_1 + (r_1+m) + 2 \le A(r+m+2)^8, \qquad A = 10^{32}c_M^8 + 2.$$

Bank starts within r_1 rows and keeps tracks 3, 4 and 15, so the archive (p) and the saved request survive, and by Lemma I.72 it ends within r' rows with the bank on tracks 9 and 10 and $t_2 + (r'+m) + 2 \le B(r_1+m+2)^3$, $B = c_B + 2$. Lemma I.73 with $x = r+m$, $y = r_1+m$, $z = r'+m$, $a = 8$ and $b = 3$ gives the bound with $K_1 = A + BA^3$ and exponent 24. □

I.8.4 Reading the bank

Many loops of the folds run a body once for each unit of a counter and count the counter down. One lemma covers them.

Lemma I.74 (Counting a counter down)**.** *For a track $k \in \{2, \dots, 15\}$ and a program P put*

$$\mathsf{down}_k(P) = \mathsf{left}^{k+1}; \mathsf{loop}(\mathsf{home}_k; P; \mathsf{retr}_k; \mathsf{left}^{k+1}); \mathsf{home}_k,$$
$$\mathsf{down}_k^{\circ}(P) = \mathsf{left}^{k+1}; \mathsf{loop}(\mathsf{home}_k; \mathsf{retr}_k; P; \mathsf{left}^{k+1}); \mathsf{home}_k.$$

Let $n \in \mathbb{N}$, let $J_0, \dots, J_n$ be properties of tapes such that every tape with J_j has origin marks up to n and carries the counter j on track k, and let $b_1, \dots, b_n \in \mathbb{N}$.

(a) *Suppose that for $1 \le j \le n$, from every tape τ with J_j, P runs from the root within b_j units to the root and a tape that agrees with τ on the cells (i,k), $i \le j$, and has J_{j-1} once $(j-1,k)$ is set to 0. Then $\mathsf{down}_k(P)$ runs from every tape with J_n to a tape with J_0 within $\sum_{j\le n}(b_j + 36n + 4k + 14) + 2k + 7$ units.*

(b) *Suppose that for $1 \le j \le n$, from every tape τ with J_j, P runs from the root and τ with $(j-1,k)$ set to 0 within b_j units to the root and a tape with J_{j-1}. Then $\mathsf{down}^\circ_k(P)$ has the same bound.*

Proof. left^{k+1} ($k+2$ units) reaches $(0,k)$, which holds 1 exactly when the counter is positive. Let I_j be the configurations at $(0,k)$ whose tape has J_{n-j}. A pass from I_j, $j < n$, tests the mark (1), returns to the root ($k+4$), runs P ($\le b_{n-j}$) and retr_k ($36(n-j-1)+2k+43$, Lemma I.66(b)) in the order of the program, and goes back to $(0,k)$ ($k+2$): at most $b_{n-j} + 36n + 4k + 14$ units, ending in I_{j+1}. At I_n the counter is 0 and the test fails (1). Lemma I.6, with the final home_k ($k+4$), gives the bound. □

The bank word $\mathrm{rec}(\mathbf{b}) = u(K)\,u(\mathbf{b}_1)\cdots u(\mathbf{b}_K)$, $K = |\mathbf{b}|$, is read on track 10 with a cursor on track 2. Put $e_0 = K$ and $e_j = e_{j-1} + \mathbf{b}_j + 1$: the word $u(\mathbf{b}_j)$ occupies the rows $e_{j-1}+1, \dots, e_j$ and ends with the blank $(e_j, 10)$. The cursor stands *after entry j* if track 2 carries e_j. With a bank view within R rows, $|\mathrm{rec}(\mathbf{b})| = e_K + 1 \le R + 1$, so every e_j is at most R. The entries are, in order: $\mathbf{b}_1 = d$; the linear shape group, for each coordinate i the count k_i^A of its A-terms followed by their k_i^A indices; the bilinear shape group, for each i the count k_i^B followed by $2k_i^B$ indices; q; the magnitudes $|z_1|, \dots, |z_q|$; and M, T, n_0, i, m. The *weight* of a list of rows of a shape group is the sum of $k+1$ over its rows; let w_A and w_B be the weights of the two groups.

Lemma I.75 (Words of the bank)**.** *Let*

$$\begin{aligned}
\mathsf{fprobe} &= \mathsf{seek}_2;\mathsf{shift}_{2\to10}, \qquad \mathsf{fstep} = \mathsf{home}_{10};\mathsf{inc}_{11};\mathsf{inc}_2;\mathsf{fprobe},\\
\mathsf{fword} &= \mathsf{clear}_{11};\mathsf{inc}_2;\mathsf{fprobe};\mathsf{loop}(\mathsf{fstep});\mathsf{home}_{10},\\
\mathsf{head} &= \mathsf{clear}_{2,11};\mathsf{fprobe};\mathsf{loop}(\mathsf{fstep});\mathsf{home}_{10}, \qquad \mathsf{dim} = \mathsf{head};\mathsf{fword},\\
\mathsf{wskip} &= \mathsf{down}_{14}(\mathsf{fword}), \qquad \mathsf{eskip}^a = \mathsf{down}_{14}(\mathsf{fword}^a),\\
\mathsf{srow}^{g,a} &= \mathsf{fword};\mathsf{gacc}^g;\mathsf{cc}_{11\to14};\mathsf{eskip}^a,\\
\mathsf{srows}^{g,a} &= \mathsf{down}_{13}(\mathsf{srow}^{g,a}), \qquad \mathsf{blk}^a = \mathsf{cc}_{6\to13};\mathsf{srows}^{0,a},
\end{aligned}$$

where fword^a is a copies of fword and skip, $\mathsf{gacc}^1 = \mathsf{acc}^6_{11\to5}$ and $\mathsf{gacc}^0 = \mathsf{skip}$. Let τ be within R rows with a bank view. All routines run from the root to the root.

(a) *If the cursor carries x, $(x+1+i, 10) = 1$ for $i < n$, $(x+1+n, 10) = 0$ and $x+n+1 \le R$, then fword leaves the value n on track 11 and the cursor at $x+n+1$, within $10000(R+1)^2$ units; only tracks 2 and 11 change. So from the cursor after entry j it reads $\mathbf{b}_{j+1}$ and leaves the cursor after entry $j+1$; fword^a reads a entries within $(a+1)\cdot 11000(R+1)^2$ units.*

(b) head *leaves the cursor after entry 0 and K on track 11 within $10000(R+1)^2$ units;* dim *leaves it after entry 1 and the value d on track 11, within $20000(R+1)^2$ units.*

(c) *If the cursor stands after entry j, track 14 carries c and $j + c \le K$, then wskip leaves the cursor after entry $j+c$ and the counter 0 on track 14, within $(c+1)\cdot 12000(R+1)^2$ units; only tracks 2, 11 and 14 change.*

(d) Let $a \in \{1,2\}$, let the entries after entry j begin with n rows of a shape group of arity a, with counts $k_1, \dots, k_n$ and weight $w = \sum_i (k_i + 1)$, let track 13 carry n, and if $g = 1$ let track 5 carry s with $s + \sum_i k_i \le R$. Then $\mathsf{srows}^{g,a}$ *leaves the cursor after these rows, the counter 0 on track 13 and, if $g = 1$, the counter $s + \sum_i k_i$ on track 5, within $(w+1)(a+1)\cdot 60000(R+1)^2$ units; only tracks 2, 11, 13, 14, and if $g = 1$ also 5 and 6, change. If track 6 carries n,* blk^a *does the same with $g = 0$, within $((w+1)(a+1)\cdot 60000 + 1000)(R+1)^2$ units.*

Proof. (a) clear_{11} $(61R+31)$, inc_2 $(36x+48)$ and fprobe $(18x+32)$ reach $(x+1, 10)$ with track 11 blank. For $i \le n$ let I_i be the configuration at $(x+1+i, 10)$ on which the cursor carries $x+1+i$ and track 11 holds i. A pass from I_i, $i < n$, scans a mark and costs $72(x+1+i) + 36i + 161$ units with its test (Lemmas I.11 and I.14), ending in I_{i+1}; at I_n the cell is blank, and home_{10} adds $18(x+1+n) + 14$. With $x + 1 + n \le R$ the total is at most $108R^2 + 294R + 126 \le 10000(R+1)^2$. The counter on track 2 ends at the blank $x+n+1$, as claimed. For fword^a, $a \cdot 10000(R+1)^2 + 1$.

(b) $\mathsf{clear}_{2,11}$ $(104R + 43)$ and fprobe (14) reach $(0, 10)$, and the same loop counts the K marks of $u(K)$; the total is at most $108R^2 + 283R + 72$. dim adds one fword.

(c) Lemma I.74(a) with $k = 14$, $n = c$, J_j the tapes with the cursor after entry $j_0 + c - j$ (j_0 the initial entry) and j on track 14, and $b_j = 10000(R+1)^2$, since fword does not write track 14: at most $c(10000(R+1)^2 + 36c + 70) + 35$ units.

(d) By (a), fword^a reads one entry of a row; eskip^a skips the k entries of a row of arity a by Lemma I.74(a), within $k((a+1)11000(R+1)^2 + 36k + 70) + 35 \le (k+1)(a+1)13000(R+1)^2$ units. $\mathsf{srow}^{g,a}$ reads the count k of a row (fword), adds it to track 5 when $g = 1$ ($\mathsf{acc}^6_{11\to 5}$, at most $500(R+1)^2$ by Lemma I.15(b), since $s + \sum_i k_i \le R$), puts it on track 14 ($\mathsf{cc}_{11\to 14}$) and skips the row's entries: at most $(k+1)(a+1)\cdot 40000(R+1)^2$ units, leaving the cursor after the row and track 13 untouched. Lemma I.74(a) with $k = 13$ and $b_i = (k_i+1)(a+1)40000(R+1)^2$ gives at most $\sum_i \big((k_i+1)(a+1)40000(R+1)^2 + 36n + 66\big) + 33$, which is at most $(w+1)(a+1)60000(R+1)^2$ because $n \le w$. blk^a adds $\mathsf{cc}_{6\to 13}$. □

A *row* of the field is a list of coefficients that the folds treat together: the A-coefficients of a coordinate j, its B-coefficients, its constant c_j, or all the constants $c_0, \dots, c_{d-1}$. Its k coefficients are consecutive in the payload: they are the integers $z_{o+1}, \dots, z_{o+2k}$, numerator and denominator in turn, with

$$o = 4d + 2P_j^A, \quad o = 4d + 2P_j^B, \quad o = 2d + 2j, \quad o = 2d$$

respectively, where $P_j^A = \sum_{i<j} k_i^A$ and $P_j^B = \sum_i k_i^A + \sum_{i<j} k_i^B$ count the coefficients that precede the row in the A-block and in the B-block. Let $h = 2 + \sum_i (1 + k_i^A) + \sum_i (1 + 2k_i^B)$ be the index of the entry q; the magnitudes of the row are the entries $h + o + 1, \dots, h + o + 2k$. The pairs of a row are *ready* if tracks 7 and 14 hold the value k and the cursor carries $e_{h+o+2k} - 1$ when $k \ge 1$ (the last mark of the word of the row's last denominator, which is positive) and e_{h+o} when $k = 0$.

Lemma I.76 (Selecting a row)**.** *Let*

$$\begin{aligned}
\mathsf{pbank} &= \mathsf{dim};\mathsf{cc}_{11\to 6};\mathsf{blk}^1;\mathsf{blk}^2;\mathsf{fword},\\
\mathsf{coff} &= \mathsf{clear}_{14};\mathsf{acc}_6^{\times 4};\mathsf{acc}_5^{\times 2}, \qquad \mathsf{twice} = \mathsf{clear}_{14};\mathsf{acc}_7^{\times 2},\\
\mathsf{pend} &= \mathsf{pbank};\mathsf{coff};\mathsf{wskip};\mathsf{twice};\mathsf{wskip},\\
\mathsf{gcur}^a &= \mathsf{cc}_{9\to 13};\mathsf{srows}^{1,a};\mathsf{fword};\mathsf{cc}_{11\to 7},\\
\mathsf{cont} &= \mathsf{branch}(\mathsf{skip},\ \mathsf{home}_{14};\mathsf{retr}_2;\mathsf{left}^{15}),\\
\mathsf{brow} &= \mathsf{cc}_{7\to 14};\mathsf{left}^{15};\mathsf{cont};\mathsf{home}_{14},
\end{aligned}$$

where $\mathsf{acc}_k^{\times n}$ *is* n *copies of* $\mathsf{acc}_{k\to 14}^{11}$ *followed by* skip*; and*

$$\begin{aligned}
\mathsf{sel}_A &= \mathsf{clear}_5;\mathsf{dim};\mathsf{gcur}^1;\mathsf{pend};\mathsf{brow},\\
\mathsf{sel}_B &= \mathsf{clear}_5;\mathsf{dim};\mathsf{cc}_{11\to 13};\mathsf{srows}^{1,1};\mathsf{gcur}^2;\mathsf{pend};\mathsf{brow},\\
\mathsf{csel}^\varepsilon &= \mathsf{clear}_5;\mathsf{pbank};\mathsf{cnt}^\varepsilon;\mathsf{clear}_{14};\mathsf{acc}_6^{\times 2};\mathsf{acc}_9^{\times 2\varepsilon};\\
&\qquad \mathsf{wskip};\mathsf{twice};\mathsf{wskip};\mathsf{brow} \qquad (\varepsilon = 0, 1),
\end{aligned}$$

with $\mathsf{cnt}^1 = \mathsf{clear}_7;\mathsf{inc}_7$ *and* $\mathsf{cnt}^0 = \mathsf{cc}_{6\to 7}$*. Let* τ *be within* R *rows with a bank view, and for* sel_A*,* sel_B *and* csel^1 *let track* 9 *carry* $j < d$*. Every routine runs from the root to the root, writes only tracks* 2*,* 5*,* 6*,* 7*,* 11*,* 13 *and* 14*, and ends within* R *rows with the bank view.*

(a) pbank *leaves the cursor after entry* h *(that of* q*) and the value* d *on track* 6*, does not change track* 5*, and takes at most* $c_p(R+1)^2$ *units,* $c_p = (w_A+1)\cdot 120000 + (w_B+1)\cdot 180000 + 40000$*.*

(b) sel_A *makes the pairs of the A-row of coordinate* j *ready and leaves* P_j^A *on track* 5*;* sel_B *does the same for the B-row, with* P_j^B *on track* 5*. Each takes at most* $(c_s + 1000)(R+1)^2$ *units,* $c_s = c_p + (w_A+1)\cdot 120000 + (w_B+1)\cdot 180000 + 12000(q+2) + 100000$*.*

(c) csel^0 *makes the pairs of the row of all constants ready, and* csel^1 *those of the row of* c_j*; each takes at most* $c_c(R+1)^2$ *units,* $c_c = c_p + 12000(q+2) + 20000$*, which is less than* c_s*.*

Proof. (a) dim puts the cursor after entry 1 and d on track 11, which $\mathsf{cc}_{11\to 6}$ copies to track 6. The two calls of blk pass the d rows of the linear and of the bilinear group (Lemma I.75(d)), and fword reads q. The costs are $20000(R+1)^2$, $300(R+1)$, $((w_A+1)120000 + 1000)(R+1)^2$, $((w_B+1)180000 + 1000)(R+1)^2$ and $10000(R+1)^2$.

(b) For sel_A: clear_5 blanks track 5; dim puts the cursor after entry 1; in gcur^1, $\mathsf{cc}_{9\to 13}$ puts j on track 13, $\mathsf{srows}^{1,1}$ passes the first j rows of the linear group and adds their counts to track 5, so that it carries P_j^A, and fword and $\mathsf{cc}_{11\to 7}$ read the count $k = k_j^A$ of row j onto track 7. Then pend resets the cursor to the entry of q; coff puts $4d + 2P_j^A = o$ on track 14 by six additions, each at most $500(R+1)^2 + 1$ units (Lemma I.15(b); all these numbers are entries of the bank or sums of them, at most R); wskip passes the o payload entries before the row, twice puts $2k$ on track 14, and wskip passes the $2k$ entries of the row: the cursor now carries e_{h+o+2k}. Finally brow copies k to track 14 and, if $k \geq 1$, steps the cursor back by one cell (cont scans $(0, 14)$, which is 1 exactly when $k \geq 1$). For sel_B the first $\mathsf{srows}^{1,1}$ passes all d rows of the linear group, adding $\sum_i k_i^A$ to track 5, and gcur^2 passes j rows of the bilinear group. The bounds of Lemma I.75 give at most $((w_A+1)120000 + (w_B+1)180000 + 50000)(R+1)^2$ for the

part before pend, at most $(c_p + 12000(o + 2k + 2) + 10000)(R + 1)^2$ for pend (with $o + 2k \le q$), and $1000(R + 1)^2$ for brow.

(c) pbank puts d on track 6 and the cursor at the entry of q; $\mathsf{cnt}^{\varepsilon}$ puts $k = 1$ or $k = d$ on track 7; the four or six additions put $o = 2d + 2\varepsilon j$ on track 14; and the rest proceeds as in (b). Besides $c_p(R + 1)^2$ for pbank and $12000(o + 2k + 2)(R + 1)^2$ for the two skips, the parts cost at most $5468(R + 1)^2$, and $o + 2k \le q$. □

I.8.5 Frame widths and unary arithmetic

The capacities s and s_0 count binary digits, and these are read off the saved request itself: the frame of an integer z has $\mathrm{bl}(|z|) + 2$ bits, so its length minus two is $\mathrm{bl}(|z|)$. A cursor on track 5 walks along ω and a flag on track 12 records whether the current cell still belongs to the same frame.

Lemma I.77 (Frame widths)**.** *Let*

$$\begin{aligned}
\mathsf{probe} &= \mathsf{branch}\big(\mathsf{left}^9; \mathsf{write}\,0; \mathsf{right}^3; \mathsf{left}^3,\\
&\qquad \mathsf{left}^{12}; \mathsf{branch}(\mathsf{right}^3; \mathsf{write}\,1,\ \mathsf{right}^3; \mathsf{write}\,0)\big),\\
\mathsf{wprobe} &= \mathsf{seek}_5; \mathsf{right}; \mathsf{right}; \mathsf{probe}, \qquad \mathsf{wstep} = \mathsf{home}_{12}; \mathsf{inc}_{13}; \mathsf{inc}_5; \mathsf{wprobe},\\
\mathsf{wspan} &= \mathsf{wprobe}; \mathsf{loop}(\mathsf{wstep}); \mathsf{home}_{12}, \qquad \mathsf{fw} = \mathsf{inc}_5; \mathsf{inc}_5; \mathsf{wspan},\\
\mathsf{rfw} &= \mathsf{fw}; \mathsf{fw}, \qquad \mathsf{skip1} = \mathsf{clear}_{13}; \mathsf{fw}, \qquad \mathsf{fskip} = \mathsf{down}_{14}(\mathsf{skip1}).
\end{aligned}$$

Let τ keep the saved request and be within R rows.

(a) From $\mathrm{cell}(x, 3)$, probe *runs in* 20 *units to* $\mathrm{cell}(x, 12)$ *and sets* $(x, 12)$ *to* 1 *if* $(x, 3) = 1$ *and* $(x, 15) = 0$*, and to* 0 *otherwise. If track* 5 *carries* x*,* wprobe *runs from the root in* $18x + 30$ *units to* $\mathrm{cell}(x, 12)$ *with the same effect.*

(b) Let track 5 *carry* x *and track* 13 *hold the value* v*. Let the rows* $x, \dots, x + n - 1$ *belong to* ω *and carry no frame-start mark, and let row* $x + n$ *carry a frame-start mark or lie after* ω*, with* $x + n \le R$ *and* $v + n \le R$*. Then* wspan *runs from the root to the root within* $(18x + 30) + n(300(R + 1) + 1) + 1 + 18(x + n) + 16$ *units, leaving the counter* $x + n$ *on track* 5 *and the value* $v + n$ *on track* 13*; only tracks* 5*,* 12 *and* 13 *change.*

(c) If the frame of an integer z *begins at row* x*, track* 5 *carries* x*, track* 13 *holds* v*,* $x + \mathrm{bl}(|z|) + 2 \le R$ *and* $v + \mathrm{bl}(|z|) \le R$*, then* fw *adds* $\mathrm{bl}(|z|)$ *to the value on track* 13 *and moves the cursor to* $x + \mathrm{bl}(|z|) + 2$*, within* $1000(R + 1)^2$ *units. For a coefficient* $q = u/e$ *whose two frames begin at row* x*,* rfw *adds* $\mathrm{cap}(q) = \mathrm{bl}(|u|) + \mathrm{bl}(e)$ *to track* 13 *and moves the cursor past both frames, within* $2000(R + 1)^2$ *units, and* $\mathsf{skip1}$ *moves the cursor past one frame within* $2000(R + 1)^2$ *units.*

(d) If track 14 *holds the value* c *and the frames of* c *payload integers begin at the cursor,* fskip *moves the cursor past them and leaves* 0 *on track* 14*, within* $(c + 1) \cdot 4000(R + 1)^2$ *units.*

Proof. (a) On a blank $(x, 3)$ the first branch reaches $(x, 12)$ in 10 units, writes 0, and moves three cells and back; on a mark it moves to $(x, 15)$ (13 units), branches on the frame-start mark

and writes the flag three cells to the right. With the tests, 20 units either way. seek_5 costs $18x + 8$ and the two moves 2.

(b) For $i \le n$ let I_i be the configuration at $(x+i, 12)$ on which track 5 carries $x+i$, track 13 holds $v+i$, and $(x+i, 12)$ holds the flag of row $x+i$. By (a) the flag is 1 for $i < n$ and 0 for $i = n$. A pass from I_i, $i < n$, returns to the root $(18(x+i)+16)$, advances tracks 13 and 5 $(36(v+i)+70$ and $36(x+i)+54)$ and probes the next row $(18(x+i+1)+30)$: $72(x+i)+36(v+i)+188 \le 300(R+1)$ units, ending in I_{i+1}. Lemma I.6 and the final home_{12} give the bound.

(c) The frame of z occupies rows $x, \dots, x+\mathrm{bl}(|z|)+1$, and only row x carries a frame-start mark. The two increments $(36x+54+36x+90$ units) put the cursor at $x+2$, and (b) with $n = \mathrm{bl}(|z|)$ counts the remaining rows of the frame. The total is at most $300R^2+427R+271 \le 1000(R+1)^2$. rfw applies this to the frame of u and then to that of e, which begins where the first ends. For $\mathsf{skip1}$ add clear_{13}.

(d) Lemma I.74(a) with $k = 14$, $n = c$ and $b_j = 2000(R+1)^2$; $\mathsf{skip1}$ does not write track 14. The bound is $c(2000(R+1)^2 + 36c + 70) + 35$. □

Division with rounding up works on unary values: the dividend is walked mark by mark, a remainder counter is advanced, and each time it reaches the divisor it is reset and the quotient grows by one mark.

Lemma I.78 (Ceilings of quotients). *Let* $\mathsf{rcol}_{k,l} = \mathsf{seek}_k; \mathsf{shift}_{k\to l}$ *and*

$$\begin{aligned}
\mathsf{dbody} &= \big[\mathsf{home}_{11}, \mathsf{inc}_2, \mathsf{inc}_5, \mathsf{rcol}_{5,12},\\
&\qquad \mathsf{branch}\big([\mathsf{home}_{12}, \mathsf{reset}_5, \mathsf{mark}_6],\ \mathsf{home}_{12}\big),\ \mathsf{rcol}_{2,11}\big],\\
\mathsf{udiv} &= \big[\mathsf{rcol}_{2,11}, \mathsf{loop}(\mathsf{dbody}), \mathsf{home}_{11}\big],\\
\mathsf{bump} &= \mathsf{left}^6; \mathsf{branch}(\mathsf{right}^6,\ \mathsf{right}^6; \mathsf{mark}_6), \qquad \mathsf{ceil} = \mathsf{udiv}; \mathsf{bump}.
\end{aligned}$$

Let τ have origin marks up to R, hold the value a on track 11 and the value b on track 12, with $1 \le b \le R$ and $a+1 \le R$, carry 0 on tracks 2, 5 and 6, and have track 7 blank. Then udiv *runs from the root to the root within $1000(a+b+1)^2$ units, leaving the value $\lfloor a/b \rfloor$ on track 7, the counter $\lfloor a/b \rfloor$ on track 6 and the counter $a \bmod b$ on track 5.* ceil *runs within $1200(a+b+1)^2$ units and leaves the value $\lceil a/b \rceil$ on track 7. Only tracks 2, 5, 6 and 7 change.*

Proof. $\mathsf{rcol}_{2,11}$ (15 units) reaches $(0, 11)$. For $j \le a$ let I_j be the configuration at $(j, 11)$ on which track 2 carries j, track 5 carries $\varrho = j \bmod b$, track 6 carries $\theta = \lfloor j/b \rfloor$ and track 7 holds the value θ. A pass from I_j, $j < a$, scans a mark of the dividend, returns to the root, advances tracks 2 and 5, and reads $(\varrho+1, 12)$, which is 1 exactly when $\varrho + 1 < b$. In that case it returns ($72j + 72\varrho + 220$ units in all). Otherwise $\varrho + 1 = b$: it resets track 5 to 0 and mark_6 writes 1 in $(\theta, 7)$ and advances track 6 ($72j + 108\varrho + 36\theta + 354$ units). Then it reads $(j+1, 11)$. Since $j + 1 = \theta b + \varrho + 1$, the pass ends in I_{j+1} (Lemmas I.11 and I.14; the origin marks reach b and $a+1$). At I_a the scanned cell is the blank $(a, 11)$. With $\varrho < b$ and $\theta \le a$ each pass costs at most $108a + 108b + 355$ units with its test, and with the final home_{11} and the list the total is at most $108a^2 + 108ab + 373a + 32 \le 1000(a+b+1)^2$. bump reads $(0, 5)$, which is 1 exactly when $a \bmod b \ne 0$, and then adds one mark to the quotient: 15 or $36\lfloor a/b \rfloor + 76$ units. So track 7 holds $\lfloor a/b \rfloor + [a \bmod b \ne 0] = \lceil a/b \rceil$. □

Lemma I.79 (Maximum of two values)**.** *Let A, B, C and F be distinct tracks with $C, F \geq 2$, and put*

$$\begin{aligned} \mathsf{em}_k(b) &= \mathsf{shift}_{k\to F}; \mathsf{write}\, b, \\ \mathsf{mprobe} &= \mathsf{shift}_{F\to A}; \mathsf{branch}\big(\mathsf{shift}_{A\to B}; \mathsf{branch}(\mathsf{em}_B(0), \mathsf{em}_B(1)),\ \mathsf{em}_A(1)\big), \\ \mathsf{max}^{F}_{A,B\to C} &= \mathsf{zero}_C; \mathsf{left}^{F+1}; \mathsf{mprobe}; \mathsf{loop}(\mathsf{shift}_{F\to C}; \mathsf{next}; \mathsf{shift}_{C\to F}; \mathsf{mprobe}); \\ &\qquad \mathsf{home}_F. \end{aligned}$$

If tracks A and B hold the values a and b and the tape has origin marks up to $R \geq \max(a,b)$, then $\mathsf{max}^{F}_{A,B\to C}$ runs from the root to the root within $400(a+b+1)$ units, after which track C carries $\max(a,b)$; only tracks C and F change, in rows $j \leq \max(a,b)$.

Proof. From (j,F), mprobe writes in (j,F) the bit $\tau(j,A) \vee \tau(j,B)$ and returns to (j,F), within 52 units: at most three shifts of at most 16 units, two tests and one write. Because A and B hold values, this bit is 1 exactly when $j < M = \max(a,b)$. zero_C $(2C+7)$ and left^{F+1} $(F+2)$ lead to $(0,F)$, where the first probe is made. Let I_j be the configuration at (j,F) with the flag of row j written and the counter j on track C. A pass from I_j, $j < M$, moves to the hole (j,C), advances it by next (20 units), returns to track F in row $j+1$ and probes it: at most 105 units with its test. At I_M the flag is 0. home_F from row M adds $18M + F + 4$. The total is at most $123M + 2C + 2F + 66 \leq 400(a+b+1)$. □

The fold routines use $\mathsf{max}^{12}_{7,13\to 11}$, $\mathsf{max}^{12}_{11,13\to 7}$ and $\mathsf{max}^{14}_{11,12\to 8}$; the last also serves the schedule tail of Section I.9.

I.8.6 One row at a time

The folds call, for one coordinate j at a time, a program that computes a number of the row of j and leaves it on track 11; track 9 holds the index j.

Definition I.80 (Row readers)**.** Let f map $\{0,\dots,d-1\}$ to $\mathbb{N}$ and let $K \in \mathbb{N}$. A program P is a *row reader* for f with constant K if, for every $j < d$ and every tape in the gather state (with any archive and any number R of rows) on which track 9 carries j, P runs from the root to the root within $K(R+1)^3$ units to a tape within R rows on which track 11 carries $f(j)$, $f(j) \leq R$, track 9 still carries j, and tracks 3, 4, 10 and 15 are unchanged.

A row reader keeps the gather state: the archive and the frame-start marks lie on tracks 3, 4 and 15, and the bank view on track 10. Every number it computes below is at most the sum of the bank entries, which is less than $|\operatorname{rec}(\mathbf{b})| \leq R+1$; so the hypotheses "value at most R" of the routines are met.

Lemma I.81 (Summing and maximizing over a row)**.** *Let*

$$\begin{aligned}
\mathsf{bword} &= \mathsf{clear}_{11};\mathsf{seek}_{2};\mathsf{left}^{8};\mathsf{tail},\\
\mathsf{pair} &= \mathsf{bword};\mathsf{cc}_{11\to12};\mathsf{retr}_{2};\mathsf{bword};\mathsf{cc}_{2\to8};\mathsf{clear}_{2,5,6,7};\mathsf{ceil},\\
\mathsf{pbody} &= \mathsf{home}_{14};\mathsf{pair};\mathsf{acc}^{2}_{7\to13};\mathsf{cc}_{8\to2};\mathsf{retr}_{14};\mathsf{left}^{15};\mathsf{cont},\\
\mathsf{prow} &= \mathsf{clear}_{13};\mathsf{left}^{15};\mathsf{loop}(\mathsf{pbody});\mathsf{home}_{14};\mathsf{cc}_{13\to11};\mathsf{cc}_{11\to13},\\
\mathsf{crow} &= \mathsf{prow};\mathsf{cc}_{13\to11}, \qquad \mathsf{pmax} = \mathsf{pair};\mathsf{max}^{12}_{7,13\to11};\mathsf{cc}_{11\to13};\mathsf{cc}_{8\to2},\\
\mathsf{gbody} &= \mathsf{home}_{14};\mathsf{pmax};\mathsf{retr}_{14};\mathsf{left}^{15};\mathsf{cont},\\
\mathsf{gam} &= \mathsf{clear}_{13};\mathsf{left}^{15};\mathsf{loop}(\mathsf{gbody});\mathsf{home}_{14}.
\end{aligned}$$

Let τ be within R rows with a bank view.

(a) Let $(x,10)=0$, $(x+i,10)=1$ for $1\le i\le a$, $(x+a+1,10)=0$ and $(x+a+1+i,10)=1$ for $1\le i\le b$, with $b\ge1$, and let the cursor carry $x+a+1+b\le R$: the words $u(a)\,u(b)$ follow the blank $(x,10)$, and the cursor is on the last mark of $u(b)$. Then pair *runs from the root to the root within $3000(R+1)^2$ units, leaving the value $\lceil a/b\rceil$ on track 7 and the value x on track 8; only tracks 2, 5, 6, 7, 8, 11 and 12 change.*

(b) Let the pairs of a row with coefficients $q_1,\dots,q_k$ be ready, with $k\le R$ and $\Sigma=\sum_l\lceil|q_l|\rceil\le R$. Then prow *and* crow *leave the value Σ on tracks 11 and 13, and* gam *leaves the value $\Gamma=\max_l\lceil|q_l|\rceil$ on track 13 ($\Gamma=0$ if $k=0$); each takes at most $22000(R+1)^3$ units and changes only tracks 2, 5, 6, 7, 8, 11, 12, 13 and 14.*

Proof. (a) The first bword blanks track 11, goes to the cursor and steps to track 10 ($61R+31+18(x+a+1+b)+5+9$ units), and Lemma I.64 with $s=x+a+1$ and $n=b$ leaves the value b on track 11 and the cursor at $x+a+1$, within $b(36R+113)+1+18R+14$ units. $\mathsf{cc}_{11\to12}$ saves b on track 12, retr_2 moves the cursor to $x+a$, and the second bword reads a in the same way, leaving the cursor at x, which $\mathsf{cc}_{2\to8}$ saves on track 8. $\mathsf{clear}_{2,5,6,7}$ ($196R+73$ units) prepares Lemma I.78, whose hypotheses hold because $b\ge1$ and $a+1\le R$; it leaves $\lceil a/b\rceil$ on track 7 within $1200(R+1)^2$ units. With the two copies and the step back the total is at most $72R^2+970R+498+1200(R+1)^2\le3000(R+1)^2$.

(b) Write $q_l=\pm a_l/b_l$, so that the magnitudes of the row are $a_1,b_1,\dots,a_k,b_k$ and $\lceil|q_l|\rceil=\lceil a_l/b_l\rceil$. Let x_l be the blank before $u(a_l)$; then x_{l+1} is the blank that ends $u(b_l)$, and $x_1=e_{h+o}$. clear_{13} and left^{15} lead to $(0,14)$. For $i\le k$ let I_i be the configuration at $(0,14)$ on which track 14 carries $k-i$, track 13 holds the value $\sum_{l>k-i}\lceil a_l/b_l\rceil$, and the cursor carries the last mark of $u(b_{k-i})$ if $i<k$ and x_1 if $i=k$. A pass from I_i, $i<k$, scans the mark $(0,14)$, returns to the root, runs pair on the pair $k-i$ by (a), adds $\lceil a_{k-i}/b_{k-i}\rceil$ to track 13 ($\mathsf{acc}^2_{7\to13}$, which uses track 2 as scratch, within $500(R+1)^2$ units), restores the cursor to x_{k-i} from track 8, counts track 14 down and returns to $(0,14)$. If pairs remain, cont steps the cursor back to $x_{k-i}-1$, the last mark of $u(b_{k-i-1})$. The pass ends in I_{i+1} within $20000(R+1)^2$ units. At I_k the counter is 0. With the clearing, the final return and the two copies the total is at most $k(20000(R+1)^2+1)+700(R+1)+120\le22000(R+1)^3$; crow adds a copy. For gam the pass runs pmax instead: after pair, Lemma I.79 with $A=7$, $B=13$, $C=11$ and $F=12$ puts $\max(\lceil a/b\rceil,\text{previous})$ on track 11 within $400(2R+1)$ units, and the copies return it to track 13 and restore the cursor; the same count applies. □

Lemma I.82 (The row readers). *Let c_s be the constant of Lemma I.76 and put*

$$\begin{aligned}
\mathsf{row}_A &= \mathsf{sel}_A;\mathsf{crow}, \qquad \mathsf{row}_B = \mathsf{sel}_B;\mathsf{crow}, \qquad \mathsf{gsum} = \mathsf{csel}^0;\mathsf{gam};\mathsf{cc}_{13\to 11},\\
\mathsf{wrow} &= \mathsf{clear}_{13};\mathsf{down}_{14}(\mathsf{rfw}),\\
\mathsf{swidth} &= \mathsf{cc}_{14\to 12};\mathsf{cc}_{12\to 14};\mathsf{load}_{0,5};\mathsf{fskip};\mathsf{cc}_{7\to 14};\mathsf{wrow},\\
\mathsf{wpart}^A &= \mathsf{sel}_A;\mathsf{pbank};\mathsf{coff};\mathsf{swidth}, \qquad \mathsf{wpart}^B = \mathsf{sel}_B;\mathsf{pbank};\mathsf{coff};\mathsf{swidth},\\
\mathsf{wpart}^c &= \mathsf{csel}^1;\mathsf{clear}_5;\mathsf{pbank};\mathsf{clear}_{14};\mathsf{acc}_6^{\times 2};\mathsf{acc}_9^{\times 2};\mathsf{swidth},\\
\mathsf{row}_W &= \mathsf{wpart}^A;\mathsf{cc}_{13\to 8};\mathsf{wpart}^B;\mathsf{acc}^{12}_{13\to 8};\mathsf{wpart}^c;\mathsf{acc}^{12}_{13\to 8};\\
&\qquad \mathsf{cc}_{8\to 13};\mathsf{cc}_{13\to 11},\\
\mathsf{row}_L &= \mathsf{sel}_A;\mathsf{cc}_{7\to 8};\mathsf{sel}_B;\mathsf{cc}_{7\to 13};\mathsf{acc}^{12}_{13\to 8};\mathsf{cc}_{8\to 13};\mathsf{cc}_{13\to 11}.
\end{aligned}$$

For a coordinate j let

$$\alpha_j = \sum_{(a,i)} \lceil |a| \rceil, \quad \beta_j = \sum_{(b,i,i')} \lceil |b| \rceil, \quad s_j = \mathrm{cap}(c_j) + \sum_{(a,i)} \mathrm{cap}(a) + \sum_{(b,i,i')} \mathrm{cap}(b), \quad L_j = k_j^A + k_j^B,$$

the sums running over the terms of coordinate j. Then row_A and row_B are row readers for $j \mapsto \alpha_j$ and $j \mapsto \beta_j$ with constant $c_s + 23000$, row_W is a row reader for $j \mapsto s_j$ and row_L one for $j \mapsto L_j$, with constants depending only on the problem. From a gather state with R rows, gsum leaves the value γ on track 11 within $(c_s + 22000)(R+1)^3$ units and keeps track 9 and the gather state.

Proof. For row_A: sel_A makes the pairs of the A-row of j ready within $(c_s + 1000)(R+1)^2$ units (Lemma I.76), and crow leaves α_j on track 11 within $22000(R+1)^3$ units (Lemma I.81). The tracks written are among 2, 5, 6, 7, 8, 11, 12, 13 and 14. row_B is the same. gsum selects the row of all constants within $c_c(R+1)^2$ units and takes the maximum of their ceilings within $22000(R+1)^3$, which is γ by Definition 15.24; with the final copy and $c_c + 300 \le c_s$ the bound follows.

For row_W consider first wpart^A. After sel_A, track 7 holds $k = k_j^A$ and track 5 carries P_j^A; pbank puts d on track 6 without changing tracks 5 and 7, and coff puts $o = 4d + 2P_j^A$ on track 14. In swidth the two copies make o a value on track 14; $\mathsf{load}_{0,5}$ puts on track 5 the first archive entry, which is the payload row p (Lemma I.68(c)); fskip passes the frames of $z_1, \dots, z_o$ (Lemma I.77(d)), so that the cursor stands at the frame of z_{o+1}, the numerator of the first coefficient of the row; $\mathsf{cc}_{7\to 14}$ puts k on track 14; and wrow adds $\mathrm{cap}(q)$ for each of the k coefficients (Lemma I.77(c) and Lemma I.74(a) with $b = 2000(R+1)^2$), leaving their sum on track 13. wpart^B is the same with the B-row, and wpart^c with the row of c_j, whose offset $2d + 2j$ is built from tracks 6 and 9. row_W adds the three sums on track 8 and copies the total s_j to track 11. For row_L, the two selections leave k_j^A and k_j^B on track 7, and their sum is formed on track 8. All parts are bounded by Lemmas I.66, I.68, I.76 and I.77 by constants depending only on the problem (the weights w_A, w_B, q and d) times $(R+1)^3$, since every loop runs at most R times. None of them writes tracks 3, 4, 9, 10 or 15, nor a header or an origin mark. □

The capacity s_0 of the initial point is a maximum over the coordinates of y_0, whose frames come first in the payload.

Lemma I.83 (The initial capacity)**.** *Let*

$$\mathsf{istep} = \mathsf{clear}_{13};\mathsf{rfw};\mathsf{max}^{12}_{11,13\to 7};\mathsf{cc}_{7\to 11}, \qquad \mathsf{iw} = \mathsf{clear}_{11};\mathsf{down}_{14}(\mathsf{istep}),$$
$$\mathsf{isum} = \mathsf{dim};\mathsf{cc}_{11\to 14};\mathsf{load}_{0,5};\mathsf{iw}.$$

From every gather state with R rows, isum *runs from the root to the root within $10^5(R+1)^3$ units to a tape that holds the value s_0 on track* 11 *and is still in the same gather state; it writes only tracks* 2, 5, 7, 11, 12, 13 *and* 14*.*

Proof. dim and $\mathsf{cc}_{11\to 14}$ put d on track 14, and $\mathsf{load}_{0,5}$ puts the payload row p on track 5: the cursor stands at the frame of the numerator of $y_{0,0}$. A pass of down_{14} measures $\mathrm{cap}(y_{0,l})$ on track 13 by Lemma I.77(c) and replaces the value on track 11 by the maximum of the two (Lemma I.79 with $A=11$, $B=13$, $C=7$, $F=12$, and a copy), within $5000(R+1)^2$ units; the cursor moves to the next coordinate. After d passes track 11 holds $\max_l \mathrm{cap}(y_{0,l}) = s_0$. By Lemma I.74(a) and the bounds of dim and $\mathsf{load}_{0,5}$, the total is at most $80000(R+1)^2 + d(5000(R+1)^2 + 36d + 70) + 400(R+1) \le 10^5(R+1)^3$, as $d \le R$. □

I.8.7 The six folds

A fold runs a row reader for every coordinate, pushing each row value onto the archive, and then replaces these d values by their maximum.

Lemma I.84 (Gathering, maximizing and collapsing)**.** *Let*

$$\mathsf{dimc}_k = \mathsf{dim};\mathsf{cc}_{11\to k}, \qquad \mathsf{gather}(P) = \mathsf{down}^{\circ}_{9}(P;\mathsf{push}),$$
$$\mathsf{amax} = \mathsf{clear}_{11};\mathsf{down}_9\big(\mathsf{rd};\mathsf{max}^{14}_{11,12\to 8};\mathsf{cc}_{8\to 11}\big),$$
$$\mathsf{collapse}_c = \mathsf{prefix}_c;\mathsf{amax};\mathsf{trunc}_c;\mathsf{push},$$
$$\mathsf{pass}(P,c) = \mathsf{dimc}_9;\mathsf{gather}(P);\mathsf{dimc}_9;\mathsf{collapse}_c.$$

(a) *With a bank view within R rows,* dimc_k *($k = 9, 14$) puts the value d on track k within $22000(R+1)^2$ units, changing only tracks* 2, 11 *and k.*

(b) *Let P be a row reader for f with constant K, and let τ be in the gather state with archive v and R rows, with d on track* 9 *and $d \le R$. Then* gather(P) *runs from the root to the root to a tape in the gather state with archive $v, f(d-1), \dots, f(0)$ and $R + \rho_f$ rows, $\rho_f = \sum_{j<d}(1+f(j))$, with* 0 *on track* 9*, within $(d+2)(K+1000200)(R+\rho_f+2)^3$ units.*

(c) *If track* 13 *carries x, the words $u(n_1)\cdots u(n_c)$ lie on track* 4 *from row x, $x + \sum_l (n_l+1) \le R$ and track* 9 *carries $c \le R$, then* amax *leaves the value $\max_l n_l$ on track* 11 *(and* 0 *if $c = 0$), the cursor on track* 13 *at $x + \sum_l (n_l+1)$ and* 0 *on track* 9*, within $16000(R+1)^3$ units; only tracks* 8, 9, 11, 12, 13 *and* 14 *change.*

(d) *If τ holds the archive $(v_1, \dots, v_c, w_1, \dots, w_d)$ within R rows and track* 9 *carries d, then* $\mathsf{collapse}_c$ *leaves the archive $(v_1, \dots, v_c, \max_l w_l)$ within $R + 1 + \max_l w_l$ rows,* 0 *on track* 9*, and tracks* 10 *and* 15 *unchanged, within $(c+5)\cdot 10^7(R+1)^3$ units.*

(e) *Let P be a row reader for f with constant K, and let τ be in the gather state with an archive v of c numbers and R rows. Then* pass(P,c) *runs from the root to the root to a tape in the*

gather state with archive v, $\max_j f(j)$ and $R+\rho_f+1+\max_j f(j)$ rows, with 0 on track 9, within

$$\big(44000+(d+2)(K+1000200)+(c+5)\cdot 10^7\big)(R+\rho_f+2)^3$$

units.

Proof. (a) Lemma I.75(b) and Lemma I.66(a).

(b) Lemma I.74(b) with $k=9$, $n=d$, and J_j the gather state with archive v, $f(d-1),\dots,f(j)$, $R_j=R+\sum_{i\ge j}(1+f(i))$ rows and the counter j on track 9. From a tape with J_j on which $(j-1,9)$ has been set to 0, the row reader leaves $f(j-1)\le R_j$ on track 11, keeps the counter $j-1$ on track 9 and the gather state, and push appends $f(j-1)$ (Lemma I.68(h)), keeping tracks 9, 10 and 15; the result has J_{j-1}. A pass costs at most $K(R_j+1)^3+10^6(R_j+f(j-1)+2)^2\le(K+10^6)(R+\rho_f+2)^3$, and the bound of Lemma I.74 is at most $d\big((K+10^6)(R+\rho_f+2)^3+36d+50\big)+25$.

(c) clear_{11} leaves the value 0. Apply Lemma I.74(a) with $k=9$, $n=c$, and J_j the tapes on which track 9 carries j, track 13 carries $x+\sum_{l\le c-j}(n_l+1)$ and track 11 holds the value $\max_{l\le c-j}n_l$. The body reads the next word onto track 12 (Lemma I.68(b)), forms the maximum on track 8 (Lemma I.79 with $A=11$, $B=12$, $C=8$, $F=14$) and copies it to track 11, within $12000(R+1)^2$ units, and does not write track 9. The total is at most $61R+31+c(12000(R+1)^2+36c+50)+25\le 16000(R+1)^3$.

(d) prefix_c puts the cursor at the first word $u(w_1)$, and amax leaves $W=\max_l w_l\le R$ on track 11 and 0 on track 9; trunc_c cuts the archive to $(v_1,\dots,v_c)$, and push appends W (Lemma I.68(h), (i)). The costs are at most $(c+3)\cdot 20000(R+1)^2$ twice, $16000(R+1)^3$ and $10^6(R+W+2)^2\le 4\cdot 10^6(R+1)^2$.

(e) dimc_9 puts d on track 9, $\mathsf{gather}(P)$ appends the d row values by (b), dimc_9 puts d on track 9 again (the bank view is unchanged), and $\mathsf{collapse}_c$ replaces the d values by their maximum $\max_j f(j)$ by (d). Add the four bounds; the gather state is kept because only tracks other than 3, 4, 10 and 15 are written, except by push, which extends the archive. □

The first and the last fold archive a single number. Put

$$\mathsf{Gam}=\mathsf{gsum};\mathsf{push},\qquad \mathsf{Ini}=\mathsf{isum};\mathsf{push}.$$

Proof of Theorem I.62(b). Let τ be in the gather state with archive (p) and r rows. clear_9 ($57r+27$ units) leaves the counter 0 on track 9. Gam computes γ (Lemma I.82) and appends it: archive (p,γ) within $r+1+\gamma$ rows, within $(c_s+22000)(r+1)^3+10^6(r+\gamma+2)^2$ units. The four passes append, by Lemma I.84(e) with the row readers of Lemma I.82, the maxima over the coordinates of α_j, β_j, s_j and L_j, which are α, β, s and L (Definition 15.24 and Section H.1); each pass starts with 0 on track 9 and ends with it, and the archive before the pass has 2, 3, 4 and 5 numbers. Ini computes s_0 (Lemma I.83) and appends it. The tape is then in the gather state with the archive $(p,\gamma,\alpha,\beta,s,L,s_0)$ and $r+\rho_S$ rows, where

$$\rho_S=(1+\gamma)+\sum_{g\in\{\alpha,\beta,s,L\}}\Big(\sum_{j<d}(1+g_j)+1+\max_j g_j\Big)+(1+s_0)$$

depends only on the problem. Each of the seven parts is bounded by a constant depending only on the problem times $(r+\rho_S+2)^3 \le (\rho_S+2)^3(r+1)^3$; so $t \le K_S(r+1)^3$ for such a constant K_S, and

$$t + (r + \rho_S + m) + 2 \le (K_S + \rho_S + 1)(r + m + 2)^3.$$

This is the bound with $K_2 = K_S + \rho_S + 1$. □

Proof of Theorem I.62(c). Let τ be a prepared request with r rows. By part (a), MetaBank ends after t_1 units within r' rows, with $t_1 + (r' + m) + 2 \le K_1(r + m + 2)^{24}$, on a tape that holds the archive (p), carries the bank, keeps the saved request, and is a workspace of r' rows with origin marks up to r'. This tape is in the gather state with archive (p) and r' rows: it is its own witness for the bank view, and track 15 carries the frame-start marks of the saved request. By part (b), Summ ends after t_2 units within $r'' = r' + \rho_S$ rows, with $t_2 + (r'' + m) + 2 \le K_2(r' + m + 2)^3$, in the gather state with the archive of the seven numbers; in particular it keeps the saved request and is a workspace of r'' rows with origin marks up to r''. Lemma I.73 with $x = r + m$, $y = r' + m$, $z = r'' + m$, $a = 24$ and $b = 3$ gives

$$t_1 + t_2 + (r'' + m) + 2 \le (K_1 + K_2K_1^3)(r + m + 2)^{72}.$$ □

The archive of the seven numbers, with the saved request, the workspace and the origin marks, is the input of the schedule tail of Section I.9, which reads M, T and n_0 from their saved frames and computes the schedule from these numbers. The exponent 72 of part (c), and the exponent 2 that Prepare adds in the composition of the whole stage (Lemma I.61), enter the exponent of the stage computed there.

I.9 The problem's constants: the schedule and the width

How does the machine find the width W_m of its answer word? Continue the example of Section I.8: the field $y' = -y^2$ with $y_0 = \frac{1}{2}$, whose archive ends as $(p, \gamma, \alpha, \beta, s, L, s_0) = (11, 0, 0, 1, 3, 1, 3)$, and take $M = 2$, $T = 1$ and $n_0 = 0$. The schedule of Definition 15.18 is

$$R = 0 + 0 + 1 \cdot 2 + 1 = 3, \quad \sigma = \mathrm{bl}(12) + 1 = 5, \quad S = 2^5 = 32, \quad G = \left\lfloor \frac{44^{32}}{2^{160}} \right\rfloor + 1 = 26650,$$

and $\kappa = \mathrm{bl}(26650 \cdot 257) + 2 = 25$. With $r = \mathrm{bl}(M) + \mathrm{bl}(R) + 1 = 5$, the coefficients of (H.1) are $q_P = 2rL + \sigma + r = 20$, $l_P = 3r + (s + 5r + 1)L + 4 = 48$ and $c_P = s_0 + 2s + 2r + (2s + 3r + 2)L + 7 = 49$. At the precision $m = 10$ the machine works at $N = \max(10, 0) + 25 = 35$ bits, the capacity is $\mathscr{B}(35) = (20 \cdot 35 + 48) \cdot 35 + 49 = 26229$ and the width is $W_{10} = 4 \cdot 26229 + 4 = 104920$. The machine computes every one of these numbers in unary, by a fixed list of fifty-one unary operations on the archive, and then evaluates the capacity by Horner's rule. On the way it writes the number 44^{32}, which has fifty-three decimal digits, as a run of marks. The number of rows this takes is enormous, but it depends only on the problem, so it enters the constant of the stage and not its exponent.

This section constructs the second half of the stage "Schedule" of Table 15.1 and composes the whole stage. Its programs are

$$\mathsf{Scalar} = \mathsf{Ctl}; \mathsf{Prog}; \mathsf{Seed}, \qquad \mathsf{Width} = \mathsf{corr}; \mathsf{max}^{14}_{11,12\to 8}; \mathsf{Code},$$
$$\mathsf{Schedule} = \big(\mathsf{Prepare}; (\mathsf{MetaBank}; (\mathsf{Summ}; \mathsf{Scalar}))\big); \mathsf{Width}.$$

Ctl reads M, T and n_0 from their saved frames and appends them to the archive (Section I.9.1). Prog is the fixed scalar program of Table I.2: each of its fifty-one instructions reads one or two entries of the archive, computes a sum, a product, a quotient, a binary length or a power in unary, and appends the result (Sections I.9.2 to I.9.4). Seed reads m and loads σ, S, κ, q_P, l_P, c_P, m and n_0 into the unary registers. Width forms $\max(m, n_0)$, then $N = \max(m, n_0) + \kappa$, $\mathscr{B}(N)$ and W_m (Section I.9.6). The routines Prepare, MetaBank, Summ, corr and $\mathsf{max}^{14}_{11,12\to 8}$ are those of Section I.8.

The registers and the contracts

As in Section I.8, "within R rows" includes that the tape is a workspace of R rows with origin marks up to R, and ω is the saved request. The archive entries are numbered from 0: an archive $(v_0, v_1, \ldots)$ holds v_k in its k-th word. Throughout, the problem has a coefficient field; σ, S, G and κ are its schedule (Definition 15.18), q_P, l_P and c_P the coefficients of its capacity $\mathscr{B}(N) = q_P N^2 + l_P N + c_P$ (H.1), $N = N(\max(m, n_0)) = \max(m, n_0) + \kappa$ and $W_m = 4\mathscr{B}(N) + 4$. Put

$$\rho(m) = \sigma + S + \kappa + m + n_0 + q_P + l_P + c_P + N + q_P N \\ + (q_P N + l_P) + (q_P N + l_P)N + \mathscr{B}(N) + W_m + 10,$$

the *tail room*: the sum of every number the last program holds, with ten to spare.

Definition I.85 (Seed and scheduled request)**.** Fix a problem with a coefficient field, n_0, i and m, and let τ be a tape and $r \in \mathbb{N}$.

(a) τ is a *seed with r rows* if it carries the saved request (Definition I.41), is a workspace of r rows with origin marks up to r, tracks 5, 6, 7, 8, 9, 10 and 13 carry the counters σ, S, κ, 0, q_P, l_P and c_P, tracks 11 and 12 hold the values m and n_0, and $\rho(m) \le r$.

(b) τ is a *scheduled request with r rows* if it carries the saved request, is a workspace of r rows with origin marks up to r, tracks 5, 6, 7, 8, 9 and 10 carry the counters

$$\sigma, \quad S, \quad \kappa, \quad N, \quad W_m, \quad \mathscr{B}(N),$$

and $\sigma + S + \kappa + N + W_m + \mathscr{B}(N) \le r$.

So a scheduled request keeps the whole request on tracks 3, 4 and 15 and adds, in unary, every number of the schedule that the later stages need. The last inequality says that each of these numbers is at most the number of rows, which is what lets the later stages count up to them with their cursors. The three results below are proved at the ends of Sections I.9.5, I.9.6 and I.9.7.

Proposition I.86 (The scalar tail)**.** *Fix a problem with a coefficient field, n_0 and a number p. There is a constant K_T, depending only on these, with the following property for every i, m, r and tape τ. If τ carries the saved request and holds the archive $(p, \gamma, \alpha, \beta, s, L, s_0)$ within r rows, where γ, α, β, s, L and s_0 are the bounds of Theorem I.62, then* Scalar *runs from the root to the root in some t units to a seed with r' rows, and*

$$t + r' + 2 \le K_T (r + m + 2)^{80 \cdot 6^{51}}.$$

Lemma I.87 (The width)**.** *From every seed with r rows,* Width *runs from the root to the root in some t units to a scheduled request with r rows, and*

$$t \le 14 \cdot 10^5 \big(\rho(m) + 1\big)^3 + 37r + 13.$$

Moreover $\rho(m) \le c_\rho X^2$ for a constant c_ρ depending only on the problem and n_0, where $X = m + n_0 + \kappa + 1$.

Theorem I.88 (The stage Schedule)**.** *Fix a problem with a coefficient field and n_0. There is a constant K_q, depending only on the problem and n_0, such that for every i, m, r and every decoded request τ with r rows (Definition I.56),* Schedule *runs from the root to the root in some t units to a scheduled request with r' rows, and*

$$t + r' + 1 \le K_q \, (X + r + 1)^{g_q}, \qquad g_q = 4\big(72 \cdot 80 \cdot 6^{51} + 3\big) = 23040 \cdot 6^{51} + 12.$$

The stage writes the headers and the new rows, the origin marks of track 1, the scratch and register tracks 2 and 5 to 14, and tracks 3 and 4 beyond the end of ω, where the archive lies. It does not change ω or track 15; the numeric record of the decoded request is used and then overwritten.

The exponent is a fixed number, the same for every problem:

$$g_q = 111736803796728983743307618367111577196298252.$$

It is the product of the exponents of the parts. The archive reads of Ctl have exponent 2 each, the fifty-one instructions of Prog exponent 6 each, and Seed exponent 10; two bounds of this kind compose by Lemma I.73 with the product of their exponents, which gives $2^3 \cdot 6^{51} \cdot 10 = 80 \cdot 6^{51}$ for Scalar. The front half contributes 72 (Theorem I.62(c)), Prepare adds 2, and Width, whose exponent is 3, turns s into $4(s + 1)$ through Lemma 15.36(a). The exponent 6 of one instruction is $2 \cdot 3$: each instruction first adds about $64(r + 1)^2$ rows, so that a product of two archived numbers, each at most r, has room, and then works at a cost at most cubic in the new number of rows.

I.9.1 Reading a control number from its frame

The numbers M, T, n_0 and m are the frames number 5, 4, 3 and 1 from the end of the saved request. The machine finds such a frame by walking back from the end of ω over the frame-start marks of track 15, and reads it as Section I.7 reads every frame: it copies the bits to the binary stream and converts them to unary. Write the saved request as in Definition I.41: its frames are $f_0, \dots, f_F$, and f_k begins at row s_k, with $s_{F+1} = |\omega|$.

Lemma I.89 (Walking back over frames)**.** *Let*

$$\mathsf{send} = \mathsf{left}^4; \mathsf{scanL}, \qquad \mathsf{mk} = \mathsf{branch}(\mathsf{right}; \mathsf{write}\,1,\ \mathsf{right}; \mathsf{write}\,0),$$
$$\mathsf{prev} = \mathsf{right}^4; \mathsf{mk}; \mathsf{loop}(\mathsf{right}^{15}; \mathsf{mk}); \mathsf{left},$$

and $\mathsf{pf}^0 = \mathsf{skip}$, $\mathsf{pf}^{k+1} = \mathsf{pf}^k; \mathsf{prev}; \mathsf{right}^{12}$.

(a) If ω lies on tracks 3 and 4 from row 0, send *runs from $(-1,\tau)$ to $(\mathrm{cell}(|\omega|,3),\tau)$ in $18|\omega|+6$ units.*

(b) mk *runs from $\mathrm{cell}(j,15)$ to $\mathrm{cell}(j,14)$ in 3 units and writes into $(j,14)$ the complement of $(j,15)$.*

(c) If $\tau(a,15)=1$ and $\tau(j,15)=0$ for $a<j\le a+n$, then prev *runs from $\mathrm{cell}(a+n+1,3)$ to $\mathrm{cell}(a,15)$ in $20n+10$ units, writing only the cells $(j,14)$, $a\le j\le a+n$.*

(d) If τ carries the saved request and $1\le k\le F$, then pf^k *runs from $\mathrm{cell}(s_{F+1},3)$ to $\mathrm{cell}(s_{F+1-k},3)$ in $20(s_{F+1}-s_{F+1-k})+3k+1$ units, writing only track 14.*

Proof. (a) left^4 (5 units) reaches $-5=\mathrm{cell}(0,3)$, and scanL walks down the presence track of ω to its blank $(|\omega|,3)$ in $18|\omega|+1$ units (Lemma I.10). (b) A test, a move from $\mathrm{cell}(j,15)$ to $\mathrm{cell}(j,15)+1=\mathrm{cell}(j,14)$ and a write. (c) right^4 (5 units) moves from $\mathrm{cell}(a+n+1,3)$ to $\mathrm{cell}(a+n+1,3)+4=\mathrm{cell}(a+n,15)$, and mk (3) writes the complement of the mark of row $a+n$ into $(a+n,14)$. For $j\le n$ let I_j be the configuration at $(a+n-j,14)$ in which the cells $(a+n-i,14)$, $i\le j$, hold the complements of the marks of their rows. The scanned cell of I_j is 1 exactly when $j<n$. A pass from I_j, $j<n$, moves by right^{15} from $(a+n-j,14)$ to $(a+n-j-1,15)$ (16 units) and runs mk: 20 units with its test, ending in I_{j+1}. Lemma I.6 gives $20n+1$ units, and left returns from $(a,14)$ to $(a,15)$: in all $5+3+20n+1+1$. (d) Induction on k. Put $e=s_{F+1-k}$ and $e'=s_{F+2-k}$. By the induction hypothesis pf^{k-1} reaches $\mathrm{cell}(e',3)$; the marks of track 15 are those of the frame starts, so $(e,15)=1$ and the rows strictly between e and e' carry no mark. Part (c) with $a=e$ and $n=e'-e-1$, and right^{12} (13 units) from $\mathrm{cell}(e,15)$ to $\mathrm{cell}(e,3)$, add $20(e'-e)+3$ units. □

Each control number is read by the program

$$\mathsf{cur}_k=\mathsf{send};\mathsf{pf}^k;\mathsf{tc},\qquad \mathsf{getc}_k=\mathsf{cur}_k;\mathsf{clear}_{8,9};\mathsf{cc}_{2\to5};\mathsf{fetch};\mathsf{reset}_8;\mathsf{toUnary}^+,$$
$$\mathsf{ctl}_k=\mathsf{getc}_k;\mathsf{cc}_{9\to11};\mathsf{push},\qquad \mathsf{Ctl}=\mathsf{ctl}_5;\mathsf{ctl}_4;\mathsf{ctl}_3;\mathsf{skip},$$

where tc, cc and push are the routines of Lemmas I.66 and I.68, and fetch and $\mathsf{toUnary}^+$ those of Lemmas I.45 and I.49. The proof of Lemma I.68(g) shows that tc runs from $\mathrm{cell}(h,3)$, on a tape with origin marks up to h, to the root in $21h+7$ units, makes track 2 carry the counter h, and changes only track 2. For a number v write $\mathrm{bits}(v)$ for the word of its $\mathrm{bl}(v)+1$ lowest binary digits, so that the frames of M, T, n_0 and m are $\mathrm{bits}(M)$, $\mathrm{bits}(T)$, $\mathrm{bits}(n_0)$ and $\mathrm{bits}(m)$.

Lemma I.90 (Reading and archiving a control number)**.** *Let τ carry the saved request and be a workspace of R rows with origin marks up to R. For $1\le k\le 5$ let c_k be the control number whose frame is the k-th from the end: $c_5=M$, $c_4=T$, $c_3=n_0$, $c_2=i$ and $c_1=m$.*

(a) cur_k *runs from the root to the root within $(100+3k)(R+1)$ units, makes track 2 carry the first row $e=s_{F+1-k}$ of that frame and changes only tracks 2 and 14.*

Let moreover $k\in\{1,3,4,5\}$, and let τ hold an archive $v=(v_0,\dots,v_{K-1})$ within R rows.

(b) getc_k *runs from the root to the root in some t units to a tape that holds the value c_k on track 9, carries the saved request and holds the archive v within $R + c_k$ rows, and*

$$t + (R + c_k) + 1 \le 55000\,(R + c_k + 1)^2.$$

(c) ctl_k *runs from the root to the root in some t units to a tape that carries the saved request and holds the archive (v, c_k) within $r' = R + 2c_k + 1$ rows, and*

$$t + r' + 2 \le 10^8 (c_k + 1)^2 (R + 2)^2.$$

(d) Ctl *runs from the root to the root in some t units to a tape that carries the saved request and holds the archive (v, M, T, n_0) within r' rows, and $t + r' + 2 \le K_C (R + 2)^8$, where K_C depends only on M, T and n_0.*

Proof. (a) The presence marks of ω lie in rows $j \le R$, so $|\omega| \le R + 1$. Lemma I.89(a) and (d) take $18|\omega| + 6 + 20(|\omega| - e) + 3k + 1$ units to reach cell$(e, 3)$, writing only track 14, and tc takes $21e + 7$; since $e \le |\omega| \le R + 1$ the sum is at most $39(R + 1) + 3k + 14 \le (100 + 3k)(R + 1)$.

(b) The archive gives $|\omega| + 1 \le R$. After (a), $\mathsf{clear}_{8,9}$ ($112R + 51$ units, Lemma I.16) leaves tracks 8 and 9 blank, and $\mathsf{cc}_{2\to5}$ (at most $300(R + 1)$) puts the value e on track 5. By Lemma I.44 the tape lays out $g = \mathrm{bits}(c_k)$, followed by the remaining frames of ω, from row e, and $n = |g|$ satisfies $e + n \le |\omega| \le R$. Lemma I.45 with $a = e$ and $o = 0$ copies g to the binary stream within $1000(R+1)^2$ units, leaving track 8 carrying n; reset_8 ($36n + 48$ units, Lemma I.14) brings it back to 0, and track 9 still holds the value 0. Lemma I.49 with $\ell = n$, $r = 0$ and the value c_k of g then writes the value c_k on track 9 and adds c_k rows, within $1000(n + R + c_k + 1)^2 \le 4000(R + c_k + 1)^2$ units. No routine writes tracks 3, 4 or 15 in the old rows, and the new rows carry 0 there, so the saved request and the archive are kept; the root and the cells $x \ge 0$ are not written. With $x = R + c_k + 1$ the total plus $R + c_k + 1$ is at most $5000x^2 + 700x \le 55000x^2$.

(c) After (b), $\mathsf{cc}_{9\to11}$ (at most $300(R + c_k + 1)$) puts c_k on track 11, and Lemma I.68(h), with $R + c_k$ rows, $n = c_k$ and the counter c_k on track 9, appends c_k to the archive within $10^6(R + 2c_k + 2)^2$ units and $R + 2c_k + 1$ rows. With $x = R + 2c_k + 2 \le (c_k + 1)(R + 2)$ the total plus $r' + 2$ is at most $55000x^2 + 300x + 10^6x^2 + x \le 10^8x^2$.

(d) The program skip takes 1 unit and keeps the rows, so $1 + R + 2 \le 2(R + 2)^1$. Lemma I.73 combines this with (c) for $k = 3$ (exponents 2 and 1), then with (c) for $k = 4$ (exponents 2 and 2), then with (c) for $k = 5$ (exponents 2 and 4). The exponent is $2 \cdot 2 \cdot 2 = 8$, and the constant is built from $10^8(n_0 + 1)^2$, $10^8(T + 1)^2$, $10^8(M + 1)^2$ and 2 alone. □

I.9.2 Rows for products

A unary number needs one row per unit, and a product of two numbers that each fit the rows need not. Before each instruction of the scalar program, the machine therefore makes the number of rows roughly quadratic: it counts the rows on track 12 and adds 64 times that many rows, once for every row. The count of track 12 is copied from the headers of track 0, and track $c \in \{13, 14\}$ is the outer cursor. Put

$$Q(R) = R + 1 + 64(R + 1)^2, \qquad \text{so that} \qquad R \le Q(R) \quad \text{and} \quad Q(R) + 1 \le 66(R + 2)^2.$$

Lemma I.91 (Quadratic growth). *Let $c \in \{13, 14\}$, let* rdc *be the routine of Lemma I.67, and put*

$$\mathsf{hc}_c = \mathsf{clear}_{12,c}; \mathsf{zero}_{12}; \mathsf{left}^1; \mathsf{loop}(\mathsf{shift}_{0\to 12}; \mathsf{next}; \mathsf{shift}_{12\to 0}); \mathsf{left}; \mathsf{home}_1,$$
$$\mathsf{more} = \mathsf{rdc}; \mathsf{loop}(\mathsf{home}_{12}; \mathsf{grow}; \mathsf{inc}_2; \mathsf{rdc}); \mathsf{home}_{12}, \qquad \mathsf{one} = \mathsf{zero}_2; \mathsf{more},$$
$$\mathsf{roc}_c = \mathsf{seek}_c; \mathsf{shift}_{c\to 12}, \qquad \mathsf{res}_c = \mathsf{roc}_c; \mathsf{loop}(\mathsf{home}_{12}; \mathsf{one}^{64}; \mathsf{inc}_c; \mathsf{roc}_c); \mathsf{home}_{12},$$
$$\mathsf{quad}_c = \mathsf{grow}; \mathsf{hc}_c; \mathsf{res}_c; \mathsf{clear}_{12,c}, \qquad \mathsf{room} = \mathsf{clear}_{5,6,7,8,9,10,11}; \mathsf{quad}_{13},$$

where one^{64} *is* 64 *copies of* one *followed by* skip. *Let τ be a workspace of R rows with origin marks up to R. Each routine runs from the root to the root, and none writes tracks* 3, 4 *or* 15 *in the rows $j \le R$.*

(a) hc_c *makes track* 12 *hold the value R and track c carry* 0, *within* $300(R+1)$ *units; only tracks* 12 *and c change.*

(b) If track 12 *holds a value $n \le R$, then* one *leaves a workspace of $R+n$ rows with origin marks up to $R+n$ and track* 12 *still holding n, within* $2000(R+n+2)^2$ *units. Besides the headers of the rows $R, \dots, R+n-1$ and the new rows, only track* 2 *is written; so every counter at most R on a track other than* 2 *is kept.*

(c) If track 12 *holds $n \le R$ and track c carries* 0, *then* res_c *leaves a workspace of $R + 64n^2$ rows with origin marks up to $R + 64n^2$, track* 12 *still holding n and track c carrying n, within* $500000(n+1)(R+64n^2+65n+2)^2$ *units; every counter at most R on a track other than* 2 *and c is kept.*

(d) quad_c *leaves a workspace of $Q(R)$ rows with origin marks up to $Q(R)$ and tracks* 12 *and c blank, within* $3 \cdot 10^9 (R+2)^5$ *units; every counter at most R on a track other than* 2, 12 *and c is kept.*

(e) If τ holds an archive within R rows, then room *leaves the same archive within $Q(R)$ rows, within* $4 \cdot 10^9 (R+2)^5$ *units.*

Proof. (a) $\mathsf{clear}_{12,c}$ costs at most $130R + 69$ units (Lemma I.16), zero_{12} costs 31 and left^1 (2) reaches the header $(0, 0)$. The loop is the copying loop of Lemma I.15(a) with the source track 0, whose headers carry the counter R: a pass from $(j, 0)$, $j < R$, advances the hole of track 12 and moves to $(j+1, 0)$ in 47 units with its test, and the loop stops at the sentinel, in $47R + 1$ units. left (1) steps to $(R, 1)$, and home_1 $(18R + 5)$ returns along the origin marks. In all at most $195R + 109 \le 300(R+1)$. Track 12 was blank, so it now holds the value R.

(b) zero_2 (11 units) makes track 2 carry 0, and more is alloc of Lemma I.67 without its first grow. The proof of that lemma, with R in place of $R+1$, shows that rdc (16 units) reaches $(0, 12)$, that the pass from row $j < n$ adds one row and advances track 2 in $72R + 144j + 198$ units with its test, and that the loop stops at the blank $(n, 12)$; the final home_{12} costs $18n + 16$. The total is $72Rn + 72n^2 + 144n + 44 \le 2000(R+n+2)^2$. A new row has 0 on track 12, and it lies beyond the hole of every counter at most R.

(c) roc_c ($c + |c-12| + 4 \le 20$ units) reaches $(0, 12)$. For $j \le n$ let I_j be the configuration at $(j, 12)$ whose tape is a workspace of $R_j = R + 64jn$ rows with origin marks up to R_j, holds n on track 12 and carries j on track c. A pass from I_j, $j < n$, scans a mark, returns to the root $(18j + 16)$, runs one sixty-four times by (b), each time adding n rows within $2000(R_{j+1} + n + 2)^2$ units,

advances track c ($36j + 2c + 44$, Lemma I.14) and reads $(j + 1, 12)$ ($18j + c + |c - 12| + 22$): at most $200000(R_j + 65n + 2)^2$ units, ending in I_{j+1}. At I_n the scanned cell is the blank $(n, 12)$. Lemma I.6 and the final home_{12} ($18n + 16$) give the bound, since $R_j \le R + 64n^2$.

(d) grow ($72R + 99$ units, Lemma I.17) gives $R + 1$ rows; (a) puts the count $R + 1$ on track 12 within $300(R + 2)$ units; (c) with $n = R + 1$ gives $R + 1 + 64(R + 1)^2 = Q(R)$ rows within $500000(R + 2)\big(64(R + 1)^2 + 66(R + 1) + 2\big)^2 \le 500000 \cdot 4096(R + 2)^5$ units; and $\mathsf{clear}_{12,c}$ on $Q(R)$ rows costs at most $131Q(R) + 69 \le 8715(R + 2)^2$. The sum is at most $3 \cdot 10^9(R + 2)^5$. A counter at most R has its hole in an old row, which none of these parts writes on its track.

(e) $\mathsf{clear}_{5,\dots,11}$ costs $385R + 169$ units; then (d) with $c = 13$. The archive lies on tracks 3 and 4, whose old rows are not written, and the new rows carry 0 there. □

I.9.3 Unary products, binary lengths and powers

The scalar program and the last program of the stage multiply unary numbers, and the scalar program also takes binary lengths and powers. All three are loops of the counter routines of Section I.4 and the division udiv of Lemma I.78: a product adds one factor once for each mark of the other, a binary length halves until nothing is left, and a power multiplies by the base once for each mark of the exponent, after reserving the rows for the next power.

Lemma I.92 (Unary products)**.** *For distinct tracks $A, B, C, s, c \in \{2, \dots, 15\}$ put*

$$\mathsf{umul}^{s,c}_{A,B\to C} = \mathsf{zero}_C; \mathsf{zero}_c; \mathsf{seek}_c; \mathsf{shift}_{c\to B}; \mathsf{loop}(\Pi); \mathsf{home}_B,$$
$$\Pi = \mathsf{shift}_{B\to c}; \mathsf{next}; \mathsf{home}_c; \mathsf{acc}^{s}_{A\to C}; \mathsf{seek}_c; \mathsf{shift}_{c\to B}.$$

If track A carries a, track B carries b, and τ has origin marks up to R with $a, b, ab \le R$, then $\mathsf{umul}^{s,c}_{A,B\to C}$ runs from the root to the root within $2000(a + b + ab + 1)^3$ units; afterwards track C carries ab and track c carries b. Only tracks C, s and c change, and only in rows $j \le R$.

Proof. zero_C, zero_c, seek_c and $\mathsf{shift}_{c\to B}$ take at most $2C+3c+|c-B|+18$ units and reach $(0, B)$ with the counters 0 on tracks C and c. For $j \le b$ let I_j be the configuration at (j, B) in which track c carries j and track C carries ja. A pass from I_j, $j < b$, scans a mark of B, moves to the hole (j, c) and advances it ($|B-c|+21$ units), returns to the root ($18(j+1)+c+4$), adds a to track C (at most $500((j+1)a+1)^2$ units by Lemma I.15(b), since $(j+1)a \le R$), and goes back to the hole of track c and over to $(j+1, B)$ ($18(j+1)+c+|c-B|+4$). It ends in I_{j+1} within $500(a+b+ab+1)^2+36b+111$ units with its test. At I_b the scanned cell is the hole of track B. Lemma I.6 and home_B ($18b+B+4$) bound the total by $500b(a+b+ab+1)^2 + 36b^2 + 129b + 126 \le 2000(a+b+ab+1)^3$. The addition writes only tracks C and s, and the advances only track c, all in rows at most $\max(ab, b) \le R$. □

Lemma I.93 (Binary length)**.** *Let*

$$\mathsf{ublen} = \mathsf{left}^{12}; \mathsf{loop}\big(\mathsf{home}_{11}; \mathsf{udiv}; \mathsf{clear}_{11}; \mathsf{copy}_{7\to 11}; \mathsf{clear}_{2,5,6,7}; \mathsf{inc}_8; \mathsf{left}^{12}\big); \mathsf{home}_{11}.$$

Let τ be a workspace of R rows with origin marks up to R that holds the value n on track 11 and the value 2 on track 12, carries 0 on tracks 2, 5 and 6, and has tracks 7 and 8 blank, with $n + 2 \le R$. Then ublen runs from the root to the root within $(2n + 1) \cdot 10^4(R + 1)^2 + 28$ units to a workspace of R rows with origin marks up to R that holds the value $\mathrm{bl}(n)$ on track 8, the value 0 on track 11 and the value 2 on track 12. Tracks 3, 4 and 15 are not written.

Proof. left^{12} (13 units) reaches $(0, 11)$. For $k \le \mathrm{bl}(n)$ let J_k be the configuration at $(0, 11)$ whose tape holds $\lfloor n/2^k \rfloor$ on track 11, k on track 8 and 2 on track 12, carries 0 on tracks 2, 5 and 6, and has track 7 blank. The scanned cell is 1 exactly when $\lfloor n/2^k \rfloor \ge 1$, that is when $k < \mathrm{bl}(n)$. A pass from J_k returns to the root (15 units); udiv leaves the value $\lfloor \lfloor n/2^k \rfloor / 2 \rfloor = \lfloor n/2^{k+1} \rfloor$ on track 7 within $1000(n+3)^2$ units (Lemma I.78, with $b = 2$); clear_{11} and $\mathsf{copy}_{7\to 11}$ move it to the blanked track 11, so that track 11 holds it as a value; $\mathsf{clear}_{2,5,6,7}$ blanks the scratch tracks; inc_8 makes track 8 hold $k+1$; and left^{12} returns to $(0, 11)$. By Lemmas I.16, I.15 and I.14 the pass costs at most $9000(R+1)^2$ units and ends in J_{k+1}. Lemma I.6 with $\mathrm{bl}(n) \le n$ passes and home_{11} (15) give the bound. Every part writes only tracks 2, 5, 6, 7, 8 and 11, in rows $j \le R$. □

Lemma I.94 (Powers with their rows)**.** *Let* one *be the routine of Lemma I.91 and put*

$$\begin{aligned}
\mathsf{rde} &= \mathsf{seek}_2; \mathsf{shift}_{2\to 6}, \qquad \mathsf{prr} = \mathsf{seek}_{14}; \mathsf{shift}_{14\to 11},\\
\mathsf{prs} &= \mathsf{prr}; \mathsf{loop}(\mathsf{home}_{11}; \mathsf{one}; \mathsf{inc}_{14}; \mathsf{prr}); \mathsf{home}_{11},\\
\mathsf{rsv} &= \mathsf{cc}_{2\to 8}; \mathsf{cc}_{5\to 12}; \mathsf{cc}_{7\to 11}; \mathsf{clear}_{13,14}; \mathsf{prs}; \mathsf{cc}_{12\to 5}; \mathsf{cc}_{8\to 2},\\
\mathsf{pbody} &= \mathsf{home}_6; \mathsf{umul}^{12,14}_{5,7\to 11}; \mathsf{clear}_7; \mathsf{copy}_{11\to 7}; \mathsf{inc}_2; \mathsf{rde},\\
\mathsf{upow} &= \mathsf{rde}; \mathsf{loop}(\mathsf{home}_6; \mathsf{rsv}; \mathsf{rde}; \mathsf{pbody}); \mathsf{home}_6.
\end{aligned}$$

Let τ be a workspace of R rows with origin marks up to R that holds the values $a \ge 1$, n and 1 on tracks 5, 6 and 7 and carries 0 on track 2, with $a, n \le R$. Then upow *runs from the root to the root within $4 \cdot 10^6 (n+1)(R + na^n + 1)^3$ units to a workspace of R_n rows with origin marks up to R_n that holds the value a^n on track 7, where*

$$R_0 = R, \qquad R_{j+1} = R_j + a^{j+1}, \qquad \text{so that} \quad R_n \le R + na^n \quad \text{and} \quad a^n \le R_n.$$

No cell of tracks 3, 4 or 15 changes.

Proof. For $j \le n$ let P_j be the configuration at $(j, 6)$ whose tape is a workspace of R_j rows with origin marks up to R_j, holds a, n and a^j on tracks 5, 6 and 7, and carries j on track 2. Every one of these numbers is at most R_j: a, n and $j \le n$ are at most $R \le R_j$, $a^0 = 1 \le a \le R$, and for $j \ge 1$ the power a^j is a summand of R_j. rde ($18j + 10$ units from the root) reaches $(j, 6)$, so the loop starts in P_0; its test is 1 exactly when $j < n$.

A pass from P_j, $j < n$, returns to the root ($18j + 10$) and runs rsv. Its first four parts save j on track 8, put the base a on track 12 and the count a^j on track 11, and blank tracks 13 and 14, within $1000(R_j + 1)$ units (Lemmas I.66 and I.16). prs runs one once for each of the a^j marks of track 11, with track 14 as its cursor; by Lemma I.91(b) each run adds a rows, so prs adds $a \cdot a^j = a^{j+1}$ rows, within $30000(a^j+1)(R_{j+1}+a+2)^2$ units, as in the proof of Lemma I.91(c). The new rows carry 1 on tracks 2, 5, 8, 11 and 14 (Lemma I.17), which keeps every counter whose hole lies in an old row; the last two copies restore the value a on track 5 and the counter j on track 2, which the runs of one used as their own counter. So rsv costs at most $10^6(R_{j+1}+1)^3$ units and leaves R_{j+1} rows. Then rde returns to $(j, 6)$, and pbody returns to the root, forms $a \cdot a^j = a^{j+1} \le R_{j+1}$ on track 11 (Lemma I.92, within $2000(a + a^j + a^{j+1} + 1)^3 \le 54000(R_{j+1}+1)^3$ units), moves it to the blanked track 7, advances track 2 to $j+1$ and reads $(j+1, 6)$. The pass costs at most $2 \cdot 10^6 (R_{j+1}+1)^3$ units with its test and ends in P_{j+1}. At P_n the scanned cell is the blank $(n, 6)$. Lemma I.6, with every pass bounded by $2 \cdot 10^6 (R + na^n + 1)^3$, and home_6

j	instruction	v_j	j	instruction	v_j
10	1	1	36	$\mathrm{bl}(v_{35})$	$\kappa-2$
11	2	2	37	$v_{36}+v_{11}$	κ
12	3	3	38	$\mathrm{bl}(v_7)$	$\mathrm{bl}(M)$
13	4	4	39	$\mathrm{bl}(v_{19})$	$\mathrm{bl}(R)$
14	$v_{10}+v_{13}$	5	40	$v_{38}+v_{39}$	$r-1$
15	$v_{12}+v_{13}$	7	41	$v_{40}+v_{10}$	r
16	$v_3\cdot v_7$	βM	42	$v_{11}\cdot v_{41}$	$2r$
17	v_1+v_2	$\gamma+\alpha$	43	$v_{42}\cdot v_5$	$2rL$
18	$v_{17}+v_{16}$	$\gamma+\alpha+\beta M$	44	$v_{43}+v_{22}$	$2rL+\sigma$
19	$v_{18}+v_{10}$	R	45	$v_{44}+v_{41}$	q_P
20	$v_{13}\cdot v_{19}$	$4R$	46	$v_{12}\cdot v_{41}$	$3r$
21	$\mathrm{bl}(v_{20})$	$\mathrm{bl}(4R)$	47	$v_{14}\cdot v_{41}$	$5r$
22	$v_{21}+v_{10}$	σ	48	v_4+v_{47}	$s+5r$
23	$v_{11}^{v_{22}}$	2^σ	49	$v_{48}+v_{10}$	$s+5r+1$
24	$v_8\cdot v_{23}$	S	50	$v_{49}\cdot v_5$	$(s+5r+1)L$
25	$v_{23}+v_{20}$	$2^\sigma+4R$	51	$v_{46}+v_{50}$	l_P-4
26	$v_{25}^{v_{24}}$	$(2^\sigma+4R)^S$	52	$v_{51}+v_{13}$	l_P
27	$v_{22}\cdot v_{24}$	σS	53	$v_{11}\cdot v_4$	$2s$
28	$v_{11}^{v_{27}}$	$2^{\sigma S}$	54	v_6+v_{53}	s_0+2s
29	$\lfloor v_{26}/v_{28}\rfloor$	$G-1$	55	$v_{54}+v_{42}$	$s_0+2s+2r$
30	$v_{29}+v_{10}$	G	56	$v_{53}+v_{46}$	$2s+3r$
31	v_7+v_{11}	$M+2$	57	$v_{56}+v_{11}$	$2s+3r+2$
32	$v_{11}\cdot v_{24}$	$2S$	58	$v_{57}\cdot v_5$	$(2s+3r+2)L$
33	$v_{32}\cdot v_{31}$	$2S(M+2)$	59	$v_{55}+v_{58}$	c_P-7
34	$v_{10}+v_{33}$	$1+2S(M+2)$	60	$v_{59}+v_{15}$	c_P
35	$v_{30}\cdot v_{34}$	$G(1+2S(M+2))$			

Table I.2. The scalar program. Instruction j computes v_j from earlier entries of the archive and appends it. The archive starts with $v_0=p$, $(v_1,\dots,v_6)=(\gamma,\alpha,\beta,s,L,s_0)$ and $(v_7,v_8,v_9)=(M,T,n_0)$; here $r=\mathrm{bl}(M)+\mathrm{bl}(R)+1$.

give the bound. Finally $R_n=R+\sum_{1\le j\le n}a^j\le R+na^n$. The routines write only the headers, tracks 2, 5, 7, 8 and 11 to 14, and the new rows, which carry 0 on tracks 3, 4 and 15. $\square$

I.9.4 The scalar program

After Ctl the archive holds ten numbers, $v_0=p$, $(v_1,\dots,v_6)=(\gamma,\alpha,\beta,s,L,s_0)$ and $(v_7,v_8,v_9)=(M,T,n_0)$. The scalar program appends fifty-one more, $v_{10},\dots,v_{60}$, each computed from earlier entries by one of six operations: a constant at most 4, a sum, a product, a floor quotient, a binary length or a power. Table I.2 lists them. The instructions refer to entries by their fixed positions, never by their values, so the program is the same for every problem.

Lemma I.95 (What the scalar program computes). *Let $v_0, \dots, v_9$ be as above and, for $10 \le j \le 60$, let v_j be the result of instruction j of Table I.2 on $v_0, \dots, v_{j-1}$. Every instruction reads only earlier entries, every constant is at most 4, the divisor v_{28} is positive and the bases v_{11} and v_{25} of the powers are at least 1. Each v_j is the number in the third column; in particular*

$$(v_{19}, v_{22}, v_{24}, v_{30}, v_{37}) = (R, \sigma, S, G, \kappa), \qquad (v_{45}, v_{52}, v_{60}) = (q_P, l_P, c_P),$$

the schedule of Definition 15.18 and the coefficients of the capacity (H.1).

Proof. Induction along the table: each row applies its operation to rows already identified. For instance $v_{19} = (\gamma + \alpha + \beta M) + 1 = R$, $v_{22} = \mathrm{bl}(4R) + 1 = \sigma$, $v_{24} = T2^{\sigma} = S$, $v_{30} = \lfloor (2^{\sigma}+4R)^{S}/2^{\sigma S} \rfloor + 1 = G$ and $v_{37} = \mathrm{bl}(G(1+2S(M+2)))+2 = \kappa$; with $r = v_{41} = \mathrm{bl}(M)+\mathrm{bl}(R)+1$ as in Section H.1, $v_{45} = 2rL + \sigma + r = q_P$, $v_{52} = 3r + (s + 5r + 1)L + 4 = l_P$ and $v_{60} = s_0 + 2s + 2r + (2s + 3r + 2)L + 7 = c_P$. The divisor is $2^{\sigma S} \ge 1$, and the bases are 2 and $2^{\sigma} + 4R \ge 1$. □

Every instruction is a program that grows the rows, loads its operands from the archive, computes, and appends the result. The loading and appending are common to all six kinds:

$$\mathsf{store}_k = \mathsf{cc}_{k \to 11}; \mathsf{start}; \mathsf{end}; \mathsf{append} \qquad (5 \le k \le 10),$$
$$\mathsf{pair}_{i,j}^{A,B} = \mathsf{room}; \mathsf{load}_{i,A}; \mathsf{load}_{j,B},$$

with the archive routines of Lemma I.68 and room of Lemma I.91.

Lemma I.96 (Loading and storing). *Let τ hold an archive $v = (v_0, \dots, v_{K-1})$ within R rows.*

(a) If track k carries $x \le R$, then store_k runs from the root to the root within $10^6(R+1)^2$ units to a tape that holds the archive (v, x) within $R + 1 + x$ rows; track 15 does not change.

(b) If $i, j < K$ and $A \ne B$ lie in $\{5, \dots, 11\}$, then $\mathsf{pair}_{i,j}^{A,B}$ runs from the root to the root within $4 \cdot 10^9(R+2)^5 + (i + j + 6) \cdot 20000\,(Q(R)+1)^2$ units to a tape that holds the same archive within $Q(R)$ rows and the values v_i and v_j on tracks A and B; track 15 does not change.

Proof. (a) $\mathsf{cc}_{k \to 11}$ (at most $300(R+1)$ units) puts the value x on track 11. By Lemma I.68(a) and (g), $\mathsf{start}; \mathsf{end}$ puts on track 13 the end $e = |\omega| + 1 + \sum_j (v_j + 1) \le R$ of the archive within $3000(R+1)$ units, and by (f) append writes the word $u(x)$ from row e and its terminator within $100000(R + e + x + 3)^2 \le 900000(R+1)^2$ units, ending within $R + 1 + x$ rows, with the rows before e of tracks 3 and 4 and all of track 15 unchanged. The sum is at most $10^6(R+1)^2$. (b) Lemma I.91(e), then Lemma I.68(c) twice with $Q(R)$ rows; the second load changes only tracks 12, 13 and B, so track A keeps v_i. □

The six instruction programs are

$$\begin{aligned}
\mathsf{op}[c] &= \mathsf{room};\mathsf{clear}_5;\mathsf{inc}_5^{c};\mathsf{store}_5,\\
\mathsf{op}[v_i+v_j] &= \mathsf{pair}_{i,j}^{5,7};\mathsf{acc}_{5\to7}^{2};\mathsf{store}_7,\\
\mathsf{op}[v_i\cdot v_j] &= \mathsf{room};\mathsf{load}_{i,5};\mathsf{load}_{j,7};\mathsf{umul}_{5,7\to11}^{12,14};\mathsf{start};\mathsf{end};\mathsf{append},\\
\mathsf{op}[\lfloor v_i/v_j\rfloor] &= \mathsf{pair}_{i,j}^{11,10};\mathsf{clear}_{2,5,6,7};\mathsf{cc}_{10\to12};\mathsf{udiv};\mathsf{store}_7,\\
\mathsf{op}[\mathrm{bl}(v_i)] &= \mathsf{room};\mathsf{load}_{i,11};\mathsf{clear}_{2,5,6,7,8};\mathsf{clear}_{12};\mathsf{inc}_{12}^{2};\mathsf{ublen};\mathsf{store}_8,\\
\mathsf{op}[v_i^{\,v_j}] &= \mathsf{pair}_{i,j}^{5,6};\mathsf{clear}_{2,7,9,10};\mathsf{clear}_7;\mathsf{inc}_7^{1};\mathsf{upow};\mathsf{store}_7.
\end{aligned}$$

Lemma I.97 (One instruction)**.** *Let τ hold an archive $v=(v_0,\dots,v_{K-1})$ within r rows, and let ι be one of the six instructions with operand positions $i,j<K$, a constant $c\le 4$, a positive divisor and a base at least* 1*. Let $\iota(v)$ be its result. Then* $\mathsf{op}[\iota]$ *runs from the root to the root in some t units to a tape that holds the archive $(v,\iota(v))$ within r' rows, with track* 15 *unchanged, and*

$$t\le C_\iota(r+2)^6,\qquad r'\le D_\iota(r+2)^6,\qquad t+r'+2\le(C_\iota+D_\iota+2)(r+2)^6,$$

where (C_ι,D_ι) is $(10^{12},132)$ for a constant, $((i+j+11)10^{12},132)$ for a sum or a quotient, $((i+j+10)10^{12},132)$ for a product, $((i+10)10^{12},132)$ for a binary length, and, for a power with $a=v_i$ and $n=v_j$,

$$C_\iota=(i+j+10)10^{12}+\big(4\cdot10^6(n+1)+1010000\big)\cdot287496\,(na^n+1)^3,$$
$$D_\iota=66\big((n+1)a^n+1\big).$$

Proof. Put $Q=Q(r)$; then $(Q+1)^2\le4356(r+2)^6$ and $(Q+1)^3\le287496(r+2)^6$, and $(r+2)^5\le(r+2)^6$. Every entry of the archive is at most r, so a sum of two entries is at most $2r\le Q$, a product at most $r^2\le Q$, and a quotient or a binary length at most r. In the first five kinds the result x is therefore at most Q, and $r'=Q+1+x\le2Q+1\le132(r+2)^6$.

Constant. room costs at most $4\cdot10^9(r+2)^5$ (Lemma I.91(e)); clear_5 and inc_5^{c} ($49Q+19$ and $18c^2+36c+1$ units) leave the value c on track 5; store_5 costs at most $10^6(Q+1)^2$. The total is at most $4\cdot10^9(r+2)^5+2\cdot10^6(Q+1)^2\le10^{12}(r+2)^6$.

Sum. By Lemma I.96(b) tracks 5 and 7 hold $a=v_i$ and $b=v_j$; $\mathsf{acc}_{5\to7}^{2}$ makes track 7 carry $a+b$ within $500(Q+1)^2$ units (Lemma I.15(b)), and store_7 appends it within $10^6(Q+1)^2$. The total is at most $4\cdot10^9(r+2)^5+(i+j+6)\cdot20000(Q+1)^2+1010000(Q+1)^2\le(i+j+11)10^{12}(r+2)^6$.

Product. After the two loads, $\mathsf{umul}_{5,7\to11}^{12,14}$ leaves the counter ab on track 11 within $2000(a+b+ab+1)^3\le54000(Q+1)^3$ units (Lemma I.92), and $\mathsf{start};\mathsf{end};\mathsf{append}$ appends it within $3000(Q+1)+100000(3Q+3)^2$ units, as in Lemma I.96(a). The total is at most $4\cdot10^9(r+2)^5+(i+j+6)20000(Q+1)^2+54000(Q+1)^3+3000(Q+1)+900000(Q+1)^2\le(i+j+10)10^{12}(r+2)^6$.

Quotient. Tracks 11 and 10 hold $a=v_i$ and $b=v_j\ge1$. The clearing and $\mathsf{cc}_{10\to12}$ prepare Lemma I.78 (the value b on track 12, tracks 2, 5, 6 and 7 blank, $a+1\le Q$), and udiv leaves the value $\lfloor a/b\rfloor$ on track 7; the three parts cost at most $10^4(Q+1)^2$ units. With store_7 the total is at most the bound of a sum.

Binary length. After the load, track 11 holds $n = v_i$; the clearing and inc_{12}^{2} give the hypotheses of Lemma I.93 with Q rows and $n + 2 \le Q$, and ublen leaves the value $\mathrm{bl}(n)$ on track 8. These parts cost at most $40000(Q+1)^3$ units, the load and room at most $4 \cdot 10^9 (r+2)^5 + (i+3)20000(Q+1)^2$, and store_8 at most $10^6(Q+1)^2$; the total is at most $(i+10)10^{12}(r+2)^6$.

Power. Tracks 5 and 6 hold $a \ge 1$ and n; the clearing and inc_{7}^{1} (at most $10^4(Q+1)^2$ units) give the hypotheses of Lemma I.94 with $R = Q$, so upow leaves the value a^n on track 7 within $4 \cdot 10^6 (n+1)(Q + na^n + 1)^3$ units and $Q_n \le Q + na^n$ rows, with $a^n \le Q_n$; store_7 costs at most $10^6(Q_n+1)^2 \le 10^6(Q+na^n+1)^3$ and leaves $r' = Q_n + 1 + a^n$ rows. Since $Q + na^n + 1 \le (na^n+1)(Q+1) \le 66(na^n+1)(r+2)^2$, the time is at most $C_\iota (r+2)^6$; and $r' \le Q + (n+1)a^n + 1 \le ((n+1)a^n+1)(Q+1) \le D_\iota (r+2)^6$.

In every kind the routines write only scratch tracks, tracks 3 and 4 from the end of the archive on, and new rows; so the archive grows by exactly $\iota(v)$ and track 15 is kept. □

Lemma I.98 (The scalar program). *Let* $\mathsf{Prog} = \mathsf{op}[\iota_{10}]; \mathsf{op}[\iota_{11}]; \cdots; \mathsf{op}[\iota_{60}]; \mathsf{skip}$, *where* ι_j *is instruction* j *of Table I.2. Fix* $v_0, \dots, v_9$. *There is a constant* K_P, *depending only on* $v_0, \dots, v_9$, *such that from every tape that carries the saved request and holds the archive* $(v_0, \dots, v_9)$ *within* r *rows,* Prog *runs from the root to the root in some* t *units to a tape that carries the saved request and holds the archive* $(v_0, \dots, v_{60})$ *of Lemma I.95 within* r' *rows, and*

$$t + r' + 2 \le K_P (r+2)^{6^{51}}.$$

Proof. For $10 \le j \le 61$ let $\mathsf{Prog}_j = \mathsf{op}[\iota_j]; \cdots; \mathsf{op}[\iota_{60}]; \mathsf{skip}$, so that $\mathsf{Prog}_{61} = \mathsf{skip}$ and $\mathsf{Prog}_{10} = \mathsf{Prog}$. We show by downward induction on j that there is a constant K_j, depending only on $v_0, \dots, v_{j-1}$, such that from every tape holding $(v_0, \dots, v_{j-1})$ within r rows, Prog_j reaches a tape holding $(v_0, \dots, v_{60})$ within r' rows with $t + r' + 2 \le K_j (r+2)^{6^{61-j}}$. For $j = 61$, skip takes 1 unit and $1 + r + 2 \le 2(r+2)$. For $j \le 60$, Lemma I.95 says that ι_j is admissible on $(v_0, \dots, v_{j-1})$ with result v_j, and Lemma I.97 gives $t + r'' + 2 \le K_{\iota_j}(r+2)^6$ for $\mathsf{op}[\iota_j]$, where $K_{\iota_j} = C_{\iota_j} + D_{\iota_j} + 2$ depends only on j and the entries ι_j reads. Lemma I.73 with the exponents 6 and 6^{60-j} gives the claim for j with $K_j = K_{\iota_j} + K_{j+1} K_{\iota_j}^{6^{60-j}}$ and the exponent $6 \cdot 6^{60-j} = 6^{61-j}$. The saved request is kept because track 15 and the rows of ω are never written. For $j = 10$ the exponent is 6^{51}. □

So the exponent 6^{51} is 6, the exponent of one instruction, raised to the number 51 of instructions. The constant K_P is astronomically large, since the power instructions write $(2^\sigma + 4R)^S$ and $2^{\sigma S}$ in unary, but it is fixed once the problem is fixed.

I.9.5 The seed

The last part of Scalar reads m, copies the eight numbers that the tail needs from the archive into the registers, and adds enough rows for every number the tail will form. The archive

reads use tracks 12 and 13 as scratch, so the order of the moves matters: m is parked on track 8, n_0 on track 14, and c_P is moved to track 13 only after the last read. Put

$$\begin{aligned}\mathsf{mv} &= \mathsf{cc}_{9\to8};\mathsf{load}_{22,5};\mathsf{load}_{24,6};\mathsf{load}_{37,7};\mathsf{load}_{45,9};\mathsf{load}_{52,10};\mathsf{load}_{9,11};\mathsf{cc}_{11\to14};\\ &\qquad \mathsf{load}_{60,11};\mathsf{cc}_{11\to13};\mathsf{cc}_{8\to11};\mathsf{cc}_{14\to12};\mathsf{clear}_{8};\mathsf{skip},\\ \mathsf{troom} &= \mathsf{cc}_{12\to8};\mathsf{quad}_{14};\mathsf{quad}_{14};\mathsf{cc}_{8\to12};\mathsf{cc}_{11\to14};\mathsf{cc}_{14\to11};\mathsf{clear}_{8,14},\\ \mathsf{Seed} &= \mathsf{getc}_{1};\mathsf{mv};\mathsf{troom}.\end{aligned}$$

Lemma I.99 (The seed)**.** *Let τ carry the saved request and hold the archive $(v_0,\dots,v_{60})$ of Lemma I.95 within R rows. Then* Seed *runs from the root to the root in some t units to a seed with $r' = Q(Q(R+m))$ rows (Definition I.85(a)), and*

$$t + r' + 1 \le 2\cdot 10^{22}\,(R+m+2)^{10}.$$

Proof. Put $R' = R + m$. By Lemma I.90(b) with $k = 1$, getc_1 leaves the value m on track 9 and the archive within R' rows, within $55000(R'+1)^2$ units.

Each of the thirteen moves of mv is a clean copy (Lemma I.66(a)) or an archive read (Lemma I.68(c), with an index at most 60), and costs at most $64\cdot 20000(R'+1)^2$ units; with the final skip, mv costs at most $17920000(R'+1)^2$. A read changes only tracks 12, 13 and its target. Following the moves, with Lemma I.95 for the entries: track 8 receives m; tracks 5, 6, 7, 9 and 10 receive $v_{22} = \sigma$, $v_{24} = S$, $v_{37} = \kappa$, $v_{45} = q_P$ and $v_{52} = l_P$; track 11 receives $v_9 = n_0$, which is copied to track 14; track 11 then receives $v_{60} = c_P$, which is copied to track 13; finally m returns to track 11, n_0 to track 12, and track 8 is blanked. So tracks 5, 6, 7, 9, 10, 11, 12 and 13 hold the values σ, S, κ, q_P, l_P, m, n_0 and c_P, and track 8 holds 0. Each is at most R', since the entries of the archive are at most R.

In troom, $\mathsf{cc}_{12\to8}$ saves n_0 on track 8. Lemma I.91(d) with $c = 14$, applied twice, grows the workspace to $Q(Q(R'))$ rows and keeps the counters on tracks 5 to 11 and 13; it blanks tracks 12 and 14. The new rows carry 1 on track 11, so the value m there has become a counter; $\mathsf{cc}_{8\to12}$, $\mathsf{cc}_{11\to14}$ and $\mathsf{cc}_{14\to11}$ make tracks 12 and 11 hold the values n_0 and m again, and $\mathsf{clear}_{8,14}$ blanks tracks 8 and 14. The two growths cost at most $3\cdot 10^9\big((R'+2)^5 + (Q(R')+2)^5\big)$ units and the copies and clearings at most $50000(Q(Q(R')) + R' + 2)$. Since $Q(R') + 2 \le 67(R'+2)^2$ and $Q(Q(R')) \le 66\cdot 67^2(R'+2)^4$, troom and the row count $Q(Q(R'))+1$ together are at most $10^{22}(R'+2)^{10}$.

The tape is now a seed with $r' = Q(Q(R'))$ rows, except for the inequality $\rho(m) \le r'$. Put $x = R' + 1$, an upper bound for each of σ, S, κ, m, n_0, q_P, l_P and c_P. Then $N \le 2x$, $q_P N \le 2x^2$, $q_P N + l_P \le 3x^2$, $(q_P N + l_P)N \le 6x^3$, $\mathscr{B}(N) \le 7x^3$ and $W_m \le 28x^3 + 4$, so $\rho(m) \le 70x^3 \le 128x^3$. And $Q(y) \ge 64(y+1)^2$ for every y, so $Q(Q(R')) \ge 64(64(R'+1)^2)^2 \ge 128(R'+1)^3$. Hence $\rho(m) \le r'$. No part writes tracks 3, 4 or 15 in the old rows, and the new rows carry 0 there, so the saved request is kept. Adding the three bounds gives $t + r' + 1 \le 2\cdot 10^{22}(R'+2)^{10}$. □

Proof of Proposition I.86. Let τ carry the saved request and hold the archive $(p,\gamma,\alpha,\beta,s,L,s_0)$ within r rows.

(i) By Lemma I.90(d), Ctl runs in t_1 units to a tape holding the archive $(v_0,\dots,v_9)$, that is

$$(p,\gamma,\alpha,\beta,s,L,s_0,M,T,n_0),$$

within r_1 rows, with $t_1 + r_1 + 2 \le K_C(r+2)^8$.

(ii) By Lemma I.98, Prog runs in t_2 units to a tape holding $(v_0, \dots, v_{60})$ within r_2 rows, with $t_2 + r_2 + 2 \le K_P(r_1+2)^{6^{51}}$.

(iii) By Lemma I.99, Seed runs in t_3 units to a seed with r' rows, with $t_3 + r' + 2 \le F(r_2 + m + 2)^{10}$ and $F = 2 \cdot 10^{22} + 1$.

Lemma I.73 with the exponents 8 and 6^{51} combines (i) and (ii): $t_1 + t_2 + r_2 + 2 \le B(r+2)^A$ with $A = 8 \cdot 6^{51}$ and $B = K_C + K_P K_C^{6^{51}}$. Since $m \le (r+m+2)^A$ and $r + 2 \le r + m + 2$,

$$t_1 + t_2 + (r_2 + m) + 2 \le (B+1)\,(r+m+2)^A.$$

Lemma I.73 with $x = r + m$, $y = r_2 + m$, $z = r'$ and the exponents A and 10 combines this with (iii):

$$t_1 + t_2 + t_3 + r' + 2 \le \big((B+1) + F(B+1)^{10}\big)(r+m+2)^{80 \cdot 6^{51}}.$$

So $K_T = (B+1) + F(B+1)^{10}$, which depends only on p, the problem and n_0. The three programs run from the root to the root. □

I.9.6 The order, the capacity and the width

From a seed, the last program forms $\max(m, n_0)$ on track 8 with $\mathsf{max}^{14}_{11,12\to 8}$ (Lemma I.79) and then runs a straight-line code of eleven register instructions. Every instruction is a foundation routine or the product of Lemma I.92 acting on counters of the tracks 5 to 13, with tracks 2 and 14 as scratch:

$$\begin{aligned}\mathsf{Code} = {} & \mathsf{acc}^2_{7\to 8}; \mathsf{umul}^{2,14}_{9,8\to 11}; \mathsf{acc}^2_{10\to 11}; \mathsf{umul}^{2,14}_{11,8\to 12}; \mathsf{acc}^2_{13\to 12}; \mathsf{copy}_{12\to 10};\\ & \mathsf{copy}_{12\to 9}; \mathsf{acc}^2_{12\to 9}; \mathsf{acc}^2_{12\to 9}; \mathsf{acc}^2_{12\to 9}; \mathsf{inc}^4_9; \mathsf{skip}.\end{aligned}$$

With $n = \max(m, n_0)$ on track 8 it computes $N = n + \kappa$ on track 8, then $q_P N$, $q_P N + l_P$ on track 11, $(q_P N + l_P)N$ and $\mathscr{B}(N) = (q_P N + l_P)N + c_P$ on track 12 (Horner's rule), copies $\mathscr{B}(N)$ to tracks 10 and 9, and adds it three more times and 4 to track 9, which then carries $4\mathscr{B}(N) + 4 = W_m$.

Lemma I.100 (Register instructions). *Let τ have origin marks up to r, let $\rho \le r$, and let k, k' and l be distinct tracks among $5, \dots, 13$. Each of the following runs from the root to the root within $10^5(\rho+1)^3$ units and changes only the tracks named, in rows $j \le \rho$.*

(a) If track k carries $x \le \rho$, $\mathsf{copy}_{k\to l}$ makes track l carry x; only track l changes.

(b) If tracks k and l carry b and a with $a + b \le \rho$, $\mathsf{acc}^2_{k\to l}$ makes track l carry $a + b$; only tracks l and 2 change.

(c) If tracks k and k' carry a and b with $a, b, ab \le \rho$, $\mathsf{umul}^{2,14}_{k,k'\to l}$ makes track l carry ab; only tracks l, 2 and 14 change.

(d) If track l carries a, $c \le 4$ and $a + c \le \rho$, inc^c_l makes track l carry $a + c$; only track l changes.

Proof. (a) and (b) are Lemma I.15, within $100(x+1)$ and $500(a+b+1)^2$ units. (c) is Lemma I.92, within $2000(a+b+ab+1)^3 \le 54000(\rho+1)^3$ units. (d) is Lemma I.14(b), within $18c^2 + (36a+2l+26)c + 1 \le 144\rho + 497$ units. The cells written are those of the counters involved, in rows at most the largest number formed, which is at most ρ. □

Proof of Lemma I.87. Let τ be a seed with r rows, write $\rho = \rho(m) \le r$ and $n = \max(m, n_0)$. By Lemma I.63, corr costs at most $37r + 13$ units and writes the origin marks up to r again, changing only track 1. Tracks 11 and 12 hold m and n_0, and $n \le \rho$, so by Lemma I.79 $\mathsf{max}^{14}_{11,12\to 8}$ makes track 8 carry n within $400(m+n_0+1) \le 10^5(\rho+1)^3$ units, changing only tracks 8 and 14 in rows $j \le n$. Then tracks 5 to 13 carry

$$\sigma,\ S,\ \kappa,\ n,\ q_P,\ l_P,\ m,\ n_0,\ c_P.$$

The eleven instructions of Code change them, in order, as follows: track 8 becomes $n+\kappa = N$; track 11 becomes $q_P N$ and then $q_P N + l_P$; track 12 becomes $(q_P N + l_P)N$ and then $(q_P N + l_P)N + c_P = \mathscr{B}(N)$; tracks 10 and 9 become $\mathscr{B}(N)$; track 9 becomes $2\mathscr{B}(N)$, $3\mathscr{B}(N)$, $4\mathscr{B}(N)$ and $4\mathscr{B}(N) + 4 = W_m$. Every number formed, and every sum and product formed on the way, is one of the summands of ρ or at most W_m, so it is at most ρ, and Lemma I.100 applies to each instruction. Each instruction changes only its destination and the scratch tracks 2 and 14, so the other registers keep their counters. The code costs at most $11 \cdot 10^5(\rho+1)^3 + 1$ units. In all,

$$t \le 37r + 13 + 10^5(\rho+1)^3 + 11 \cdot 10^5(\rho+1)^3 + 1 \le 14 \cdot 10^5(\rho+1)^3 + 37r + 13.$$

Tracks 5 to 10 now carry σ, S, κ, N, W_m and $\mathscr{B}(N)$, and $\sigma + S + \kappa + N + W_m + \mathscr{B}(N) \le \rho \le r$. Tracks 3, 4 and 15 are not written, so the saved request is kept; every write is to a track $k \ge 1$ in a row $j \le r$, so the tape is still a workspace of r rows, and track 1 is not written after corr. This is a scheduled request with r rows.

For the bound on ρ: $N \ge \kappa \ge 2$, so $q_P N$, $q_P N + l_P$ and $(q_P N + l_P)N$ are at most $\mathscr{B}(N)$, and

$$\rho(m) \le (F+10) + m + N + 4\mathscr{B}(N) + W_m, \qquad F = \sigma + S + \kappa + n_0 + q_P + l_P + c_P.$$

Now $m \le X$ and $N \le X$, so N^2, N, 1 are at most X^2 and $\mathscr{B}(N) \le (q_P + l_P + c_P)X^2$; and $W_m \le (4(q_P + l_P + c_P) + 4)X^2$ by Proposition I.32(e). With $X \ge 1$ this gives $\rho(m) \le c_\rho X^2$ for $c_\rho = F + 16 + 8(q_P + l_P + c_P)$, which depends only on the problem and n_0. □

I.9.7 The whole stage

The stage is assembled from four bounds: Prepare in the rows, the front half and the scalar tail in the rows plus the precision, and Width in the tail room. Only the exponents are tracked; every constant that appears depends on the problem and n_0 alone.

Proof of Theorem I.88. Let τ be a decoded request with r rows, and put $Y = X + r + 1$, so that $Y \ge 1$, $m + 1 \le X$ and $r + 2 \le Y$.

(i) By Lemma I.61, Prepare runs in $t_0 \le 4000(r+2)^2 \le 4000Y^2$ units to a prepared request with $r+1$ rows.

(ii) By Theorem I.62(c), MetaBank; Summ then runs in t_1 units to a tape that carries the saved request and holds the archive $(p, \gamma, \alpha, \beta, s, L, s_0)$ within r_1 rows, with $t_1 + (r_1 + m) + 2 \le E((r+1) + m + 2)^{72}$ and $E = K_1 + K_2 K_1^3$.

(iii) By Proposition I.86, for the payload row p, which depends only on the problem, Scalar then runs in t_2 units to a seed with r_2 rows, with $t_2 + r_2 + 2 \le K_T(r_1 + m + 2)^b$ and $b = 80 \cdot 6^{51}$.

(iv) By Lemma I.87, Width then runs in $t_3 \le 14 \cdot 10^5(\rho(m) + 1)^3 + 37 r_2 + 13$ units to a scheduled request with $r' = r_2$ rows, and $\rho(m) \le r_2$.

Lemma I.73 with $x = r + 1 + m$, $y = r_1 + m$, $z = r_2$, $a = 72$ and the exponent b combines (ii) and (iii):

$$t_1 + t_2 + r_2 + 2 \le K_P(r + m + 3)^{g_P}, \qquad K_P = E + K_T E^b, \quad g_P = 72 \cdot 80 \cdot 6^{51}.$$

Since $r + m + 3 \le X + r + 2 \le 2Y$, this gives $t_1 + t_2 + r_2 + 1 \le K_P 2^{g_P} Y^{g_P}$, and with (i)

$$t_0 + t_1 + t_2 + r_2 + 1 \le K_S Y^{g_P + 2} \le K_S Y^{g_P + 3}, \qquad K_S = K_P 2^{g_P} + 16000.$$

For the last part, $\rho(m) + 1 \le Y + r_2 + 1$ and $38 r_2 + 14 \le 100(Y + r_2 + 1)^3$, so

$$t_3 + r_2 + 1 \le 1400100\,(Y + r_2 + 1)^3.$$

Lemma 15.36(a) with $N = Y$, $s = t_0 + t_1 + t_2$, $r = r' = r_2$, $A = K_S$, $e = g_P + 3$, $K = 1400100$ and $g = 3$ gives

$$(t_0 + t_1 + t_2 + t_3) + r' + 1 \le \left(K_S + 1400100(K_S + 1)^3\right) Y^{4(g_P + 3)}.$$

This is the claim with $K_q = K_S + 1400100(K_S + 1)^3$ and $g_q = 4(72 \cdot 80 \cdot 6^{51} + 3)$. The four programs run from the root to the root, so their sequence does. □

The exponent g_q enters the exponent λ of Theorem 15.33 as the exponent of the stage that follows the decoding (Section I.16). It is large because the composition rule multiplies exponents and the schedule is a long fixed sequence of polynomial steps; it is fixed, because the program Schedule and every count in this section are the same for every problem.

I.10 The first vector

How does the machine turn the initial point $y_0 = (\frac{1}{3}, -2)$ into its first stored vector? For illustration let $N = 3$. Lemma 15.35 asks for $Z_0 = \mathrm{near}(2^3 y_0)$, coordinate by coordinate, and $\mathrm{near}(u/v) = \lfloor (2u + v)/(2v) \rfloor$ for $v > 0$. For the first coordinate the request supplies the numerator 1 and the denominator 3. The machine forms the dividend $2 \cdot 1 \cdot 2^3 + 3 = 19$ and the divisor $2 \cdot 3 = 6$ and divides once: $\lfloor 19/6 \rfloor = 3 = \mathrm{near}(8/3)$. For the second coordinate, $-2 = -2/1$, the dividend is $2 \cdot (-2) \cdot 2^3 + 1 = -31$ and the divisor 2. The machine divides the magnitude, $31 = 2 \cdot 15 + 1$, and because the dividend is negative and the remainder is not zero it corrects the quotient to $-16 = \lfloor -31/2 \rfloor = \mathrm{near}(-16)$. It then writes the signed words $\mathrm{sw}_W(3)\,\mathrm{sw}_W(-16)$ on the binary stream, $2(W + 2)$ bits. Everything else in this section

is the bookkeeping around these divisions: the machine stores its controls, reserves the rows that every later stage will use, finds the frames of y_0 in the saved request, and after the last coordinate puts back the controls that the steps read.

This section constructs that stage, proves its contract and proves its step bound, with exponent 190674. The stored vector it leaves is the state between Taylor steps, defined first; the step loop that starts from it, the selection of the answer coordinate and the exponent of the whole trajectory are in Section I.13.

I.10.1 The stored vector and the stage

Throughout this section and Section I.13 a problem with a coefficient field of dimension d is fixed, with rational initial point y_0, radius $M \geq 1$, natural end time T, a level n_0 at which the region certificate holds, and the constants σ, S and κ of Definition 15.18. Write $y_{0,j} = u_j/e_j$ with the integer u_j and the positive integer e_j that the request carries. For a request with coordinate $i < d$ and precision m put $n = \max(m, n_0)$, $N = N(n)$, $W = W_m$ and $X = m + n_0 + \kappa + 1$, and let $Z_0, \dots, Z_S$ be the integer vectors of Lemma 15.35 at the level n, so that $Z_0 = \mathrm{near}(2^N y_0)$. Let c_{req} be the constant of Theorem I.42(c) and put

$$\rho_{\mathrm{traj}} = S + 100(d+1)(N+1)(W+2)^2 + (c_{\mathrm{req}} + 2)X + 1.$$

Definition I.101 (The stored vector)**.** Let $Z = (Z_0, \dots, Z_{d-1})$ be a list of d integers and $r \in \mathbb{N}$. A tape τ *holds the stored vector Z on r rows* if

(a) τ carries the saved request (Definition I.41);

(b) tracks 5, 8, 9 and 10 carry the counters m, N, W and σ;

(c) the binary stream holds the word $\mathrm{sw}_W(Z) = \mathrm{sw}_W(Z_0) \cdots \mathrm{sw}_W(Z_{d-1})$ of $d(W+2)$ bits;

(d) τ is a workspace of r rows with origin marks up to r;

(e) $r \geq \rho_{\mathrm{traj}}$ (the *reserve*).

In (c), $\mathrm{sw}_W(z)$ is the bit $[z < 0]$ followed by the $W+1$ lowest binary digits of $|z|$, as in Definition I.3(c); for every entry that fits W this is its signed word, and every vector the machine stores fits W (Proposition I.32(a)). The definition says nothing about tracks 2, 11 to 14, or about the cells of tracks 3 and 4 after the end of the saved request: the step loop keeps its counter on track 2 and the step count S on track 14, the coefficient stage uses the output buffer, and this stage leaves its archive of controls after the saved request. The reserve exceeds S, $d(W+2)$ and $100(d+1)(N+1)(W+2)^2$; the later stages use it for the step counter, the stored vector and the coefficient table.

The stage is

$$\mathsf{Init} = \mathsf{Arch};\mathsf{Res};\mathsf{Pos};\mathsf{Pack};\mathsf{Rest}, \qquad \mathsf{Pack} = \mathsf{Res};\mathsf{Pos};\mathsf{Begin};\mathsf{Coords},$$
$$\mathsf{Coords} = \mathsf{left}^{15};\mathsf{loop}(\mathsf{home}_{14};\mathsf{Coord};\mathsf{left}^{15});\mathsf{home}_{14},$$
$$\mathsf{Coord} = [\mathsf{Prep},\ \mathsf{Load},\ \mathsf{Copy},\ \mathsf{Scale},\ \mathsf{Add},\ \mathsf{Div},\ \mathsf{Emit},\ \mathsf{pred}_{14}],$$

with pred_{14} of Lemma I.127. Its parts do the following.

(1) Arch appends the five numbers S, σ, N, W and m to the saved request as an archive (Definition I.60), reading m from the last frame of the request (Lemma I.103).

(2) Res adds rows; afterwards there are at least $512(r+1)^4 \geq \rho_{\mathrm{traj}}$ of them (Lemma I.104).

(3) Pos finds the first payload frame and loads N and W from the archive (Lemma I.105). The reservation and the positioning run twice, once in Init and once at the start of Pack; the second pair acts on more rows and leaves the tape in the same form.

(4) Begin empties the binary stream, and Coords runs Coord once for each of the d marks of a counter on track 14. Each run reads the frames of u_j and e_j, computes $\mathrm{near}(2^N u_j/e_j)$ with one division, appends its signed word to the binary stream and removes one mark (Lemmas I.115 and I.116).

(5) Rest loads S, σ, N, W and m back from the archive onto the tracks the loop reads, puts S on track 14 and sets the step counter on track 2 to 0 (Lemma I.106).

Theorem I.102 (The first vector)**.** *There is a constant K_{i}, depending only on the problem and n_0, with the following property. Let $r \in \mathbb{N}$, $i < d$ and m, and let τ be a scheduled request with r rows (Definition I.85). Then* Init *runs from $(-1,\tau)$ to $(-1,\tau')$ in some number t of units, where for some r'*

(a) τ' holds the stored vector $Z_0 = \mathrm{near}(2^N y_0)$ on r' rows;

(b) track 2 of τ' carries the counter 0, and track 14 carries a retained word of length S from row 0;

(c) $t + r' + 1 \leq K_{\mathrm{i}}(X + r + 1)^{190674}$.

The exponent is built from the exponents of the parts by Lemma 15.36(a):

$$190674 = 66\big(8\big(90(2+2)+1\big)+1\big).$$

One coordinate costs at most $250000(r+1)^2$ units (exponent 2), so the loop over the d coordinates has exponent 3; the reservation contributes the exponent 10 and the positioning 8, which makes 90 for the first three parts of Pack and $90 \cdot 4 = 360$ for Pack; positioning in front of it gives $8 \cdot 361 = 2888$; the archive, of exponent 2, followed by the reservation gives $2 \cdot 11 = 22$, then $22 \cdot 2889$; and the restoration, of exponent 2, gives the factor 3. The constants are recorded in the proof, which closes the section.

I.10.2 Routines from the schedule

Two kinds of routine of this stage were already needed by the schedule. The precision m is the last frame of the request and the coordinate i the one before it; the machine finds such a frame by walking back from the end of the saved request over the frame-start marks of track 15, and reads it as Section I.7 reads every frame. With the routines send and pf^k of Lemma I.89,

$$\mathsf{cur}_k = \mathsf{send};\mathsf{pf}^k;\mathsf{tc}, \qquad \mathsf{getc}_k = \mathsf{cur}_k;\mathsf{clear}_{8,9};\mathsf{cc}_{2\to 5};\mathsf{fetch};\mathsf{reset}_8;\mathsf{toUnary}^+$$

Part	Bound	Reference
Arch	$t + r' + 1 \le 10^{10}(X + r + 1)^2$	Lemma I.103
Res	$t + r' + 1 \le 10^{22}(r + 2)^{10}$	Lemma I.104
Pos	$t + r' + 1 \le C_{\mathrm{pos}}(X + r + 1)^8$	Lemma I.105
Begin	48	Lemma I.116
Coord	$250000(r + 1)^2$, rows fixed	Lemma I.115
Coords	$t + r + 1 \le (250080(d + 1) + 2)(r + 1)^3$	Lemma I.116
Pack	$t + r' + 1 \le K_{\mathrm{pk}}(X + r + 1)^{360}$	Lemma I.117
Pos; Pack	$t + r' + 1 \le K_{\mathrm{pp}}(X + r + 1)^{2888}$	Lemma I.117
Rest	$t + r + 1 \le 2 \cdot 10^6(r + 1)^2$, rows fixed	Lemma I.106
Init	$t + r' + 1 \le K_{\mathrm{i}}(X + r + 1)^{190674}$	Theorem I.102

Table I.3. The bounds of the first-vector stage; r is the number of rows at the start of the part and r' at its end.

are the programs of Lemma I.90: cur_k puts on track 2 the first row of the k-th frame from the end, and getc_k reads that frame into unary on track 9. That lemma states getc_k for $k \in \{1, 3, 4, 5\}$ and cur_k for every $k \le 5$. The selection of Section I.13 uses cur_2: by part (a) of that lemma with $k = 2$, if τ carries the saved request and is a workspace of R rows with origin marks up to R, then cur_2 runs from the root to the root within $106(R + 1)$ units, makes track 2 carry the first row s_{F-1} of the frame of i, and changes only tracks 2 and 14, in rows $j \le R$. The reserve is made by the program $\mathsf{room} = \mathsf{clear}_{5,\dots,11}; \mathsf{quad}_{13}$ of Lemma I.91, which takes a tape that holds an archive within R rows to one that holds the same archive within $Q(R) = R + 1 + 64(R + 1)^2$ rows, within $4 \cdot 10^9(R + 2)^5$ units, writing tracks 3, 4 and 15 only in the new rows, where it leaves 0.

I.10.3 The controls, the reserve and the frames of y_0

The archive of Definition I.60(b) is built on the saved word $\omega = w_u$: a tape *holds the archive* $(v_1, \dots, v_k)$ *within R rows* if w_u lies on tracks 3 and 4 from row 0, the unary words $1^{v_1}0 \cdots 1^{v_k}0$ follow it from row $|w_u| + 1$, and $|w_u| + 1 + \sum_j (v_j + 1) \le R$. A program that writes no cell of tracks 3, 4 and 15 keeps the saved request and every archive. For a track k put $\mathsf{psh}_k = \mathsf{cc}_{k \to 11}; \mathsf{push}$, with push of Lemma I.68, and

$$\mathsf{Arch} = \mathsf{cc}_{5\to 10}; \mathsf{init}; \mathsf{psh}_6; \mathsf{psh}_{10}; \mathsf{psh}_8; \mathsf{psh}_9; \mathsf{grow}; \mathsf{getc}_1; \mathsf{psh}_9.$$

Lemma I.103 (The archive of the controls)**.** *Let τ be a scheduled request with r rows. Then* Arch *runs from $(-1, \tau)$ to $(-1, \tau_{\mathrm{a}})$ in some t units, where τ_{a} carries the saved request, holds the archive (S, σ, N, W, m) within r_{a} rows, is a workspace of r_{a} rows with origin marks up to r_{a}, $r \le r_{\mathrm{a}}$, and*

$$t + r_{\mathrm{a}} + 1 \le 10^{10}(X + r + 1)^2.$$

Proof. By Definition I.85, τ carries the saved request, is a workspace of r rows with origin marks up to r, and carries the counters σ, S, κ, N, W and $\mathscr{B}(N)$ on tracks 5 to 10, whose sum is at most r. Put $Y = X + r + 1$. The parts run from the root to the root, and each writes no cell of tracks 3, 4 and 15 except in the rows of the archive it extends.

(1) $\mathsf{cc}_{5\to 10}$ (at most $300(r+1)$ units, Lemma I.66) makes track 10 hold σ.

(2) init (Lemma I.68(e)) leaves the empty archive within $R_0 = r+2$ rows, keeps the counters S, N, W and σ on tracks 6, 8, 9 and 10, and takes at most $3000(r+3)$ units.

(3) Each psh_k copies the counter n of track k onto track 11 (at most $300(R+1)$ units) and appends it (Lemma I.68(h)): from an archive within R rows to the extended archive within $R{+}1{+}n$ rows, within $10^6(R{+}n{+}2)^2$ units. Its parts start and end write only tracks 2 and 13, and append keeps every counter at most R on tracks 8 to 11 (parts (a), (g) and (f) of that lemma); so the counters still to be archived survive. After psh_6, psh_{10}, psh_8 and psh_9 the tape holds (S,σ,N,W) within $R_4 = R_0+(S+1)+(\sigma+1)+(N+1)+(W+1)$ rows.

(4) grow ($72R_4+99$ units) adds a row, which keeps the archive, and getc_1 reads the last frame, that of m (Lemma I.90(b) with $k=1$): track 9 holds m, the tape carries the saved request and holds the archive within $R_5 = R_4+1+m$ rows, and the two parts take at most $72R_4+99+55000(R_4+m+2)^2$ units.

(5) The last psh_9 appends m: the archive (S,σ,N,W,m) within $r_{\mathrm{a}} = R_5+1+m$ rows.

Since $S+\sigma+N+W \le r$ and $r+m+1 \le Y-1$, every row count R above satisfies $R+m+2 \le r_{\mathrm{a}}+2 \le 2r+2m+10 \le 8Y$. So the five pushes cost at most $5(2400Y+64\cdot 10^6Y^2)$, the read $56000\cdot 64Y^2$, and the copies and init at most $9300Y$; with $r_{\mathrm{a}}+1 \le 8Y$ the total is below $10^{10}Y^2$. □

Lemma I.104 (The reserve). *Let* $\mathsf{Res} = \mathsf{room};\mathsf{room}$. *Let τ carry the saved request, hold the archive (S,σ,N,W,m) within R rows and be a workspace of R rows with origin marks up to R. Then* Res *runs from $(-1,\tau)$ to $(-1,\tau')$ in some t units, where τ' has the same properties for $R' = Q(Q(R))$ rows,*

$$\rho_{\mathrm{traj}} \le 512(R+1)^4 \le R', \qquad t+R'+1 \le 10^{22}(R+2)^{10}.$$

Proof. By Lemma I.91(e) twice, Res ends on a workspace of $Q(Q(R))$ rows with origin marks up to it and keeps the saved request and the archive: tracks 3, 4 and 15 are written only in new rows, where they get 0. Since $Q(R)+2 \le 66(R+2)^2$,

$$t \le 4\cdot 10^9(R+2)^5 + 4\cdot 10^9\,66^5(R+2)^{10}, \qquad R'+1 \le 66(Q(R)+2)^2 \le 66^3(R+2)^4,$$

and the sum is below $10^{22}(R+2)^{10}$. For the reserve, let $r_1 = R+1$. The entries of the archive are at most R and $|w_u|+1 \le R$. The payload holds the $2d$ nonempty frames of the u_j and e_j, so $d \le |w_u|$. The request without its last two frames has at most $3|w_u|$ cells, because the frame of a word b has $2|b|+1 \le 3|b|$ cells; so $c_{\mathrm{req}} \le 5|w_u|+7$ and $c_{\mathrm{req}}+2 \le 5r_1$. Finally $X \le 2N+1 \le 2r_1$, since $m, n_0 \le n$ and $N = n+\kappa$. Hence

$$\rho_{\mathrm{traj}} \le r_1 + 100\,r_1\cdot r_1\cdot(2r_1)^2 + 10r_1^2 \le 411\,r_1^4 \le 512\,r_1^4.$$

And $Q(R)+1 \ge 64r_1^2$, so $Q(Q(R)) \ge 64(Q(R)+1)^2 \ge 64\cdot 4096\,r_1^4$. □

The positioning uses the payload address paddr of Lemma I.71 and the loads of Lemma I.68(c):

$$\mathsf{Pos} = \mathsf{paddr}; \mathsf{load}_{2,8}; \mathsf{load}_{3,9}.$$

Let c_{num} be the constant with $V(u) \le c_{\mathrm{num}} X$ of Section I.7, and put $A_{\mathrm{p}} = 4 \cdot 10^{12}(c_{\mathrm{num}}+1)^4$ and $C_{\mathrm{pos}} = A_{\mathrm{p}} + 220000 A_{\mathrm{p}}^2$.

Lemma I.105 (The first frame of y_0)**.** *Let τ carry the saved request, hold the archive (S, σ, N, W, m) within R rows and be a workspace of R rows with origin marks up to R. Then Pos runs from $(-1, \tau)$ to $(-1, \tau')$ in some t units, where for some $R' \ge R$*

(a) τ' carries the saved request, holds the same archive within R' rows, and is a workspace of R' rows with origin marks up to R';

(b) track 5 carries the row of the frame of u_0, tracks 8 and 9 carry the counters N and W, and track 14 carries the counter d;

(c) $t + R' + 1 \le C_{\mathrm{pos}}(X + R + 1)^8$.

Proof. Put $Y = X+R+1$. By Lemma I.71(a), paddr runs in some t_1 units to a tape within R' rows that keeps the saved request, carries on track 5 the row p of the first payload frame, which is the frame of u_0, and holds the value d on track 14; and $t_1 + R' + 1 \le 4 \cdot 10^{12}(R+V(u)+1)^4 \le A_{\mathrm{p}} Y^4$, since $R + V(u) + 1 \le (c_{\mathrm{num}}+1)Y$ (the number D of that lemma is $V(u)$). By the same lemma $R \le R'$, and no cell of tracks 3, 4 and 15 changes, so the archive is kept. In the archive (S, σ, N, W, m), numbered from 0, N is entry 2 and W entry 3. By Lemma I.68(c), $\mathsf{load}_{2,8}$ and $\mathsf{load}_{3,9}$ make tracks 8 and 9 hold N and W, within $(5+6) \cdot 20000(R'+1)^2 \le 220000 A_{\mathrm{p}}^2 Y^8$ units, and change only tracks 8, 9, 12 and 13. So tracks 5 and 14 keep p and d, and $t + R' + 1 \le A_{\mathrm{p}} Y^4 + 220000 A_{\mathrm{p}}^2 Y^8 \le C_{\mathrm{pos}} Y^8$. □

After the last coordinate the controls go back where the loop reads them:

$$\mathsf{Rest} = \left[\mathsf{load}_{0,11}, \mathsf{load}_{1,10}, \mathsf{load}_{2,8}, \mathsf{load}_{3,9}, \mathsf{load}_{4,5}\right]; \mathsf{cc}_{11\to 14}; \mathsf{zero}_2.$$

Lemma I.106 (Restoring the controls)**.** *Let τ carry the saved request, hold the archive (S, σ, N, W, m) within r rows, be a workspace of r rows with origin marks up to r, let $r \ge \rho_{\mathrm{traj}}$, and let the binary stream hold $\mathrm{sw}_W(Z)$ for an integer vector Z of length d. Then Rest runs from $(-1, \tau)$ to $(-1, \tau')$ in some t units, where τ' holds the stored vector Z on the same r rows, track 2 carries the counter 0, track 14 carries a retained word of length S from row 0, and $t + r + 1 \le 2 \cdot 10^6 (r+1)^2$.*

Proof. The five loads make tracks 11, 10, 8, 9 and 5 hold S, σ, N, W and m; each changes only tracks 12, 13 and its destination (Lemma I.68(c)), so the binary stream and the earlier destinations survive, and they take at most $(3+4+5+6+7) \cdot 20000(r+1)^2 + 1$ units. $\mathsf{cc}_{11\to 14}$ (at most $300(r+1)$) makes track 14 hold the value S, which is a retained word of length S from row 0, and zero_2 (11 units) sets $(0, 2)$ to 0. Tracks 3, 4, 15, 0 and 1 are not written, so the saved request, the workspace and the origin marks are kept, and the reserve holds by hypothesis. So τ' satisfies (a) to (e) of Definition I.101, and $t + r + 1 \le 500000(r+1)^2 + 300(r+1) + 12 + (r+1) \le 2 \cdot 10^6 (r+1)^2$. □

I.10.4 One coordinate

The loop over the coordinates keeps the following state between two runs of Coord. Write $Z_0 = (Z_{0,0}, \dots, Z_{0,d-1})$, so that $Z_{0,j} = \mathrm{near}(2^N u_j / e_j)$. In the payload the frame of u_j is followed by that of e_j, and the frame of e_{d-1} by the frame of the first coefficient of the field.

Definition I.107 (The prefix state)**.** Let $k \le d$ and $r \in \mathbb{N}$. A tape τ is in the *prefix state k with r rows* if

(a) τ carries the saved request, holds the archive (S, σ, N, W, m) within r rows, is a workspace of r rows with origin marks up to r, and $r \ge \rho_{\mathrm{traj}}$;

(b) track 5 carries the row of the frame of u_k (for $k = d$, of the frame after that of e_{d-1});

(c) tracks 8 and 9 carry the counters N and W, and track 14 carries the counter $d - k$;

(d) the binary stream holds $\mathrm{sw}_W(Z_{0,0}) \cdots \mathrm{sw}_W(Z_{0,k-1})$, and track 11 carries its length $k(W+2)$.

For the rest of this subsection fix $k < d$ and a tape in the prefix state k with r rows, and write $u = u_k$, $e = e_k$, $z = Z_{0,k}$,

$$a = \lfloor r/2 \rfloor, \qquad \rho = 400(W+2)^2, \qquad D = 2u2^N + e.$$

Three facts are used throughout. First, $2\rho \le r$, so that $\rho \le a$ and $a + \rho \le r$: since $d \ge 1$ and $N = n + \kappa \ge 3$ (the definition of κ gives $\kappa \ge 3$), $(d+1)(N+1) \ge 8$ and $r \ge \rho_{\mathrm{traj}} \ge 800(W+2)^2$. Second, $y_{0,k}$ has width s_0 (Section H.1), so $|u| \le 2^{s_0}$ and $e \le 2^{s_0}$, and $s_0 + N + 3 \le \mathscr{B}(N)$ by (H.1), so $s_0 + N + 2 \le W$. Hence

$$|D| \le 2^{s_0+N+1} + 2^{s_0} < 2^W, \qquad 0 < 2e \le 2^W, \qquad 2^{N+1}|u| < 2^W,$$

and $|u|$ and e have at most W binary digits. Third, $z = \mathrm{near}(2^N u/e) = \lfloor D/(2e) \rfloor$.

The arithmetic area. The arithmetic of one coordinate runs on track 13 around the row a. The *view at a* of a tape τ is the virtual tape $\sigma_\tau(j) = \tau(\mathrm{cell}(a, 13) + 16j)$, $j \in \mathbb{Z}$; for $|j| \le a$ its cell j is the cell $(a - j, 13)$. Twelve-cell rows at the virtual position 1 occupy the cells $(a - 1 - 12t - c, 13)$: column c of row t. The virtual cell 0, that is $(a, 13)$, holds a sign. Put

$$\mathsf{call}(P) = \mathsf{seek}_{12}; \mathsf{left}^1; P^{\uparrow 16}; \mathsf{home}_{13}.$$

Lemma I.108 (Running on the arithmetic area)**.** *Let τ satisfy (a), (c) and (d) of Definition I.107 with r rows, and let track 12 carry a. Let a program P run on σ_τ from the position 0 in t units to a position f with $0 \le f \le \rho$ and a virtual tape σ' that agrees with σ_τ outside $[-\rho, \rho]$. Then* call(P) *runs from $(-1, \tau)$ to $(-1, \tau')$ within $17t + 36r + 40$ units, where $\tau'(\mathrm{cell}(a, 13) + 16j) = \sigma'(j)$ for every j, τ' agrees with τ off track 13, and τ' again satisfies (a), (c) and (d).*

Proof. seek_{12} ($18a + 15$ units) reaches $(a, 12)$ and left^1 (2) the cell $o = \mathrm{cell}(a, 13)$. The configuration carries $(0, \sigma_\tau)$ for the stride 16 and the origin o, so by Lemma I.8 $P^{\uparrow 16}$ runs within $17t$

units to the cell $o + 16f = \mathrm{cell}(a - f, 13)$ and a tape that carries σ' and agrees with τ off the cells $o + 16j$. The changed cells are $(a - j, 13)$ with $|j| \le \rho$, which lie in the rows $a - \rho, \dots, a + \rho$ of $[0, r]$, on track 13; so the workspace and the origin marks are kept, and the other conditions do not involve track 13. home_{13} from row $a - f$ takes $18(a - f) + 17$ units. In all at most $17t + 36a + 34$. □

The centre and the marks. Put $\mathsf{flag}_b = \mathsf{left}^{11}; \mathsf{write}\, b; \mathsf{home}_{10}$, $\mathsf{up} = \mathsf{left}^{1}; \mathsf{home}_{1}$, $\mathsf{wbit}_b = \mathsf{seek}_{12}; \mathsf{left}^{1}; \mathsf{write}\, b; \mathsf{home}_{13}$, and

$$\begin{aligned}
\mathsf{half} &= \mathsf{zero}_2; \mathsf{zero}_{12}; \mathsf{flag}_0; \mathsf{peek}_{2\to 0}; \mathsf{loop}(\mathsf{up}; \mathsf{par}; \mathsf{inc}_2; \mathsf{peek}_{2\to 0}); \mathsf{up},\\
\mathsf{par} &= \mathsf{left}^{11}; \mathsf{branch}(\mathsf{write}\,1; \mathsf{home}_{10}; \mathsf{skip},\ \mathsf{write}\,0; \mathsf{home}_{10}; \mathsf{inc}_{12}),\\
\mathsf{marks} &= \mathsf{left}^{3}; \mathsf{loop}(\mathsf{home}_2; \mathsf{wbit}_1; \mathsf{pred}_{12}^{12}; \mathsf{pred}_2; \mathsf{left}^{3}); \mathsf{home}_2,\\
\mathsf{Prep} &= [\mathsf{clear}_{13}, \mathsf{half}, \mathsf{dup}_{12\to 10}, \mathsf{dup}_{9\to 2}, \mathsf{inc}_2, \mathsf{inc}_2, \mathsf{pred}_{12},\\
&\qquad \mathsf{marks}, \mathsf{dup}_{10\to 12}, \mathsf{dup}_{12\to 2}],
\end{aligned}$$

where pred_k^c is c copies of pred_k followed by skip; when track k carries $b + c$ and there are origin marks up to $b + c$, it takes $\sum_{j<c}(36(b + j) + 2k + 43) + 1$ units and leaves the counter b (Lemma I.127).

Lemma I.109 (The centre and the marks)**.** *Let τ be in the prefix state k with r rows.*

(a) half *runs from $(-1, \tau)$ to $(-1, \tau')$ within $1000(r + 1)^2$ units; track 2 of τ' carries r, track 12 carries $\lfloor r/2 \rfloor$, and τ' differs from τ only on tracks 2, 10 and 12, in rows $j \le r$; so it is still in the prefix state k.*

(b) If track 12 carries b, track 2 carries n, $12n \le b \le r$ and track 13 is blank, then marks *runs from $(-1, \tau)$ to $(-1, \tau')$ within $1500(n + 1)(r + 1)$ units, where $\tau'(b - 12j, 13) = 1$ for $j < n$, every other cell of track 13 is 0, track 12 carries $b - 12n$, track 2 carries 0, and only tracks 2, 12 and 13 have changed.*

(c) Prep *runs from $(-1, \tau)$ within $50000(r + 1)^2$ units to a tape τ_1 in the prefix state k with r rows on which tracks 2 and 12 carry a, $(a - 1 - 12t, 13) = 1$ for $t < W + 2$, and every other cell of track 13 is 0. In the view at a, the twelve-cell rows at 1 have $W + 2$ marked rows, are closed on the left, and all their other columns hold 0. Only tracks 2, 10, 12 and 13 have changed.*

Proof. (a) The first three parts take $11 + 31 + 27 = 69$ units and leave the counter 0 on tracks 2 and 12 and 0 in $(0, 10)$; $\mathsf{peek}_{2\to 0}$ (8) reaches the header $(0, 0)$. For $j \le r$ let I_j be the configuration at $(j, 0)$ on which track 2 carries j, track 12 carries $\lfloor j/2 \rfloor$, $(0, 10)$ holds $[j \text{ odd}]$, and nothing else has changed. For $j < r$ the header is 1; the pass returns to the root $(18j + 7)$, flips the parity cell and, if j is odd, advances track 12 (at most $18j + 96$), advances track 2 $(36j + 48)$ and peeks at the header of row $j + 1$ $(18j + 26)$: at most $400(r + 1)$ units with its test, ending in I_{j+1}, since $\lfloor (j + 1)/2 \rfloor = \lfloor j/2 \rfloor + [j \text{ odd}]$. At I_r the sentinel header stops the loop, and up takes $18r + 7$. In all at most $69 + 8 + r(400(r + 1) + 1) + 1 + 18r + 7 \le 1000(r + 1)^2$.

(b) Let J_j, $j \le n$, be the configuration at $(0, 2)$ on which track 2 carries $n - j$, track 12 carries $b - 12j$, and the marks $(b - 12j', 13)$, $j' < j$, have been written; left^3 (4 units) reaches J_0. The scanned

cell $(0,2)$ is 1 exactly when $j < n$. A pass from J_j returns to the root (6), writes the mark at row $b-12j$ ($36(b-12j)+35$), steps track 12 back twelve times (at most $12(36r+67)+1$), steps track 2 back ($36(n-j-1)+47$) and returns to $(0,2)$ (4): at most $1500(r+1)$ units with its test. With the final test and home_2 (6), Lemma I.6 gives at most $n \cdot 1500(r+1) + 11 \le 1500(n+1)(r+1)$.

(c) clear_{13} takes $65r+35$ units with its skip, and half leaves a on track 12. The two copies of Lemma I.126(b) (at most $102r+76$ units each) put a on track 10 and W on track 2; two increments ($72W+132$ units) make track 2 carry $W+2$, and pred_{12} ($36a+31$) makes track 12 carry $a-1$. Since $12(W+2)+1 \le \rho \le a$, part (b) applies with $b = a-1$ and $n = W+2$, within $1500(W+3)(r+1)$ units. The last two copies put a back on track 12 and then on track 2. The prefix state involves none of the tracks 2, 10, 12 and 13. In the view at a, the presence cell of row t is $(a-1-12t, 13)$, which is 1 for $t < W+2$ and 0 for $t \ge W+2$ and for $t = -1$, the cell $(a+11, 13)$. Since $W+3 \le r+1$ the total is at most $50000(r+1)^2$. □

Loading the numerator and the denominator. With src and probe of Lemma I.45, put

$$\begin{aligned}
\mathsf{sbit} &= \mathsf{seek}_5; \mathsf{right}^1; \mathsf{branch}(\mathsf{home}_4; \mathsf{wbit}_0,\ \mathsf{home}_4; \mathsf{wbit}_1),\\
\mathsf{lcol} &= \mathsf{src}; \mathsf{probe}; \mathsf{loop}(\mathsf{home}_{10}; \mathsf{sbit}; \mathsf{inc}_5; \mathsf{pred}_{12}^{12}; \mathsf{src}; \mathsf{probe}); \mathsf{home}_{10},\\
\mathsf{lsign} &= \mathsf{sbit}; \mathsf{inc}_5; \mathsf{pred}_{12}^{2}; \mathsf{lcol},\\
\mathsf{Load} &= \big[\mathsf{lsign}, \mathsf{dup}_{2\to12}, \mathsf{pred}_{12}^{3}, \mathsf{lsign}, \mathsf{dup}_{2\to12}\big].
\end{aligned}$$

Lemma I.110 (Loading a rational)**.** *Let τ be a workspace of r rows with origin marks up to r.*

(a) If track 5 carries $s \le r$ and track 12 carries $b \le r$, then sbit *runs from the root to the root in $36s+36b+54$ units and sets $(b,13)$ to $\tau(s,4)$; nothing else changes.*

(b) Let the saved tracks lay out the frame of an integer x, the bit $[x<0]$ followed by the $\ell = \mathrm{bl}(|x|)+1$ lowest binary digits $x_0,\dots,x_{\ell-1}$ of $|x|$, possibly followed by further frames, from row s (Definition I.43), with $s+\ell+1 \le r$; let track 5 carry s and track 12 carry b, with $2+12\ell \le b \le r$. Then lsign *runs from the root to the root within $12000(r+1)^2$ units; it sets $(b,13)$ to $[x<0]$ and $(b-2-12t,13)$ to x_t for $t<\ell$, leaves the counter $s+\ell+1$ on track 5 and $b-2-12\ell$ on track 12, and changes nothing else outside tracks 10 and 12.*

(c) From the tape τ_1 of Lemma I.109(c), Load *runs within $50000(r+1)^2$ units to a tape τ_2 on which, in the view at a, the cell 0 holds $[u<0]$, column 1 of the twelve-cell rows at 1 holds the $W+2$ lowest binary digits of $|u|$, column 4 holds those of e, and the other columns are as in τ_1; tracks 2 and 12 carry a; track 5 carries the row of the frame of u_{k+1} (for $k+1=d$, of the frame after that of e_{d-1}); and τ_2 agrees with τ_1 off tracks 5, 10, 12 and 13.*

Proof. (a) seek_5 ($18s+8$) and right^1 (2) reach $(s,4)$; the branch (1) reads it, home_4 ($18s+8$) returns, and wbit writes $(b,13)$ in $36b+35$ units.

(b) By (a), sbit copies the sign cell $(s,4) = [x<0]$ into $(b,13)$; inc_5 ($36s+54$) and pred_{12}^2 (at most $72r+135$) leave $s+1$ on track 5 and $b-2$ on track 12. In lcol, src ($18(s+1)+11$) reaches $(s+1,3)$ and probe (22) writes into $(s+1,10)$ whether the frame goes on at row $s+1$, which it does. For $t \le \ell$ let K_t be the configuration at $(s+1+t,10)$ on which track 5 carries $s+1+t$,

track 12 carries $b-2-12t$, the digits $x_0,\dots,x_{t-1}$ have been written, and the flag holds $[t<\ell]$: by the layout the rows $s+1,\dots,s+\ell-1$ continue the frame and row $s+1+\ell$ does not. A pass from K_t, $t<\ell$, returns to the root, copies $x_t=\tau(s+1+t,4)$ into $(b-2-12t,13)$, advances track 5, steps track 12 back twelve times and probes the next row: at most $1000(r+1)$ units with its test, ending in K_{t+1}. With $\ell\le r$, the last test and home_{10}, lcol takes at most $1100(r+1)^2$ units and lsign at most $12000(r+1)^2$. The probes write track 10; nothing else outside track 12 and the named cells of track 13 is written.

(c) By Definition I.107(b) track 5 carries the row of the frame of u, and the frame of e follows it. Both have at most W digits, so $2+12W+3\le\rho\le a$ gives the room for (b), and the frames end before $|w_u|\le r$. The first lsign, with $b=a$, writes $[u<0]$ into $(a,13)$, the virtual cell 0, and the digits of $|u|$ into $(a-2-12t,13)$, which is column 1 of row t. Track 2 still carries a, so $\mathsf{dup}_{2\to12}$ puts a back on track 12, and pred^3_{12} makes it $a-3$. The second lsign, with $b=a-3$, writes $[e<0]=0$ into column 2 of row 0 and the digits of e into $(a-5-12t,13)$, column 4 of row t; the last copy puts a back on track 12. Prep left these columns 0, so above the digits they hold 0, and the columns hold the $W+2$ lowest digits. Track 5 now carries the row after the frame of e. The total is at most $24000(r+1)^2+204(r+1)+3(36r+67)+2\le50000(r+1)^2$. □

The dividend. Let $\overline{\mathsf{cp}}_{s\to t}$ be $\mathsf{cp}_{s\to t}$ of Lemma I.25 with the two parts of its branch exchanged; it writes the complement of column s into column t, in the same number of units. Put $\mathsf{c}_b=\mathsf{right}^6;\mathsf{write}\,b;\mathsf{left}^6$ and

$$\mathsf{Copy}=\mathsf{call}(\mathsf{right};\mathsf{cp}_{1\to2};\mathsf{back}),$$
$$\mathsf{Scale}=\big[\mathsf{zero}_{10},\ \mathsf{rep}_{10,8}(\mathsf{twice}),\ \mathsf{twice}\big],\qquad \mathsf{twice}=\mathsf{call}(\mathsf{right};\mathsf{z}_{10};\mathsf{dbl};\mathsf{back});\mathsf{inc}_{10},$$
$$\mathsf{Add}=\mathsf{call}(\mathsf{right};\mathsf{madd};\mathsf{back}),$$
$$\mathsf{madd}=\mathsf{left};\mathsf{branch}\big(\mathsf{right};\mathsf{c}_0;\mathsf{add}^{\uparrow2},\ \mathsf{right};\overline{\mathsf{cp}}_{2\to2};\mathsf{back};\mathsf{c}_1;\mathsf{add}^{\uparrow2}\big),$$

with rep of Lemma I.126.

Lemma I.111 (The dividend). *From the tape τ_2 of Lemma I.110(c), the programs* Copy, Scale *and* Add, *run in this order, take at most $100000(r+1)^2$ units and lead to a tape τ_3 on which, in the view at a, column 8 of the $W+2$ marked rows holds a word x with $\mathrm{sval}_{W+2}(x)=D$, column 4 still holds the digits of e, the cell 0 holds $[u<0]$, and the marks are kept; τ_3 agrees with τ_2 off tracks 10 and 13.*

Proof. Write $n=W+2$ and $h=1$. Each virtual program starts at the cell 0, steps to the presence cell h of row 0 and ends there after back; it moves and writes only in $[-12,12(n+1)]\subseteq[-\rho,\rho]$, so Lemma I.108 applies.

Copy: the virtual program takes $1+(26n+1)+(14n+27)$ units (Lemma I.25) and copies the digits of $|u|$ into column 2; with the call, at most $1000(r+1)$ units, since $n\le r$.

Scale: zero_{10} (27 units) makes track 10 carry 0, and track 8 carries N. For $j\le N+1$ let R_j say that column 2 holds the n lowest digits of $2^j|u|$ and track 10 carries j. The virtual program of twice clears column 10 of row 0 and doubles column 2 with the incoming bit 0 (Lemma I.26), in $1+23+(62n+1)+(14n+27)$ units. Since $2^{j+1}|u|\le2^{N+1}|u|<2^n$ for $j\le N$, there is no overflow, and inc_{10} $(36j+64)$ makes track 10 carry $j+1$: twice runs from R_j to R_{j+1} within $2000(r+1)$ units. By Lemma I.126(c) with $c=10$ and $g=8$, $\mathsf{rep}_{10,8}(\mathsf{twice})$ runs from R_0 to

R_N within $(N+1)\cdot 2200(r+1) \le 2200(r+1)^2$ units, and the last twice reaches R_{N+1}: column 2 holds $2^{N+1}|u|$.

Add: the virtual program steps back to the cell 0, reads $[u<0]$ and returns to h. The adder $\mathsf{add}^{\uparrow 2}$ works on the even columns, so its operands are columns 2 and 4, its carries column 6, its sum column 8, and c_b sets its carry-in, column 6 of row 0, to b. If $u \ge 0$, the sum of $2^{N+1}|u|$ and e is $D < 2^{n-1}$, so it is $\mathrm{val} = \mathrm{sval}_n$ of the sum (Lemma I.21). If $u<0$, column 2 is first complemented and the carry-in is 1; addition is symmetric in its two operand columns, so Corollary I.22(b) gives $\mathrm{sval}_n = e - 2^{N+1}|u| = D$, since e and $2^{N+1}|u|$ are below 2^{n-1} and $|D| < 2^{n-1}$. The virtual program takes at most $3+(28n+1)+(14n+27)+15+3(20n+1)+(14n+27) \le 120n+80$ units; with the call, at most $3000(r+1)$.

The three parts write only track 13 inside the area and track 10; the total is at most $1000(r+1)+2200(r+1)^2+2000(r+1)+28+3000(r+1) \le 100000(r+1)^2$. □

The division. Put $\mathsf{c}^- = \mathsf{left}^6;\mathsf{write}\,0;\mathsf{right}^6$ and, for twelve-cell rows,

$$\begin{aligned}
\mathsf{abs} &= \mathsf{cp}_{8\to 6};\mathsf{left}^6;\mathsf{branch}(A_0,A_1),\\
A_0 &= \mathsf{right}^6;\mathsf{back};\mathsf{skip};\mathsf{cp}_{6\to 1};\mathsf{back};\mathsf{left};\mathsf{write}\,0;\mathsf{right},\\
A_1 &= \mathsf{right}^6;\mathsf{back};\mathsf{neg}^{\uparrow 6};\mathsf{back};\mathsf{cp}_{6\to 1};\mathsf{back};\mathsf{left};\mathsf{write}\,1;\mathsf{right},\\
\mathsf{trim} &= \mathsf{fwd};\mathsf{left}^{12};\mathsf{write}\,0;\mathsf{back}, \qquad \mathsf{ddiv} = \mathsf{cp}_{4\to 2};\mathsf{back};\mathsf{z}_{10};\mathsf{dbl};\mathsf{back};\overline{\mathsf{cp}}_{2\to 4};\mathsf{back},\\
\mathsf{Div} &= \mathsf{call}(\mathsf{right};\mathsf{c}^-;\mathsf{abs};\mathsf{trim};\mathsf{ddiv};\mathsf{sdiv}),
\end{aligned}$$

with neg of Lemma I.24 and sdiv of Lemma I.30.

Lemma I.112 (Magnitude, trimming and the divisor)**.** *Let the twelve-cell rows at h have $n \ge 1$ marked rows and be closed on the left.*

(a) If column 8 holds a word x with $\mathrm{sval}_n(x) = Z$ and $\sigma(h-6) = 0$, then abs runs from h to h within $400(n+1)$ units, leaves in column 1 the n lowest binary digits of $|Z|$ and in the cell $h-1$ the bit $[Z<0]$, and writes only columns 1 and 6 and the cell $h-1$.

(b) trim runs from h to h in $28n+28$ units and sets column 0 of row $n-1$ to 0, so that $n-1$ rows remain marked; nothing else changes.

(c) If column 4 holds a word y with $2\,\mathrm{val}(y) < 2^n$, then ddiv runs from h to h within $200(n+1)$ units and leaves in column 4 the complement of the n lowest digits of $2\,\mathrm{val}(y)$; it writes only columns 2, 4 and 10.

Proof. (a) $\mathsf{cp}_{8\to 6}$ ($48n+1$ units) copies x into column 6 and ends at the end row; left^6 (7) reaches column 6 of row $n-1$, the top bit of x, which is 1 exactly when $Z<0$, because $\mathrm{sval}_n(x) = \mathrm{val}(x) - 2^n x_{n-1}$ and $0 \le \mathrm{val}(x) < 2^n$. After the branch, right^6 and back return to h ($7+14n+27$). If $Z<0$, the cells $h+6i$ form a tape of stride 6 whose two-cell rows at h are columns 0 and 6, and the cell before them, $h-6$, holds 0; by Lemmas I.24 and I.8, $\mathsf{neg}^{\uparrow 6}$ replaces column 6, within $7\cdot 30(n+1)$ units, by a word of value $2^n - \mathrm{val}(x) = -Z = |Z|$; then back. If $Z \ge 0$, column 6 already holds $\mathrm{val}(x) = |Z|$. Then $\mathsf{cp}_{6\to 1}$ ($34n+1$), back ($14n+27$),

and the last three primitives write $[Z < 0]$ into $h - 1$. In all at most $334n + 311 \le 400(n + 1)$ units.

(b) fwd $(14n + 1)$ reaches the end row, left^{12} (13) the presence cell of row $n - 1$, which is set to 0 (1), and back over the $n - 1$ remaining rows takes $14(n - 1) + 27$.

(c) $\mathsf{cp}_{4\to 2}$ $(32n+1)$ and back $(14n+27)$ copy y into column 2; z_{10} (23), dbl $(62n+1)$ and back double it without overflow; $\overline{\mathsf{cp}}_{2\to 4}$ $(32n + 1)$ and back write its complement into column 4: in all $168n + 107$ units. □

Lemma I.113 (The rounded quotient)**.** *From the tape τ_3 of Lemma I.111,* Div *runs within $20000(r + 1)^2$ units to a tape τ_4 on which, in the view at a, the cell 0 holds $[z < 0]$ and column 8 of the first $W + 1$ twelve-cell rows at 1 holds the $W + 1$ binary digits of $|z|$; so the cell 0 followed by column 8 is $\mathrm{sw}_W(z)$. Track 12 carries a, and τ_4 agrees with τ_3 off track 13.*

Proof. The virtual program steps to $h = 1$ and clears the cell $h-6$ (15 units). By Lemma I.112(a) with $n = W + 2$ and $Z = D$, column 1 then holds the digits of $|D|$ and the cell 0 the bit $[D < 0]$; column 4 still holds e. trim leaves $n' = W + 1$ marked rows, and since $|D| < 2^W$ and $e \le 2^{s_0}$, the first n' rows of columns 1 and 4 still hold $|D|$ and e. By Lemma I.112(c), column 4 then holds the complement of $y = 2e$, with $0 < y \le 2^W = 2^{n'-1}$. These are the hypotheses of Lemma I.30 with n' rows and $|D| < 2^{n'}$: sdiv leaves in the cell $h - 1$, the virtual cell 0, the bit $[\lfloor D/y \rfloor < 0]$ and in column 8 the n' digits of $|\lfloor D/y \rfloor|$, and $\lfloor D/(2e) \rfloor = z$. The virtual program takes at most $1 + 15 + 400(W + 3) + (28W + 84) + 200(W + 2) + 945(W + 2)^2 \le 2000(W + 3)^2$ units and stays in $[-12, 12(W + 3)]$. With $(W + 3)^2 \le 4(W + 2)^2 \le r/100$, Lemma I.108 bounds Div by $340r + 36r + 40 \le 20000(r + 1)^2$ units. □

Appending the word. Put

$$\begin{aligned}
\mathsf{app}_{11}(b) &= \big[\mathsf{seek}_{11}, \mathsf{right}^5, \mathsf{write}\,1, \mathsf{left}, \mathsf{write}\,b, \mathsf{left}^4, \mathsf{next},\\
&\qquad \mathsf{right}^5, \mathsf{write}\,0, \mathsf{left}^5, \mathsf{home}_{11}\big],\\
\mathsf{obit} &= \mathsf{seek}_{12}; \mathsf{left}^1; \mathsf{branch}\big(\mathsf{home}_{13}; \mathsf{app}_{11}(0),\\
&\qquad \mathsf{home}_{13}; \mathsf{app}_{11}(1)\big),\\
\mathsf{Emit} &= \mathsf{obit}; \mathsf{pred}^9_{12}; \mathsf{zero}_{10}; \mathsf{rep}_{10,9}(\mathsf{ebit}); \mathsf{ebit},\\
\mathsf{ebit} &= \mathsf{obit}; \mathsf{pred}^{12}_{12}; \mathsf{inc}_{10}.
\end{aligned}$$

Lemma I.114 (Appending the signed word)**.** *Let τ be a workspace of r rows with origin marks up to r.*

(a) If track 11 carries ℓ with $\ell + 1 \le r$, then $\mathsf{app}_{11}(b)$ runs from the root to the root in $36\ell + 95$ units; it sets $(\ell, 6)$ to 1, $(\ell, 7)$ to b and $(\ell + 1, 6)$ to 0, and advances the counter on track 11 to $\ell + 1$. So if the binary stream holds a word w of length ℓ, it then holds wb.

(b) If moreover track 12 carries $b' \le r$, then obit runs from the root to the root in $36b' + 36\ell + 130$ units and appends $\tau(b', 13)$ in the same way.

(c) From the tape τ_4 of Lemma I.113, Emit *runs within* $20000(r+1)^2$ *units to a tape on which the binary stream holds* $\mathrm{sw}_W(Z_{0,0})\cdots\mathrm{sw}_W(Z_{0,k})$ *and track* 11 *carries its length* $(k+1)(W+2)$*; only tracks* 6, 7, 10, 11 *and* 12 *change.*

Proof. (a) seek_{11} $(18\ell+14)$ reaches $(\ell,11)$; right^5 (6) the presence cell $(\ell,6)$; the writes and moves (3) set the presence and value cells; left^4 (5) returns to $(\ell,11)$; next (20) advances the counter; right^5, the write of the new terminator $(\ell+1,6)$ and left^5 take 13; and home_{11} from row $\ell+1$ takes $18\ell+33$; the list adds 1. (b) seek_{12} $(18b'+15)$, left^1 (2), the branch (1) and home_{13} $(18b'+17)$, then (a).

(c) Let $\ell = k(W+2)$; then $\ell+W+3 \le d(W+2)+1 < r$. In the view at a the cell 0 is $(a,13)$ and column 8 of row t is $(a-9-12t,13)$. The first obit appends $[z<0]$; pred^9_{12} (at most $9(36r+67)+1$) leaves $a-9$ on track 12, and zero_{10} (27) the counter 0 on track 10. For $t \le W+1$ let R_t say that the stream holds the old word, the sign and the digits $|z|_0,\dots,|z|_{t-1}$, track 11 carries $\ell+1+t$, track 12 carries $a-9-12t$, and track 10 carries t. Then ebit runs from R_t to R_{t+1} within $2500(r+1)$ units: it appends $|z|_t$, steps track 12 back twelve rows and advances track 10. Track 9 carries W, so by Lemma I.126(c) $\mathsf{rep}_{10,9}(\mathsf{ebit})$ runs from R_0 to R_W within $(W+1)\cdot 2700(r+1)$ units, and the last ebit reaches R_{W+1}: $W+2$ bits, $\mathrm{sw}_W(z)$, have been appended. In all at most $72r+130+9(36r+67)+1+27+2700(r+1)^2+2500(r+1) \le 20000(r+1)^2$ units. □

Lemma I.115 (One coordinate)**.** *Let $k<d$. From every tape in the prefix state k with r rows,* Coord *runs from the root to the root within* $250000(r+1)^2$ *units to a tape in the prefix state* $k+1$ *with the same r rows.*

Proof. Lemmas I.109(c), I.110(c), I.111, I.113 and I.114(c) take the tape through $\tau_1,\dots,\tau_4$ to a tape whose binary stream holds $\mathrm{sw}_W(Z_{0,0})\cdots\mathrm{sw}_W(Z_{0,k})$ with its length on track 11. Then pred_{14} $(36(d-k-1)+71$ units) leaves the counter $d-k-1$ on track 14. None of the parts writes tracks 0, 1, 3, 4, 8, 9 or 15, and only Load writes track 5, which then carries the row of the frame of u_{k+1}. So the tape is in the prefix state $k+1$. The parts take at most $(50000+50000+100000+20000+20000)(r+1)^2+36r+72 \le 250000(r+1)^2$ units. □

Lemma I.116 (All coordinates)**.**

(a) Let $\mathsf{Begin} = \mathsf{zero}_6;\mathsf{zero}_{11}$*. From the tape of Lemma I.105 with $R' \ge \rho_{\mathrm{traj}}$,* Begin *runs in* 48 *units to a tape in the prefix state* 0 *with R' rows.*

(b) From every tape in the prefix state 0 *with r rows,* Coords *runs from the root to the root in some t units to a tape in the prefix state d with r rows, and*

$$t+r+1 \le \big(250080(d+1)+2\big)(r+1)^3.$$

Proof. (a) zero_6 (19 units) writes 0 into $(0,6)$, so the binary stream is empty, and zero_{11} (29) makes track 11 carry 0; nothing else changes, and Lemma I.105 supplies the rest of the prefix state 0.

(b) For $k \le d$ let I_k be the set of configurations at $(0,14)$ whose tape is in the prefix state k with r rows; left^{15} (16 units) goes from the root to $(0,14)$. The scanned cell $(0,14)$ holds 1

exactly when the counter $d-k$ is positive. A pass from I_k, $k<d$, returns to the root (18 units), runs Coord (at most $250000(r+1)^2$ units, Lemma I.115) and steps back to $(0,14)$ (16). By Lemma I.6 the loop takes at most $d(250000(r+1)^2+35)+1$ units, and home_{14} adds 18. So $t+r+1 \le 250000d(r+1)^2+35d+36+(r+1) \le (250080(d+1)+2)(r+1)^3$. □

Lemma I.117 (Packing the first vector)**.** *Let τ carry the saved request, hold the archive (S,σ,N,W,m) within r rows and be a workspace of r rows with origin marks up to r. Put $A'=2^{10}\cdot 10^{22}$, $B'=A'+(C_{\mathrm{pos}}+48)(A'+1)^8$, $K_{\mathrm{c}}=250080(d+1)+2$,*

$$K_{\mathrm{pk}}=B'+K_{\mathrm{c}}(B'+1)^3, \qquad K_{\mathrm{pp}}=C_{\mathrm{pos}}+K_{\mathrm{pk}}(C_{\mathrm{pos}}+1)^{360}.$$

(a) Pack *runs from $(-1,\tau)$ to $(-1,\tau')$ in some t units, where for some $r'\ge r$ the tape τ' carries the saved request, holds the archive within r' rows, is a workspace of r' rows with origin marks up to r', $r'\ge\rho_{\mathrm{traj}}$, the binary stream holds $\mathrm{sw}_W(Z_0)$, and*

$$t+r'+1 \le K_{\mathrm{pk}}(X+r+1)^{360}.$$

(b) Pos; Pack *runs to a tape with the same properties, and $t+r'+1\le K_{\mathrm{pp}}(X+r+1)^{2888}$.*

Proof. (a) Put $Y=X+r+1$. By Lemma I.104, Res ends on $r_1\ge\rho_{\mathrm{traj}}$ rows, and $t_1+r_1+1\le 10^{22}(r+2)^{10}\le A'Y^{10}$, since $r+2\le 2Y$. By Lemmas I.105 and I.116(a), Pos; Begin runs in t_2+48 units to a tape in the prefix state 0 with $r_2\ge r_1$ rows, and $(t_2+48)+r_2+1\le(C_{\mathrm{pos}}+48)(Y+r_1+1)^8$. Lemma 15.36(a) with Y for its N, $e=10$ and $g=8$ bounds the first three parts and r_2+1 by $B'Y^{90}$. By Lemma I.116(b), Coords ends in the prefix state d with r_2 rows, whose binary stream holds $\mathrm{sw}_W(Z_0)$, and $t_3+r_2+1\le K_{\mathrm{c}}(Y+r_2+1)^3$. Lemma 15.36(a) with $e=90$ and $g=3$ gives the bound with $r'=r_2$ and the exponent $90\cdot 4=360$.

(b) By Lemma I.105, Pos runs in t_0 units to a tape with the hypotheses of (a) on $R\ge r$ rows, and $t_0+R+1\le C_{\mathrm{pos}}Y^8$; by (a) from there, $t+r'+1\le K_{\mathrm{pk}}(Y+R+1)^{360}$. Lemma 15.36(a) with $e=8$ and $g=360$ gives the exponent $8\cdot 361=2888$. □

Proof of Theorem I.102. Put $Y=X+r+1$. By Lemma I.103, Arch runs in t_1 units to a tape with the hypotheses of Lemma I.104 on $r_1\ge r$ rows, and $t_1+r_1+1\le 10^{10}Y^2$. By Lemma I.104, Res runs in t_2 units to $r_2\ge\rho_{\mathrm{traj}}$ rows, and $t_2+r_2+1\le 10^{22}(Y+r_1+1)^{10}$; Lemma 15.36(a) with $e=2$ and $g=10$ gives $t_1+t_2+r_2+1\le L_1Y^{22}$ with $L_1=10^{10}+10^{22}(10^{10}+1)^{10}$. By Lemma I.117(b), Pos; Pack runs in t_3 units to $r_3\ge r_2\ge\rho_{\mathrm{traj}}$ rows with $\mathrm{sw}_W(Z_0)$ on the binary stream, and $t_3+r_3+1\le K_{\mathrm{pp}}(Y+r_2+1)^{2888}$; with $e=22$ and $g=2888$, $t_1+t_2+t_3+r_3+1\le L_2Y^{63558}$, $L_2=L_1+K_{\mathrm{pp}}(L_1+1)^{2888}$, since $22\cdot 2889=63558$. By Lemma I.106, Rest runs in t_4 units to a tape that holds the stored vector Z_0 on $r'=r_3$ rows, with the counter 0 on track 2 and the retained word of length S on track 14, and $t_4+r'+1\le 2\cdot 10^6(Y+r_3+1)^2$. With $e=63558$ and $g=2$ the total satisfies

$$t+r'+1\le K_{\mathrm{i}}Y^{190674}, \qquad K_{\mathrm{i}}=L_2+2\cdot 10^6(L_2+1)^2,$$

because $63558\cdot 3=190674$. The constants depend only on the problem and n_0: on d, and on c_{num} through C_{pos}. □

I.11 One coefficient of the Taylor table

How does the machine find the coefficient of t^3 in the Taylor series of $\sin t$? Take the oscillator $y_0' = y_1$, $y_1' = -y_0$ as a coefficient field: $c = 0$, the coordinate 0 has the single term $(1, 1)$ of A and the coordinate 1 the single term $(-1, 0)$, and there are no terms of B. For illustration let $N = 3$, so $s = 2^N = 8$, and let the stored vector be $Z = (0, 8)$, the point $(0, 1)$ at precision 2^{-3}. The rows of the table are $r_0 = (0, 8)$, $r_1 = (8, 0)$, $r_2 = (0, -4)$, and the stage is about to append the first entry of r_3. The output buffer then holds six signed words, $0, 8, 8, 0, 0, -4$. From its length the machine finds the order $l = 2$ and the coordinate $j = 0$. It reads from the saved request that the coordinate 0 has the constant $0 = 0/1$ and the coefficient $1 = 1/1$, so the common denominator is $D = 1 \cdot 1 = 1$; it reads the index 1 of the term and finds $r_{2,1} = -4$ as the word number $2 \cdot 2 + 1 = 5$ of the buffer. The integer numerator of $(Af_2)_0 = f_{2,1}$ over Ds^2 is $U = D \cdot 1 \cdot r_{2,1} \cdot s = -32$. One division with rounding finishes the entry:

$$r_{3,0} = \left\lfloor \frac{2U + Ds(l+1)}{2Ds(l+1)} \right\rfloor = \left\lfloor \frac{-40}{48} \right\rfloor = -1 = \operatorname{near}\Big(\frac{-4}{3}\Big),$$

and $-1/8$ is the coefficient $-\frac{1}{6}$ of t^3 in $\sin t$, rounded to precision 2^{-3}. The machine appends the signed word of -1 to the buffer and sets two flags that tell the two enclosing loops whether the row and the table are complete.

This section constructs that computation for every coefficient field. It gives the three programs of the stage (one entry, one row, the whole table), proves their contracts, and proves the step bound of each with its exponent. The exponent of one entry is 118, of one row 118 and of the whole table 120. Section I.11.1 states the results and proves the two loops; the routines follow; the proof of the bound for one entry closes the section.

I.11.1 The table, the three programs and their contracts

Throughout this section a problem with a coefficient field of dimension d is fixed, with level n_0, radius M and the constants of Section 15.9. For a request with coordinate $i < d$ and precision m put $n = \max(m, n_0)$, $N = N(n)$, $s = 2^N$ and $W = W_m$, and let $X = m + n_0 + \kappa + 1$. Since $N = n + \kappa$, we have $N < X$.

Definition I.118 (The table in the buffer)**.** Let $Z = (Z_0, \dots, Z_{d-1})$ be an integer vector. Put $r_0 = Z$ and, for $l < N$,

$$r_{l+1,j} = \operatorname{near}\big(s\,\operatorname{next}(f, l)_j\big), \qquad f_i = r_i/s,$$

the rows of Lemma 15.35. For $l < N$ and $j \le d$ the *table prefix* $T_{l,j}(Z)$ is the list of integers

$$r_{0,0}, \dots, r_{0,d-1},\ r_{1,0}, \dots, r_{l,d-1},\ r_{l+1,0}, \dots, r_{l+1,j-1}\ :$$

the rows $r_0, \dots, r_l$ followed by the first j entries of r_{l+1}. So $T_{l,d}(Z) = T_{l+1,0}(Z)$ for $l + 1 < N$, and the *table* is $T(Z) = T_{N-1,d}(Z)$, the rows $r_0, \dots, r_N$. We say that the output buffer *holds the table prefix* (l, j) of Z if it holds the word $\operatorname{sw}_W(T_{l,j}(Z))$ with capacity r, the number of rows in use (Definition I.3(b)). This word has $((l + 1)d + j)(W + 2)$ bits.

If $\|Z/s\| \le M$, every entry of every r_l fits W, by Proposition I.32(a), so the signed words are defined. The stored vector is kept between steps as in Definition I.101: the saved request on

tracks 3, 4 and 15, the counters m, N, W_m and σ on tracks 5, 8, 9 and 10, the word $\mathrm{sw}_W(Z)$ on the binary stream, origin marks and a workspace of r rows, and a reserve of rows; in particular $r \geq 100(d+1)(N+1)(W+2)^2$. The coefficient stage uses tracks 11, 12 and 13 of every row and, as scratch that it restores, the tracks 2, 5 to 10 and 14 and the positive half of the tape.

The programs. The stage is three nested programs. Let $\mathsf{sent} = \mathsf{left}^1; \mathsf{scanL}; \mathsf{left}^{13}$; from the root of a tape whose track 0 carries the counter r it walks along the headers to the sentinel row and stops at cell$(r, 13)$, in $2 + (18r + 1) + 14 = 18r + 17$ units (Lemma I.10). The two sentinel cells $(r, 13)$ and $(r, 11)$ are the flags of the two loops: the buffer occupies the rows below its length, which is less than r, so these cells are free. Put

$$\begin{aligned} \mathsf{Build} &= \mathsf{seed}; \mathsf{sent}; \mathsf{write}\,1; \mathsf{loop}(\mathsf{Row}); \mathsf{home}_{13}, \\ \mathsf{Row} &= \mathsf{right}^2; \mathsf{write}\,1; \mathsf{loop}(\mathsf{Append}); \mathsf{left}^2, \\ \mathsf{Append} &= \big[\mathsf{home}_{11},\ \mathsf{Archive},\ \mathsf{Indices},\ \mathsf{dup}_{5\to10},\ \mathsf{dup}_{8\to7},\ \mathsf{pred}_7,\ \mathsf{Width}, \\ &\qquad \mathsf{BankInit},\ \mathsf{zero}_{14},\ \mathsf{Den},\ \mathsf{Num},\ \mathsf{Ops},\ \mathsf{Quot},\ \mathsf{Finish}\big]. \end{aligned}$$

seed copies the stored vector into the empty output buffer as row r_0 (Lemma I.125). Build then repeats Row while the flag $(r, 13)$ holds 1, and Row repeats Append while the flag $(r, 11)$ holds 1. Each run of Append appends one entry and rewrites both flags. Its parts are constructed in the rest of the section; here is what they do.

(1) home_{11} returns from the flag to the root. $\mathsf{Archive}$ copies every cell of rows $0, \dots, r$ into the positive half of the tape (Lemma I.130), so that the stage can use the registers freely and restore them at the end.

(2) $\mathsf{Indices}$ recovers $l+1$ and j from the length $((l+1)d+j)(W+2)$ of the buffer, by counting words and dividing by d in unary (Lemma I.142). The track copies, the decrement and Width put l on track 7, j on track 10 and W on track 9.

(3) $\mathsf{BankInit}$ lays out an arithmetic area of $H = 3W$ twelve-cell rows on track 13, below the sentinel row and above the buffer (Lemma I.139).

(4) Den forms the common denominator D_j of the coefficients of coordinate j, Num the integer numerator

$$U = [l = 0]\, D_j c_j s^2 + \sum_{(a,u)} D_j a\, r_{l,u}\, s + \sum_{(b,u,v)} \sum_{p \leq l} D_j b\, r_{p,u}\, r_{l-p,v}$$

of Proposition I.32(b), reading every coefficient from the saved request and every $r_{p,u}$ from the buffer, and Ops the two operands $2U + D_j s(l+1)$ and $2D_j s(l+1)$ (Lemmas I.151, I.153 and I.154).

(5) Quot divides once, rounding down (Lemma I.155); by Proposition I.32(b) the quotient is $r_{l+1,j}$.

(6) Finish computes the two flags, appends the signed word of the quotient to the buffer, restores every borrowed track from the archive, clears the archive and writes the flags (Lemma I.156).

Theorem I.119 (One entry of the table). *Put $K_{\mathrm{a}} = 2\,000\,000 + 114\,744\,020 \cdot 17^{118}$. Let $r, l < N$ and $j < d$, and let Z be an integer vector with $\|Z/s\| \le M$. Let τ be a tape that holds the stored vector Z on r rows (Definition I.101), on which track 2 carries a counter $c < S$ and track 14 a retained word of length S from row 0, whose output buffer holds the table prefix (l, j) of Z, and with $\tau(r, 11) = 1$. Then* Append *runs from $(\mathrm{cell}(r, 11), \tau)$ to $(\mathrm{cell}(r, 11), \tau')$ in some number t of units, where*

(a) τ' holds the stored vector Z on the same r rows;

(b) the output buffer of τ' holds the table prefix $(l, j + 1)$ of Z;

(c) $\tau'(r, 11) = [j + 1 < d]$ and $\tau'(r, 13) = [l + 1 < N]$;

(d) $\tau'(q, 2) = \tau(q, 2)$ and $\tau'(q, 14) = \tau(q, 14)$ for every $q \le S$;

(e) $t \le K_{\mathrm{a}}(X + r + 1)^{118}$.

The constant K_{a} is a numeral: it depends on nothing, not even on the problem. The exponent is $118 = g_D + g_F + g_R + g_Q + 3$ with the exponents $g_D = 24$, $g_F = 78$, $g_R = 8$ and $g_Q = 5$ of the four arithmetic phases; Table I.4 lists how each arises. The proof is at the end of the section.

Theorem I.120 (One row of the table). *Let $r, l < N$, Z, c and τ be as in Theorem I.119, except that the output buffer holds the table prefix $(l, 0)$ of Z, that is, the rows $r_0, \dots, r_l$, and that $\tau(r, 13) = 1$. Then* Row *runs from $(\mathrm{cell}(r, 13), \tau)$ to $(\mathrm{cell}(r, 13), \tau')$ in t units, where τ' holds the stored vector Z on r rows, its output buffer holds the rows $r_0, \dots, r_{l+1}$, $\tau'(r, 13) = [l + 1 < N]$, τ' agrees with τ on the cells $(q, 2)$ and $(q, 14)$, $q \le S$, and*

$$t \le (dK_{\mathrm{a}} + d + 8)(X + r + 1)^{118}.$$

Proof. Put $Y = X + r + 1$ and $b = K_{\mathrm{a}} Y^{118}$. For $j \le d$ let I_j be the set of configurations $(\mathrm{cell}(r, 11), \sigma)$ such that σ holds the stored vector Z on r rows, its buffer holds the table prefix (l, j), $\sigma(r, 11) = [j < d]$, $\sigma(r, 13)$ is 1 if $j = 0$ and $[l + 1 < N]$ if $j \ge 1$, and σ agrees with τ on the cells $(q, 2)$ and $(q, 14)$, $q \le S$. The program right^2 takes 3 units from $\mathrm{cell}(r, 13)$ to $\mathrm{cell}(r, 13) + 2 = \mathrm{cell}(r, 11)$, and write 1 one unit; since $d \ge 1$, the configuration is in I_0. Let $j < d$ and let a configuration in I_j be given. It scans $\sigma(r, 11) = 1$. Track 2 still carries c and track 14 the retained word, because these depend only on the cells $(q, 2)$, $q \le c < S$, and $(q, 14)$, $q \le S$. So Theorem I.119 applies: Append runs within b units to a configuration in I_{j+1}. Each configuration in I_d scans $[d < d] = 0$. By Lemma I.6, loop(Append) runs within $d(b + 1) + 1$ units from I_0 to I_d, and left^2 returns to $\mathrm{cell}(r, 13)$ in 3 units. In I_d the buffer holds the prefix (l, d), the rows $r_0, \dots, r_{l+1}$, and since $d \ge 1$ the flag $(r, 13)$ holds $[l + 1 < N]$. In all $t \le 3 + 1 + d(b + 1) + 1 + 3 = dK_{\mathrm{a}} Y^{118} + d + 8 \le (dK_{\mathrm{a}} + d + 8)Y^{118}$, because $Y \ge 1$. □

Theorem I.121 (The whole table). *Let r and Z with $\|Z/s\| \le M$ be given, and let τ hold the stored vector Z on r rows, with a counter $c < S$ on track 2 and a retained word of length S on track 14 from row 0. Then* Build *runs from $(-1, \tau)$ to $(-1, \tau')$ in t units, where τ' holds the stored vector Z on the same r rows, its output buffer holds the word $\mathrm{sw}_W(r_0, \dots, r_N)$ of the whole table with capacity r, τ' agrees with τ on the cells $(q, 2)$ and $(q, 14)$, $q \le S$, and*

$$t \le (dK_{\mathrm{a}} + d + 1008)(X + r + 1)^{120}.$$

The rows in use do not change, so in the sense of Definition I.7 this is a contract whose bound counts the units alone, with exponent 120 *and the constant* $dK_{\mathrm{a}} + d + 1008$*, which depends on the problem only through* d*.*

Proof. Put $Y = X + r + 1$ and $K_{\mathrm{r}} = dK_{\mathrm{a}} + d + 8$. By Lemma I.125, seed runs from $(-1, \tau)$ within $600(r+1)^2$ units to $(-1, \tau_1)$, where the buffer of τ_1 holds $\mathrm{sw}_W(Z)$, the prefix $(0, 0)$, and τ_1 differs from τ only on tracks 11, 12 and 13 of rows $j \le r$; the hypothesis $d(W+2) \le r$ of that lemma is part of the reserve. So τ_1 still holds the stored vector Z, which involves none of these tracks. sent takes $18r + 17$ units to $\mathrm{cell}(r, 13)$, and $\mathsf{write}\,1$ one unit; the buffer is not affected, since its length $d(W+2)$ is less than r.

For $l \le N$ let J_l be the set of configurations $(\mathrm{cell}(r, 13), \sigma)$ such that σ holds the stored vector Z on r rows, its buffer holds the rows $r_0, \dots, r_l$, $\sigma(r, 13) = [l < N]$, and σ agrees with τ on the cells $(q, 2)$ and $(q, 14)$, $q \le S$. The configuration reached is in J_0, because $N \ge 1$. From J_l with $l < N$ the scanned cell holds 1, and Theorem I.120 gives a run of Row within $K_{\mathrm{r}} Y^{118}$ units to a configuration in J_{l+1}; each configuration in J_N scans 0. By Lemma I.6 the loop takes at most $N(K_{\mathrm{r}} Y^{118} + 1) + 1$ units, and home_{13} from row r takes $18r + 17$ units (Lemma I.11). In J_N the buffer holds the whole table.

Since $N \le X \le Y$ and $r + 1 \le Y$,

$$t \le 600(r+1)^2 + 2(18r+17) + 1 + N(K_{\mathrm{r}} Y^{118} + 1) + 1 \le 600Y^2 + 36Y + K_{\mathrm{r}} Y^{119} + Y \le (K_{\mathrm{r}} + 1000) Y^{120},$$

and $K_{\mathrm{r}} + 1000 = dK_{\mathrm{a}} + d + 1008$. □

I.11.2 Packed lists, the output buffer and natural numbers

The stage keeps one more kind of record, in the positive half of the tape: a list of bits written as pairs of cells.

Definition I.122 (Packed lists)**.** A tape τ holds the *packed list* $w = w_0 \cdots w_{n-1}$ at h if $\tau(h+2i) = 1$ and $\tau(h + 2i + 1) = w_i$ for $i < n$, and $\tau(h + 2n) = 0$. The cell $h + 2i$ is the *presence* cell and $h + 2i + 1$ the *value* cell of the entry i.

Lemma I.123 (Packed lists)**.** *Let* τ *hold the packed list* w *of length* n *at* h*. Put*

$$\mathsf{scanP} = \mathsf{loop}(\mathsf{right}; \mathsf{right}), \qquad \mathsf{emitP}(b) = \mathsf{write}\,1; \mathsf{right}; \mathsf{write}\,b; \mathsf{right}; \mathsf{write}\,0,$$
$$\mathsf{appP}(b) = \mathsf{scanP}; \mathsf{emitP}(b), \qquad \mathsf{rewP} = \mathsf{left}^2; \mathsf{loop}(\mathsf{left}^2).$$

(a) scanP *runs from* (h, τ) *to* $(h + 2n, \tau)$ *in* $3n + 1$ *units.*

(b) $\mathsf{emitP}(b)$ *runs from* (x, τ) *in* 5 *units to* $x + 2$*, writing* 1*,* b *and* 0 *into the cells* x*,* $x + 1$ *and* $x + 2$*.*

(c) $\mathsf{appP}(b)$ *runs from* (h, τ) *in* $3n + 6$ *units to* $(h + 2n + 2, \tau')$*, where* τ' *holds the packed list* wb *at* h *and differs from* τ *only in the cells* $h + 2n$*,* $h + 2n + 1$ *and* $h + 2n + 2$*.*

(d) If $\tau(h - 2) = 0$*,* rewP *runs from* $(h + 2n, \tau)$ *to* $(h - 2, \tau)$ *in* $4n + 4$ *units.*

Routine	Bound	Reference
mul	$400(H+1)^2$	Lemma I.27
sdiv	$945(H+1)^2 \le 945(H+1)^3$	Lemma I.30
bank multiplication bmul	$441(H+1)^3$	Lemma I.138
denominator product, in the bank	$641(H+1)^4$, called: $11\,297(r+1)^5$	Lemma I.151
reading one coefficient	$6\,300\,000\,Y^4$	Lemma I.147
one denominator factor	$6\,311\,297\,Y^9$	Lemma I.151
one loop over the terms	$11\,311\,297\,Y^{14}$	Lemma I.151
Den	$K_D\,Y^{24}$, $K_D = 28\,935\,892$	Lemma I.151
quotient D_j/e, in the bank	$1\,245(H+1)^4$, called: $21\,565(r+1)^5$	Lemma I.152
weight product, in the bank	$541(H+1)^4$, called: $9\,597(r+1)^5$	Lemma I.152
one weight $D_j q$	$12\,631\,163\,Y^{14}$	Lemma I.152
constant part	$12\,681\,184\,Y^{16}$	Lemma I.153
one linear term	$21\,974\,448\,Y^{22}$	Lemma I.153
one Cauchy pair	$18\,627\,652\,Y^{12}$	Lemma I.153
one bilinear term	$36\,258\,815\,Y^{29}$	Lemma I.153
Num	$K_F\,Y^{78}$, $K_F = 80\,917\,448$	Lemma I.153
the factor $l+1$	$29\,798(r+1)^6$	Lemma I.154
Ops	$K_R(r+1)^8$, $K_R = 91\,799$	Lemma I.154
Quot	$K_Q(r+1)^5$, $K_Q = 798\,880$	Lemma I.155
start and Finish	$2\cdot 10^6(r+1)^3$ and $4\cdot 10^6(r+1)^3$	Lemmas I.150, I.156
Append	$K_{\mathrm{a}}(X{+}r{+}1)^{118}$, $118 = 24{+}78{+}8{+}5{+}3$	Theorem I.119
Row	$(dK_{\mathrm{a}} + d + 8)(X+r+1)^{118}$	Theorem I.120
Build	$(dK_{\mathrm{a}} + d + 1008)(X+r+1)^{120}$	Theorem I.121

Table I.4. The bounds of the coefficient stage. Here $H = 3W$ is the width of the arithmetic area and $Y = 17(r+1)$. A routine's exponent is the one its proof composes; none is the least possible.

Proof. (a) Lemma I.6 with $I_i = \{(h+2i, \tau)\}$: a pass costs 3 units with its test, and the cell $h+2n$ stops the loop. (b) Five primitives. (c) is (a) and (b) with $x = h+2n$; the new end cell $h+2n+2$ receives 0. (d) left^2 (3 units) reaches $h+2n-2$; the loop then passes over the presence cells $h+2n-2, \dots, h$, 4 units each with its test, and stops at $h-2$ (1 unit). □

The output buffer grows one bit at a time.

Lemma I.124 (Appending to the buffer)**.** *Let*

$$\mathsf{put}_{11}(b) = [\mathsf{seek}_{11}, \mathsf{left}^2, \mathsf{write}\, b, \mathsf{right}, \mathsf{write}\, 1, \mathsf{right}, \mathsf{next}, \mathsf{home}_{11}].$$

If track 11 *carries the counter* ℓ *and* τ *has origin marks up to* $\ell+1$*, then* $\mathsf{put}_{11}(b)$ *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *in* $36\ell+75$ *units, where* τ' *is* τ *with* $(\ell,13)$ *set to* b*,* $(\ell,12)$ *and* $(\ell,11)$ *set to* 1 *and* $(\ell+1,11)$ *set to* 0*. In particular, if the output buffer holds* w *with capacity* $R > |w| = \ell$*, then it holds* wb *with capacity* R *afterwards.*

Proof. seek_{11} ($18\ell+14$) reaches $(\ell,11)$, left^2 (3) the cell $(\ell,13)$; the writes and moves (4) set $(\ell,13)$ and $(\ell,12)$ and return to $(\ell,11)$; next (20) advances the counter (Lemma I.13); home_{11}

from row $\ell+1$ takes $18\ell+33$; and the list adds 1. The sum is $36\ell+75$. The presence cells $(j,12)$ with $\ell<j\le R$ are unchanged, so they still hold 0. □

Lemma I.125 (Emptying the buffer, and the seed row)**.** *Let τ be a workspace of r rows with origin marks up to r. Put* $\mathsf{empty}=\mathsf{clear}_{11,12}$ *and*

$$\mathsf{seed}=\mathsf{empty};\mathsf{rep}_{11,6}(\mathsf{cbit}),$$
$$\mathsf{cbit}=\mathsf{peek}_{11\to7};\mathsf{branch}\big(\mathsf{home}_7;\mathsf{put}_{11}(0),\ \mathsf{home}_7;\mathsf{put}_{11}(1)\big),$$

with peek *and* rep *as in Lemma I.126 below.*

(a) empty *runs from $(-1,\tau)$ to $(-1,\tau')$ in $124r+63$ units; τ' is τ with tracks 11 and 12 set to 0 in rows $j\le r$, so its output buffer holds the empty word with capacity r.*

(b) If the binary stream of τ holds w with $|w|\le r$, then seed *runs from $(-1,\tau)$ to $(-1,\tau'')$ within $600(r+1)^2$ units, where the output buffer of τ'' holds w with capacity r, τ'' is a workspace of r rows, and τ'' differs from τ only on tracks 11, 12 and 13 of rows $j\le r$.*

Proof. (a) By Lemma I.16 the two clears take $(61r+30)+(63r+32)$ units, and the list adds 1. (b) After empty, let R_q ($q\le|w|$) say that the buffer holds $w_0\cdots w_{q-1}$ and that the tape is otherwise as after empty. Track 11 then carries q, and the guard track 6 carries the counter $|w|$, the presence marks of the stream. A run of cbit from R_q, $q<|w|$, goes to $(q,11)$ and over to $(q,7)$ ($18q+14+5$ units), branches on w_q, returns along the origin marks ($18q+11$) and appends w_q with Lemma I.124 ($36q+75$): at most $200(r+1)$ units, ending in R_{q+1}. By Lemma I.126(c), $\mathsf{rep}_{11,6}(\mathsf{cbit})$ takes at most $(|w|+1)(200(r+1)+200(r+1))\le400(r+1)^2$ units, and with (a) the total is at most $600(r+1)^2$. Only tracks 11, 12 and 13 of rows below $|w|+1\le r+1$ were written. □

Several stages move a whole track, or repeat a body while a column holds marks. The next lemma gives these routines, and the one after it the decrement of a counter.

Lemma I.126 (Tracks and columns)**.** *Let $k,l,c,g<16$. Put* $\mathsf{peek}_{c\to g}=\mathsf{seek}_c;\mathsf{shift}_{c\to g}$,

$$\mathsf{dup}_{k\to l}=\mathsf{left}^1;\mathsf{loop}(\mathsf{row}_{k\to l};\mathsf{left}^{16});\mathsf{row}_{k\to l};\mathsf{left}^1;\mathsf{home}_1,$$
$$\mathsf{row}_{k\to l}=\mathsf{left}^k;\mathsf{branch}(E_0,E_1),\qquad E_b=\mathsf{right}^k;\mathsf{left}^l;\mathsf{write}\,b;\mathsf{right}^l,$$
$$\mathsf{rep}_{c,g}(P)=\mathsf{peek}_{c\to g};\mathsf{loop}(\mathsf{home}_g;P;\mathsf{peek}_{c\to g});\mathsf{home}_g.$$

(a) If track c carries the counter q, $\mathsf{peek}_{c\to g}$ *runs from $(-1,\tau)$ to $(\mathrm{cell}(q,g),\tau)$ in $18q+c+4+|c-g|$ units.*

(b) If track 0 of τ carries the counter r, τ has origin marks up to r and $2\le l$, then $\mathsf{dup}_{k\to l}$ *runs from $(-1,\tau)$ to $(-1,\tau')$ in exactly $(2k+2l+42)r+2k+2l+16$ units, where τ' is τ with (j,l) set to $\tau(j,k)$ for every $j\le r$.*

(c) Let $1\le g,n\le r$, and let $\mathrm{R}_0,\dots,\mathrm{R}_n$ be properties of tapes such that on every tape with R_q track c carries q, track g carries n, and there are origin marks up to r. If for every $q<n$ the program P runs from $(-1,\sigma)$ within b units to some $(-1,\sigma')$ with R_{q+1} whenever σ has R_q, then $\mathsf{rep}_{c,g}(P)$ *runs from every tape with R_0 to a tape with R_n within $(n+1)(b+200(r+1))$ units.*

Proof. (a) Lemma I.11(a) and (c). (b) left^1 (2 units) reaches the header $(0,0)$. For $j < r$ the pass at the header $(j,0) = 1$ reads (j,k), writes it into (j,l), returns to the header and moves to $(j+1,0)$: $1+(k+1)+1+(k+1)+(l+1)+1+(l+1)+17 = 2k+2l+24$ units with its test. At the sentinel $(r,0) = 0$ the test fails (1 unit), $\mathsf{row}_{k\to l}$ copies the cell of row r ($2k+2l+6$), left^1 (2) reaches $(r,1)$ and home_1 takes $18r+5$. The sum is the stated count. (c) Let I_q be the set of configurations at cell(q,g) whose tape has R_q. The first peek ($18\cdot 0 + c + 4 + |c-g| \le 36$ units) reaches I_0. A pass from I_q, $q < n$, scans $(q,g) = 1$ (1 unit), returns ($18q+g+4$), runs P (at most b) and peeks again ($18(q+1)+c+4+|c-g|$): at most $b + 100(r+1)$ units, ending in I_{q+1}. At I_n the cell (n,g) is the hole of track g. With the final test and home_g ($18n+g+4$), Lemma I.6 gives at most $36 + n(b+100(r+1)+1) + 1 + 18n + 20 \le (n+1)(b+200(r+1))$ units. □

Lemma I.127 (Decrementing a counter). *Let* $2 \le k < 16$ *and*

$$\mathsf{pred}_k = \mathsf{seek}_k; \mathsf{right}^{16}; \mathsf{write}\ 0; \mathsf{home}_k.$$

If track k *carries the counter* $a+1$ *and* τ *has origin marks up to* $a+1$*, then* pred_k *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *in* $36a+2k+43$ *units, where* τ' *is* τ *with* (a,k) *set to* 0*; track* k *then carries* a*.*

Proof. seek_k ($18(a+1)+k+3$) reaches the hole in row $a+1$, right^{16} (17) the cell (a,k), which is set to 0 (1), and home_k from row a takes $18a+k+4$. □

The coefficients of the field are read from the saved request, one framed number at a time, with the frame reader of Section I.7. By Lemma I.44 the saved request lays out the frames $f_0, f_1, \dots, f_F$ of the request from row 0 (Definition I.43); so every frame f_k, followed by the later frames, is laid out from its first row s_k, and we call s_k the *row of* f_k. The program skipf of Lemma I.46 moves the counter on track 5 from the row of a frame g to the row of the next one, in $36a + 70|g| + 16$ units if track 5 carried a, and fetch of Lemma I.45 copies the frame whose row track 5 carries onto tracks 6 and 7 from the row o that track 8 carries, within $1000(a+o+|g|+1)^2$ units. Both write only tracks 5 to 8 and 10. The test probe used there writes into $(a,10)$ whether the frame goes on at row a.

Lemma I.128 (Reading a natural number). *Let* $\mathsf{nat} = \mathsf{fetch}; \mathsf{reset}_8; \mathsf{toUnary}$*, with* $\mathsf{toUnary}$ *of Lemma I.31. Let* τ *have origin marks up to* r*, let track* 5 *carry* s*, track* 8 *the counter* 0*, and track* 9 *hold* 0 *in every row, and let the binary digits* $\beta(v)$ *of a natural number* v *(as the request writes them) be the frame whose row is* s*, with* $s + |\beta(v)| \le r$ *and* $v+1 \le r$*. Then* nat *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *within* $10\,000(r+1)^2$ *units, where track* 8 *of* τ' *carries* v*, track* 9 *holds the value* v*, track* 5 *carries* $s+|\beta(v)|$*, only tracks* 5 *to* 10 *have changed, and every cell* $x \ge -1$ *and every cell of a row beyond* r *is kept. With* $\mathsf{clear}_{8,9}$ *in front, the same holds without the hypotheses on tracks* 8 *and* 9*, within* $11\,000(r+1)^2$ *units, provided track* 0 *carries the counter* r *and every cell of a row beyond* r *holds* 0*.*

Proof. By Lemma I.45 with $o = 0$, fetch puts $\beta(v)$ on the binary stream within $1000(s + |\beta(v)| + 1)^2 \le 1000(r+1)^2$ units, and track 8 then carries $|\beta(v)|$. reset_8 ($36|\beta(v)| + 48$ units, Lemma I.14(c)) makes it carry 0 again. By Lemma I.31 with the counter 0, $\mathsf{toUnary}$ takes at most $1000(|\beta(v)| + v + 1)^2 \le 4000(r+1)^2$ units and leaves the counter and the value v on tracks 8 and 9. The sum is at most $10\,000(r+1)^2$. The clears of tracks 8 and 9 take $(55r+24) + (57r+26) + 1 \le 1000(r+1)^2$ units. □

I.11.3 Snapshots of the rows

The coefficient stage needs more registers than the sixteen tracks offer while the stored vector, the step counter and the saved request must survive it. It therefore copies every cell of the rows in use into the positive half of the tape, which is blank between stages, uses the tracks freely, and restores them from that copy at the end. The same mechanism, nested, gives subroutines with call frames.

Definition I.129 (The archive of the rows)**.** Let σ be a tape and r a row count. The *archive* $A_r(\sigma)$ is the word of $16(r+1)$ bits that lists, for $j = 0, 1, \dots, r$ in turn, the bit $[j < r]$ followed by $\sigma(j,1), \dots, \sigma(j,15)$. Its entry $16j + k$ records the cell (j,k), $k \geq 1$, and the entry $16j$ the header of row j in a workspace of r rows.

Lemma I.130 (Taking a snapshot)**.** *Let*

$$\begin{aligned}
\mathsf{root} &= \mathsf{left}; \mathsf{home}_1, \qquad \mathsf{push}(b) = \mathsf{root}; \mathsf{right}^3; \mathsf{appP}(b); \mathsf{rewP}; \mathsf{left}^1; \mathsf{seek}_0, \\
\mathsf{Archive} &= \mathsf{left}^1; \mathsf{loop}(\mathsf{write}\,0; \mathsf{crow}(1); \mathsf{write}\,1; \mathsf{left}^{16}); \mathsf{crow}(0); \mathsf{root}, \\
\mathsf{crow}(b) &= \mathsf{push}(b); \mathsf{pk}_1; \mathsf{pk}_2; \cdots; \mathsf{pk}_{15}, \\
\mathsf{pk}_k &= \mathsf{left}^k; \mathsf{branch}\big(\mathsf{right}^k; \mathsf{push}(0),\ \mathsf{right}^k; \mathsf{push}(1)\big).
\end{aligned}$$

Let τ have origin marks up to r, $\tau(j,0) = 1$ for $j < r$, $\tau(r,0) = 0$, $\tau(0) = 0$, and let τ hold a packed list w at 2. Then Archive *runs from $(-1,\tau)$ to $(-1,\tau')$ within $10^6(|w| + r + 1)^2$ units, where τ' holds the packed list $wA_r(\tau)$ at 2 and differs from τ only in the cells x with $2 + 2|w| \leq x \leq 2 + 2(|w| + 16(r+1))$.*

Proof. The program keeps the header of the row it is copying at 0, so that seek_0 finds that row again after each trip to the positive half. Consider a configuration at cell$(j,0)$ in which the header of row j holds 0, the headers of rows $q < j$ hold 1, and the tape holds a packed list v at 2. Then $\mathsf{push}(b)$ returns to the root along the origin marks ($18j + 6$ units), steps to cell 2 (4), appends b ($3|v| + 6$, Lemma I.123), rewinds to the blank cell 0 ($4|v| + 8$), steps to the root (2) and finds the row again ($18j + 3$, Lemma I.11): at most $100(j + |v| + 1)$ units, and the list is now vb. The program pk_k reads (j,k) and pushes it, so $\mathsf{crow}(b)$ appends $b, \tau(j,1), \dots, \tau(j,15)$ within $16 \cdot 100(j + |v| + 17) + 15 \cdot 33 \leq 2000(j + |v| + 17)$ units. A pass of the loop at the header $(j,0) = 1$, $j < r$, sets it to 0, copies the row with the true header bit 1, restores the header and moves to row $j + 1$; at the sentinel $(r,0) = 0$ the loop stops and $\mathsf{crow}(0)$ copies row r with its header bit 0. So the list becomes $wA_r(\tau)$. At row j the list has length $|w| + 16j \leq |w| + 16r$, so each row costs at most $2000(17r + |w| + 17) + 20 \leq 34\,020(|w| + r + 1)$ units, and with left^1, the final test and root the total is at most $(r+1) \cdot 34\,020(|w| + r + 1) + 18r + 9 \leq 10^6(|w| + r + 1)^2$. Every header is restored, and only the cells of the new entries and the end cell of the list are written. $\square$

Lemma I.131 (Restoring from a snapshot). *For a set $\mathcal{S}$ of tracks with $1 \notin \mathcal{S}$ put*

$$\mathsf{pop}_k^{\mathcal{S}} = \mathsf{root}; \mathsf{right}^3; \mathsf{scanP}; \mathsf{left}^1; \mathsf{branch}\big(\mathsf{pa}_k(0), \mathsf{pa}_k(1)\big),$$
$$\mathsf{pa}_k(b) = \mathsf{write}\,0; \mathsf{left}; \mathsf{write}\,0; \mathsf{rewP}; \mathsf{left}^1; \mathsf{seek}_0; \mathsf{left}^k; W_k(b); \mathsf{right}^k,$$

where $W_k(b)$ is $\mathsf{write}\,b$ *if $k = 0$ or $k \in \mathcal{S}$ and* skip *otherwise. Let* $\mathsf{rrow}^{\mathcal{S}}$ *be* $\mathsf{pop}_{15}^{\mathcal{S}}; \cdots; \mathsf{pop}_1^{\mathcal{S}}; \mathsf{pop}_0^{\mathcal{S}}$ *and*

$$\mathsf{Restore}_{\mathcal{S}} = \mathsf{seek}_0; \mathsf{rrow}^{\mathcal{S}}; \mathsf{left}; \mathsf{loop}(\mathsf{right}^{17}; \mathsf{write}\,0; \mathsf{rrow}^{\mathcal{S}}; \mathsf{left}); \mathsf{right}^2.$$

Let τ have origin marks up to r, $\tau(j,0) = 1$ for $j < r$, $\tau(r,0) = 0$, $\tau(0) = 0$, and let τ hold the packed list $wA_r(\sigma)$ at 2 for some tape σ. Then $\mathsf{Restore}_{\mathcal{S}}$ *runs from $(-1,\tau)$ to $(-1,\tau')$ within $2 \cdot 10^6(|w| + r + 1)^2$ units, where $\tau'(j,k) = \sigma(j,k)$ for $j \le r$ and $k \in \mathcal{S}$, the cells x with $2 + 2|w| \le x < 2 + 2(|w| + 16(r+1))$ hold 0, and τ' agrees with τ on every other cell. In particular τ' holds the packed list w at 2. For the set $\mathcal{B} = \{2,5,6,7,8,9,10,14\}$ of the tracks the stage borrows we write* $\mathsf{Restore} = \mathsf{Restore}_{\mathcal{B}}$*; it takes at most $10^6(|w| + r + 1)^2$ units.*

Proof. $\mathsf{pop}_k^{\mathcal{S}}$, started at cell$(j,0)$ with the header of row j at 0 and the list v at 2, goes to the end of the list, reads the value b of its last entry, erases that entry, comes back to row j and writes b into (j,k) if $k = 0$ or $k \in \mathcal{S}$; as in the previous proof this costs at most $150(j + |v| + 1)$ units. The entries of row j are the last sixteen of the list, track 15 last, so $\mathsf{rrow}^{\mathcal{S}}$ at row j restores the selected cells of that row and finally its header, from the entry $16j$. The first seek_0 finds the sentinel row r, whose header is 0, and the archive ends with row r. After a row j is restored, left reads the origin mark $(j,1)$: it is 1 for $j \ge 1$, and then the pass moves to $(j-1,0)$, sets that header to 0 so that seek_0 finds the row, restores the row and steps to $(j-1,1)$; at row 0 the mark is 0 and right^2 returns to the root. The rows are thus restored from r down to 0, sixteen pops each, with lists of length at most $|w| + 16(r+1)$: at most $(r+1)(16 \cdot 150(17r + |w| + 18) + 40) + 18r + 6$ units, which is less than $10^6(|w| + r + 1)^2$; the bound $2 \cdot 10^6(|w| + r + 1)^2$ covers every $\mathcal{S}$. Erased entries hold 0 in both cells. □

Lemma I.132 (Call frames). *Let $\mathcal{S}$ be a set of tracks, $1 \notin \mathcal{S}$, $0 \notin \mathcal{S}$, and* $\mathsf{Call}_{\mathcal{S}}(P) = \mathsf{Archive}; P; \mathsf{Restore}_{\mathcal{S}}$*. Let τ be as in Lemma I.130, and let τ_A be the tape it produces. Suppose P runs from $(-1,\tau_A)$ in t_P units to some $(-1,\sigma)$ such that σ has origin marks up to r, $\sigma(j,0) = 1$ for $j < r$, $\sigma(r,0) = 0$, $\sigma(0) = 0$, and σ holds the list $wA_r(\tau)$ at 2. Then* $\mathsf{Call}_{\mathcal{S}}(P)$ *runs from $(-1,\tau)$ to $(-1,\tau')$ within $t_P + 3 \cdot 10^6(|w| + r + 1)^2$ units, where τ' agrees with τ on the cells (j,k), $j \le r$, $k \in \mathcal{S}$, holds 0 in the cells x with $2 + 2|w| \le x < 2 + 2(|w| + 16(r+1))$, and agrees with σ on every other cell.*

Proof. Lemmas I.130 and I.131 with the archived tape τ; the costs add. □

The *registers* of the stage are the tracks $\mathcal{R} = \{2,5,6,7,8,9,10,11,12,14\}$. For a set O of output tracks we write $\mathsf{frame}_O(P) = \mathsf{Call}_{\mathcal{R}\setminus O}(P)$: a subroutine P run in this frame may use every register as scratch, and afterwards every register except those in O is as before; the arithmetic area on track 13 and the outputs in O are as P left them.

The snapshot also serves as a record of the registers at the start of the stage, from which the stage reads back the order N and the width W when it needs them.

Lemma I.133 (Reading a counter from the snapshot)**.** *Let $1 \le u < 16$, $2 \le k < 16$, and put*

$$\mathsf{arch}_{k\leftarrow u} = \mathsf{right}^{3};\mathsf{write}\,0;\mathsf{right}^{1+2u};\mathsf{loop}(\mathsf{abody});\mathsf{left}^{1+2u};\mathsf{write}\,1;\mathsf{rewP};\mathsf{left}^{1},$$
$$\mathsf{abody} = \mathsf{left}^{1+2u};\mathsf{write}\,1;\mathsf{right}^{32};\mathsf{write}\,0;\mathsf{rewP};\mathsf{left}^{1};\mathsf{inc}_{k};\mathsf{right}^{3};\mathsf{scanP};\mathsf{right}^{1+2u}.$$

Let τ hold the packed list $A_r(\sigma)$ at 2, with $\tau(0) = 0$ and origin marks up to r, let track u of σ carry a counter $n < r$, and let track k of τ carry 0. Then $\mathsf{arch}_{k\leftarrow u}$ runs from $(-1,\tau)$ to $(-1,\tau')$ within $10^4(r+1)^2$ units, where track k of τ' carries n, and τ' agrees with τ on every cell $x \ge 0$ and on every cell $x < 0$ that is not on track k.

Proof. The entry of (q,u) in the archive has its value in cell $2+2(16q+u)+1 = (2+32q)+(1+2u)$. The program marks its place by setting the presence cell $2+32q$ of the header entry of row q to 0 (a bookmark). The first three parts set the bookmark for $q = 0$ and move to the value of $(0,u)$. A pass of the loop at the value of $(q,u) = 1$, that is $q < n$, clears the bookmark, sets the bookmark of row $q+1$ (which exists, since $q+1 \le n < r$), rewinds over the $16(q+1)$ entries before it to the blank cell 0 (Lemma I.123(d)), steps to the root, increments track k ($36q+2k+44$ units, Lemma I.14), and scans from cell 2 to the new bookmark and then to the value of $(q+1,u)$. With the rewind ($64(q+1)+4$ units) and the scan ($48(q+1)+1$) a pass costs at most $148q+2k+4u+207 \le 1000(r+1)$ units. At the value of $(n,u) = 0$, the hole, the loop stops, and the last four parts clear the bookmark and return to the root ($64n+2u+10$ units). By Lemma I.6 the total is at most $1000n(r+1)+64n+4u+50 \le 10^4(r+1)^2$. The positive half is as before, and of the scratch area only track k was written. □

With $\mathsf{Width} = \mathsf{zero}_9;\mathsf{arch}_{9\leftarrow 9}$ the width W returns to track 9 from the snapshot, within $25 + 10^4(r+1)^2 \le 11\,000(r+1)^2$ units; likewise $\mathsf{Order} = \mathsf{zero}_8;\mathsf{arch}_{8\leftarrow 8}$ brings back N.

I.11.4 The arithmetic area

The arithmetic of one entry is done in twelve-cell rows (Section I.5) laid along track 13, between the end of the buffer and the sentinel row, and read upwards, towards row 0.

Definition I.134 (The bank)**.** Let τ be a tape and a a row. The *bank view* of τ at a is the tape $\beta(x) = \tau(\mathrm{cell}(a,13)+16x)$, $x \in \mathbb{Z}$; for $x \le a$ this is the cell $(a-x,13)$. The *bank of width H* at a is the block of twelve-cell rows at 1 of β; it is *laid out* if it has H marked rows and is closed on the left (Definition I.20). The cell $\beta(0)$, that is $(a,13)$, is the *sign cell.* Column c of the bank *reads* the integer z if the word $\beta(1+12q+c)$, $q < H$, has a value congruent to z modulo 2^H. The *support* of the bank is the set of cells x with $-11 \le x \le 12H+24$ of β, that is, the cells of track 13 in the rows $a-12H-24,\dots,a+11$.

Reading modulo 2^H is what the ripple adder and the multiplier compute (Lemma I.21, Corollary I.22, Lemma I.27), so sums and products of columns are read correctly whatever the signs. Only two readings need more: the common denominator, which is read as a natural number, and the two final operands, which are read in two's complement; Proposition I.32(b) puts both below 2^{H-1} for $H = 3W$. The columns have fixed roles: 1 holds the operand just loaded, 3 a second factor, 7 the common denominator, 9 the weight of the current coefficient and 11 the numerator; 2, 4, 5, 6, 8 and 10 are the working columns of the routines of Section I.5.

Definition I.135 (The working state)**.** Fix r, $l < N$, $j < d$ and Z as in Theorem I.119, let τ_0 be the tape with which Append starts, $H = 3W$, $a = r - 12$ and $\ell = ((l+1)d + j)(W+2)$. A tape τ is in the *working state with indices* (q, p) if

(i) $12H + 25 \le a$ and $\ell \le a - 12H - 25$;

(ii) track 0 carries the counter r, there are origin marks up to r, and the output buffer holds the table prefix (l, j) of Z with capacity r;

(iii) the root holds 1, the cells 0 and 1 hold 0, τ holds the packed list $A_r(\tau_0)$ at 2, and every cell beyond it holds 0;

(iv) tracks 7, 10, 8, 14, 9 and 6 carry the counters l, j, p, q, W and H;

(v) the bank of width H at a is laid out;

(vi) τ agrees with τ_0 on tracks 0, 1, 3, 4 and 15, and every cell of a row $j' > r$ holds 0.

The index q counts the terms of a sum, p the pairs of a Cauchy sum.

By (vi) the saved request of τ_0 is kept, and by (i) and (ii) the buffer lies in rows below $a - 12H - 25$, outside the support of the bank. The first condition holds for every tape the stage meets.

Lemma I.136 (Room for the bank)**.** *If $r \ge 100(d+1)(N+1)(W+2)^2$ and $W \ge 4$, then for $l < N$ and $j \le d$ we have $((l+1)d + j)(W+2) + 12(3W+4) \le r$; hence $12H + 25 \le r - 12$ and $\ell < r - 12 - (12H + 25)$.*

Proof. Put $F = (d+1)(N+1)$. Then $(l+1)d + j \le (N+1)d + d \le F$, so the prefix has at most $F(W+2) \le F(W+2)^2$ bits, and $12(3W+4) \le 10(W+2)^2 \le 10F(W+2)^2$. The sum is at most $11F(W+2)^2 \le r$. Since $12(3W+4) = 12H + 48$, the remaining claims follow. Also $W = 4\mathscr{B}(N) + 4 \ge 4$. □

Lemma I.137 (Calling the bank)**.** *Let*

$$\mathsf{bcall}(P) = \mathsf{sent}; \mathsf{right}^{192}; P^{\uparrow 16}; \mathsf{home}_{13}, \qquad \mathsf{bank}(P) = \mathsf{bcall}(\mathsf{right}; P; \mathsf{left}).$$

Let τ have origin marks up to r, $\tau(j', 0) = 1$ for $j' < r$, $\tau(r, 0) = 0$, and let $a = r - 12$ and $12H + 25 \le a$. Let β be the bank view of τ at a, and suppose P runs on the tape β from $(0, \beta)$ in t units to (f, β'), where $0 \le f \le a$ and β' agrees with β outside the support. Then $\mathsf{bcall}(P)$ runs from $(-1, \tau)$ to $(-1, \tau')$ within $17t + 36r + 300$ units, where the bank view of τ' at a is β' and τ' agrees with τ on every cell that is not in the support. If P runs from $(1, \beta)$ to $(1, \beta')$ in t units, $\mathsf{bank}(P)$ does the same within $17t + 36r + 334$ units.

Proof. sent $(18r+17)$ reaches $\mathrm{cell}(r, 13)$ and right^{192} (193) the cell $\mathrm{cell}(r, 13) + 192 = \mathrm{cell}(a, 13)$. With $o = \mathrm{cell}(a, 13)$ and stride 16 we have $o + 16x = \mathrm{cell}(a - x, 13)$, so the configuration carries $(0, \beta)$ in the sense of Lemma I.8, and $P^{\uparrow 16}$ runs within $17t$ units to a configuration that carries (f, β') and has changed no cell off track 13; its head is on $\mathrm{cell}(a - f, 13)$, in row $a - f \le a$. home_{13} then takes $18(a - f) + 17 \le 18r$ units. For $\mathsf{bank}(P)$ the two extra moves add 2 units to t. □

Lemma I.138 (Bank routines)**.** *Let the twelve-cell rows at h have $H \geq 1$ marked rows and be closed on the left, and let $1 \leq s, t, c < 12$. Put*

$$\begin{aligned} \mathsf{bcp}_{s\to t} &= \mathsf{cp}_{s\to t};\mathsf{back}, \qquad \mathsf{bz}_t = \mathsf{cl}_t;\mathsf{back}, \qquad \mathsf{bmul} = \mathsf{mul};\mathsf{back},\\ \mathsf{bone}_t &= \mathsf{bz}_t;\mathsf{right}^t;\mathsf{write}\,1;\mathsf{left}^t,\\ \mathsf{bsign}_c &= \mathsf{loop}(\mathsf{right}^{12});\mathsf{left}^{12-c};\mathsf{branch}(E_0,E_1),\\ E_b &= \mathsf{right}^{12-c};\mathsf{back};\mathsf{left};\mathsf{write}\,b;\mathsf{right}, \end{aligned}$$

and, with $\Sigma = \mathsf{z}_6;\mathsf{add}^{\uparrow 2};\mathsf{back};\mathsf{cp}_{8\to 4};\mathsf{back}$ of Lemma I.27,

$$\begin{aligned} \mathsf{bacc} &= [\mathsf{bcp}_{4\to 2},\mathsf{bcp}_{11\to 4},\Sigma,\mathsf{bcp}_{4\to 11}], \qquad \mathsf{bdbl}_c = [\mathsf{bcp}_{c\to 2},\mathsf{bcp}_{c\to 4},\Sigma,\mathsf{bcp}_{4\to c}],\\ \mathsf{binc} &= [\mathsf{bcp}_{1\to 2},\mathsf{bcp}_{9\to 4},\Sigma,\mathsf{bcp}_{4\to 9}]. \end{aligned}$$

Each runs from h to h, keeps the presence marks and writes only the columns named below, the cell $h-1$ for bsign*, and cells of the end row.*

(a) $\mathsf{bcp}_{s\to t}$ *copies column s into column t in $(2s+2t+20)H+1+(14H+27)$ units;* bz_t *clears column t in $(2t+17)H+1+(14H+27)$ units;* bone_t *makes column t read 1 within $(2t+31)H+2t+31$ units.*

(b) bsign_c *copies the last bit of column c, its two's complement sign, into the cell $h-1$, within $28H+80$ units.*

(c) If columns 1 and 3 read x and y, bmul *makes column 4 read xy within $441(H+1)^3$ units, writing columns 2, 4, 5, 6, 8 and 10.*

(d) Σ makes column 4 read the sum of what columns 2 and 4 read, within $132H+73$ units, writing columns 4, 6 and 8. Hence bacc *adds column 4 to column 11 within $400H+160$ units,* bdbl_c *doubles column c ($c \neq 2, 4$) and* binc *adds column 1 to column 9, each within $400H+161$ units.*

Proof. (a) Lemma I.25(a) and (b); bone_t adds $(t+1)+1+(t+1)$ units and sets the least significant bit. (b) The loop walks to the end row ($14H+1$ units), left^{12-c} reaches column c of row $H-1$ and the branch reads it; E_b returns to the end row, walks back to row 0 ($14H+27$) and writes b into $h-1$: at most $28H+58$ units. (c) Lemma I.27 gives column 4 the H lowest bits of $\mathrm{val}(x)\,\mathrm{val}(y)$ within $400(H+1)^2$ units, and back adds $14H+27$; the low H bits of a product are determined by the factors modulo 2^H, and $400(H+1)^2+14H+27 \leq 441(H+1)^3$. (d) The count of Σ is part of the proof of Lemma I.27; by Lemma I.21 the sum is correct modulo 2^H. The three copies of each composite cost at most $(2\cdot 11+2\cdot 4+34)H+28$ each, and the sums are as stated. □

Lemma I.139 (Laying out the bank). *Put* $\mathsf{BankInit} = \mathsf{ext}; \mathsf{lay}$, *where*

$$\mathsf{ext} = \mathsf{zero}_6; \mathsf{acc}^{14}_{9\to 6}; \mathsf{acc}^{14}_{9\to 6}; \mathsf{acc}^{14}_{9\to 6}; \mathsf{reset}_{14},$$
$$\mathsf{lay} = \mathsf{dup}_{0\to 2}; \mathsf{pred}_2; \mathsf{zc}^{12}; \mathsf{rep}_{14,6}(\mathsf{wr}(1); \mathsf{pred}_2; \mathsf{zc}^{11}; \mathsf{inc}_{14}); \mathsf{zc}^{12},$$

$\mathsf{wr}(b) = \mathsf{seek}_2; \mathsf{left}^{11}; \mathsf{write}\, b; \mathsf{home}_{13}$ *writes* b *into* $(a', 13)$ *when track* 2 *carries* a', *and* zc^n *is* n *copies of* $\mathsf{wr}(0); \mathsf{pred}_2$. *Let track* 0 *of* τ *carry* r, *track* 9 *carry* W, *let the output buffer hold a word* v *with capacity* r, $36W + 37 \le r$ *and* $|v| \le r - 25 - 36W$. *Then* $\mathsf{BankInit}$ *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $12\,000(r+1)^2$ *units, where the bank of width* $H = 3W$ *at* $a = r - 12$ *is laid out in* τ' *with its sign cell and all its columns* 0, *tracks* 6 *and* 14 *carry* H, *track* 2 *carries* $r - 25 - 12H$, *track* 9 *still carries* W, *the buffer still holds* v, *and only tracks* 2, 6, 14 *and the rows* $r - 25 - 12H < j' < r$ *of track* 13 *have changed.*

Proof. ext adds the counter W three times to the empty track 6, using track 14 as scratch (Lemma I.15(b), at most $500(3W+1)^2$ units each), and resets track 14: at most $2000(r+1)^2$ units. In lay, the copy of the headers puts the counter r on track 2 ($46r + 20$ units), and one decrement makes it $r - 1$, the row of the cell $x = -11$ of the bank view. The program $\mathsf{wr}(b)$ takes $36a' + 35$ units and pred_2 at most $100(r+1) + 1$, so zc^{12} clears the cells $x = -11, \dots, 0$ within $2400(r+1) + 1$ units. A pass of the repetition, with track 14 carrying $q < H$ and track 2 carrying the row of the cell $x = 1 + 12q$, writes the presence mark of the bank row q, clears its eleven columns and advances track 14, within $3000(r+1)$ units; by Lemma I.126(c) with the guard track 6 carrying H, the H passes take at most $(H+1) \cdot 3200(r+1)$ units. The last zc^{12} clears the end row. Since $H \le r$, lay takes at most $10\,000(r+1)^2$ units. The written cells of track 13 are those of the rows $r - 1$ down to $r - 24 - 12H$, all above the buffer. □

I.11.5 Reading the indices and the operands

The stage receives no numbers from outside: it reads l and j off the buffer, every coefficient off the saved request and every entry of the table off the buffer. This subsection gives the readers.

Dividing in unary. The order and the coordinate come from a division of the number of words in the buffer by d. The divider below reads its dividend on track 4, which in this stage belongs to the saved request; it is therefore run with tracks 4 and 7 exchanged, by a transformation of the program like the dilation of Lemma I.8.

Lemma I.140 (Exchanging two tracks). *Let* $\tau^{\times}$ *be* τ *with tracks* 4 *and* 7 *exchanged:* $\tau^{\times}(j,4) = \tau(j,7)$, $\tau^{\times}(j,7) = \tau(j,4)$, *and* $\tau^{\times} = \tau$ *elsewhere. Let* P *be a program in which both parts of every branch end on the same track and every loop body ends on the track on which it starts, so that the track of the head before each primitive is determined by the program text when* P *starts at the root. Let* $P^{\times}$ *replace each move of* P *by the moves that the exchange requires: a move from track* k *to track* $k \pm 1$ *becomes the move, of at most four cells, between the images of the two cells. If* P *runs from* $(-1, \tau^{\times})$ *to* $(-1, \sigma)$ *in* t *units, then* $P^{\times}$ *runs from* $(-1, \tau)$ *to* $(-1, \sigma^{\times})$ *within* $5t$ *units.*

Proof. The exchange maps the cell of (j,k) to that of (j,k'), where k' is k with 4 and 7 swapped; it fixes the root and every cell off tracks 4 and 7, and it is its own inverse. Induction on the execution of P: a test or a write at the image cell sees or changes the image of the cell that P

sees or changes, so the same branch is taken, and a move of P becomes at most four moves and a skip, at most 5 units. The static track is what makes the replacement of a move a fixed piece of program text. □

Lemma I.141 (Unary division)**.** *Let*

$$\mathsf{udiv} = [\mathsf{val}_2,\ \mathsf{loop}(D),\ \mathsf{home}_4],$$
$$D = [\mathsf{home}_4, \mathsf{inc}_2, \mathsf{inc}_5, \mathsf{pres}_5, \mathsf{branch}(\mathsf{home}_6; \mathsf{reset}_5; \mathsf{mark}_8,\ \mathsf{home}_6), \mathsf{val}_2].$$

Let track 4 *hold the value* u*, track* 6 *the value* $b \geq 1$*, tracks* 2, 5 *and* 8 *carry* 0*, track* 9 *hold the value* 0*, and let* τ *have origin marks up to* R *with* $u + 1 \leq R$ *and* $b \leq R$*. Then* udiv *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $1000(u + b + 1)^2$ *units; in* τ' *track* 8 *carries* $\lfloor u/b \rfloor$*, track* 9 *holds the value* $\lfloor u/b \rfloor$*, track* 5 *carries* $u \bmod b$ *and track* 2 *carries* u*, and only tracks* 2, 5, 8 *and* 9 *have changed. The program* $\mathsf{udiv}^{\times}$ *does the same with the dividend on track* 7 *instead of* 4*, within* $5000(u + b + 1)^2$ *units.*

Proof. Let $I_{j'}$ be the configuration at $(j', 4)$ in which tracks 2, 5 and 8 carry j', ρ and q with $j' = qb + \rho$ and $\rho < b$, and track 9 holds the value q. val_2 reaches I_0. A pass from $I_{j'}$, $j' < u$, reads the mark $(j', 4) = 1$, returns, advances tracks 2 and 5 and reads $(\rho + 1, 6)$. If $\rho + 1 < b$ this is 1 and the pass returns; if $\rho + 1 = b$ it is the hole of track 6, and the pass resets track 5 to 0 and appends a mark to the quotient (Lemma I.14(c) and (e)). Then val_2 moves to $(j' + 1, 4)$. By Lemmas I.11 and I.14 a pass costs at most $72j' + 108\rho + 36q + 334 \leq 108(u + b) + 334$ units, and I_u scans the hole of track 4. Lemma I.6 and the two moves to and from the root give at most $u(108(u + b) + 335) + 18u + 40 \leq 1000(u + b + 1)^2$ units. The second claim is Lemma I.140: every loop body and branch of udiv ends on the track it starts from. □

The order and the coordinate. The buffer holding the prefix (l, j) has $\ell = ((l+1)d+j)(W+2)$ bits, so the number of its words is $w = (l + 1)d + j$, with $j < d$, and $l + 1 = \lfloor w/d \rfloor$, $j = w \bmod d$. The dimension d is the first frame of the saved request.

Lemma I.142 (Recovering l and j)**.** *Let*

$$\mathsf{Dim} = \mathsf{clear}_{5,8,9}; \mathsf{nat}, \qquad \mathsf{skipW} = \mathsf{zero}_6; \mathsf{rep}_{6,9}(\mathsf{inc}_2; \mathsf{inc}_6); \mathsf{inc}_2; \mathsf{inc}_6; \mathsf{inc}_2; \mathsf{inc}_6,$$
$$\mathsf{Words} = \mathsf{zero}_2; \mathsf{zero}_5; \mathsf{peek}_{2\to 11}; \mathsf{loop}(\mathsf{home}_{11}; \mathsf{skipW}; \mathsf{inc}_5; \mathsf{peek}_{2\to 11}); \mathsf{home}_{11},$$
$$\mathsf{Prep} = \mathsf{clear}_{6,7,9}; \mathsf{copy}_{8\to 6}; \mathsf{copy}_{5\to 7}; \mathsf{zero}_2; \mathsf{zero}_5; \mathsf{zero}_8,$$
$$\mathsf{Indices} = \mathsf{dup}_{9\to 2}; \mathsf{Dim}; \mathsf{dup}_{2\to 9}; \mathsf{Words}; \mathsf{Prep}; \mathsf{udiv}^{\times}.$$

Let τ *hold the saved request (Definition I.41), have origin marks up to* r*, track* 0 *carrying the counter* r *and every cell of a row beyond* r *equal to* 0*, let its buffer hold the table prefix* (l, j) *of* Z *with capacity* r*,* $l < N$*,* $j < d$*, let track* 9 *carry* W*, and let the reserve of Definition I.101 hold. Then* $\mathsf{Indices}$ *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $40\,000(r + 1)^3$ *units, where track* 8 *of* τ' *carries* $l + 1$*, track* 5 *carries* j*, track* 6 *holds the value* d*, and* τ' *differs from* τ *only on tracks* 2 *and* 5 *to* 10 *of the scratch area.*

Proof. $\mathsf{dup}_{9\to 2}$ saves W on track 2; each dup costs at most $102(r + 1)$ units (Lemma I.126(b)). Dim reads the first frame of the saved request, the binary digits of d at row 0, onto track 8

(Lemma I.128, within $11\,000(r+1)^2$ units), and $\mathsf{dup}_{2\to 9}$ puts W back on track 9. In Words, skipW advances the counter on track 2 by $W+2$: the repetition, guarded by track 9 which carries W, advances tracks 2 and 6 W times, and two more steps follow; it costs at most $500(r+1)^2$ units. The loop of Words runs while the cell of track 11 in the row given by track 2 holds 1, that is, while the counter p on track 2 is less than ℓ; since ℓ is a multiple of $W+2$, it stops with $p=\ell$ after w passes, each of which also advances track 5. So track 5 carries w, within $2000(r+1)^3$ units. Prep puts the values d and w on tracks 6 and 7 and empties the registers of the divider, within $1000(r+1)^2$ units, and $\mathsf{udiv}^{\times}$ (Lemma I.141) leaves $\lfloor w/d\rfloor = l+1$ on track 8 and $w \bmod d = j$ on track 5, within $5000(w+d+1)^2 \le 20\,000(r+1)^2$ units, since $w+1\le r$ and $d\le r$ by the reserve. The total is at most $14\,000(r+1)^3+1000(r+1)^2+20\,000(r+1)^2 \le 40\,000(r+1)^3$. □

At the end of the stage the same computation yields the two flags. For tracks p, b and f let

$$\mathsf{cmp}_{p,b,f} = \mathsf{peek}_{p\to b};\mathsf{branch}(\mathsf{cf}(0),\mathsf{cf}(1)),$$
$$\mathsf{cf}(x) = \mathsf{home}_b;\mathsf{sent};\mathsf{shift}_{13\to f};\mathsf{write}\,x;\mathsf{home}_f.$$

If track p carries v and track b carries $\beta \ge v$, the cell (v,b) holds $[v<\beta]$, so $\mathsf{cmp}_{p,b,f}$ writes $[v<\beta]$ into the sentinel cell (r,f), within $200(r+1)$ units, and writes nothing else.

Lemma I.143 (The flags)**.** *Let*

$$\mathsf{Flags} = \mathsf{dup}_{8\to 14};\mathsf{Indices};\mathsf{dup}_{14\to 9};\mathsf{inc}_5;\mathsf{cmp}_{5,6,7};\mathsf{pred}_5;\mathsf{cmp}_{8,9,6}.$$

Under the hypotheses of Lemma I.142, if moreover track 8 *carries N, then* Flags *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *within* $42\,000(r+1)^3$ *units, where* $\tau'(r,7)=[j+1<d]$, $\tau'(r,6)=[l+1<N]$, *tracks* 8, 5, 6 *and* 9 *carry* $l+1$, *j, d and N, and only tracks* 2, 5 *to* 10 *and* 14 *have changed.*

Proof. The order N is kept on track 14 while $\mathsf{Indices}$ runs and then moved to track 9. Then inc_5 makes track 5 carry $j+1\le d$, and $\mathsf{cmp}_{5,6,7}$ writes $[j+1<d]$; pred_5 returns to j, and $\mathsf{cmp}_{8,9,6}$ compares $l+1\le N$ with N. The four last parts take at most $1000(r+1)$ units, the copies $204(r+1)$, and Lemma I.142 the rest. □

Where a coefficient is. By Definition 15.32 the saved request holds, after the frame of d: for each coordinate a *linear row*, the number of its A-terms followed by their indices; for each coordinate a *bilinear row*, the number of its B-terms followed by their index pairs; the payload length; and the payload, the numerator and the denominator of every coordinate of y_0, every c_i, every A-coefficient (coordinate by coordinate, term by term) and every B-coefficient, each a framed integer. So the rational with *ordinal* o in the payload is

$$o = \begin{cases} d+j & \text{for } c_j,\\ 2d+L_A(j)+q & \text{for the term } q \text{ of } A_j,\\ 2d+L_A+L_B(j)+q & \text{for the term } q \text{ of } B_j, \end{cases} \tag{I.1}$$

where $L_A(j)$ and $L_B(j)$ count the terms of A and of B of the coordinates before j and $L_A = L_A(d)$; its numerator and denominator are the frames number $2o$ and $2o+1$ of the payload.

Lemma I.144 (Addresses in the saved request)**.** *There are fixed programs* $\mathsf{coefA}(\kappa, e)$*, for the three kinds* κ *of coefficient (c, A, B) and* $e \in \{0, 1\}$*, and* $\mathsf{idxA}(\kappa, e)$ *for* $\kappa \in \{A, B\}$*, with the following property. Let* τ *hold the saved request, have origin marks up to r, track* 0 *carrying r, every cell of a row beyond r equal to* 0*, and the reserve; let track* 5 *carry* $j < d$ *and track* 14 *carry q, less than the number of terms of the kind* κ *of coordinate j (no condition on q for* $\kappa = c$*).*

(a) $\mathsf{coefA}(\kappa, e)$ *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $150\,000(r+1)^4$ *units, where track* 5 *of* τ' *carries a row* s' *which is the row of the frame of the numerator (*$e = 0$*) or the denominator (*$e = 1$*) of the selected coefficient, with* $s' + (\text{its length}) \le r$*.*

(b) $\mathsf{idxA}(\kappa, e)$ *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *within* $80\,000(r+1)^4$ *units, where track* 8 *of* τ' *carries the index of the term q: the index u of* (a, u) *for* $\kappa = A$*, and u (*$e = 0$*) or v (*$e = 1$*) of* (b, u, v) *for* $\kappa = B$*.*

(c) For $\kappa \in \{A, B\}$ *the program* $\mathsf{rowA}(\kappa)$ *below runs within* $60\,000(r+1)^4$ *units and leaves on track* 8 *the number of terms of the kind* κ *of coordinate j.*

In each case only tracks 2 *and* 5 *to* 12 *change, the saved request and every cell* $x \ge -1$ *and every cell below row* $r+1$ *are kept, and track* 14 *still carries q.*

Proof. All programs are built from nat, skipf, counter copies and additions and the following countdown. For a track p let

$$\mathsf{cp1}_p = \mathsf{peek}_{p\to 1}; \mathsf{branch}(\mathsf{shift}_{1\to 10}; \mathsf{write}\,0,\ \mathsf{shift}_{1\to 10}; \mathsf{write}\,1),$$
$$\mathsf{down}_p(P) = \mathsf{cp1}_p; \mathsf{loop}(\mathsf{home}_{10}; P; \mathsf{pred}_p; \mathsf{cp1}_p); \mathsf{home}_{10}.$$

With track p carrying n, $\mathsf{cp1}_p$ writes into $(n, 10)$ the origin mark $(n, 1) = [n \ge 1]$ and stops there, and $\mathsf{down}_p(P)$ runs P exactly n times, counting track p down to 0, within $(n+1)(b+400(r+1))$ units if each run of P; pred_p takes at most b (Lemma I.6 with the invariant that track p carries $n - i$ after i passes).

Write $\mathsf{natC} = \mathsf{clear}_{8,9}; \mathsf{nat}$; it reads the natural number at the row on track 5 onto track 8 and moves track 5 past its frame (Lemma I.128, $11\,000(r+1)^2$ units). Put

$$\mathsf{terms}_p(\alpha, x) = \mathsf{down}_p(\mathsf{skipf}^{\alpha}; [\mathsf{inc}_{11} \text{ if } x = 1]), \qquad \mathsf{srow}(\alpha, x) = \mathsf{natC}; \mathsf{terms}_8(\alpha, x),$$

where skipf^{α} is α copies of skipf: $\mathsf{terms}_p(\alpha, x)$ skips as many terms of arity α as track p counts, adding one to track 11 per term if $x = 1$, within $2000(r+1)^3$ units, and $\mathsf{srow}(\alpha, x)$ skips one whole row of the shape whose terms have α indices, within $13\,000(r+1)^3$ units (Lemma I.46 for each frame; at most r frames are skipped in all). Next,

$$\mathsf{rowA}(A) = \mathsf{pos}; \mathsf{down}_2(\mathsf{srow}(1, 1)); \mathsf{natC}, \qquad \mathsf{pos} = \mathsf{dup}_{5\to 2}; \mathsf{zero}_{11}; \mathsf{Dim}; \mathsf{dup}_{8\to 12},$$
$$\mathsf{rowA}(B) = \mathsf{pos}; \mathsf{down}_{12}(\mathsf{srow}(1, 1)); \mathsf{down}_2(\mathsf{srow}(2, 1)); \mathsf{natC}.$$

pos puts j on track 2, 0 on track 11, d on track 12 and moves track 5 past the frame of d. $\mathsf{rowA}(A)$ skips the linear rows of the coordinates before j, counting their terms on track 11, and reads the term count of the linear row of j; $\mathsf{rowA}(B)$ first skips all d linear rows, counting their terms, and then the bilinear rows before j. So track 11 carries $L_A(j)$, respectively $L_A + L_B(j)$,

track 8 the number of terms of row j, and track 5 the row of the first index frame of that row. The bounds of (c) follow from the parts: at most $12\,000(r+1)^2$ for pos and, for each countdown, at most $(r+1)(14\,000(r+1)^3 + 400(r+1)) \le 15\,000(r+1)^4$.

For the term q put $\mathsf{tail}(\alpha) = \mathsf{dup}_{14\to12};\mathsf{terms}_{12}(\alpha,1)$, which skips the first q terms of the row and adds q to track 11 (within $2200(r+1)^3$ units), and $\mathsf{termA}(\kappa) = \mathsf{rowA}(\kappa);\mathsf{tail}(\alpha_\kappa)$ with $\alpha_A = 1$, $\alpha_B = 2$ (within $65\,000(r+1)^4$). Then $\mathsf{idxA}(\kappa,e) = \mathsf{termA}(\kappa);[\mathsf{skipf}\text{ if }\kappa = B, e = 1];\mathsf{natC}$ reads the first or second index of the term onto track 8, within $65\,000(r+1)^4 + 1000(r+1)^2 + 11\,000(r+1)^2 \le 80\,000(r+1)^4$ units, which is (b).

For (a), the ordinal (I.1) is formed on track 11: for $\kappa = c$ by $\mathsf{dup}_{5\to11}$ and one addition of d, and for $\kappa = A, B$ by $\mathsf{termA}(\kappa)$ and two additions of d; the additions are Dim followed by one or two copies of $\mathsf{acc}^{12}_{8\to11}$, within $13\,000(r+1)^2$ units. Call this part $\mathsf{ordA}(\kappa)$; it takes at most $80\,000(r+1)^4$ units. Then

$$\mathsf{payA}(e) = \mathsf{pos}';\mathsf{down}_{12}(\mathsf{srow}(1,0));\mathsf{dup}_{2\to12};\mathsf{down}_{12}(\mathsf{srow}(2,0));\mathsf{skipf};\\ \mathsf{zero}_2;\mathsf{acc}^{12}_{11\to2};\mathsf{acc}^{12}_{11\to2};\mathsf{down}_2(\mathsf{skipf});[\mathsf{skipf}\text{ if }e=1],$$

with $\mathsf{pos}' = \mathsf{Dim};\mathsf{dup}_{8\to2};\mathsf{dup}_{8\to12}$, skips the d linear and the d bilinear rows and the payload length, then $2o$ integer frames, and for $e = 1$ one more: track 5 then carries the row of the numerator or the denominator of the rational o. Its parts cost at most $50\,000(r+1)^4$ (the position of the payload) and $3000(r+1)^3 + 1000(r+1)^2$ (the skips), so at most $60\,000(r+1)^4$, and $\mathsf{coefA}(\kappa,e) = \mathsf{ordA}(\kappa);\mathsf{payA}(e)$ takes at most $150\,000(r+1)^4$ units. Every routine used writes only the tracks named, and the frames read lie below row r by the reserve. □

Where an entry of the table is. The entry $r_{p,u}$ is word number $pd+u$ of the buffer, so its signed word starts at row $(pd+u)(W+2)$ of track 13. Let

$$\mathsf{adv} = \mathsf{inc}_5;\mathsf{inc}_{10}, \qquad \mathsf{lw} = \mathsf{zero}_{10};\mathsf{rep}_{10,9}(\mathsf{adv});\mathsf{adv};\mathsf{adv},\\ \mathsf{lws}_g = \mathsf{zero}_8;\mathsf{rep}_{8,g}(\mathsf{lw};\mathsf{inc}_8),\\ \mathsf{Look} = \mathsf{dup}_{5\to2};\mathsf{dup}_{8\to12};\mathsf{dup}_{9\to14};\mathsf{Dim};\mathsf{dup}_{8\to6};\mathsf{dup}_{14\to9};\\ \mathsf{zero}_5;\mathsf{zero}_7;\mathsf{rep}_{7,12}(\mathsf{lws}_6;\mathsf{inc}_7);\mathsf{lws}_2.$$

Lemma I.145 (Locating an entry)**.** *Let τ be as in Lemma I.144, with track 5 carrying $u < d$, track 8 carrying $p \le r$, track 9 carrying W, and $(pd+u)(W+2) \le r$. Then* Look *runs from $(-1,\tau)$ to $(-1,\tau')$ within $20\,000(r+1)^4$ units, where track 5 of τ' carries $(pd+u)(W+2)$; it writes only tracks 2 and 5 to 14 other than 11 and 13.*

Proof. The copies keep u, p and W while Dim reads d, which is moved to track 6. lw advances track 5 by $W+2$ (as skipW above, within $1000(r+1)^2$ units); lws_g advances it by $W+2$ as many times as track g counts, within $2000(r+1)^3$ units; and the outer repetition, guarded by track 12 carrying p, advances it by $d(W+2)$ p times, within $3000(r+1)^4$ units. The last lws_2 adds $u(W+2)$. With the copies and Dim the total is at most $20\,000(r+1)^4$. □

Loading an operand. An operand enters the bank as a magnitude in column 1 and a sign in the sign cell, and is then converted to two's complement. Track 2 serves as the write pointer

into the bank: $\mathsf{bptr} = \mathsf{dup}_{0\to 2};\mathsf{pred}_2^{12}$ makes it carry $r-12=a$, the row of the sign cell, within $1500(r+1)$ units. With $\mathsf{wr}(b)$ of Lemma I.139, put

$$\begin{aligned}\mathsf{ldS} &= \mathsf{cb}_4;\mathsf{inc}_5;\mathsf{pred}_2;\mathsf{sv};\mathsf{loop}(\mathsf{home}_7;\mathsf{wr}(1);\mathsf{pred}_2;\mathsf{cb}_4;\mathsf{inc}_5;\mathsf{pred}_2^{11};\mathsf{inc}_{10};\mathsf{sv});\\ &\qquad \mathsf{home}_7,\\ \mathsf{ldT} &= \mathsf{cb}_{13};\mathsf{inc}_5;\mathsf{pred}_2;\mathsf{rep}_{10,9}(\mathsf{mg});\mathsf{mg};\mathsf{wr}(0);\mathsf{peek}_{10\to 6};\\ &\qquad \mathsf{branch}(\mathsf{home}_6;\mathsf{wr}(0),\ \mathsf{home}_6;\mathsf{wr}(1)),\\ \mathsf{mg} &= \mathsf{wr}(1);\mathsf{pred}_2;\mathsf{cb}_{13};\mathsf{inc}_5;\mathsf{pred}_2^{11};\mathsf{inc}_{10},\end{aligned}$$

where $\mathsf{cb}_k = \mathsf{peek}_{5\to k};\mathsf{branch}(\mathsf{home}_k;\mathsf{wr}(0),\ \mathsf{home}_k;\mathsf{wr}(1))$ copies the cell of track k in the row given by track 5 into the cell of track 13 in the row given by track 2, and $\mathsf{sv} = \mathsf{seek}_5;\mathsf{right}^2;\mathsf{probe}'$, where probe' is the test probe of Lemma I.45 with its flag written into track 7 instead of 10 (left^4 and right^8 in place of left^7 and right^5). Finally let $\mathsf{sgn} = \mathsf{bcall}([\mathsf{right},\mathsf{bcp}_{9\to 5},\mathsf{left},\mathsf{neg}',\mathsf{back},\mathsf{bcp}_{5\to 9},\mathsf{bcp}_{8\to 1},\mathsf{left}])$, where neg' branches on the sign cell and either copies column 1 into column 8 or writes into column 8 the complement of column 1 plus one (through column 9 and the ripple adder).

Lemma I.146 (Loading operands)**.** *Let τ be in the working state with indices (q,p), and let*

$$\mathsf{loadS} = \mathsf{frame}_{\varnothing}([\mathsf{bank}(\mathsf{bz}_1),\mathsf{bptr},\mathsf{zero}_{10},\mathsf{ldS},\mathsf{sgn}]),$$
$$\mathsf{loadT} = \mathsf{frame}_{\varnothing}([\mathsf{bank}(\mathsf{bz}_1),\mathsf{bptr},\mathsf{zero}_{10},\mathsf{ldT},\mathsf{sgn}]).$$

Put $Y = 17(r+1)$.

(a) If track 5 carries the row s' of the frame of an integer z (its sign bit followed by the binary digits of $|z|$, at most H of them), with $s'+(\text{its length})\le r$, then loadS runs from $(-1,\tau)$ to $(-1,\tau')$ within $3\,100\,000\,Y^2$ units, and column 1 of the bank of τ' reads z.

(b) If track 5 carries the first row of the signed word $\mathrm{sw}_W(z)$ of an entry z of the buffer, then loadT does the same within $3\,100\,000\,Y^2$ units.

In both cases τ' is in the working state with indices (q,p), columns 3, 7, 9 and 11 of the bank are unchanged, and τ' differs from τ only on track 13, inside the support of the bank.

Proof. Inside the frame, $\mathsf{bank}(\mathsf{bz}_1)$ clears column 1 (at most $2000(r+1)$ units, Lemmas I.137 and I.138). (a) cb_4 copies the sign bit, the first bit of the frame, into the sign cell $(a,13)$, and the loop of ldS copies the following bits of the frame, one per bank row, into column 1: a pass rewrites the presence mark of the bank row, copies the bit into column 1 of that row ($x=2+12i$), advances the read pointer, moves the write pointer down twelve rows and counts on track 10; sv sets the flag $(s'',7)$ to 1 exactly while the frame continues. A pass costs at most $4500(r+1)$ units, and the frame has at most r bits, so ldS takes at most $50\,000(r+1)^2$ units. The rest of column 1 is 0, so column 1 holds $|z|$, and the sign cell holds $[z<0]$. (b) ldT copies the sign cell of the word and its $W+1$ magnitude bits in the same way (the repetition guarded by track 9 copies W bits and mg one more), writes a presence 0 after them, and then restores that presence mark from the comparison $[W+1<H]=1$ of track 10 with track 6; within $51\,000(r+1)^2$ units. In both cases sgn (within $25\,000(r+1)$ units, since neg' takes at most $210H+112$ virtual units) leaves in column 1 the H-bit word $|z|$ or $2^H-|z|$ according

to the sign, which reads z; it saves and restores column 9, which neg' uses. So the body of the frame takes at most $85\,000(r+1)^2$ units, and by Lemma I.132 with the archive of length $16(r+1)$ the frame adds at most $3 \cdot 10^6(17(r+1))^2$: in all at most $3\,100\,000\,Y^2$. The frame restores every register, and the bank lies on track 13, which no frame restores. □

The operand programs of the arithmetic combine an address with a load:

$$\begin{aligned}
\mathsf{coef}(\kappa,e) &= \mathsf{frame}_{\{5\}}(\mathsf{dup}_{10\to5};\mathsf{coefA}(\kappa,e));\mathsf{loadS},\\
\mathsf{idx}(\kappa,e) &= \mathsf{frame}_{\{5\}}(\mathsf{dup}_{10\to5};\mathsf{idxA}(\kappa,e);\mathsf{dup}_{8\to5}),\\
\mathsf{at} &= \mathsf{frame}_{\{5\}}(\mathsf{Look}), \qquad \mathsf{atL} = \mathsf{frame}_{\{5\}}(\mathsf{dup}_{7\to8};\mathsf{Look}),\\
\mathsf{atC} &= \mathsf{frame}_{\{5\}}([\mathsf{dup}_{5\to10},\mathsf{dup}_{8\to2},\mathsf{dup}_{7\to8},\mathsf{zero}_5,\\
&\qquad \mathsf{rep}_{5,2}(\mathsf{pred}_8;\mathsf{inc}_5),\mathsf{dup}_{10\to5},\mathsf{Look}]).
\end{aligned}$$

Lemma I.147 (Operands of the arithmetic)**.** *Let τ be in the working state with indices (q,p) and $Y = 17(r+1)$.*

(a) $\mathsf{coef}(\kappa,e)$ *runs within $6\,300\,000\,Y^4$ units to a tape in the same working state in which column 1 reads the numerator ($e=0$) or the denominator ($e=1$) of the coefficient of kind κ of coordinate j (the term q for $\kappa \neq c$), with columns 3, 7, 9 and 11 unchanged.*

(b) $\mathsf{idx}(\kappa,e)$, at, atL *and* atC *run within $3\,100\,000\,Y^4$ units each and change only the counter on track 5:* $\mathsf{idx}(\kappa,e)$ *makes it the index of Lemma I.144(b); if track 5 carries an index $u<d$, then* at, atL *and* atC *make it the first row of the word of $r_{p,u}$, $r_{l,u}$ and $r_{l-p,u}$ respectively ($p \le l$ for* atC*).*

Proof. Each frame restores every register but track 5 and runs its body on the archived tape, which holds the saved request, carries the counter r on track 0, has origin marks up to r and holds 0 in every row beyond r, so that the lemmas of this subsection apply to it. (a) The body $\mathsf{dup}_{10\to5};\mathsf{coefA}(\kappa,e)$ takes at most $102(r+1)+150\,000(r+1)^4$ units (Lemma I.144(a)), so the address costs at most $3\,200\,000\,Y^4$ with the frame, and Lemma I.146(a) adds $3\,100\,000\,Y^2$. Every coefficient of the field has width s (Section H.1), so its numerator and denominator have at most $s+1$ binary digits, and $s+1 \le \mathscr{B}(N) < W \le H$ by (H.1). (b) The bodies take at most $81\,000(r+1)^4$ (index) and $21\,000(r+1)^4$ or $22\,000(r+1)^4$ units (lookups; Lemmas I.144(b) and I.145); the repetition of atC decrements the copy of l on track 8 p times, within $1000(r+1)^2$ units. With the frame each is at most $3\,100\,000\,Y^4$. The words lie in the buffer since $p \le l$ and $u<d$. □

Loops over the terms. Two guards drive the loops of the arithmetic; both write the flag into the sentinel cell $(r,13)$, which is free while the stage runs, and stop there:

$$\begin{aligned}
\mathsf{tg}(\kappa) &= \mathsf{frame}_{\varnothing}([\mathsf{dup}_{10\to5},\mathsf{rowA}(\kappa),\mathsf{cmp}_{14,8,13}]);\mathsf{sent},\\
\mathsf{cg} &= \mathsf{frame}_{\varnothing}([\mathsf{dup}_{7\to5},\mathsf{inc}_5,\mathsf{cmp}_{8,5,13}]);\mathsf{sent}.
\end{aligned}$$

For a program Q put

$$\begin{aligned}
\mathsf{terms}(\kappa,Q) &= [\mathsf{zero}_{14},\mathsf{tg}(\kappa),\mathsf{loop}([\mathsf{home}_{13},Q,\mathsf{inc}_{14},\mathsf{tg}(\kappa)]),\mathsf{home}_{13}],\\
\mathsf{pairs}(Q) &= [\mathsf{zero}_8,\mathsf{cg},\mathsf{loop}([\mathsf{home}_{13},Q,\mathsf{inc}_8,\mathsf{cg}]),\mathsf{home}_{13}].
\end{aligned}$$

Lemma I.148 (Term loops). *Let $Y = 17(r+1)$ and $\kappa \in \{A, B\}$, and let n_κ be the number of terms of the kind κ of coordinate j.*

(a) $\mathsf{tg}(\kappa)$*, from the working state with indices (q, p), $q \le n_\kappa$, ends at* $\mathrm{cell}(r, 13)$ *within $4\,000\,000\,Y^4$ units, with $(r, 13)$ set to $[q < n_\kappa]$ and nothing else changed;* cg*, with $p \le l + 1$, does the same within $4\,000\,000\,Y^2$ units with the flag $[p < l + 1]$.*

(b) Let $\mathrm{I}_0, \dots, \mathrm{I}_{n_\kappa}$ be properties of tapes that do not depend on track 14 or on the cell $(r, 13)$. Suppose that for every $q < n_\kappa$, Q runs from every tape in the working state with term index q that has I_q, within b units, to a tape in the working state with the same term index q that has I_{q+1}. Then $\mathsf{terms}(\kappa, Q)$ *runs from every tape in the working state that has I_0 to a tape in the working state with term index n_κ that has I_{n_κ}, within $(n_\kappa + 1)(b + 5\,000\,000\,Y^4)$ units. The same holds for* $\mathsf{pairs}(Q)$ *with the Cauchy index on track 8 in place of the term index, $l + 1$ in place of n_κ, and the bound $(l + 2)(b + 5\,000\,000\,Y^2)$.*

Proof. (a) Inside the frame, $\mathsf{rowA}(\kappa)$ puts n_κ on track 8 (Lemma I.144(c), with j copied to track 5), and $\mathsf{cmp}_{14,8,13}$ compares the term index with it; for cg, track 5 receives $l + 1$ and the Cauchy index is compared with it. The bodies take at most $61\,000(r + 1)^4$ and $1000(r + 1)$ units, the frame at most $3 \cdot 10^6 Y^2$ more, and sent $18r + 17$. The frame restores the registers; the flag is on track 13, which it does not restore, and it lies outside the buffer and the bank. (b) Lemma I.6 with the invariant that the head is at $\mathrm{cell}(r, 13)$, the tape is in the working state with term index q, has I_q, and $(r, 13) = [q < n_\kappa]$: a pass returns to the root $(18r + 17)$, runs Q, increments the term index $(36q + 72)$ and runs the guard, within $b + 5 \cdot 10^6 Y^4$ units with its test; the first guard, the reset of track 14 and the final return add at most another $5 \cdot 10^6 Y^4$. For pairs the guard is the smaller one. □

I.11.6 The five phases of one entry

Throughout this subsection r, $l < N$, $j < d$, Z with $\|Z/s\| \le M$ and the tape τ_0 are as in Theorem I.119, $H = 3W$, $a = r - 12$, and $Y = 17(r + 1)$, the length of the archive plus $r + 1$. We write $\mathsf{rpt}(P) = [\mathsf{zero}_2, \mathsf{arch}_{2\leftarrow 8}, \mathsf{zero}_5, \mathsf{rep}_{5,2}(P; \mathsf{inc}_5)]$; it reads N from the snapshot onto track 2 and runs P exactly N times.

Lemma I.149 (Doubling N times). *Let τ be in the working state and let column $c \notin \{0, 2, 4\}$ of the bank read z. Then* $\mathsf{rpt}(\mathsf{bank}(\mathsf{bdbl}_c))$ *runs within $22\,000(r + 1)^2$ units to a tape in the same working state in which column c reads $2^N z$; only columns 2, 4, 6, 8 and c of the bank and tracks 2 and 5 change.*

Proof. $\mathsf{arch}_{2\leftarrow 8}$ copies N from the snapshot, where track 8 carries $N < r$ (Definition I.101), within $10^4(r+1)^2$ units (Lemma I.133). A pass of the repetition doubles column c (Lemmas I.137 and I.138(d), at most $17(400H+161)+36r+334$ units) and advances track 5: at most $10\,000(r+1)$ units. By Lemma I.126(c) the N passes take at most $(N + 1) \cdot 10\,200(r + 1)$, and with the two zeroings the total is at most $22\,000(r + 1)^2$, since $N \le r$. □

Lemma I.150 (The start)**.** *The list*

$$\mathsf{Start} = [\mathsf{home}_{11}, \mathsf{Archive}, \mathsf{Indices}, \mathsf{dup}_{5\to10}, \mathsf{dup}_{8\to7}, \mathsf{pred}_7, \mathsf{Width}, \mathsf{BankInit}, \mathsf{zero}_{14}]$$

of the first nine parts of Append *runs from* $(\mathrm{cell}(r, 11), \tau_0)$ *within* $2 \cdot 10^6(r+1)^3$ *units to a tape in the working state with indices* $(0, l+1)$ *in which every column of the bank reads* 0 *and every cell below row* $r+1$ *holds* 0*.*

Proof. home_{11} takes $18r + 15$ units. Archive with the empty list takes at most $10^6(r+1)^2$ (Lemma I.130); Indices at most $40\,000(r+1)^3$ and leaves $l+1$ on track 8 and j on track 5 (Lemma I.142); the two copies and the decrement put j on track 10 and l on track 7, within $300(r+1)$; Width restores W on track 9 within $11\,000(r+1)^2$; BankInit lays out the bank within $12\,000(r+1)^2$ (Lemma I.139, whose hypotheses hold by Lemma I.136); and zero_{14} sets the term index to 0. The sum is at most $2 \cdot 10^6(r+1)^3$. None of these parts writes track 0, 1, 3, 4 or 15, a cell below row $r+1$, or the buffer. □

Lemma I.151 (The common denominator)**.** *Let*

$$\begin{aligned} \mathsf{mulD} &= \mathsf{bank}([\mathsf{bcp}_{7\to3}, \mathsf{bmul}, \mathsf{bcp}_{4\to7}]), \\ \mathsf{Den} &= [\mathsf{bank}(\mathsf{bone}_7),\ \mathsf{coef}(c, 1),\ \mathsf{mulD},\ \mathsf{terms}(A, \mathsf{coef}(A, 1); \mathsf{mulD}), \\ &\qquad \mathsf{terms}(B, \mathsf{coef}(B, 1); \mathsf{mulD})]. \end{aligned}$$

From a tape in the working state, Den *runs within* $K_D Y^{24}$ *units,* $K_D = 28\,935\,892$*, to a tape in the working state with the same Cauchy index in which column* 7 *holds the binary digits of the product* D_j *of the denominators of* c_j*, of the A-coefficients and of the B-coefficients of coordinate* j*, and columns* 9 *and* 11 *are unchanged.*

Proof. $\mathsf{bank}(\mathsf{bone}_7)$ sets column 7 to 1 within $17(45H+45)+36r+334 \le 2000(r+1)$ units. The body of mulD takes $(54H+28)+441(H+1)^3+(56H+28)+1 \le 641(H+1)^4$ units and replaces what column 7 reads by its product with column 1; with Lemma I.137 and $H \le r$, mulD takes at most $17 \cdot 641(H+1)^4 + 36r + 334 \le 11\,297(r+1)^5$ units. So one factor, $\mathsf{coef}(\kappa, 1); \mathsf{mulD}$, takes at most $6\,300\,000\,Y^4 + 11\,297(r+1)^5 \le 6\,311\,297\,Y^9$ (Lemma I.147(a)). Each loop over terms runs at most r times, so by Lemma I.148(b) with the invariant "column 7 reads the product of the first q denominators, after the earlier ones" it takes at most $(r+1)(6\,311\,297\,Y^9 + 5 \cdot 10^6 Y^4) \le 11\,311\,297\,Y^{14}$ units. The total is at most $2000(r+1)+6\,311\,297\,Y^9+2 \cdot 11\,311\,297\,Y^{14}+1 \le K_D Y^{24}$. Every denominator is positive, and $D_j < 2^s \le 2^H$ (the proof of Proposition I.32), so the word that reads D_j modulo 2^H is its binary expansion. □

Lemma I.152 (The weight of a coefficient)**.** *Let*

$$\begin{aligned} \mathsf{fac} &= \mathsf{bcall}([\mathsf{right}, \overline{\mathsf{cp}}_{1\to4}, \mathsf{back}, \mathsf{bcp}_{7\to1}, \mathsf{sdiv}, \mathsf{back}, \mathsf{bcp}_{8\to3}, \mathsf{left}]), \\ \mathsf{wgt}(\kappa) &= [\mathsf{coef}(\kappa, 1),\ \mathsf{fac},\ \mathsf{coef}(\kappa, 0),\ \mathsf{bank}(\mathsf{bmul}; \mathsf{bcp}_{4\to9})], \end{aligned}$$

where $\overline{\mathsf{cp}}_{1\to4}$ *is* $\mathsf{cp}_{1\to4}$ *with the two parts of its branch exchanged, so that it writes the complement of column* 1 *into column* 4 *in the same number of units. Let* τ *be in the working state with column* 7 *holding the binary digits of* D_j*. Then* $\mathsf{wgt}(\kappa)$ *runs within* $12\,631\,163\,Y^{14}$ *units to a tape in the same working state in which column* 9 *reads* $D_j q$*, where* q *is the selected coefficient of kind* κ*, and columns* 7 *and* 11 *are unchanged.*

Proof. After $\mathsf{coef}(\kappa, 1)$, column 1 holds the denominator e of q, a positive integer below $2^s \le 2^{H-1}$, and the sign cell holds 0. In fac, column 4 receives $\bar{e}$ and column 1 the binary digits of D_j, and sdiv (Lemma I.30) leaves $\lfloor D_j/e \rfloor = D_j/e$ in column 8, which is copied to column 3; e divides D_j. The virtual count is $1 + (30H + 1) + (14H + 27) + (50H + 28) + 945(H+1)^3 + (14H+27) + (56H+28) + 2 \le 1245(H+1)^4$, so fac takes at most $21\,565(r+1)^5$ units. Then column 1 receives the numerator of q, and the last part, whose body takes $1+441(H+1)^3+(60H+28)+1 \le 541(H+1)^4$ units, puts the product $(D_j/e)\cdot(\text{numerator}) = D_j q$ into column 9 within $9597(r+1)^5$ units. With the two reads ($6\,300\,000\,Y^4$ each) the total is at most $(12\,600\,001 + 21\,565 + 9597)Y^{14} = 12\,631\,163\,Y^{14}$. □

Lemma I.153 (The numerator)**.** *Let*

$$\begin{aligned}
\mathsf{cst} &= [\mathsf{wgt}(c), \mathsf{bank}(\mathsf{bcp}_{9\to 11}), \mathsf{rpt}(\mathsf{bank}(\mathsf{bdbl}_{11})), \mathsf{rpt}(\mathsf{bank}(\mathsf{bdbl}_{11}))],\\
\mathsf{cbr} &= \mathsf{left}^8; \mathsf{branch}(\mathsf{home}_7; \mathsf{cst},\ \mathsf{home}_7),\\
\mathsf{lin} &= [\mathsf{wgt}(A), \mathsf{idx}(A,0), \mathsf{atL}, \mathsf{loadT}, \mathsf{bank}([\mathsf{bcp}_{9\to 3}, \mathsf{bmul}, \mathsf{bcp}_{4\to 9}]),\\
&\qquad \mathsf{rpt}(\mathsf{bank}(\mathsf{bdbl}_9)), \mathsf{bank}([\mathsf{bcp}_{9\to 4}, \mathsf{bacc}])],\\
\mathsf{pr} &= [\mathsf{idx}(B,0), \mathsf{at}, \mathsf{loadT}, \mathsf{bank}([\mathsf{bcp}_{9\to 3}, \mathsf{bmul}, \mathsf{bcp}_{4\to 3}]),\\
&\qquad \mathsf{idx}(B,1), \mathsf{atC}, \mathsf{loadT}, \mathsf{bank}([\mathsf{bmul}, \mathsf{bacc}])],\\
\mathsf{bil} &= \mathsf{wgt}(B); \mathsf{pairs}(\mathsf{pr}),\\
\mathsf{Num} &= [\mathsf{bank}(\mathsf{bz}_{11}), \mathsf{cbr}, \mathsf{terms}(A, \mathsf{lin}), \mathsf{terms}(B, \mathsf{bil})].
\end{aligned}$$

From a tape in the working state whose column 7 holds the binary digits of D_j, Num *runs within $K_F Y^{78}$ units, $K_F = 80\,917\,448$, to a tape in the working state in which column 7 is unchanged and column 11 reads the numerator*

$$U = [l = 0]\, D_j c_j s^2 + \sum_{(a,u)} D_j a\, r_{l,u}\, s + \sum_{(b,u,v)} \sum_{p \le l} D_j b\, r_{p,u}\, r_{l-p,v}$$

of Proposition I.32(b), the sums running over the terms of A and of B of coordinate j in their order.

Proof. The constant part. Track 7 carries l, so the cell $(0,7)$ holds $[l \ge 1]$: cbr runs cst exactly when $l = 0$. cst puts $D_j c_j$ into column 9 (Lemma I.152), copies it into column 11 (at most $3000(r+1)$ units) and doubles column 11 $2N$ times (Lemma I.149): column 11 reads $D_j c_j s^2$. It takes at most $12\,631\,163\,Y^{14} + 3000(r+1) + 44\,000(r+1)^2 + 1 \le 12\,681\,163\,Y^{16}$ units, and cbr adds 21: at most $12\,681\,184\,Y^{16}$.

A linear term (a,u), the term q of A_j: $\mathsf{wgt}(A)$ puts $D_j a$ into column 9; $\mathsf{idx}(A,0)$, atL and loadT put $r_{l,u}$ into column 1 (Lemmas I.147 and I.146, at most $9\,300\,000\,Y^4$ units together); the product block replaces column 9 by $D_j a\, r_{l,u}$, within $17((58H+28) + 441(H+1)^3 + (60H + 28) + 1) + 36r + 334 \le 11\,284(r+1)^4$ units; rpt multiplies it by $s = 2^N$ (at most $22\,000(r+1)^2$); and the last block adds it to column 11 (at most $10\,000(r+1)$). So lin takes at most $(12\,631\,163 + 11\,284 + 9\,332\,001)Y^{22} = 21\,974\,448\,Y^{22}$ units.

A bilinear term (b,u,v): $\mathsf{wgt}(B)$ puts $D_j b$ into column 9. A run of pr with Cauchy index $p \le l$ loads $r_{p,u}$, puts $D_j b\, r_{p,u}$ into column 3, loads $r_{l-p,v}$ into column 1, and adds the product $D_j b\, r_{p,u} r_{l-p,v}$ to column 11; with the two bank blocks ($11\,284(r+1)^4$ and, for $\mathsf{bmul}; \mathsf{bacc}$, $17(441(H+1)^3 + 400H + 161) + 36r + 334 \le 16\,367(r+1)^4$) it takes at most $(11\,284 + 16\,367 +$

$18\,600\,001)Y^{12} = 18\,627\,652\,Y^{12}$ units. By Lemma I.148(b) with the invariant "column 11 has received the products for $p' < p$", $\mathsf{pairs}(\mathsf{pr})$ takes at most $(l+2)(18\,627\,652\,Y^{12} + 5\cdot 10^6 Y^2)$ units, and since $l+2 \le Y$, bil takes at most $(12\,631\,163+18\,627\,652+5\cdot 10^6)Y^{29} = 36\,258\,815\,Y^{29}$.

The sum. $\mathsf{bank}(\mathsf{bz}_{11})$ clears column 11 within $3000(r+1)$ units. Each loop over terms runs at most r times, and by Lemma I.148(b) with the invariant "column 11 reads the part of U formed by the earlier terms" the two loops take at most $(21\,974\,448+5\cdot 10^6)Y^{27}$ and $(36\,258\,815+5\cdot 10^6)Y^{34}$ units. Every part adds a term of U to column 11 in the order of the formula, modulo 2^H, and keeps column 7; the total is at most $(12\,681\,184 + 21\,974\,448 + 36\,258\,815 + 10\,003\,001)Y^{78} = K_F Y^{78}$. □

Lemma I.154 (The two operands)**.** *Let*

$$\begin{aligned}\mathsf{ordf} &= [\mathsf{bank}([\mathsf{bone}_1, \mathsf{bz}_9]), \mathsf{dup}_{7\to 2}, \mathsf{inc}_2, \mathsf{zero}_5, \mathsf{rep}_{5,2}(\mathsf{bank}(\mathsf{binc}); \mathsf{inc}_5),\\ &\qquad \mathsf{bank}([\mathsf{bcp}_{7\to 3}, \mathsf{bcp}_{9\to 1}, \mathsf{bmul}, \mathsf{bcp}_{4\to 7}])],\\ \mathsf{Ops} &= [\mathsf{rpt}(\mathsf{bank}(\mathsf{bdbl}_7)),\ \mathsf{ordf},\ \mathsf{bank}([\mathsf{bdbl}_{11}, \mathsf{bcp}_{7\to 4}, \mathsf{bacc}, \mathsf{bdbl}_7])].\end{aligned}$$

If columns 7 *and* 11 *read* D *and* U*, then* Ops *runs within* $K_R(r+1)^8$ *units,* $K_R = 91\,799$*, to a tape in the same working state in which column* 11 *reads* $2U + Ds(l+1)$ *and column* 7 *reads* $2Ds(l+1)$.

Proof. rpt makes column 7 read Ds (at most $22\,000(r+1)^2$ units). In ordf, columns 1 and 9 become 1 and 0, track 2 receives $l+1$, and the repetition adds column 1 to column 9 $l+1$ times, so column 9 reads $l+1$; this costs at most $15\,000(r+1)^2$ units. The last block multiplies column 7 by column 9, within $17((54H+28)+(54H+28)+441(H+1)^3+(56H+28)+1)+36r+334 \le 14\,797(r+1)^4$ units; so ordf takes at most $29\,798(r+1)^6$, and column 7 reads $Ds(l+1)$. The final block doubles column 11, adds column 7 to it and doubles column 7, within $17((400H+161)+(56H+28)+(400H+160)+(400H+161)+1)+36r+334 \le 40\,000(r+1)$ units. The total is at most $22\,000(r+1)^2+29\,798(r+1)^6+40\,000(r+1)+1 \le (62\,001+29\,798)(r+1)^8$. □

Lemma I.155 (The quotient)**.** *Let*

$$\begin{aligned}\mathsf{Quot} = \mathsf{bcall}([&\mathsf{right}, \mathsf{bcp}_{11\to 1}, \mathsf{bsign}_1, \mathsf{left}, \mathsf{neg}', \mathsf{back}, \mathsf{bcp}_{8\to 1},\\ &\overline{\mathsf{cp}}_{7\to 4}, \mathsf{back}, \mathsf{sdiv}, \mathsf{back}, \mathsf{left}]).\end{aligned}$$

If column 11 *reads an integer* u *and column* 7 *an integer* v *with* $|u| < 2^{H-1}$ *and* $0 < v < 2^{H-1}$*, then* Quot *runs within* $K_Q(r+1)^5$ *units,* $K_Q = 798\,880$*, to a tape in the same working state in which the sign cell holds* $[\lfloor u/v \rfloor < 0]$ *and column* 8 *holds the* H *binary digits of* $|\lfloor u/v \rfloor|$.

Proof. After the copy, column 1 holds the H-bit two's complement word of u, whose last bit is $[u < 0]$ because $|u| < 2^{H-1}$; bsign_1 puts it into the sign cell. neg' then writes $|u|$ into column 8 (Lemma I.146), which is copied into column 1, and column 4 receives $\bar{v}$. Now Lemma I.30 applies with $z = u$ and $y = v \le 2^{H-1}$. The virtual count is at most $945(H+1)^3 + 500(H+1) \le 1445(H+1)^4$, the parts other than sdiv being $1 + (58H+28) + (28H+80)+1+(210H+112)+(14H+27)+(52H+28)+(42H+1)+3(14H+27)+2$. Lemma I.137 gives at most $17\cdot 1445(H+1)^4 + 36r + 300 \le 24\,965(H+r+1)^5$, and $H+r+1 \le 2(r+1)$ gives $24\,965\cdot 2^5(r+1)^5 = K_Q(r+1)^5$. □

Lemma I.156 (The finish). *Let*

$$\begin{aligned}
\mathsf{rcp} &= \mathsf{peek}_{2\to 13};\mathsf{branch}(\mathsf{home}_{13};\mathsf{put}_{11}(0),\ \mathsf{home}_{13};\mathsf{put}_{11}(1)),\\
\mathsf{rmag} &= \mathsf{rcp};\mathsf{pred}_2^{12};\mathsf{inc}_5, \qquad \mathsf{Ret} = \mathsf{rcp};\mathsf{pred}_2^{9};\mathsf{rep}_{5,9}(\mathsf{rmag});\mathsf{rmag},\\
\mathsf{patch}_k &= \mathsf{seek}_0;\mathsf{left}^k;\mathsf{branch}(\mathsf{pt}_k(0),\mathsf{pt}_k(1)),\\
\mathsf{pt}_k(b) &= \mathsf{home}_k;\mathsf{right}^3;\mathsf{scanP};\mathsf{left}^{31-2k};\mathsf{write}\,b;\mathsf{right}^{31-2k};\mathsf{rewP};\mathsf{left}^1,\\
\mathsf{FRestore} &= \mathsf{patch}_6;\mathsf{patch}_7;\mathsf{Restore},\\
\mathsf{setfl} &= \mathsf{right}^4;\mathsf{branch}(\mathsf{left}^4;\mathsf{write}\,0,\ \mathsf{left}^4;\mathsf{write}\,1);\mathsf{right}^5;\\
&\qquad \mathsf{branch}(\mathsf{left}^7;\mathsf{write}\,0;\mathsf{right}^2,\ \mathsf{left}^7;\mathsf{write}\,1;\mathsf{right}^2),\\
\mathsf{Finish} &= [\mathsf{Order},\mathsf{Width},\mathsf{Flags},\mathsf{Width},\mathsf{bptr},\mathsf{zero}_5,\mathsf{Ret},\mathsf{FRestore},\\
&\qquad \mathsf{sent},\mathsf{right}^2,\mathsf{setfl}].
\end{aligned}$$

Let τ be in the working state, with the sign cell holding $[z<0]$ and column 8 holding the binary digits of $|z|$, where $z = r_{l+1,j}$. Then Finish *runs from $(-1,\tau)$ to $(\mathrm{cell}(r,11),\tau')$ within $4\cdot 10^6(r+1)^3$ units, and τ' has the properties (a) to (d) of Theorem I.119.*

Proof. Order and Width bring N and W back to tracks 8 and 9 from the snapshot (within $11\,000(r+1)^2$ each). The buffer still holds the prefix (l,j), so Flags writes $[j+1<d]$ into $(r,7)$ and $[l+1<N]$ into $(r,6)$, within $42\,000(r+1)^3$ units (Lemma I.143), and Width restores W again: at most $80\,000(r+1)^3$ units for these four parts. bptr puts a, the row of the sign cell, on track 2, and zero_5 empties track 5. In Ret, rcp appends to the buffer the bit of track 13 in the row given by track 2 (Lemma I.124): first the sign cell, then, after nine rows up, column 8 of the bank row 0 ($x=9$), and rmag moves twelve rows up after each bit. The repetition, guarded by track 9 which carries W, and the last rmag append the $W+1$ lowest binary digits of $|z|$, least significant first; since z fits W (Proposition I.32(a)), the higher digits are 0, and the buffer now holds $\mathrm{sw}_W(T_{l,j}(Z))\,\mathrm{sw}_W(z) = \mathrm{sw}_W(T_{l,j+1}(Z))$. This takes at most $5000(r+1)^2$ units, and the buffer and the bank are disjoint by Lemma I.136.

patch_k finds the sentinel row with seek_0, reads (r,k), goes to the end of the packed list, steps back $31-2k$ cells to the value of the entry $16r+k$ of the archive, the entry of (r,k), and writes the bit read there; within $150(17(r+1)+1)$ units. So after the two patches the archive records the flags in the cells $(r,6)$ and $(r,7)$. Restore (Lemma I.131, with the archive as the whole packed list) then restores the borrowed tracks 2, 5 to 10 and 14 of every row from the archive, within $10^6(r+1)^2$ units, and clears the archive; FRestore takes at most $2\cdot 10^6(r+1)^2$ units. The resulting tape agrees with τ_0 on tracks 0 to 10, 14 and 15 of every row, except the cells $(r,6)$ and $(r,7)$, which hold the flags, and the root and the positive half are as in τ_0. Tracks 6 and 7 of τ_0 hold the binary stream $\mathrm{sw}_W(Z)$, which ends before row r; so the tape holds the stored vector Z, which is (a), and (d) holds. sent and right^2 reach $\mathrm{cell}(r,11)$ ($18r+20$ units), and setfl (31 units) copies $(r,7)$ into $(r,11)$ and $(r,6)$ into $(r,13)$ and returns to $\mathrm{cell}(r,11)$, which is (c). The buffer occupies the rows below its length, so these writes leave it holding the prefix $(l,j+1)$, which is (b). The total is at most $80\,000(r+1)^3+1500(r+1)+17+5000(r+1)^2+2\cdot 10^6(r+1)^2+18r+52\le 4\cdot 10^6(r+1)^3$. □

I.11.7 Proof of the bound for one entry

The five phases are composed by adding their bounds after raising each exponent to a common one.

Lemma I.157 (Sums of bounds). *Let $Y \geq 1$, $t_i \leq K_i Y^{e_i}$ for $i \leq k$, and $e \geq \max_i e_i$. Then $t_1 + \cdots + t_k + 1 \leq (K_1 + \cdots + K_k + 1) Y^{e}$.*

Proof. $Y^{e_i} \leq Y^{e}$ and $1 \leq Y^{e}$. □

All composite bounds of this section are of this kind, with the exponent e that Table I.4 records; for a loop over at most r terms the factor $n + 1 \leq Y$ is absorbed by raising the exponent.

Proof of Theorem I.119. Put $Y = 17(r + 1)$. By Lemma I.150, the first nine parts of Append run from $(\mathrm{cell}(r, 11), \tau_0)$ within $t_S \leq 2 \cdot 10^6 (r + 1)^3$ units to a tape in the working state whose bank columns read 0. Then, by Lemmas I.151, I.153 and I.154, the parts Den, Num and Ops run within $t_D \leq K_D Y^{24}$, $t_F \leq K_F Y^{78}$ and $t_R \leq K_R (r + 1)^8$ units to a tape in the working state whose columns 11 and 7 read

$$u = 2U + D_j s(l + 1) \quad \text{and} \quad v = 2 D_j s(l + 1).$$

By Proposition I.32(b), $|u| < 2^{3W-1} = 2^{H-1}$ and $0 < v < 2^{H-1}$, and

$$\left\lfloor \frac{u}{v} \right\rfloor = \mathrm{near}\Big(\frac{U}{D_j s(l + 1)} \Big) = r_{l+1,j}.$$

So Quot (Lemma I.155) runs within $t_Q \leq K_Q (r + 1)^5$ units to a tape in which the sign cell and column 8 hold the sign and the binary digits of $|r_{l+1,j}|$, and Finish (Lemma I.156) runs within $t_E \leq 4 \cdot 10^6 (r + 1)^3$ units to $(\mathrm{cell}(r, 11), \tau')$ with (a) to (d). The hypotheses of every lemma used along the way are kept by the working state, which each part preserves.

For (e): the list adds one unit, so $t = t_S + t_D + t_F + t_R + t_Q + t_E + 1$. Since $r + 1 \leq Y$, Lemma I.157 with $e = 24 + 78 + 8 + 5 + 3 = 118$ gives

$$t_D + t_F + t_R + t_Q + t_E + 1 \leq (K_D + K_F + K_R + K_Q + 4\,000\,001)\, Y^{118} = 114\,744\,020\, Y^{118}.$$

Finally $Y = 17(r + 1) \leq 17(X + r + 1)$ and $(r + 1)^3 \leq (X + r + 1)^{118}$, so

$$t \leq \big(2\,000\,000 + 114\,744\,020 \cdot 17^{118} \big) (X + r + 1)^{118} = K_{\mathrm{a}} (X + r + 1)^{118}.$$

The number of rows is r throughout. □

The exponent 118 is the sum $g_D + g_F + g_R + g_Q + 3$ of the exponents of the four arithmetic phases and of the finish, and each of these is itself a sum along the nested routines, as Table I.4 shows: for the denominator $24 = 9 + 14 + 1$, where $9 = 5 + 4$ is one factor (the bank multiplication, called, has exponent $5 = 3 + 1 + 1$, and reading a coefficient 4) and $14 = 9 + 5$ the loop over the terms; for the numerator $78 = 16 + 22 + 29 + 11$, the constant part, one linear term and one bilinear term, with 5 added for each loop over terms and 1 for the sum. None of these exponents depends on the problem, on i or on m, so the exponent 120 of the whole table in Theorem I.121 is a fixed number, as Theorem 15.33 requires.

I.12 The Taylor sum of a step

How does the machine add up a column of the Taylor table? Take one coordinate with $N = 2$, step $2^{-\sigma}$ with $\sigma = 1$, and table entries $z_0 = 5$, $z_1 = 3$, $z_2 = -7$. The new stored integer is $\mathrm{near}(z_0 + z_1/2 + z_2/4) = \mathrm{near}(4.75) = 5$. The machine never forms a fraction. It runs Horner's rule on the integers, $T = (5 \cdot 2 + 3) \cdot 2 - 7 = 19$, which is the sum times $2^{\sigma N} = 4$, and then divides by 4 with one rounding: it halves 19 twice, keeping the last bit it drops, $19 \mapsto 9$ (drops 1) $\mapsto 4$ (drops 1), and adds that bit, $4 + 1 = 5$. Only this one rounding is made, so the result is the nearest integer to the exact sum, as Lemma 15.35 requires.

This section constructs that computation on the tape for every coordinate of every stored vector, proves its contract and its step bound, and composes it with the coefficient stage of Section I.11 into one stored step of the trajectory. The Taylor sum has exponent 8 in its own size measure, and the stored step has exponent $g + 18 = 136$, where $g = 118$ is the exponent of one coefficient entry (Theorem I.119). Section I.12.1 proves the arithmetic; Section I.12.2 states the programs and contracts and proves the two theorems from the bound for one coordinate; the routines follow, and Section I.12.7 proves that bound.

I.12.1 One rounding of the whole sum

Lemma I.158 (The single rounding of a Taylor sum)**.** *Let $\sigma, N \in \mathbb{N}$, let $z_0, \dots, z_N$ be integers, put $E = \sigma N$, and define $T_0 = z_0$, $T_{k+1} = 2^{\sigma} T_k + z_{k+1}$, so that*

$$T = T_N = \sum_{l=0}^{N} 2^{\sigma(N-l)} z_l.$$

(a) $\mathrm{near}\big(\sum_{l \le N} 2^{-\sigma l} z_l\big) = \big\lfloor (2T + 2^E)/(2 \cdot 2^E) \big\rfloor$.

(b) The reverse doubled floor. *Put $R_N = 2z_N$ and $R_l = 2z_l + \lfloor R_{l+1}/2^{\sigma} \rfloor$ for $l < N$. Then $R_0 = \lfloor 2T/2^E \rfloor$ and $\lfloor (R_0 + 1)/2 \rfloor = \lfloor (2T + 2^E)/(2 \cdot 2^E) \rfloor$.*

(c) Rounding by shifts. *Put $q_k = \lfloor T/2^k \rfloor$. Then $q_{k+1} = \lfloor q_k/2 \rfloor$. If $E \ge 1$ and $g \in \{0, 1\}$ is $q_{E-1} \bmod 2$, then $\lfloor (2T + 2^E)/(2 \cdot 2^E) \rfloor = q_E + g$; if $E = 0$, it is T.*

(d) Sizes. *Let every z_l fit W and put $H = W + E + N + 4$. Then $|T_k| < (k+1)2^{W+1+\sigma k}$ for $k \le N$, and $|2^q T_k| < 2^{H-2}$ whenever $k \le N$ and $\sigma k + q \le E$. Moreover, with g as in (c) (and $g = 0$ if $E = 0$), $-2^{H-1} \le q_E + g < 2^{H-1}$, and $q_E + g$ fits $H - 1$.*

Proof. For integers u and $v > 0$, $\mathrm{near}(u/v) = \lfloor u/v + 1/2 \rfloor = \lfloor (2u + v)/(2v) \rfloor$, and $\lfloor \lfloor x/a \rfloor / b \rfloor = \lfloor x/(ab) \rfloor$ for integers x and $a, b > 0$.

(a) Multiplying by 2^E gives $\sum_l 2^{-\sigma l} z_l = T/2^E$.

(b) Put $T^{(l)} = \sum_{k=l}^{N} 2^{\sigma(N-k)} z_k$. We show $R_l = \lfloor 2T^{(l)}/2^{\sigma(N-l)} \rfloor$ by downward induction on l. For $l = N$ both sides are $2z_N$. For $l < N$, $T^{(l)} = 2^{\sigma(N-l)} z_l + T^{(l+1)}$, so

$$\left\lfloor \frac{2T^{(l)}}{2^{\sigma(N-l)}} \right\rfloor = 2z_l + \left\lfloor \frac{2T^{(l+1)}}{2^{\sigma(N-l-1)} 2^{\sigma}} \right\rfloor = 2z_l + \left\lfloor \frac{R_{l+1}}{2^{\sigma}} \right\rfloor = R_l.$$

With $l = 0$, $R_0 = \lfloor 2T/2^E \rfloor$, and $\lfloor (R_0 + 1)/2 \rfloor = \lfloor \lfloor (2T + 2^E)/2^E \rfloor / 2 \rfloor = \lfloor (2T + 2^E)/(2 \cdot 2^E) \rfloor$.

(c) The first claim is the rule for nested floors. If $E \geq 1$, then $q_{E-1} = 2q_E + g$ and $T = 2^{E-1}q_{E-1} + \rho$ with $0 \leq \rho < 2^{E-1}$, so $T = 2^E q_E + 2^{E-1}g + \rho$. Since $\lfloor (2T + 2^E)/(2 \cdot 2^E) \rfloor = \mathrm{near}(T/2^E)$, Corollary I.22(c) with $k = E - 1$ gives $q_E + g$. If $E = 0$, $\lfloor (2T + 1)/2 \rfloor = T$.

(d) $|T_0| = |z_0| < 2^{W+1}$, and

$$|T_{k+1}| \leq 2^{\sigma}|T_k| + |z_{k+1}| < (k+1)2^{W+1+\sigma(k+1)} + 2^{W+1} \leq (k+2)2^{W+1+\sigma(k+1)}.$$

If $\sigma k + q \leq E$, then $|2^q T_k| < (k+1)2^{W+1+E} \leq 2^N 2^{W+1+E} = 2^{H-3}$, because $k + 1 \leq N + 1 \leq 2^N$. For integers, $|\lfloor x/2^E \rfloor| \leq |x|$, so $|q_E| \leq |T| \leq 2^{H-2} - 1$ and $|q_E + g| \leq 2^{H-2}$, which gives both claims. □

The machine of this section computes the integer of (a) in the form (c): Horner's rule in a word of H bits, E halvings that record the last bit dropped, and one addition. By (d) every integer it forms fits that word (Proposition I.32(c)), and $H = W + \sigma N + N + 4$ is at most a constant depending on the problem times X^2. The reverse recurrence (b) gives the same integer; the machine does not use it.

By Lemma 15.35, the step from the stored vector Z_k takes the table rows $r_0 = Z_k, r_1, \ldots, r_N$ and returns $Z_{k+1} = \mathrm{near}(2^N S_N(f))$ with $f_l = r_l/2^N$, that is,

$$(Z_{k+1})_i = \mathrm{near}\Big(\sum_{l \leq N} 2^{-\sigma l} r_{l,i}\Big) \qquad (i < d).$$

So Lemma I.158(a) with $z_l = r_{l,i}$ is exactly the coordinate i of the next stored vector.

I.12.2 The stage, its contract and the stored step

The stage works on any table of integers, independently of where it came from. Let $d, \sigma, N, W, r \in \mathbb{N}$ and let $z_0, \ldots, z_N$ be integer vectors of length d (the rows of a table) whose entries all fit W. For $i < d$ put

$$V_i = \mathrm{near}\Big(\sum_{l \leq N} 2^{-\sigma l} z_{l,i}\Big), \qquad V = (V_0, \ldots, V_{d-1}),$$

and write $\mathrm{sw}_W(z_0, \ldots, z_N)$ for the signed words of the rows, row after row; its length is $\ell = (N+1)d(W+2)$. For an integer y that need not fit W, $\mathrm{sw}_W(y)$ denotes the same rule as in Definition I.3: the bit $[y < 0]$ followed by the $W + 1$ lowest binary digits of $|y|$. The size measure of the stage is

$$Y = d + \sigma + N + W + r + 1.$$

Definition I.159 (The stage after j coordinates)**.** Let Z be an integer vector of length d and $j \leq d$. A tape τ is in the *state* C_j if

(i) track 11 carries the counter $j(W+2)$;

(ii) the binary stream holds $\mathrm{sw}_W(Z)$, and tracks 8, 9 and 10 carry the counters N, W and σ;

(iii) $\tau(k, 12) = [k < \ell]$ for $k \leq r$, and $\ell \leq r$;

(iv) for $k < \ell$, $\tau(k, 13)$ is the bit k of $\mathrm{sw}_W(V)$ if $k < j(W+2)$, and the bit k of $\mathrm{sw}_W(z_0, \dots, z_N)$ otherwise;

(v) τ has origin marks up to r and is a workspace of r rows.

So in C_j the first j words of the first table row have been replaced by the first j coordinates of the answer. Since the first row has d words and the answer has d coordinates, the replacement never touches a word that a later coordinate still needs: coordinate j reads the words $kd + j$, $k \le N$, and only the word j of the first row is replaced when coordinate j is done. The *protected tracks* are 2, 3, 4, 5, 8, 9, 10, 14 and 15.

The programs. Let $\mathsf{rep}_{c,g}$ and $\mathsf{dup}_{k\to l}$ be the programs of Lemma I.126. The stage is

$$\mathsf{Sum} = \mathsf{zero}_{11}; \mathsf{rep}_{11,6}(\mathsf{Coord}); \mathsf{dup}_{13\to 7},$$
$$\mathsf{Coord} = \mathsf{Prep}; \mathsf{Horner}; \mathsf{Round}; \mathsf{zero}_2; \mathsf{cpre}; \mathsf{Wback}; \mathsf{Restore}.$$

Sum resets the cursor on track 11, runs Coord once for each coordinate, using the presence marks of the stored vector on track 6 as the guard, and finally copies the first row of the buffer, which now holds the answer, onto the value track 7 of the binary stream. One run of Coord treats the coordinate j whose word starts at the row $a = j(W+2)$ that track 11 carries:

(1) Prep saves a and the borrowed tracks on a stack in the positive half of the tape, adds rows to the workspace, computes $E = \sigma N$ and $H = W + E + N + 4$ in unary, and lays out an arithmetic area of H twelve-cell rows along track 7 (Lemma I.181);

(2) Horner computes $T = \sum_l 2^{\sigma(N-l)} z_{l,j}$ in that area by Horner's rule, reading each $z_{l,j}$ from the buffer (Lemma I.178);

(3) Round halves T E times, adds the last bit dropped and writes the signed word of the result V_j (Lemma I.179);

(4) zero_2 and cpre put a back on track 2, and Wback writes the first $W+2$ bits of that word, which are $\mathrm{sw}_W(V_j)$, into the buffer at row a (Lemmas I.169 and I.180);

(5) Restore clears the arithmetic area, restores every borrowed track from the stack, returns the workspace to its r rows and advances track 11 to $a + W + 2$ (Lemma I.182).

Theorem I.160 (One coordinate). *Let $j < d$ and let τ be in the state C_j. Then* Coord *runs from $(-1, \tau)$ to $(-1, \tau')$ in some t units, where τ' is in the state C_{j+1}, agrees with τ on every cell of the protected tracks, and*

$$t \le 10^{31}\, Y^6.$$

The proof is in Section I.12.7. The exponent 6 comes from words of $H \le 7Y^2$ bits handled by loops whose length and cost per pass are both at most a constant times Y^2.

Theorem I.161 (The Taylor sum). *Put $K_\Sigma = 10^{31}+1000$. Let d, σ, N, W, r and the table $z_0, \dots, z_N$ be as above, and let Z be an integer vector of length d. Let τ be a tape whose binary stream holds $\mathrm{sw}_W(Z)$, on which tracks 8, 9 and 10 carry the counters N, W and σ, whose output buffer holds $\mathrm{sw}_W(z_0, \dots, z_N)$ with capacity r, and which is a workspace of r rows. Then* Sum *runs from $(-1, \tau)$ to $(-1, \tau')$ in some number t of units, where*

(a) the binary stream of τ' holds $\mathrm{sw}_W(V)$;

(b) τ' has origin marks up to r and is a workspace of r rows;

(c) τ' agrees with τ on every cell of the tracks 2, 3, 4, 5, 8, 9, 10, 14 and 15;

(d) $t \le K_\Sigma\, Y^8$.

The rows in use do not change, so in the sense of Definition I.7 this is a contract whose bound counts the units alone, with exponent 8 in the size measure Y and a constant that does not depend on anything.

Proof. Put $K = 10^{31}$ and $\ell = (N+1)d(W+2)$. Since the buffer holds a word of length ℓ with capacity r, $\ell \le r$, and so $d(W+2) \le r$.

Start. By Lemma I.14(d), zero_{11} takes 29 units and sets $(0, 11)$ to 0; it changes no other cell. The tape is then in the state C_0: track 11 carries 0, conditions (ii) and (v) are inherited, and (iii) and (iv) with $j = 0$ restate that the buffer holds the table word.

The loop over the coordinates. Recall that

$$\mathsf{rep}_{11,6}(\mathsf{Coord}) = \mathsf{peek}_{11\to6}; \mathsf{loop}(\mathsf{home}_6; \mathsf{Coord}; \mathsf{peek}_{11\to6}); \mathsf{home}_6.$$

For $j \le d$ let I_j be the set of configurations $(\mathrm{cell}(j(W+2), 6), \sigma)$ such that σ is in the state C_j and agrees with τ on the protected tracks. The binary stream holds d words of $W+2$ bits, so the presence cell $(j(W+2), 6)$ holds 1 for $j < d$ and 0 for $j = d$. The first $\mathsf{peek}_{11\to6}$ takes 20 units (Lemma I.126(a)) and ends in I_0. From a configuration in I_j with $j < d$, home_6 returns to the root along the origin marks in $18j(W+2) + 10$ units (Lemma I.11(b)), Theorem I.160 runs Coord within KY^6 units to a tape in C_{j+1} that keeps the protected tracks, and $\mathsf{peek}_{11\to6}$ goes to $\mathrm{cell}((j+1)(W+2), 6)$ in $18(j+1)(W+2)+20$ units. As $(j+1)(W+2) \le r$, a pass ends in I_{j+1} within $KY^6 + 100(r+1)$ units. By Lemma I.6 the loop takes at most $d(KY^6 + 100(r+1) + 1) + 1$ units, and the final home_6 at most $18r + 10$. So $\mathsf{rep}_{11,6}(\mathsf{Coord})$ ends in C_d within $d(KY^6 + 100(r+1) + 1) + 100(r+1)$ units, with the protected tracks unchanged.

The copy. In C_d the cells $(k, 13)$, $k < d(W+2)$, hold $\mathrm{sw}_W(V)$, and track 6 still holds the presence marks of a stream of $d(W+2)$ bits. By Lemma I.126(b), $\mathsf{dup}_{13\to7}$ copies track 13 onto track 7 in the rows $k \le r$ in exactly $82r + 56$ units and changes nothing else. Afterwards the binary stream holds $\mathrm{sw}_W(V)$; the origin marks and the workspace are kept, because track 7 holds 0 beyond row r before and after; and track 7 is not protected. This proves (a)–(c).

The count. Since $Y \ge 1$, $d \le Y$ and $r + 1 \le Y$,

$$t \le 29 + d\big(KY^6 + 100(r+1) + 1\big) + 100(r+1) + 82r + 56 \le KY^7 + 100Y^2 + Y + 185Y \le (K+1000)Y^8. \qquad \square$$

The stored vector between steps is described by Definition I.101: the saved request on tracks 3, 4 and 15, the counters m, N, W and σ on tracks 5, 8, 9 and 10, the word $\mathrm{sw}_W(Z)$ on the binary stream, origin marks up to r, a workspace of r rows and the reserve of rows. The contract of one stored step is Definition I.183. The coefficient stage Build of Theorem I.121 fills the output buffer with the table; the Taylor sum then consumes it.

Theorem I.162 (One stored step)**.** *Put* Step $=$ Build; Sum*, let* $g = 118$ *be the exponent of Theorem I.119,* K_{a} *its constant,* $c_W = 4(q_P + l_P + c_P) + 4$ *the constant of Proposition I.32(e), and*

$$K_{\mathrm{s}} = dK_{\mathrm{a}} + d + 1008 + K_\Sigma c_\Sigma^8, \qquad c_\Sigma = d + \sigma + c_W + 2.$$

Let a problem with a coefficient field and a level n_0 *be given. Let* $i < d$*,* m*,* r *and* $c < S$*, let* Z *be an integer vector of length* d *with* $\|Z/2^N\| \le M$*, and let* τ *hold the stored vector* Z *on* r *rows (Definition I.101), with the counter* c *on track* 2 *and a retained word of length* S *on track* 14 *from row* 0*. Then* Step *runs from* $(-1, \tau)$ *to* $(-1, \tau')$ *in* t *units, where* τ' *holds the stored vector* Z^+ *of Definition I.183 on the same* r *rows,*

$$Z_i^+ = \mathrm{near}\Big(\sum_{l \le N} 2^{-\sigma l} r_{l,i}\Big) \qquad (i < d)$$

for the table rows $r_0 = Z, r_1, \dots, r_N$ *of Lemma 15.35 (so* $Z^+ = Z_{k+1}$ *when* $Z = Z_k$*),* τ' *agrees with* τ *on the cells* $(q, 2)$ *and* $(q, 14)$*,* $q \le S$*, and*

$$t \le K_{\mathrm{s}}\,(X + r + 1)^{g+18} = K_{\mathrm{s}}\,(X + r + 1)^{136}.$$

So Step *meets the contract of Definition I.183 with exponent* $g_{\mathrm{s}} = g + 18 = 136$*; its constant depends on the problem only through* d*,* σ *and* c_W*.*

Proof. Put $X' = X + r + 1$. By Theorem I.121, Build runs from $(-1, \tau)$ within $(dK_{\mathrm{a}} + d + 1008)X'^{120}$ units to a tape τ_1 that holds the stored vector Z on the same r rows, whose output buffer holds $\mathrm{sw}_W(r_0, \dots, r_N)$ with capacity r, and that agrees with τ on the cells $(q, 2)$ and $(q, 14)$, $q \le S$. By Proposition I.32(a) every entry of every r_l fits W. So τ_1 satisfies the hypotheses of Theorem I.161 with the head vector Z and the table $r_0, \dots, r_N$: its binary stream holds $\mathrm{sw}_W(Z)$, its tracks 8, 9 and 10 carry N, W and σ, and it is a workspace of r rows. Theorem I.161 gives a run of Sum from $(-1, \tau_1)$ to $(-1, \tau')$ within $K_\Sigma Y^8$ units, where the binary stream of τ' holds $\mathrm{sw}_W(Z^+)$, τ' has origin marks up to r, is a workspace of r rows and agrees with τ_1 on tracks $2, 3, 4, 5, 8, 9, 10, 14, 15$. These tracks carry the saved request (3, 4, 15), the counters m, N, W, σ (5, 8, 9, 10), the step counter and the retained step word (2, 14). The reserve concerns the row count r only. So τ' holds the stored vector Z^+ on r rows and agrees with τ on the cells $(q, 2)$ and $(q, 14)$. The formula for Z^+ is the last display of Section I.12.1.

For the count, $N < X$ and $W \le c_W X^2$ by Proposition I.32(e), and $X' \ge 1$, so

$$Y = d + \sigma + N + W + r + 1 \le (d + \sigma)X'^2 + c_W X'^2 + 2X'^2 = c_\Sigma X'^2.$$

Hence Sum takes at most $K_\Sigma c_\Sigma^8 X'^{16}$ units, and

$$t \le (dK_{\mathrm{a}} + d + 1008)X'^{120} + K_\Sigma c_\Sigma^8 X'^{16} \le K_{\mathrm{s}} X'^{136},$$

because $120 \le 136$ and $16 \le 136$. The exponent is $(g+2)+2\cdot 8$: the coefficient stage contributes $g + 2 = 120$, and the Taylor sum its exponent 8 doubled by the passage from Y to X'^2. □

I.12.3 Repetition, rows and a stack

A coordinate needs more room than the r rows it is given and more registers than the free tracks offer. It takes both for a while and gives them back: it adds rows to the workspace and removes them at the end, and it saves the borrowed tracks on a stack in the positive half of the tape. This subsection gives these tools. The repetition $\mathsf{rep}_{c,g}(P) = \mathsf{peek}_{c\to g};\mathsf{loop}(\mathsf{home}_g;P;\mathsf{peek}_{c\to g});\mathsf{home}_g$ of Lemma I.126 is used throughout, also when its cursor does not start at 0.

Lemma I.163 (Repeating from a start, and $W+2$ times)**.** *Let $c,g<16$ with $1\le g$, and let $s\le n\le r$.*

(a) Let $\mathrm{R}_s,\dots,\mathrm{R}_n$ be properties of tapes such that on every tape with R_q track c carries q, track g carries n and there are origin marks up to r. If for every q with $s\le q<n$ the program P runs from $(-1,\sigma)$ within b units to some $(-1,\sigma')$ with R_{q+1} whenever σ has R_q, then $\mathsf{rep}_{c,g}(P)$ runs from every tape with R_s to a tape with R_n within $(n-s+1)(b+200(r+1))$ units.

(b) Put $\mathsf{wp}(P)=\mathsf{rep}_{11,9}(P);P;P$. Let $W+2\le r$ and let $\mathrm{R}_0,\dots,\mathrm{R}_{W+2}$ be properties such that on every tape with R_q track 11 carries q, track 9 carries W and there are origin marks up to r. If P runs from every tape with R_q, $q<W+2$, within b units to a tape with R_{q+1}, then $\mathsf{wp}(P)$ runs from every tape with R_0 to a tape with R_{W+2} within $(W+3)(b+200(r+1))$ units.

Proof. (a) As in the proof of Lemma I.126(c), with the configurations I_q ($s\le q\le n$) at cell(q,g) whose tape has R_q. The first $\mathsf{peek}_{c\to g}$ takes $18s+c+4+|c-g|\le 18r+34$ units, each of the $n-s$ passes at most $b+100(r+1)$ units with its test, and the final test and home_g at most $18r+20$. Since $36r+55\le b+200(r+1)$, the total is at most $(n-s+1)(b+200(r+1))$. (b) By (a) with $s=0$ and $n=W$, $\mathsf{rep}_{11,9}(P)$ reaches R_W within $(W+1)(b+200(r+1))$ units; the two further runs of P reach R_{W+2} within $2b$ more. □

The program $\mathsf{wp}(P)$ runs P once for each bit of a signed word of width W; this section uses it to append a word's worth of marks and to write one word back.

While a coordinate is computed the workspace has more rows than usual and the positive half is not blank.

Definition I.164 (Working workspace)**.** A tape τ is a *working workspace of c rows* if $\tau(-1)=1$, track 0 carries the counter c, τ has origin marks up to c, and every cell of every row $j>c$ holds 0.

A workspace of r rows with origin marks up to r is a working workspace of r rows. Every counter on such a tape is at most $c+1$, as in Lemma I.4.

Lemma I.165 (Moving a counter back, and adding rows)**.** *Let $2 \le k < 16$ and put*

$$\begin{aligned} \mathsf{retr}_k &= \mathsf{seek}_k; \mathsf{write}\ 1; \mathsf{right}^{16}; \mathsf{write}\ 0; \mathsf{home}_k, \\ \mathsf{addrow} &= \mathsf{seek}_0; \mathsf{write}\ 1; \mathsf{left}^{16}; \mathsf{write}\ 0; \mathsf{left}; \mathsf{write}\ 1; \mathsf{home}_1, \\ \mathsf{ginc}_k &= \mathsf{addrow}; \mathsf{inc}_k. \end{aligned}$$

Each runs from $(-1, \tau)$ to $(-1, \tau')$.

(a) If track k carries $a+1$ and τ has origin marks up to $a+1$, retr_k takes $36a + 2k + 44$ units, and τ' is τ with $(a+1, k)$ set to 1 and (a, k) set to 0; track k carries a.

(b) If τ is a working workspace of c rows, addrow takes $36c + 47$ units, and τ' is τ with $(c, 0)$ and $(c+1, 1)$ set to 1 and $(c+1, 0)$ set to 0, a working workspace of $c+1$ rows.

(c) If moreover track k carries $a \le c$, ginc_k takes $36c + 47 + 36a + 2k + 44$ units and leaves a working workspace of $c+1$ rows on which track k carries $a+1$.

No other cell is written.

Proof. (a) seek_k ($18(a+1) + k + 3$) reaches the hole $(a+1, k)$, the two writes and right^{16} (19) set it and the cell (a, k) above it, and home_k from row a takes $18a + k + 4$ units (Lemma I.11). (b) seek_0 ($18c + 3$) reaches the sentinel header $(c, 0)$; the writes and moves (21 units) promote it, blank the header of row $c+1$ and set the origin mark $(c+1, 1)$; and home_1 from row $c+1$ takes $18c + 23$. Rows beyond $c+1$ are untouched, so they hold 0. (c) is (b) followed by Lemma I.14(a), which needs origin marks up to $a + 1 \le c + 1$. □

The stack lies on the cells $15 + 16i$, $i \ge 0$, of the positive half; these cells are blank between stages and are never used by any other routine of the stage.

Definition I.166 (The stack)**.** For a tape τ the *stack view* is the tape $\varsigma(i) = \tau(15 + 16i)$, $i \in \mathbb{Z}$; so $\varsigma(-1)$ is the root and $\varsigma(i)$, $i \ge 0$, are positive cells. The tape *holds the stack* $w = w_0 \cdots w_{n-1}$ if ς holds the packed list w at 2 (Definition I.122), $\varsigma(0) = \varsigma(1) = 0$, and $\varsigma(x) = 0$ for every $x \ge 2 + 2n$. The cells $\varsigma(2 + 2i)$ and $\varsigma(3 + 2i)$ are the cells of the *entry* i; the last entry is the *top*.

Lemma I.167 (Pushing and popping)**.** *Put*

$$\begin{aligned} \mathsf{spush}(b) &= \big(\mathsf{right}; \mathsf{right}; \mathsf{right}; \mathsf{appP}(b); \mathsf{left}; \mathsf{left}; \mathsf{loop}(\mathsf{left}; \mathsf{left}); \mathsf{left}\big)^{\uparrow 16}, \\ \mathsf{spop} &= \big(\mathsf{right}^3; \mathsf{scanP}; \mathsf{left}^2; \mathsf{write}\ 0; \mathsf{right}; \mathsf{write}\ 0; \mathsf{left}; \mathsf{rewP}; \mathsf{left}\big)^{\uparrow 16}, \\ \mathsf{top}(P_0, P_1) &= \big(\mathsf{right}^3; \mathsf{scanP}; \mathsf{left}\big)^{\uparrow 16}; \mathsf{branch}\big(\mathsf{bk}; P_0,\ \mathsf{bk}; P_1\big), \\ \mathsf{bk} &= (\mathsf{right}; \mathsf{rewP}; \mathsf{left})^{\uparrow 16}, \end{aligned}$$

with scanP, appP *and* rewP *of Lemma I.123. Let τ hold the stack w of length n.*

(a) $\mathsf{spush}(b)$ runs from $(-1, \tau)$ within $17(6n + 16)$ units to $(-1, \tau')$, where τ' holds the stack wb and differs from τ only in the cells $47 + 32n$, $63 + 32n$ and $79 + 32n$.

(b) If $w = w'b$, then spop *runs from $(-1,\tau)$ within $17(7n+13)$ units to $(-1,\tau')$, where τ' holds the stack w' and is τ with the cells $15+32n$ and $31+32n$ of the top set to 0.*

(c) If $w = w'b$ and P_b runs from $(-1,\tau)$ in t units to a configuration γ, then $\mathsf{top}(P_0,P_1)$ *runs from $(-1,\tau)$ to γ within $17(7n+12)+1+t$ units.*

Proof. We count on the stack view and apply Lemma I.8 with origin 15 and stride 16: the root is $\varsigma(-1)$, each unit of the virtual run becomes at most 17, and only cells $15+16x$ are written. (a) Three moves reach $\varsigma(2)$, $\mathsf{appP}(b)$ appends b in $3n+6$ units and stops at $2n+4$ (Lemma I.123(c)), two moves reach the presence cell $2n+2$ of the new entry, the loop passes the $n+1$ presence cells $2n+2,\dots,2$ (3 units each) and stops at $\varsigma(0)=0$, and a last move returns to -1: $6n+16$ units. The cells written are $2+2n$, $3+2n$ and $4+2n$ of the view. (b) right^3 and scanP $(4+3n+1)$ reach $2+2n$, left^2 (3) the presence cell $2n$ of the top; four units clear its two cells and return; rewP over the $n-1$ entries below $(4n)$ reaches 0, and left the root: $7n+13$ units. (c) The first part ($3n+6$ units) stops at the value cell $2n+1$ of the top, the branch reads b, and bk ($4n+6$ units) returns to the root without writing. □

For a tape σ, a track k and a row count c, the *snapshot* $\theta^c_k(\sigma)$ is the word $\sigma(0,k)\sigma(1,k)\cdots\sigma(c,k)$ of $c+1$ bits. The stage saves a track by pushing its snapshot and restores it by popping the snapshot back, last row first.

Lemma I.168 (Saving and restoring a track)**.** *Let $2\le k<16$, $k\neq 11$, and put*

$$\begin{aligned}
\mathsf{brd}_k &= \mathsf{peek}_{11\to k};\mathsf{branch}\big(\mathsf{home}_k;\mathsf{spush}(0),\ \mathsf{home}_k;\mathsf{spush}(1)\big),\\
\mathsf{save}_k &= \mathsf{peek}_{11\to 0};\mathsf{loop}(\mathsf{left};\mathsf{home}_1;\mathsf{brd}_k;\mathsf{inc}_{11};\mathsf{peek}_{11\to 0});\\
&\qquad \mathsf{left};\mathsf{home}_1;\mathsf{brd}_k;\mathsf{zero}_{11},\\
\mathsf{wcl}_k(b) &= \mathsf{peek}_{11\to k};\mathsf{write}\,b;\mathsf{home}_k,\\
\mathsf{rone}_k &= \mathsf{top}\big(\mathsf{spop};\mathsf{wcl}_k(0),\ \mathsf{spop};\mathsf{wcl}_k(1)\big),\\
\mathsf{rest}_k &= \mathsf{dup}_{3\to 11};\mathsf{left}^{12};\mathsf{loop}(\mathsf{right}^{12};\mathsf{rone}_k;\mathsf{retr}_{11};\mathsf{left}^{12});\\
&\qquad \mathsf{right}^{12};\mathsf{rone}_k.
\end{aligned}$$

(a) Let track 0 carry the counter c, let τ have origin marks up to c, let track 11 carry 0, and let τ hold the stack w. Then save_k *runs from $(-1,\tau)$ to $(-1,\tau')$ within $11\,000(c+|w|+1)^2$ units, where τ' holds the stack $w\theta^c_k(\tau)$, track 11 carries 0, and τ' agrees with τ on every cell of a row that is not on track 11. For a list $k_1,\dots,k_s$ of such tracks,* $\mathsf{save}_{k_1};\cdots;\mathsf{save}_{k_s};\mathsf{skip}$ *pushes $\theta^c_{k_1}(\tau)\cdots\theta^c_{k_s}(\tau)$ within $(11\,000s+1)(c+|w|+s(c+1)+1)^2$ units.*

(b) Let τ be a working workspace of R rows on which track 3 carries $c\le R$, and let τ hold the stack $w\theta$, where θ is $\theta^c_k(\sigma)$ for some tape σ. Then rest_k *runs from $(-1,\tau)$ to $(-1,\tau')$ within $30\,000(R+|w|+c+1)^2$ units, where τ' holds the stack w, track 11 carries 0, $\tau'(q,k)=\sigma(q,k)$ for $q\le c$, τ' agrees with τ on every cell of a row that is not on track 11 or k, and τ' is a working workspace of R rows. For a list $k_1,\dots,k_s$ of such tracks, all different from 3,* $\mathsf{skip};\mathsf{rest}_{k_s};\cdots;\mathsf{rest}_{k_1}$ *restores them from the stack $w\theta^c_{k_1}(\sigma)\cdots\theta^c_{k_s}(\sigma)$ within $(30\,000s+1)(R+|w|+s(c+1)+c+1)^2$ units.*

No stack cell outside the entries pushed or popped is written.

Proof. (a) When track 11 carries $q \le c$, $\mathsf{peek}_{11\to 0}$ reaches the header $(q,0)$, which holds 1 exactly when $q < c$. A pass at the header of row $q < c$ steps to $(q,1)$ and returns along the origin marks ($18q+6$ units), reads (q,k) and pushes it (brd_k: at most $36q+2k+30$ units for the walks and $17(6(|w|+q)+16)$ for the push, Lemma I.167(a)), advances track 11 to $q+1$ ($36q+66$) and peeks again ($18q+44$): at most $1000(c+|w|+1)$ units with its test. At the sentinel header $(c,0)=0$ the loop stops; the program pushes (c,k) the same way and zero_{11} (29 units) makes track 11 carry 0. By Lemma I.6, the c passes and the rest cost at most $11\,000(c+|w|+1)^2$ units. Only track 11 and the stack cells are written. The list follows by adding the bounds, the stack growing by $c+1$ entries each time.

(b) $\mathsf{dup}_{3\to 11}$ copies track 3 onto track 11 in the rows $q \le R$ ($70R+44$ units, Lemma I.126(b)), so track 11 carries c. left^{12} reaches cell$(0,11)$, which holds 1 exactly when the counter on track 11 is positive. Let the counter be $q \ge 1$, the stack $wb_0 \cdots b_q$, and the rows $q+1,\dots,c$ of track k already restored. A pass returns to the root, reads the top b_q and pops it (Lemma I.167(b),(c)), writes it into (q,k) with wcl_k ($36q+50$ units at most), moves the counter back to $q-1$ with retr_{11} (Lemma I.165(a)) and steps to cell$(0,11)$ again: at most $238(|w|+c+1)+72q+842$ units. When the counter is 0, the loop stops, right^{12} returns to the root and rone_k writes b_0 into $(0,k)$. So rows $c, c-1,\dots,0$ of track k receive $\sigma(c,k),\dots,\sigma(0,k)$, and the stack is w. By Lemma I.6 the total is at most $30\,000(R+|w|+c+1)^2$. The tracks 0 and 1 are not written (track $k \ne 1$, and $k=0$ is excluded), so the working workspace is kept. For the list, the snapshots come off in the reverse order of the pushes, and the restored tracks differ from 3, so track 3 still carries c for each of them. □

The address of the coordinate, the counter $a = j(W+2)$ on track 11, is kept on the stack too, as a entries 1 followed by a separator 0.

Lemma I.169 (The address)**.** *Put*

$$\begin{aligned}
\mathsf{sav} &= \mathsf{left}^{12};\mathsf{loop}(\mathsf{right}^{12};\mathsf{retr}_{11};\mathsf{spush}(1);\mathsf{left}^{12});\mathsf{right}^{12};\mathsf{spush}(0),\\
\mathsf{cpre} &= \mathsf{right}^{48};\mathsf{write}\,0;\mathsf{right}^{16};\mathsf{loop}(G^{\uparrow 16};\mathsf{inc}_2;F^{\uparrow 16});J^{\uparrow 16},\\
G &= \mathsf{left};\mathsf{rewP};\mathsf{left}, \qquad F = \mathsf{right}^{3};\mathsf{scanP};\mathsf{write}\,1;\mathsf{right}^{2};\mathsf{write}\,0;\mathsf{right},\\
J &= \mathsf{left};\mathsf{write}\,1;\mathsf{loop}(\mathsf{left};\mathsf{left});\mathsf{left},\\
\mathsf{fin} &= \mathsf{wp}(\mathsf{spush}(1);\mathsf{inc}_{11});\mathsf{zero}_{11};\mathsf{right}^{48};\\
&\qquad \mathsf{loop}(\mathsf{left}^{48};\mathsf{spop};\mathsf{inc}_{11};\mathsf{right}^{48});\mathsf{left}^{48}.
\end{aligned}$$

Let τ have origin marks up to c.

(a) If track 11 carries $n \le c$ and τ holds the stack w, then sav *runs from $(-1,\tau)$ to $(-1,\tau')$ within $5000(c+|w|+n+1)^2$ units; τ' holds the stack $w1^n0$, track 11 carries 0, and of the rows only track 11 was written.*

(b) If track 2 carries 0, τ holds a stack 1^n0v for some word v, and $n \le c$, then cpre *runs from $(-1,\tau)$ to $(-1,\tau')$ within $2000(c+n+1)^2$ units, where track 2 of τ' carries n, the stack and every cell $x \ge 0$ are unchanged, and of the rows only track 2 was written.*

(c) If track 11 carries 0, track 9 carries W, τ holds the stack 1^a and $a+W+2 \le c$, then fin *runs from $(-1,\tau)$ to $(-1,\tau')$ within $7000(c+a+W+3)^2$ units, where track 11 of τ' carries*

$a + W + 2$, every cell $\varsigma(x)$, $x \geq 0$, of the stack view holds 0, and of the rows only track 11 was written.

Proof. (a) left^{12} reaches cell$(0, 11)$, which holds 1 while the counter on track 11 is positive. A pass with the counter $q+1$ returns to the root (13), moves the counter back to q ($36q+66$), pushes 1 ($\leq 17(6|w'|+16)$ with $|w'| \leq |w|+n$) and steps back (13): at most $36q+102(|w|+n)+364$ units. After n passes the counter is 0 and the loop stops; right^{12} and $\mathsf{spush}(0)$ finish. Lemma I.6 bounds the whole by $5000(c + |w| + n + 1)^2$.

(b) right^{48} reaches $\varsigma(2)$, the presence cell of the entry 0, and sets it to 0, a *hole*; right^{16} reaches its value cell $\varsigma(3)$. Let I_q ($q \leq n$) be the configuration at the value cell $\varsigma(3 + 2q)$ in which the presence cell of the entry q is a hole, the other cells of the stack are as before, and track 2 carries q. The scanned value is 1 for $q < n$ and the separator 0 for $q = n$. A pass from I_q, $q < n$, steps to the hole and rewinds over the q entries before it to $\varsigma(0)$ and the root (G: $4q + 6$ virtual units), increments track 2 ($36q + 48$), then walks right over those q entries, which stop at the hole, fills the hole, makes a hole of the next presence cell and steps to its value (F: $3q + 11$ virtual units). By Lemma I.8 a pass costs at most $155q + 337$ units and ends in I_{q+1}. At I_n the loop stops and J fills the last hole, rewinds to $\varsigma(0)$ and returns to the root ($3n + 7$ virtual units). The total is at most $2000(c + n + 1)^2$.

(c) By Lemma I.163(b), with the property that the stack is 1^{a+q} and track 11 carries q, $\mathsf{wp}(\mathsf{spush}(1); \mathsf{inc}_{11})$ pushes $W + 2$ entries 1 within $(W + 3) \cdot 1200(c + a + W + 3)$ units (a run of the body costs at most $102(a + W + 2) + 272 + 36c + 66$), and zero_{11} resets track 11. Then right^{48} reaches $\varsigma(2)$, which holds 1 while the stack is not empty. A pass returns to the root, pops the top, increments track 11 and returns to $\varsigma(2)$; after $a + W + 2$ passes the stack is empty, track 11 carries $a + W + 2$, and the loop stops on $\varsigma(2) = 0$. Every popped entry was cleared, and $\varsigma(0)$, $\varsigma(1)$ and the cells beyond the stack were already 0. The bound follows from Lemmas I.167 and I.6. □

The parameters of the arithmetic are computed in unary, each mark of a growing counter paid for by one new row, so that every counter always fits the workspace.

Lemma I.170 (The parameters and the rows)**.** *For tracks $2 \leq g, k, u < 16$, all different, and $p \in \mathbb{N}$ put*

$$
\begin{aligned}
\mathsf{gadd}^{u}_{g\to k} &= \mathsf{zero}_u; \mathsf{rep}_{u,g}(\mathsf{ginc}_k; \mathsf{inc}_u),\\
\mathsf{addrow}^{(p)} &= \mathsf{addrow}; \cdots; \mathsf{addrow}; \mathsf{skip} \quad (p \text{ copies}),\\
\mathsf{Exp} &= \mathsf{zero}_{12}; \mathsf{zero}_{14}; \mathsf{rep}_{14,8}(\mathsf{gadd}^{11}_{10\to 12}; \mathsf{inc}_{14}),\\
\mathsf{Wid} &= \mathsf{zero}_4; \mathsf{gadd}^{11}_{9\to 4}; \mathsf{gadd}^{11}_{12\to 4}; \mathsf{gadd}^{11}_{8\to 4}; \mathsf{ginc}_4; \mathsf{ginc}_4; \mathsf{ginc}_4; \mathsf{ginc}_4; \mathsf{skip},\\
\mathsf{Res} &= \mathsf{zero}_{11}; \mathsf{rep}_{11,4}(\mathsf{addrow}^{(1000)}; \mathsf{inc}_{11}); \mathsf{addrow}^{(1020)},\\
\mathsf{Par} &= \mathsf{dup}_{0\to 3}; \mathsf{addrow}; \mathsf{Exp}; \mathsf{Wid}; \mathsf{Res},\\
\mathsf{shrink} &= \mathsf{seek}_3; \mathsf{right}^3; \mathsf{write}\,0; \mathsf{left}^{16}; \mathsf{loop}(\mathsf{put}(0^{16})); \mathsf{put}(0^{16});\\
&\qquad \mathsf{left}; \mathsf{findR}^{\uparrow 16}; \mathsf{home}_1,
\end{aligned}
$$

with put *of Lemma I.12 and* findR *of Lemma I.34.*

(a) *If τ is a working workspace of C rows on which track k carries $a \le C$ and track g carries $n \le C$, then* $\mathsf{gadd}^{u}_{g\to k}$ *runs from $(-1,\tau)$ within $10\,000(C+a+n+1)^2$ units to a working workspace of $C+n$ rows on which tracks k, u and g carry $a+n$, n and n; only tracks 0, 1, k and u are written.*

(b) *Let τ be a working workspace of r rows on which tracks 8, 9 and 10 carry N, W and σ, all at most $r+1$. Put $E=\sigma N$, $H=W+E+N+4$ and $R=r+1+E+H+1000(H+1)+20$. Then* Par *runs from $(-1,\tau)$ to $(-1,\tau')$ within*

$$\Gamma_{\mathrm{par}} = (48r+22)+(36r+47)+10^6(r+1+\sigma(N+1)+N+1)^3 + 10^6(r+1+2E+W+N+5)^2+10^9(R+H+1)^2$$

units, where τ' is a working workspace of R rows, tracks 3, 12 and 4 carry r, E and H, tracks 8, 9 and 10 still carry N, W and σ, only tracks 0, 1, 3, 4, 11, 12 and 14 are written, and every cell $x \ge 0$ is unchanged.

(c) *If τ is a working workspace of R rows on which track 3 carries r with $1 \le r < R$, then* shrink *runs from $(-1,\tau)$ within $1000(R+1)$ units to a working workspace of r rows that agrees with τ on the cells $x \ge 0$ and on every cell (q,k) with $q \le r$ and $k \ne 0$, and holds 0 in every cell of the rows $q > r$.*

Proof. (a) Let R_q ($q \le n$) say that the tape is a working workspace of $C+q$ rows on which tracks k, u and g carry $a+q$, q and n. After zero_u ($2u+7$ units) R_0 holds. From R_q, ginc_k (Lemma I.165(c)) and inc_u take at most $36(C+q)+36(a+q)+36q+200 \le 200(C+a+n+1)$ units and reach R_{q+1}; the origin marks up to C suffice for the repetition, whose cursor and guard stay in rows at most $n \le C$. By Lemma I.126(c), with r there equal to C, the repetition takes at most $(n+1)\cdot 400(C+a+n+1)$ units, and the total is at most $10\,000(C+a+n+1)^2$.

(b) $\mathsf{dup}_{0\to 3}$ ($48r+22$ units) copies the headers onto track 3, which then carries r; addrow ($36r+47$) makes the workspace have $r+1$ rows. Exp repeats N times, with track 14 as cursor and the guard 8, the addition of σ to track 12. Put $D = r+1+\sigma(N+1)+N+1$. The pass $q < N$ starts with $r+1+q\sigma$ rows and $q\sigma$ on track 12, so by (a) it costs at most $40\,000D^2+36D+66 \le 50\,000D^2$ units, and by Lemma I.126(c) the repetition leaves E on track 12 and $r+1+E$ rows within $(N+1)(50\,000D^2+200D)$ units; so Exp takes at most 10^6D^3. Wid adds W, E and N to the empty track 4 and four more marks, each mark with a new row (by (a) and Lemma I.165(c)): track 4 carries H and there are $r+1+E+H$ rows. With $D' = r+1+2E+W+N+5$, each of the three additions costs at most $10\,000(2D')^2$ by (a), and each ginc_4 at most $200D'$, so Wid takes at most $10^6D'^2$ units. Res adds 1000 rows for each of the H marks of track 4 and then 1020 rows, $1000(H+1)+20$ in all. Each addrow costs at most $36R+47$ units, so a pass of the repetition costs at most $50\,000(R+H+1)$, the repetition at most $(H+1)\cdot 50\,200(R+H+1)$ (Lemma I.126(c); there are at least H rows), and Res at most $10^9(R+H+1)^2$. The counters on tracks 8, 9 and 10 are only read.

(c) seek_3 ($18r+6$) and right^3 (4) reach the header $(r,0)$, which is set to 0; left^{16} reaches $(r+1,0)$. The headers of the rows $r+1,\dots,R-1$ hold 1 and that of row R holds 0, so the loop clears these rows with $\mathsf{put}(0^{16})$ (34 units a row with the test, Lemma I.12(c)), and the last put clears row R. The head is then at $(R+1,0)$; left reaches $(R+1,1)$. Along track 1 the rows $R+1, R, \dots, r+1$ now hold 0 and row $r \ge 1$ holds its origin mark 1, so $\mathsf{findR}^{\uparrow 16}$ stops at $(r,1)$

within $17(6(R+1-r)+5)$ units (Lemmas I.34(a) and I.8), and home_1 ($18r+5$) returns to the root. In all at most $36r+34R+102(R+1)+153 \le 1000(R+1)$ units. The headers now carry r, the origin marks reach r, and the rows beyond r are blank. □

I.12.4 The arithmetic area of a coordinate

The arithmetic of a coordinate is done in twelve-cell rows laid along track 7, whose value cells hold the binary stream between stages and are saved on the stack while a coordinate is computed.

Definition I.171 (The bank of a coordinate)**.** For a tape τ the *bank view* is the tape $\beta(x) = \tau(-9-16x)$, $x \in \mathbb{Z}$; for $x \ge 0$ this is the cell $(x,7)$, and for $x<0$ it is the positive cell $-9-16x$. The *bank of width H* is the block of twelve-cell rows at 12 of β (Definition I.20); it is *laid out* if it has H marked rows and is closed on the left, that is, $\beta(0)=0$ and $\beta(12+12k)=[k<H]$ for $k \le H$. Column c of the bank *reads* the integer z if the word $\beta(12+12k+c)$, $k<H$, has $\mathrm{sval}_H = z$ (Definition I.3(d)). The cell $\beta(10)$, column 10 of the row before row 0, is the *guard*.

The bank view and the stack view use the cells $\equiv 7$ and $\equiv 15$ modulo 16; they share no cell with each other or with any other track.

Lemma I.172 (Calling the bank)**.** *Let* $\mathsf{bcall}(P) = \mathsf{left}^{200}; P^{\downarrow 16}; \mathsf{home}_7$. *Let track 0 of τ carry the counter c, let τ have origin marks up to c, let β be its bank view, and let P run on β from $(12,\beta)$ in t units to (f,β') with $0 \le f \le c$. Then* $\mathsf{bcall}(P)$ *runs from $(-1,\tau)$ to $(-1,\tau')$ within $17t+18c+212$ units, where the bank view of τ' is β' and τ' agrees with τ on every cell that is not of the form $-9-16x$.*

Proof. left^{200} takes 201 units from -1 to $-201 = -9-16\cdot 12 = \mathrm{cell}(12,7)$. By Lemma I.9 with $o=-9$ and stride 16, $P^{\downarrow 16}$ runs within $17t$ units to $-9-16f = \mathrm{cell}(f,7)$, writing only cells $-9-16x$, and leaves the bank view β'. home_7 from row f takes $18f+11 \le 18c+11$ units (Lemma I.11(b)). □

The bank programs below run on β and are written for twelve-cell rows at h with $n \ge 1$ marked rows, closed on the left; they are used with $h=12$ and $n=H$. Here $\mathsf{bcp}_{s\to t}$ and bz_t are the copy and clear of Lemma I.138(a) (copy column s into column t in $(2s+2t+34)n+28$ units; set column t to 0 in $(2t+31)n+28$ units; both return to h), and the six-cell adder add of Lemma I.21 runs on the even cells as $\mathsf{add}^{\uparrow 2}$: its columns 1, 2, 3, 4 are the columns 2, 4, 6, 8 of the twelve-cell rows, so it adds columns 2 and 4 into column 8 with the carries in column 6 and the carry-in $h+6$.

Lemma I.173 (Bank routines)**.** *Let the twelve-cell rows at h have $n \geq 1$ marked rows and be closed on the left, let column 2 hold the word x and put $u = \mathrm{sval}_n(x)$. Define*

$$\begin{aligned}
\mathsf{badd} &= \mathsf{add}^{\uparrow 2};\mathsf{back};\mathsf{bcp}_{8\to 2}, \qquad \mathsf{bshl} = \mathsf{z}_{10};\mathsf{dbl};\mathsf{back},\\
\mathsf{bshr} &= \mathsf{fwd};\mathsf{left}^{10};\mathsf{branch}(S_0,S_1);\mathsf{loop}(D);\mathsf{right}^{12};\mathsf{bcp}_{8\to 2},\\
S_b &= \mathsf{right}^{8};\mathsf{write}\,b;\mathsf{left}^{10}, \qquad D = \mathsf{right}^{2};\mathsf{branch}(O_0,O_1),\\
O_b &= \mathsf{left}^{4};\mathsf{write}\,b;\mathsf{right}^{12};\mathsf{branch}(F_0,F_1),\\
F_b &= \mathsf{left}^{2};\mathsf{write}\,b;\mathsf{left}^{20},\\
\mathsf{bround} &= \mathsf{bz}_4;\mathsf{left}^{2};\mathsf{branch}(G_0,G_1);\mathsf{badd}, \qquad G_b = \mathsf{right}^{8};\mathsf{write}\,b;\mathsf{left}^{6}.
\end{aligned}$$

Each runs from h to h and keeps the presence marks.

(a) *If column 4 holds y and the cell $h+6$ holds c_0,* badd *leaves in column 2 the n lowest bits of $\mathrm{val}(x)+\mathrm{val}(y)+c_0$, in $128n+58$ units, writing only columns 2, 6 and 8.*

(b) bshl *leaves in column 2 the n lowest bits of $2\,\mathrm{val}(x)$, in $76n+51$ units, writing only columns 2 and 10; if $-2^{n-1} \leq 2u < 2^{n-1}$, column 2 then has $\mathrm{sval}_n = 2u$.*

(c) bshr *leaves in column 2 the word $x_1 \cdots x_{n-1}x_{n-1}$, whose sval_n is $\lfloor u/2 \rfloor$, and in the cell $h-2$ the bit $x_0 = u \bmod 2$, in $118n+76$ units, writing only columns 2, 8 and 10 (the cell $h-2$ is column 10 of row -1).*

(d) *If the cell $h-2$ holds g and $-2^{n-1} \leq u+g < 2^{n-1}$,* bround *leaves column 2 with $\mathrm{sval}_n = u+g$, in $167n+107$ units, writing only columns 2, 4, 6 and 8.*

Proof. (a) $\mathsf{add}^{\uparrow 2}$ takes at most $3(20n+1)$ units (Lemmas I.21 and I.8) and stops at the end row, back takes $14n+27$ (Lemma I.25(a)) and $\mathsf{bcp}_{8\to 2}$ takes $54n+28$. The sum is $128n+58$, and the value is Lemma I.21. (b) z_{10} (23 units) clears column 10 of row 0; by Lemma I.26 with $e=0$, dbl ($62n+1$) shifts column 2 by one place, and back ($14n+27$) returns. The word $0x_0 \cdots x_{n-2}$ has $\mathrm{sval}_n \equiv 2u$ modulo 2^n, so it is $2u$ in the stated range. (c) fwd ($14n+1$) walks to the end row, and left^{10} (11) to column 2 of row $n-1$, the top bit x_{n-1}; $S_{x_{n-1}}$ ($1+21$ with the test) writes it into column 10 of row $n-1$ and moves to the presence cell of row $n-1$. A pass of the loop at row k (presence 1) reads x_k in column 2, writes it into column 10 of row $k-1$, reads column 10 of row k, which holds x_{k+1} for $k < n-1$ and x_{n-1} for $k = n-1$, writes that bit into column 8 of row k and moves to the presence cell of row $k-1$: $1+3+1+5+1+13+1+3+1+21 = 50$ units. After the passes for $k = n-1, \ldots, 0$ the scanned cell is the closed left end, and right^{12} (13) returns to h: $64n+48$ units so far. Column 8 now holds $x_1 \cdots x_{n-1}x_{n-1}$ and the cell $h-2$ the bit x_0; the copy into column 2 adds $54n+28$. For the value, write $u = x_0 + 2v$ with $v = \sum_{1 \leq i < n-1} x_i 2^{i-1} - x_{n-1}2^{n-2}$; this v is the sval_n of the new word, and $u \equiv x_0$ modulo 2, so $v = \lfloor u/2 \rfloor$ (for $n=1$ the new word is x_0, and both sides are $-x_0$). (d) bz_4 ($39n+28$) clears column 4; left^2, the branch on g and G_g (21 units) copy g into the carry-in $h+6$ and return; badd ($128n+58$) adds. The value $\mathrm{val}(x)+g$ modulo 2^n has $\mathrm{sval}_n = u+g$ in the stated range, as in Corollary I.22(c). □

The last bank routine writes the signed word of the result in a form that can be read off along the column 6. It uses the stride-6 view of the bank, in which the twelve-cell rows become

two-cell rows whose presence cells are the presence cells of the bank and whose value cells are its column 6.

Lemma I.174 (Emitting the signed word)**.** *Put*

$$\begin{aligned}\mathsf{bemit} &= \mathsf{bcp}_{2\to 6};\mathsf{fwd};\mathsf{left}^6;\mathsf{branch}(P_0,P_1);\mathsf{left}^{18};\mathsf{sem}^{\uparrow 6},\\ P_b &= \mathsf{right}^6;\mathsf{back};\mathsf{left}^{18};\mathsf{write}\,b;\mathsf{right}^6;\mathsf{write}\,0;\mathsf{right}^6;\mathsf{write}\,0;\mathsf{right}^6,\\ \mathsf{sem} &= \mathsf{branch}(Q_0,Q_1),\qquad Q_b = \mathsf{right}^3;M_b;\mathsf{rewP};\mathsf{write}\,1;\mathsf{right};\mathsf{write}\,b;\mathsf{left},\end{aligned}$$

with $M_0 = \mathsf{scanP}$ *and* $M_1 = \mathsf{neg}$ *of Lemma I.24. Let the bank of width* $n \ge 1$ *be laid out on* β *(rows at* 12*), let column* 2 *read* z*, and let* z *fit* $n-1$*. Then* bemit *runs on* β *from* 12 *to* 0 *within* $428n+483$ *units, and afterwards the cells* $\beta(6+12k)$*,* $k \le n$*, hold the* $n+1$ *bits of* $\mathrm{sw}_{n-1}(z)$*.*

Proof. $\mathsf{bcp}_{2\to 6}$ ($50n+28$ units) copies the two's complement word of z into column 6. fwd ($14n+1$) and left^6 (7) reach column 6 of row $n-1$, its top bit $b = [z<0]$; the branch (1) and P_b ($14n+77$) return to the end row, walk back to 12, and write b into $\beta(-6)$ and 0 into $\beta(0)$ and $\beta(6)$, ending at 12. left^{18} (19) reaches $\beta(-6)$. Now consider the view $v(i) = \beta(6i)$. It holds b at -1 and 0 at 0 and 1, and since $\beta(12+12k) = [k<n]$ for $k \le n$ and $\beta(18+12k)$ is column 6 of row k, it holds at 2 the packed list of the word in column 6 (Definition I.122). On v, sem reads b at -1. If $b=0$, then $z \ge 0$ and the word is the n binary digits of z; Q_0 walks to 2 (4 units), over the list ($3n+1$), back to 0 (rewP, $4n+4$), and writes 1 at 0 and b at 1 (4 units). If $b=1$, the word has value $2^n+z>0$, and neg (within $30(n+1)$ units) replaces it by the word of value $2^n-(2^n+z) = |z|$, which has n digits because $|z|<2^n$; then Q_1 finishes in the same way. Either way v holds at 0 the packed list $\mathrm{sw}_{n-1}(z)$, so its value cells $v(1+2k) = \beta(6+12k)$, $k \le n$, hold the bits of $\mathrm{sw}_{n-1}(z)$, and the head is at $v(0) = \beta(0)$. On v, sem costs at most $34n+43$ units, so $\mathsf{sem}^{\uparrow 6}$ costs at most $7(34n+43)$ (Lemma I.8). The total is at most $316n+434 \le 428n+483$. □

Lemma I.175 (Laying out the bank)**.** *Put*

$$\begin{aligned}\mathsf{binit} &= \mathsf{clear}_7;\mathsf{zero}_5;\mathsf{inc}_5^{12};\mathsf{zero}_{11};\mathsf{rep}_{11,4}(\mathsf{bmark};\mathsf{inc}_5^{12};\mathsf{inc}_{11}),\\ \mathsf{bmark} &= \mathsf{seek}_5;\mathsf{left}^2;\mathsf{write}\,1;\mathsf{home}_7.\end{aligned}$$

Let τ *be a working workspace of* R *rows on which track* 4 *carries* H*, with* $12+12H \le R$*, and let* $\beta(x) = 0$ *for every* $x<0$*. Then* binit *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *within* $10^6(R+1)^2$ *units, where the bank of width* H *is laid out on* τ' *and every other cell of its bank view holds* 0*; tracks* 5*,* 11 *and* 4 *carry* $12+12H$*,* H *and* H*; only tracks* 5*,* 7 *and* 11 *are written; and the cells* $x \ge 0$ *are unchanged.*

Proof. clear_7 ($53R+22$ units, Lemma I.16) sets $(q,7)$ to 0 for $q \le R$; the rows beyond R are blank, and β vanishes on the negative indices, so now $\beta = 0$. zero_5 and inc_5^{12} put the counter 12 on track 5 ($17+3025$ units, Lemma I.14), and zero_{11} resets track 11. Let R_q say that track 5 carries $12+12q$, track 11 carries q, and $\beta(12+12k) = 1$ for $k<q$ while every other cell of β is 0. A pass from R_q, $q<H$: bmark goes to the hole $(12+12q,5)$, two cells left to $(12+12q,7) = \beta(12+12q)$, writes 1 and returns ($36(12+12q)+23$ units); inc_5^{12} and inc_{11} advance the two counters; at most $10\,000(R+1)$ units. By Lemma I.126(c) with the guard track 4, the H passes take at most $(H+1)\cdot 10\,200(R+1)$ units and end in R_H, in which the bank is laid out: $\beta(0)=0$, $\beta(12+12k)=1$ for $k<H$ and $\beta(12+12H)=0$. □

I.12.5 Horner's rule, the rounding and the write-back

The table entries are read bit by bit from the buffer, one signed word at a time, at the row that track 2 carries. Cells are copied between two cursors by short trips to the root.

Lemma I.176 (Reading and writing at a cursor)**.** *For tracks c and $v \geq 1$ put* $\mathsf{wrc}_{c,v}(b) = \mathsf{peek}_{c\to v}; \mathsf{write}\, b; \mathsf{home}_v$ *and* $\mathsf{rdc}_{c,v}(P_0, P_1) = \mathsf{peek}_{c\to v}; \mathsf{branch}(\mathsf{home}_v; P_0, \mathsf{home}_v; P_1)$*, and for a map $f : \{0,1\} \to \{0,1\}$ let* $\mathsf{cpb}^f_{c,v\to c',v'} = \mathsf{rdc}_{c,v}(\mathsf{wrc}_{c',v'}(f(0)), \mathsf{wrc}_{c',v'}(f(1)))$ *and* $\mathsf{lit}_k(p) = \mathsf{zero}_k; \mathsf{inc}^p_k$*. Let τ have origin marks up to q, and let track c carry q.*

(a) $\mathsf{wrc}_{c,v}(b)$ *runs from $(-1,\tau)$ to $(-1,\tau')$ in $36q + c + v + 9 + |c - v|$ units, τ' being τ with (q,v) set to b ($v \geq 2$). If $P_{\tau(q,v)}$ runs from $(-1,\tau)$ in t units to γ, then* $\mathsf{rdc}_{c,v}(P_0,P_1)$ *runs from $(-1,\tau)$ to γ in $36q + c + v + 9 + |c - v| + t$ units.*

(b) If also track c' carries q' and there are origin marks up to q', then $\mathsf{cpb}^f_{c,v\to c',v'}$ *sets (q', v') to $f(\tau(q,v))$ within $72\max(q,q') + 110$ units.*

(c) If $2 \leq k < 16$, $p \leq R$ and τ has origin marks up to R, $\mathsf{lit}_k(p)$ *makes track k carry p, writing only track k, within $100(p+1)(R+1)$ units.*

Proof. (a) $\mathsf{peek}_{c\to v}$ takes $18q + c + 4 + |c - v|$ units (Lemma I.126(a)), the write or the test one, and home_v from row q takes $18q + v + 4$. (b) is (a) twice, with all tracks below 16. (c) zero_k takes $2k + 7$ units and inc^p_k from 0 takes $18p^2 + (2k + 26)p + 1$ (Lemma I.14). □

Lemma I.177 (Loading a table entry)**.** *Put*

$$\mathsf{load} = \mathsf{dup}_{2\to 15}; \mathsf{rdc}_{2,13}(\mathsf{lb}(0), \mathsf{lb}(1)); \mathsf{dup}_{15\to 2},$$
$$\mathsf{lb}(b) = \mathsf{lit}_5(18); \mathsf{wrc}_{5,7}(b); \mathsf{inc}_2; \mathsf{lit}_5(16); \mathsf{lit}_{11}(0); \mathsf{rep}_{11,9}(\mathsf{lbit}(f_b)); \mathsf{lbit}(f_b); \mathsf{rep}_{11,4}(\mathsf{lbit}(e_b)),$$
$$\mathsf{lbit}(f) = \mathsf{cpb}^f_{2,13\to 5,7}; \mathsf{inc}_2; \mathsf{inc}^{12}_5; \mathsf{inc}_{11},$$

where $f_b(x) = x \oplus b$ and $e_b(x) = b$. Let τ be a working workspace of R rows on which tracks 2, 4 and 9 carry p, H and W, with $W + 1 \leq H$, $p + 1 + H \leq R$ and $16 + 12H \leq R$, and let the cells $(p + i, 13)$, $i < W + 2$, hold $\mathrm{sw}_W(z)$ for an integer z that fits W. Then load *runs from $(-1,\tau)$ to $(-1,\tau')$ within $20\,000(H+1)(R+1)$ units, where column 4 of the bank of τ' holds the H lowest binary digits of $|z|$, each complemented if $z < 0$, the cell $\beta(18)$ holds $[z < 0]$, track 2 carries p again, τ' is a working workspace of R rows, and τ' differs from τ only on tracks 2, 5, 11 and 15 and in the columns 4 and 6 of the twelve-cell rows at 12 of the bank view; no cell $x \geq 0$ is written.*

Proof. $\mathsf{dup}_{2\to 15}$ ($76R + 50$ units) saves track 2 on track 15. The test reads the sign bit $b = \tau(p, 13) = [z < 0]$ (Lemma I.176(a)). In $\mathsf{lb}(b)$, the first two parts write b into $(18, 7) = \beta(18)$, the carry-in of bank row 0; inc_2 moves the read cursor to the first digit $p + 1$; and the two literals point track 5 at $(16, 7) = \beta(16)$, column 4 of bank row 0, and reset track 11: at most $5000(R + 1)$ units. Let L_i ($i \leq H$) say that tracks 2, 5 and 11 carry $p + 1 + i$, $16 + 12i$ and i, and that column 4 of the bank rows $k < i$ holds the bits $\beta(16 + 12k) = \phi_k$, where ϕ_k is the digit k of $|z|$ (which is 0 for $k > W$) complemented if $b = 1$. From L_i, $\mathsf{lbit}(f)$ copies f of the cell $(p + 1 + i, 13)$ into $(16 + 12i, 7)$ and advances the three counters by 1, 12 and 1: at most $3000(R + 1)$ units (Lemma I.176(b) and Lemma I.14). With $f = f_b$ and $i \leq W$ the copied bit

is the digit i of $|z|$ xor b, which is ϕ_i; with $f = e_b$ and $W < i < H$ it is b, which is ϕ_i because the digit is 0. So $\mathsf{rep}_{11,9}(\mathsf{lbit}(f_b))$ (guard 9, Lemma I.126(c)) leads from L_0 to L_W, one more $\mathsf{lbit}(f_b)$ to L_{W+1}, and $\mathsf{rep}_{11,4}(\mathsf{lbit}(e_b))$ (guard 4, Lemma I.163(a) from $s = W+1$) to L_H, within $(H+2)\cdot 3200(R+1)$ units in all. Finally $\mathsf{dup}_{15\to 2}$ $(76R+50)$ restores track 2, which carried p. The total is at most $20\,000(H+1)(R+1)$. □

Lemma I.178 (Adding a table entry, and the Horner loop)**.** *Put*

$$\begin{aligned}\mathsf{addc} &= \mathsf{load};\mathsf{bcall}(\mathsf{badd}), \qquad \mathsf{lsh} = \mathsf{zero}_{11};\mathsf{rep}_{11,10}(\mathsf{bcall}(\mathsf{bshl});\mathsf{inc}_{11}),\\ \mathsf{str} &= \mathsf{zero}_{11};\mathsf{rep}_{11,6}(\mathsf{inc}_2;\mathsf{inc}_{11}),\\ \mathsf{Horner} &= \mathsf{zero}_{14};\mathsf{addc};\mathsf{rep}_{14,8}(\mathsf{lsh};\mathsf{str};\mathsf{addc};\mathsf{inc}_{14}).\end{aligned}$$

Let τ be a working workspace of R rows with $12 + 1000(H+1) \le R$, on which the bank of width H is laid out.

(a) Under the hypotheses of Lemma I.177, if column 2 reads x and $-2^{H-1} \le x+z < 2^{H-1}$, then addc *runs within $50\,000(H+1)(R+1)$ units to a working workspace of R rows on which column 2 reads $x+z$, the bank is still laid out and track 2 carries p; only tracks 2, 5, 7, 11 and 15 are written, no positive cell off the bank view, and no bank cell $\beta(x)$ with $x < -R$.*

(b) If track 10 carries $\sigma \le R$, column 2 reads x and $-2^{H-1} \le 2^q x < 2^{H-1}$ for $q \le \sigma$, then lsh *runs within $5000(\sigma+1)(R+H+1)$ units to a working workspace of R rows on which column 2 reads $2^\sigma x$ and the bank is still laid out; only tracks 7 and 11 are written, no positive cell off the bank view, and no bank cell $\beta(x)$ with $x < -R$.*

(c) If track 6 carries L and track 2 carries p with $p + L \le R$, str *runs within $1000(L+1)(R+1)$ units to a tape on which track 2 carries $p+L$; only tracks 2 and 11 are written, and no cell $x \ge 0$.*

(d) Let $z_0, \dots, z_N$ fit W, put $E = \sigma N$, $H = W + E + N + 4$ and $T_k = \sum_{l\le k} 2^{\sigma(k-l)} z_l$, let column 2 hold 0, and let tracks 2, 4, 6, 8, 9 and 10 carry $j(W+2)$, H, $d(W+2)$, N, W and σ, where $j < d$, $(kd+j+1)(W+2)+H \le R$ for $k \le N$ and $\sigma \le R$. Let the cells $((kd+j)(W+2)+i, 13)$, $i < W+2$, hold $\mathrm{sw}_W(z_k)$ for $k \le N$. Then Horner *runs from $(-1,\tau)$ within*

$$\Gamma_{\mathrm{H}} = 200\,000(N+1)\big(\sigma + d(W+2) + H + 1\big)(R+H+1)$$

units to a working workspace of R rows on which column 2 reads $T = T_N$ and the bank is still laid out; only tracks 2, 5, 7, 11, 14 and 15 are written, no positive cell off the bank view, and no bank cell $\beta(x)$ with $x < -R$.

Proof. (a) After load, column 4 holds y with $\mathrm{val}(y) = |z|$ if $z \ge 0$ and $2^H - 1 - |z|$ if $z < 0$, and the carry-in $\beta(18)$ holds $[z < 0]$. By Lemmas I.173(a) and I.172, bcall(badd) makes column 2 hold the H lowest bits of $\mathrm{val}(x) + |z|$ or of $\mathrm{val}(x) + 2^H - |z|$, a word whose value is congruent to $x+z$ modulo 2^H; in the stated range its sval_H is $x+z$. The call costs at most $17(128H+58)+18R+212$ units, so addc costs at most $50\,000(H+1)(R+1)$. The load writes no cell $x \ge 0$, and badd runs at most $128H + 58 \le 1000(H+1)$ units from 12; a unit moves the head by at most one cell, so it writes only cells $\beta(x)$ with $|x-12| \le 1000(H+1)$, which lie

between $-R$ and R. So the working workspace is kept and no bank cell below $-R$ is written. The same holds for every bank program of (b).

(b) Let R_q ($q \le \sigma$) say that column 2 reads $2^q x$ and track 11 carries q. A pass calls bshl (at most $17(76H+51)+18R+212$ units, Lemma I.173(b)) and increments track 11, reaching R_{q+1} within $4000(R+H+1)$ units. Lemma I.126(c) with the guard track 10 gives the bound.

(c) σ is replaced by the L marks of track 6: each pass increments tracks 2 and 11 ($36(p+q)+48+36q+66$ units), and Lemma I.126(c) gives at most $(L+1)\cdot 1000(R+1)$.

(d) zero_{14} resets track 14. Let H_k ($k \le N$) say that column 2 reads T_k, tracks 2 and 14 carry $(kd+j)(W+2)$ and k, the bank is laid out, and the other conditions of the lemma hold. Since $T_0 = 0 + z_0$ and $|T_0| < 2^{H-2}$ (Lemma I.158(d)), (a) with $p = j(W+2)$ leads to H_0. A pass from H_k, $k < N$: by (b), lsh makes column 2 read $2^\sigma T_k$, the range holding because $\sigma k + q \le E$ for $q \le \sigma$ (Lemma I.158(d)); by (c), str moves track 2 to $((k+1)d+j)(W+2)$; by (a), addc adds z_{k+1}, and $2^\sigma T_k + z_{k+1} = T_{k+1}$ is in range; inc_{14} advances track 14. Each pass reaches H_{k+1} within $100\,000(\sigma + d(W+2) + H + 1)(R+H+1)$ units, and Lemma I.126(c) with the guard track 8 gives the bound Γ_{H}. The written cells are those of (a)–(c). □

Lemma I.179 (The rounding)**.** *Put*

$$\mathsf{Round} = \mathsf{zero}_{14}; \mathsf{bcall}(\mathsf{left}^2; \mathsf{write}\,0; \mathsf{right}^2); \mathsf{rep}_{14,12}(\mathsf{bcall}(\mathsf{bshr}); \mathsf{inc}_{14}); \mathsf{bcall}(\mathsf{bround}); \mathsf{bcall}(\mathsf{bemit}).$$

Let τ be a working workspace of R rows with $12 + 1000(H+1) \le R$, on which the bank of width H is laid out, tracks 4 and 12 carry H and $E \le R$, and column 2 reads an integer T with $|T| < 2^{H-2}$. Put $V = \lfloor (2T + 2^E)/(2 \cdot 2^E) \rfloor$. Then Round *runs from $(-1,\tau)$ to $(-1,\tau')$ within*

$$\Gamma_{\mathrm{R}} = 20\,000(E+1)(R+H+1)$$

units to a working workspace of R rows, where the cells $\beta(6+12k)$, $k \le H$, of the bank view of τ' hold the bits of $\mathrm{sw}_{H-1}(V)$; only tracks 7 and 14 are written, no positive cell off the bank view, and no bank cell $\beta(x)$ with $x < -R$.

Proof. zero_{14} (35 units) and the first call (within $17 \cdot 7 + 18R + 212$ units) reset track 14 and the guard $\beta(10)$. Let S_k ($k \le E$) say that column 2 reads $q_k = \lfloor T/2^k \rfloor$, track 14 carries k, the guard holds 0 if $k = 0$ and $q_{k-1} \bmod 2$ if $k \ge 1$, and the bank is laid out. By Lemmas I.173(c) and I.158(c), a pass of the repetition, $\mathsf{bcall}(\mathsf{bshr})$ (at most $17(118H+76)+18R+212$ units) and inc_{14}, leads from S_k to S_{k+1} within $4000(R+H+1)$ units, and Lemma I.126(c) with the guard track 12 bounds the E passes by $(E+1)\cdot 4200(R+H+1)$. In S_E column 2 reads q_E and the guard holds g, with $g = q_{E-1} \bmod 2$ if $E \ge 1$ and $g = 0$ if $E = 0$. By Lemma I.158(c),(d), $q_E + g = V$ and $-2^{H-1} \le V < 2^{H-1}$, so $\mathsf{bcall}(\mathsf{bround})$ makes column 2 read V (Lemma I.173(d), within $17(167H+107)+18R+212$ units). Since V fits $H-1$, $\mathsf{bcall}(\mathsf{bemit})$ writes $\mathrm{sw}_{H-1}(V)$ into the cells $\beta(6+12k)$ (Lemma I.174, within $17(428H+483)+18R+212$ units). Each bank program called here runs at most $428H + 483 \le 1000(H+1)$ units from 12 on the bank view, so, as in the proof of Lemma I.178(a), it writes only cells $\beta(x)$ with $-R \le x \le R$. The sum of the bounds is at most Γ_{R}. □

Lemma I.180 (Writing the word back). *Put* $\mathsf{wb} = \mathsf{cpb}^{\mathrm{id}}_{5,7\to 2,13};\mathsf{inc}_2;\mathsf{inc}^{12}_5;\mathsf{inc}_{11}$ *and* $\mathsf{Wback} = \mathsf{lit}_5(6);\mathsf{zero}_{11};\mathsf{wp}(\mathsf{wb})$. *Let τ be a working workspace of R rows on which tracks 2 and 9 carry a and W, with $W+1\le H$, $a+W+2\le R$ and $6+12(W+2)\le R$, and let the cells $\beta(6+12k)$, $k\le H$, hold $\mathrm{sw}_{H-1}(V)$. Then* Wback *runs from $(-1,\tau)$ to $(-1,\tau')$ within $\Gamma_{\mathrm{W}}=10\,000(W+3)(R+1)$ units, where the cells $(a+k,13)$, $k<W+2$, hold the bits of $\mathrm{sw}_W(V)$, every other cell of track 13 is unchanged, track 2 carries $a+W+2$, only tracks 2, 5, 11 and 13 are written, and the bank view and the cells $x\ge 0$ are unchanged.*

Proof. The first $W+2$ bits of $\mathrm{sw}_{H-1}(V)$ and of $\mathrm{sw}_W(V)$ agree: both are $[V<0]$ followed by the lowest binary digits of $|V|$, and $W+1\le H$. $\mathsf{lit}_5(6)$ points track 5 at $(6,7)=\beta(6)$ and zero_{11} resets track 11. Let Q_k ($k\le W+2$) say that tracks 2, 5 and 11 carry $a+k$, $6+12k$ and k and that the cells $(a+i,13)$, $i<k$, hold the first k bits. A run of wb from Q_k copies $\beta(6+12k)$ into $(a+k,13)$ and advances the counters, within $3000(R+1)$ units (Lemma I.176(b)). By Lemma I.163(b) the $W+2$ runs take at most $(W+3)\cdot 3200(R+1)$ units, and with the first two parts at most Γ_{W}. □

I.12.6 Preparing and restoring the workspace

For the rest of the section fix $j<d$, a tape τ in the state C_j, and put

$$a=j(W+2),\qquad E=\sigma N,\qquad H=W+E+N+4,\qquad R=r+1+E+H+1000(H+1)+20,$$

and $\theta_k=\theta^r_k(\tau)$ for the snapshots of τ. Since the buffer word has $\ell=(N+1)d(W+2)\le r$ bits, $a+W+2\le d(W+2)\le r$; also $r\ge 2$, $12+1000(H+1)\le R$, and $R\ge r+H+1$. The *saved word* is

$$\Omega = 1^a\,0\,\theta_3\theta_2\theta_4\theta_5\theta_7\theta_{12}\theta_{14}\theta_{15}\,0^R,$$

of length $|\Omega|=a+1+8(r+1)+R$: the address, a separator, the eight borrowed tracks and R entries of padding. For $\epsilon=+$ and $-$ put $\mathsf{pad}^\epsilon=\mathsf{dup}_{0\to 14};\mathsf{zero}_{11};\mathsf{rep}_{11,14}(\mathsf{m}^\epsilon;\mathsf{inc}_{11})$ with $\mathsf{m}^+=\mathsf{spush}(0)$ and $\mathsf{m}^-=\mathsf{spop}$; on a working workspace of R rows it pushes, or pops, R entries 0 within $10\,000(R+|w|+1)^2$ units, w being the stack before (Lemmas I.126 and I.167; track 14 carries R after the copy of the headers).

Lemma I.181 (Preparation). *Put* $\mathsf{Prep} = \mathsf{sav};\mathsf{bkp};\mathsf{Par};\mathsf{pad}^+;\mathsf{binit};\mathsf{zero}_2;\mathsf{cpre}$ *with* $\mathsf{bkp}=\mathsf{save}_3;\mathsf{save}_2;\mathsf{save}_4;\mathsf{save}_5;\mathsf{save}_7;\mathsf{save}_{12};\mathsf{save}_{14};\mathsf{save}_{15};\mathsf{skip}$. *Then* Prep *runs from $(-1,\tau)$ to $(-1,\tau_1)$ within*

$$\begin{aligned}\Gamma_{\mathrm{P}} &= 5000(r+a+1)^2+88\,001(r+a+1+8(r+1)+1)^2+\Gamma_{\mathrm{par}}\\ &\quad+10\,000(R+a+1+8(r+1)+1)^2+10^6(R+1)^2+11+2000(R+a+1)^2\end{aligned}$$

units (Γ_{par} of Lemma I.170(b)), where

(i) τ_1 is a working workspace of R rows on which tracks 3, 12, 4, 8, 9, 10 and 2 carry r, E, H, N, W, σ and a;

(ii) the bank of width H is laid out on τ_1, and every other cell of its bank view holds 0;

(iii) τ_1 holds the stack Ω;

(iv) τ_1 *agrees with* τ *on tracks* 6, 8, 9, 10 *and* 13, *and every cell* $x \geq 0$ *that is not of the form* $15 + 16i$ *holds* 0.

Proof. τ is a workspace of r rows with origin marks up to r, so it is a working workspace of r rows, its positive half is blank and it holds the empty stack. Track 11 carries $a \leq r$. By Lemma I.169(a), sav leaves the stack 1^a0 and track 11 at 0; it wrote only track 11 among the rows, which is not saved, so by Lemma I.168(a) bkp pushes $\theta_3\theta_2\theta_4\theta_5\theta_7\theta_{12}\theta_{14}\theta_{15}$, the snapshots of τ itself. The counters on tracks 8, 9 and 10 are at most $r + 1$ (Lemma I.4), so Lemma I.170(b) applies to Par: R rows, and r, E, H on tracks 3, 12, 4. pad^{+} pushes 0^R. Every positive cell $-9 - 16x$ is still blank, so Lemma I.175 lays out the bank and clears the rest of the bank view. zero_2 (11 units) and, by Lemma I.169(b), cpre put a on track 2. The bounds of the seven parts are the seven terms of Γ_{P}, with $|w| = 0$, $a + 1$, $a + 1 + 8(r + 1)$ for the three uses of the stack lemmas. The tracks written are 0, 1, 2, 3, 4, 5, 7, 11, 12 and 14, and no positive cell outside the stack and the bank. □

Lemma I.182 (Restoration)**.** *Put*

$$\begin{aligned}
\mathsf{Restore} &= \mathsf{clear}_7; \mathsf{cpos}; \mathsf{pad}^{-}; \mathsf{zero}_{11}; \mathsf{rtail}; \mathsf{shrink}; \mathsf{rest}_3; \mathsf{spop}; \mathsf{fin},\\
\mathsf{rtail} &= \mathsf{skip}; \mathsf{rest}_{15}; \mathsf{rest}_{14}; \mathsf{rest}_{12}; \mathsf{rest}_7; \mathsf{rest}_5; \mathsf{rest}_4; \mathsf{rest}_2,\\
\mathsf{cpos} &= \mathsf{right}^8; \mathsf{write}\,0; \mathsf{right}^{16}; \mathsf{write}\,0; \mathsf{right}^{24};\\
&\qquad \mathsf{loop}(\mathsf{left}^8; \mathsf{write}\,0; \mathsf{right}^{16}; \mathsf{write}\,0; \mathsf{right}^{24}); \mathsf{left}^{32}; \mathsf{loop}(\mathsf{left}^{32}); \mathsf{left}^{16}.
\end{aligned}$$

Let τ_4 *be a working workspace of* R *rows on which tracks* 3 *and* 9 *carry* r *and* W, *which holds the stack* Ω, *whose positive cells not of the forms* $15 + 16i$ *and* $-9 - 16x$ *hold* 0, *and whose bank view vanishes below* $-R$. *Then* $\mathsf{Restore}$ *runs from* $(-1, \tau_4)$ *to* $(-1, \tau_5)$ *within*

$$\Gamma_{\mathrm{X}} = 10^8(R + a + W + r + 1)^2$$

units, where τ_5 *is a workspace of* r *rows with origin marks up to* r, *track* 11 *carries* $a + W + 2$, $\tau_5(q, k) = \tau(q, k)$ *for* $q \leq r$ *and* $k \in \{2, 3, 4, 5, 7, 12, 14, 15\}$, *and* $\tau_5(q, k) = \tau_4(q, k)$ *for* $q \leq r$ *and every other track* $k \neq 0, 11$.

Proof. clear_7 ($53R + 22$ units) clears track 7 in the rows $q \leq R$. cpos goes to the cells $7 = \beta(-1)$ and $23 = \beta(-2)$ and clears them, and reaches $\varsigma(2) = 47$. While the presence cell $\varsigma(2 + 2k) = 47 + 32k$ of the entry k holds 1, a pass clears the bank cells $39 + 32k = \beta(-3 - 2k)$ and $55 + 32k = \beta(-4 - 2k)$ and moves to the next presence cell (54 units with its test). After the $L = |\Omega|$ passes it returns over the presence cells to $\varsigma(0) = 15$ and to the root: $88L + 105$ units in all, and $\beta(x) = 0$ now for $-2L - 2 \leq x \leq -1$, hence for all $x < 0$, since $L \geq R$. No stack cell is written. pad^{-} pops the padding, zero_{11} resets track 11, and by Lemma I.168(b) rtail restores tracks 15, 14, 12, 7, 5, 4 and 2 in the rows $q \leq r$ from their snapshots, track 3 carrying r throughout. By Lemma I.170(c), shrink returns to a working workspace of r rows, blank beyond row r; then rest_3 (Lemma I.168(b) with $R = r$) restores track 3, spop removes the separator, and by Lemma I.169(c), fin leaves $a + W + 2$ on track 11 and empties the stack view on the cells $x \geq 0$. These last three write only rows $q \leq r$ and stack cells. Now every positive cell holds 0: the stack cells by fin, the bank cells by cpos and the others by hypothesis; the root still holds 1; so τ_5 is a workspace of r rows with origin marks up to r. Tracks other than 0,

11 and the borrowed ones were written only in rows beyond r (by shrink). The bounds of the nine parts, with $L + R + 1 \le 10(R + a + W + r + 1)$ and $R + a + 9r + 10 \le 10(R + a + W + r + 1)$, add up to at most $10^6 + 2.1 \cdot 10^7 + 2 \cdot 10^5$ times $(R + a + W + r + 1)^2$, which is less than Γ_{X}. □

I.12.7 Proof of Theorem I.160

Proof of Theorem I.160. Keep the notation of Section I.12.6, and let $z_{k,j}$ be the entry j of the row k of the table. By Lemma I.181, Prep runs from $(-1, \tau)$ within Γ_{P} units to $(-1, \tau_1)$.

Horner. On τ_1 the bank is laid out with column 2 equal to 0, and tracks 2, 4, 6, 8, 9, 10 carry $a = j(W + 2)$, H, $d(W + 2)$, N, W, σ: track 6 is the presence track of the stored vector, kept from τ. By (iv) of Definition I.159 and Lemma I.181(iv), the cells $((kd + j)(W + 2) + i, 13)$, $i < W + 2$, hold $\mathrm{sw}_W(z_{k,j})$ for $k \le N$: only the words $0, \dots, j - 1$ of the first row have been replaced, and $kd + j \ge j$. Also $(kd + j + 1)(W + 2) + H \le \ell + H \le R$ and $\sigma \le r + 1 \le R$. So Lemma I.178(d) runs Horner within Γ_{H} units to τ_2, on which column 2 reads $T = \sum_l 2^{\sigma(N-l)} z_{l,j}$.

Rounding. $|T| < 2^{H-2}$ by Lemma I.158(d), and tracks 4 and 12 still carry H and $E \le R$. By Lemma I.179, Round runs within Γ_{R} units to τ_3, on which the cells $\beta(6 + 12k)$, $k \le H$, hold $\mathrm{sw}_{H-1}(V_j)$, where by Lemma I.158(a)

$$V_j = \left\lfloor \frac{2T + 2^E}{2 \cdot 2^E} \right\rfloor = \mathrm{near}\Big(\sum_{l \le N} 2^{-\sigma l} z_{l,j} \Big).$$

Write-back. Horner and Round wrote no stack cell and no row cell outside tracks 2, 5, 7, 11, 14 and 15. zero_2 (11 units) and cpre (Lemma I.169(b), within $2000(R + a + 1)^2$ units, the stack being $\Omega = 1^a 0 \cdots$) put a on track 2 again. By Lemma I.180, Wback runs within Γ_{W} units to τ_4, on which the cells $(a + k, 13)$, $k < W + 2$, hold $\mathrm{sw}_W(V_j)$ and the rest of track 13 is as on τ.

Restoration. τ_4 satisfies the hypotheses of Lemma I.182: it is a working workspace of R rows (tracks 0 and 1 were not written after Prep), tracks 3 and 9 carry r and W, the stack is Ω, the positive cells outside the stack and the bank were never written, and no bank cell $\beta(x)$ with $x < -R$ was written after Prep, so these still hold 0. So Restore runs within Γ_{X} units to $\tau_5 = \tau'$.

The state C_{j+1}. τ' is a workspace of r rows with origin marks up to r, and track 11 carries $a + W + 2 = (j + 1)(W + 2)$. The borrowed tracks 2, 3, 4, 5, 7, 12, 14, 15 agree with τ in the rows $q \le r$, and beyond row r both tapes are blank. The tracks 6, 8, 9 and 10 were never written, so they agree with τ too; with tracks 3, 4, 5, 2, 14 and 15 this proves that the protected tracks are kept, and with tracks 6, 7, 12 that (ii) and (iii) of Definition I.159 hold. Track 13 is that of τ_4: the word j of the first row now holds $\mathrm{sw}_W(V_j)$, which is the word j of $\mathrm{sw}_W(V)$, and every other cell is as on τ. So τ' is in the state C_{j+1}.

The count. The units add up to

$$\Gamma = \Gamma_{\mathrm{P}} + \Gamma_{\mathrm{H}} + \Gamma_{\mathrm{R}} + 11 + 2000(R + a + 1)^2 + \Gamma_{\mathrm{W}} + \Gamma_{\mathrm{X}}.$$

Put $Y = d + \sigma + N + W + r + 1$ and $U = 20\,000 Y^2$. Since $\sigma, N, W, r, d < Y$,

$$E \le Y^2, \qquad H \le 7Y^2, \qquad R \le r + 1 + E + 1001H + 1020 \le 10^4 Y^2,$$

$a \le r$, $W + 2 \le 2Y$ and $\sigma(N+1) \le Y^2$. Hence every factor in the terms of Γ — such as $R + a + W + r + 1$, $R + H + 1$, $\sigma + d(W+2) + H + 1$, $r + 1 + \sigma(N+1) + N + 1$ and $N + 1$ — is at most U, and each term is a constant times a product of at most three such factors. The constants of the terms add up to less than $1.2 \cdot 10^9$, and the linear terms of Γ_{par} and the two terms 11 are less than $100U$. So

$$\Gamma \le 2 \cdot 10^9 U^3 = 2 \cdot 10^9 \cdot 8 \cdot 10^{12}\, Y^6 < 10^{31}\, Y^6. \qquad \square$$

The exponent 6 is that of U^3: every term of the count is a constant times at most three factors, and each factor is at most quadratic in Y, because the exponent $E = \sigma N$ of the rounding, the width H of the arithmetic and the row count R of a coordinate are.

I.13 The step loop and the selected word

After the first vector the machine repeats one Taylor step S times and then looks for the answer. Where is it? After S steps the binary stream holds $\mathrm{sw}_W(Z_S)$, the d signed words of the last vector, and the request asks for coordinate i. The machine reads i from the next-to-last frame of the saved request, walks to bit $i(W+2)$ of the stream, copies the $W+2$ bits of word i to the front of the stream and ends it there. For the oscillator of Section I.11, with $d = 2$ and $i = 1$, the stream $\mathrm{sw}_W(Z_{S,0})\,\mathrm{sw}_W(Z_{S,1})$ becomes $\mathrm{sw}_W(Z_{S,1})$. This section proves the loop rule for the steps, constructs the selection and the emptying of the buffer, and composes the whole trajectory: the first vector, the S steps and the selection, with the exponent $762700(g+20)$.

I.13.1 The step contract and the loop

The notation is that of Section I.10.1. For an integer vector Z of length d write Z^+ for the vector that Lemma 15.35, with $w = N$, obtains from $Z_k = Z$: the rows $r_0 = Z$ and $r_{l+1} = \mathrm{near}\big(2^N \mathrm{next}(f, l)\big)$ with $f_j = r_j / 2^N$ for $l < N$, and then $Z^+ = \mathrm{near}\big(2^N S_N(f)\big)$. So $Z_{k+1} = Z_k^+$.

Definition I.183 (The step contract)**.** A program P *meets the step contract with exponent* g if for every problem with a coefficient field and every level n_0 there is a constant K, depending only on the problem and n_0, with the following property. Let $i < d$, m, r and $c < S$, let Z be an integer vector of length d with $\|Z/2^N\| \le M$, and let τ hold the stored vector Z on r rows, with the counter c on track 2 and a retained word of length S from row 0 on track 14. Then P runs from $(-1, \tau)$ to $(-1, \tau')$ in some number t of units, where

(a) τ' holds the stored vector Z^+ on the same r rows;

(b) $\tau'(q, 2) = \tau(q, 2)$ and $\tau'(q, 14) = \tau(q, 14)$ for every $q \le S$;

(c) $t \le K(X + r + 1)^g$.

The rows do not change, so the bound counts the units alone (Definition I.7). The stored step of Section I.12, the coefficient table followed by the Taylor sum, meets this contract (Theorem I.162).

Theorem I.184 (The step loop)**.** *Let P meet the step contract with exponent g, and let K_{s} be its constant for the problem and n_0. Let r, $i < d$ and m, and let τ hold the stored vector Z_0 on r rows, with the counter 0 on track 2 and a retained word of length S from row 0 on track 14. Then $\mathsf{repeat}_2(P)$ of Lemma I.18 runs from $(-1,\tau)$ to $(-1,\tau')$ in some number t of units, where τ' holds the stored vector Z_S on the same r rows and*

$$t + r + 1 \le \big(S(36S + K_{\mathrm{s}} + 58) + 18S + 40\big)(X + r + 1)^{g+1}.$$

Proof. Apply Lemma I.18 with $c = 2$, $r_0 = 0$, $n = S$, $b = K_{\mathrm{s}}(X + r + 1)^g$ and R_k, for $k \le S$, the property of holding the stored vector Z_k on r rows. Every tape with R_k has origin marks up to r, and $r \ge \rho_{\mathrm{traj}} \ge S$, so it has origin marks up to S. Let $k < S$ and let σ be a tape with R_k on which track 2 carries k and track 14 the retained word. Since $n \ge n_0$ and $k \le S$, Lemma 15.35 and Theorem 15.20 give $\|Z_k/2^N\| = \|y_k^{(n)}\| \le M$. So the step contract applies to σ with $c = k$: P runs from $(-1,\sigma)$ within b units to some $(-1,\sigma_1)$ that holds the stored vector $Z_k^+ = Z_{k+1}$ on r rows and agrees with σ on the cells $(q,2)$ and $(q,14)$, $q \le S$. The tape σ_2 obtained from σ_1 by setting $(k,2)$ to 1 and $(k+1,2)$ to 0 still has R_{k+1}: Definition I.101 does not mention track 2, and the two cells lie in rows $k+1 \le S \le r$, so the workspace and the origin marks are kept. This is the hypothesis of Lemma I.18, which gives a run of $\mathsf{repeat}_2(P)$ to a tape with R_S. Every tape with some R_k is a workspace of r rows and $S \le r$, so Lemma I.19 with $Y = X + r + 1 \ge r + 1$ gives the bound. □

So the loop adds 1 to the exponent of the step and puts the number S of steps into the constant only, however large the precision.

I.13.2 The selected word and the empty buffer

The selection is

$$\begin{aligned}
\mathsf{Select} &= \mathsf{Read};\mathsf{Slice},\\
\mathsf{Read} &= \mathsf{cur}_2;\mathsf{dup}_{5\to11};\mathsf{dup}_{8\to12};\mathsf{dup}_{7\to13};\mathsf{dup}_{9\to14};\mathsf{skip};\mathsf{clear}_{8,9};\mathsf{dup}_{2\to5};\mathsf{nat};\\
&\qquad \mathsf{dup}_{8\to2};\mathsf{dup}_{13\to7};\mathsf{dup}_{11\to5};\mathsf{dup}_{12\to8};\mathsf{dup}_{2\to13};\mathsf{dup}_{14\to9};\mathsf{skip},\\
\mathsf{Slice} &= \mathsf{Locate};\mathsf{Cut},
\end{aligned}$$

with cur_2 of Section I.10.2, dup of Lemma I.126, nat of Lemma I.128, and Locate and Cut of Lemma I.186. Read finds the frame of i, saves the counters m, N, W and the values of the stored vector on tracks 11 to 14, decodes i in unary, and copies everything back, with i on track 13. Slice moves a cursor on track 2 to row $i(W+2)$ and copies the $W+2$ bits of word i to the rows $0,\dots,W+1$ of the binary stream. The empty buffer is the program $\mathsf{empty} = \mathsf{clear}_{11,12}$ of Lemma I.125(a).

A word of the stream has $W+2$ bits, and the register W on track 9 counts all but two of them; the program $\mathsf{wp}(P) = \mathsf{rep}_{11,9}(P);P;P$ of Lemma I.163(b) runs P once per bit. Moving a cursor and copying bits use

$$\begin{aligned}
\mathsf{sk} &= \mathsf{inc}_2;\mathsf{inc}_{11}, \qquad \mathsf{Locate} = \mathsf{zero}_2;\mathsf{zero}_{12};\mathsf{rep}_{12,13}\big(\mathsf{zero}_{11};\mathsf{wp}(\mathsf{sk});\mathsf{inc}_{12}\big),\\
\mathsf{out}_b &= \mathsf{seek}_{11};\mathsf{shift}_{11\to6};\mathsf{write}\,1;\mathsf{left};\mathsf{write}\,b;\mathsf{home}_7,\\
\mathsf{xbit} &= \mathsf{peek}_{2\to7};\mathsf{branch}\big(\mathsf{home}_7;\mathsf{out}_0;\mathsf{sk},\ \mathsf{home}_7;\mathsf{out}_1;\mathsf{sk}\big),\\
\mathsf{Cut} &= \mathsf{zero}_{11};\mathsf{wp}(\mathsf{xbit});\mathsf{seal}_6, \qquad \mathsf{seal}_6 = \mathsf{seek}_{11};\mathsf{shift}_{11\to6};\mathsf{write}\,0;\mathsf{home}_6.
\end{aligned}$$

Lemma I.185 (Reading the coordinate). *Let τ hold the stored vector Z on r rows, for an integer vector Z of length d. Then* Read *runs from $(-1,\tau)$ to $(-1,\tau_1)$ within $20000(r+1)^2$ units, where track 13 of τ_1 carries i, tracks 5, 8 and 9 carry m, N and W, $\tau_1(q,7)=\tau(q,7)$ for every $q\le r$, and τ_1 is a workspace of r rows with origin marks up to r.*

Proof. The last two frames of the request are those of i and m, so by Section I.10.2 cur_2 leaves on track 2 the row s of the frame of i, within $106(r+1)$ units, and changes only tracks 2 and 14. The next four copies (Lemma I.126(b), at most $102r+76$ units each) put the cells of rows $j\le r$ of tracks 5, 8, 7 and 9 onto tracks 11, 12, 13 and 14. $\mathsf{clear}_{8,9}$ ($112r+51$ units) blanks tracks 8 and 9, and $\mathsf{dup}_{2\to5}$ makes track 5 carry s. The frame of i is followed by that of m, so $s+\mathrm{bl}(i)+1<|w_u|\le r+1$, and $i+1\le r$ by the reserve. So Lemma I.128 applies: nat runs within $10000(r+1)^2$ units, leaves the counter i on track 8, changes only tracks 5 to 10, and keeps every cell $x\ge -1$ and every row beyond r. Tracks 11 to 14 have not been written since the backup. The last six copies put the counter i on track 2, the old track 7 back on track 7, m on track 5, N on track 8, i on track 13 and W on track 9. Copies write only rows $j\le r$ of tracks other than 0 and 1, so the workspace and the origin marks are kept. The total is at most $106(r+1)+10(102r+76)+112r+53+10000(r+1)^2\le 20000(r+1)^2$. □

Lemma I.186 (Cutting out a word). *Let τ be a workspace of r rows with origin marks up to r on which track 13 carries i and track 9 carries W, with $(i+1)(W+2)\le r$, and let $b_j=\tau(i(W+2)+j,7)$ for $j<W+2$.*

(a) Locate *runs from $(-1,\tau)$ to $(-1,\tau')$ within $4000(i+1)(W+3)(r+1)$ units; track 2 of τ' carries $i(W+2)$, and only tracks 2, 11 and 12 change.*

(b) Slice = Locate; Cut *runs from $(-1,\tau)$ to $(-1,\tau'')$ within $7000(i+1)(W+3)(r+1)$ units; the binary stream of τ'' holds $b_0\cdots b_{W+1}$, τ'' is a workspace of r rows with origin marks up to r, and only tracks 2, 6, 7, 11 and 12 change.*

Proof. (a) zero_2 and zero_{12} take 42 units. For $q\le i$ let R_q say that track 2 carries $q(W+2)$, track 12 carries q, and tracks 9 and 13 are as in τ. From R_q, $q<i$, the body sets the counter of track 11 to 0 (29 units); by Lemma I.163(b), with R'_j saying that tracks 2 and 11 carry $q(W+2)+j$ and j, wp(sk) advances both $W+2$ times, each run of sk taking at most $100(r+1)$ units, so within $300(W+3)(r+1)$ units; and inc_{12} ($36q+68$) reaches R_{q+1}. So the body takes at most $400(W+3)(r+1)$ units, and Lemma I.126(c) with $c=12$, $g=13$ and $n=i$ bounds the loop by $(i+1)(400(W+3)(r+1)+200(r+1))$.

(b) After Locate, zero_{11} (29 units) sets track 11 to 0. For $j\le W+2$ let C_j say that tracks 2 and 11 carry $i(W+2)+j$ and j, that $(q,6)=1$ and $(q,7)=b_q$ for $q<j$, and that the cells $(i(W+2)+q,7)$, $j\le q<W+2$, still hold b_q. From C_j, xbit reads $(i(W+2)+j,7)=b_j$, returns, writes 1 and b_j into $(j,6)$ and $(j,7)$, and advances both counters: at most $500(r+1)$ units. Since $j\le i(W+2)+j$ and every later read is of a row above j, no unread cell is overwritten, also when $i=0$; so C_{j+1} holds. By Lemma I.163(b), wp(xbit) reaches C_{W+2} within $700(W+3)(r+1)$ units, and seal_6 ($36(W+2)+31$) writes the terminator $(W+2,6)=0$. Now the stream on tracks 6 and 7 from row 0 is $b_0\cdots b_{W+1}$. All writes are in rows $j\le(i+1)(W+2)\le r$ of tracks 2, 6, 7, 11 and 12. With (a) the total is at most $7000(i+1)(W+3)(r+1)$. □

Theorem I.187 (The selected word). *Let $r, i < d$ and m, let Z be an integer vector of length d whose entry Z_i fits W, and let τ hold the stored vector Z on r rows.*

(a) Select *runs from $(-1, \tau)$ to $(-1, \tau')$ in some number t of units, where the binary stream of τ' holds $\mathrm{sw}_W(Z_i)$, tracks 5, 8 and 9 carry the counters m, N and W, τ' is a workspace of the same r rows with origin marks up to r, and*

$$t + r + 1 \le 100000(r+1)^3 \le 100000(X + r + 1)^3.$$

(b) Select; empty *runs from $(-1, \tau)$ to $(-1, \tau'')$ in some number t of units, where τ'' has all the properties of τ' in (a), its output buffer holds the empty word with capacity r, and*

$$t + r + 1 \le 18\,800\,000\,(X + r + 1)^3.$$

The exponent of the stage is 3.

Proof. (a) By Lemma I.185, Read runs within $20000(r+1)^2$ units to a tape τ_1 with i on track 13, m, N, W on tracks 5, 8, 9, the cells $(q, 7)$, $q \le r$, of τ, and the same workspace and origin marks. By Definition I.101(c) the bits $b_0, \dots, b_{W+1}$ of $\mathrm{sw}_W(Z_i)$ are the cells $(i(W+2)+j, 7)$ of τ, and $(i+1)(W+2) \le d(W+2) < r$ by the reserve. So Lemma I.186 applies to τ_1: Slice runs within $7000(i+1)(W+3)(r+1)$ units to a tape τ' whose binary stream holds $b_0 \cdots b_{W+1} = \mathrm{sw}_W(Z_i)$, which keeps the workspace and the origin marks and changes only tracks 2, 6, 7, 11 and 12, so that tracks 5, 8 and 9 still carry m, N and W. Since $i + 1 \le r + 1$ and $W + 3 \le 3(r+1)$ (by the reserve, $W < r$),

$$t + r + 1 \le 20000(r+1)^2 + 21000(r+1)^3 + (r+1) \le 100000(r+1)^3,$$

and $r + 1 \le X + r + 1$.

(b) By Lemma I.125(a), empty takes $124r + 63 \le 187(r+1)$ units on the workspace of r rows, empties the output buffer with capacity r and changes only tracks 11 and 12, so the properties of (a) survive. Since $r + 1 \le 100000(X + r + 1)^3$, the total is at most $100000(X + r + 1)^3 + 187 \cdot 100000(X + r + 1)^3 = 18\,800\,000\,(X + r + 1)^3$. □

I.13.3 The whole trajectory

Let Step = Build; Sum be the stored step of Section I.12, with Build of Theorem I.121, and let $g = 118$ be the exponent of one entry of the coefficient table (Theorem I.119). By Theorem I.162, Step meets the step contract with exponent $g_{\mathrm{s}} = g + 18$. The trajectory stage is

$$\mathsf{Traj} = \mathsf{Init}; \mathsf{repeat}_2(\mathsf{Step}); \mathsf{Select}; \mathsf{empty}.$$

Theorem I.188 (The trajectory). *There is a constant K_{t}, depending only on the problem and n_0, with the following property. Let r, $i < d$ and m, and let τ be a scheduled request with r rows (Definition I.85). Then* Traj *runs from $(-1,\tau)$ to $(-1,\tau')$ in some number t of units, where for some r' the binary stream of τ' holds $\mathrm{sw}_W\big((Z_S)_i\big)$, tracks 5, 8 and 9 carry the counters m, N and W, the output buffer holds the empty word with capacity r', τ' is a workspace of r' rows, and*

$$t + r' + 1 \le K_{\mathrm{t}}(X + r + 1)^{762700(g+20)}.$$

With $g = 118$ the exponent is $762700 \cdot 138 = 105\,252\,600$.

Proof. Put $Y = X + r + 1 \ge 1$. By Theorem I.102, Init runs in t_1 units to a tape τ_1 that holds the stored vector Z_0 on r_1 rows, with the counter 0 on track 2 and the retained word of length S on track 14, and $t_1 + r_1 + 1 \le K_{\mathrm{i}} Y^{190674} \le K_{\mathrm{i}} Y^{190675}$. By Theorem I.184 with $P =$ Step and its constant K_{s}, $\mathsf{repeat}_2(\mathsf{Step})$ runs from τ_1 in t_2 units to a tape τ_2 that holds the stored vector Z_S on the same r_1 rows, and

$$t_2 + r_1 + 1 \le K_{\mathrm{l}}(X + r_1 + 1)^{g_{\mathrm{s}}+1} \le K_{\mathrm{l}}(Y + r_1 + 1)^{g_{\mathrm{s}}+1}, \qquad K_{\mathrm{l}} = S(36S + K_{\mathrm{s}} + 58) + 18S + 40.$$

Lemma 15.36(a) with Y for its N, $e = 190675$ and $g = g_{\mathrm{s}} + 1$ gives $t_1 + t_2 + r_1 + 1 \le LY^D$ with $L = K_{\mathrm{i}} + K_{\mathrm{l}}(K_{\mathrm{i}} + 1)^{g_{\mathrm{s}}+1}$ and $D = 190675(g_{\mathrm{s}} + 2)$. Every entry of Z_S fits W (Proposition I.32(a)), so Theorem I.187(b) applies to τ_2: Select; empty runs in t_3 units to a tape τ' with the stated properties on $r' = r_1$ rows, and $t_3 + r' + 1 \le K_{\mathrm{p}}(X + r_1 + 1)^3 \le K_{\mathrm{p}}(Y + r_1 + 1)^3$ with $K_{\mathrm{p}} = 18\,800\,000$. Lemma 15.36(a) again, with $e = D$ and $g = 3$, gives

$$t_1 + t_2 + t_3 + r' + 1 \le K_{\mathrm{t}} Y^{4D}, \qquad K_{\mathrm{t}} = L + K_{\mathrm{p}}(L + 1)^3.$$

Finally $4D = 4 \cdot 190675\,(g + 20) = 762700(g + 20)$, since $g_{\mathrm{s}} + 2 = g + 20$. □

The trajectory stage hands to the rounding stage of Section I.14 the stored integer $(Z_S)_i$ at precision N, the precision m and the width W; by Lemma 15.35 the rounding of $2^m (Z_S)_i / 2^N$ is the answer a_m.

I.14 Rounding and emitting the answer

When the step loop and the selection have run, the binary stream holds the signed word of one integer $z = (Z_S)_i$ at the working precision $w = N$, and the answer is $a_m = \mathrm{near}(2^m z / 2^w)$ (Lemma 15.35). Take $m = 1$, $w = 3$ and $z = -6$. Then $2^m z / 2^w = -3/2$, and $\mathrm{near}(-3/2) = \lfloor -1 \rfloor = -1$: a half is rounded upward, to -1 and not to -2. The machine never forms $2^m z$. It counts $w - m = 2$ on its unary registers and shifts the magnitude 6, with digits $0, 1, 1$ from the least significant, two places to the right. The digit that remains gives 1. The last digit shifted out, the *guard*, is 1; the OR of the digits shifted out before it, the *sticky* bit, is 0. For a nonnegative number a guard 1 always raises the magnitude by one. For a negative number it does so only when the sticky bit is 1 as well, because only then is the dropped part more than one half. Here the number is negative and the sticky bit is 0, so the magnitude stays 1 and the answer is -1. With $z = -7$ the guard and the sticky bit are both 1, the magnitude becomes 2, and the answer is $-2 = \mathrm{near}(-7/4)$.

This section writes that procedure as one program, proves that it leaves the signed word of $\mathrm{near}(2^m z / 2^w)$ in the binary stream, and proves its bound with exponent 7 (Theorem I.198). The arithmetic on magnitudes comes first, then the routines, then the stage.

I.14.1 Rounding a magnitude

For natural numbers n and k and a bit e put

$$\mu_e(n,k) = \Big\lfloor \frac{n}{2^k} \Big\rfloor + \begin{cases} [\,2^k \le 2(n \bmod 2^k)\,] & \text{if } e = 0, \\ [\,2^k < 2(n \bmod 2^k)\,] & \text{if } e = 1, \end{cases}$$

and write $(-1)^e n$ for the integer with magnitude n that is negative exactly when $e = 1$ and $n > 0$.

Lemma I.189 (Rounding on magnitudes). *Let $m \le w$, $k = w - m$, e a bit, $n \in \mathbb{N}$ and $z = (-1)^e n$. Then* $\mathrm{near}(2^m z / 2^w) = (-1)^e \mu_e(n,k)$.

Proof. By definition $\mathrm{near}(2^m z/2^w) = \lfloor (2^{m+1}z + 2^w)/2^{w+1} \rfloor = \lfloor (2z + 2^k)/2^{k+1} \rfloor$. Write $n = 2^k q + \rho$ with $0 \le \rho < 2^k$. If $e = 0$, then $(2n + 2^k)/2^{k+1} = q + (2\rho + 2^k)/2^{k+1}$, and the second term lies in $[1/2, 3/2)$; its floor is $[2\rho \ge 2^k]$. If $e = 1$, then $(-2n + 2^k)/2^{k+1} = -q + (2^k - 2\rho)/2^{k+1}$, and the second term lies in $(-1/2, 1/2]$; its floor is 0 when $2\rho \le 2^k$ and -1 when $2\rho > 2^k$. In both cases the value is $(-1)^e \mu_e(n,k)$, also when the magnitude is 0. □

The machine reaches $\mu_e(n,k)$ by k single shifts. For a word $x = x_0 x_1 \cdots x_{\ell-1}$ with $\ell \ge 1$ and bits g and s put

$$\theta(x, g, s) = \big(x_1 \cdots x_{\ell-1} 0,\ x_0,\ s \vee g \big)\ :$$

the word loses its lowest digit and gains a 0 on top, the guard becomes the digit that was lost, and the sticky bit absorbs the old guard. Write $(x^{(k)}, g_k, s_k) = \theta^k(x, 0, 0)$.

Lemma I.190 (Guard and sticky bit). *Let x be a word of length $\ell \ge 1$ and $n = \mathrm{val}(x)$. For every k, $|x^{(k)}| = \ell$ and, for each bit e,*

$$\mathrm{val}(x^{(k)}) + [\, g_k \wedge (\neg e \vee s_k) \,] = \mu_e(n,k).$$

Proof. The length is kept by every step. For $k = 0$ the guard is 0 and $\mu_e(n,0) = n$. For $k \ge 1$ we show by induction that $n = 2^k \mathrm{val}(x^{(k)}) + 2^{k-1} g_k + \rho_k$ with $0 \le \rho_k < 2^{k-1}$ and $s_k = [\rho_k > 0]$. For $k = 1$, $g_1 = x_0$, $s_1 = 0$, $\rho_1 = 0$ and $n = 2\,\mathrm{val}(x_1 \cdots x_{\ell-1}0) + x_0$. For the step write $x^{(k)} = bu$; then $\mathrm{val}(x^{(k)}) = 2\,\mathrm{val}(u0) + b$, $x^{(k+1)} = u0$, $g_{k+1} = b$, $s_{k+1} = s_k \vee g_k$, and with $\rho_{k+1} = 2^{k-1} g_k + \rho_k < 2^k$,

$$n = 2^{k+1}\,\mathrm{val}(x^{(k+1)}) + 2^k g_{k+1} + \rho_{k+1}, \qquad s_{k+1} = [\rho_k > 0] \vee g_k = [\rho_{k+1} > 0].$$

Hence $\lfloor n/2^k \rfloor = \mathrm{val}(x^{(k)})$ and $n \bmod 2^k = 2^{k-1} g_k + \rho_k$. Since $2\rho_k < 2^k$, the condition $2^k \le 2(n \bmod 2^k)$ holds exactly when $g_k = 1$, and $2^k < 2(n \bmod 2^k)$ exactly when $g_k = 1$ and $\rho_k > 0$, that is, when $g_k \wedge s_k$. So the bracket in $\mu_e(n,k)$ is g_k for $e = 0$ and $g_k \wedge s_k$ for $e = 1$, which is $[g_k \wedge (\neg e \vee s_k)]$ in both cases. □

I.14.2 The routines

Throughout this subsection the magnitude of the stored word is read from row 1: if the binary stream holds $\mathrm{sw}_W(z)$, then the word x of the $W + 1$ binary digits of $|z|$ lies on tracks 6 and 7 from row 1, and the cell $(0, 7)$ holds $[z < 0]$. Three cells of row 0 serve as flags: $(0, 10)$ for the

guard and later for the sign, $(0,3)$ for the sticky bit, and $(0,6)$, the presence cell of the sign, which is set to 0 while the magnitude is shifted. For a track k and a bit b let

$$\mathsf{set}_k(b) = \mathsf{left}^{k+1}; \mathsf{write}\, b; \mathsf{right}^{k+1};$$

it sets $(0,k)$ to b and returns to the root in $2k+5$ units (Lemma I.5). The decrement retr_k of Lemma I.66(b) moves the hole of a counter one row back.

Lemma I.191 (Consuming the precision)**.** *Let*

$$\mathsf{diff} = \mathsf{left}^6; \mathsf{loop}(\mathsf{home}_5; \mathsf{retr}_8; \mathsf{retr}_5; \mathsf{left}^6); \mathsf{home}_5.$$

Let τ be a workspace of r rows whose output buffer holds some word with capacity r, and on which track 5 carries m and track 8 carries $w \geq m$. Then diff *runs from $(-1,\tau)$ to $(-1,\tau')$ in $36mw + 93m + 17 \leq 150(m+w+1)^2$ units, where track 5 of τ' carries 0, track 8 carries $w-m$, and τ' differs from τ only in cells of tracks 5 and 8 in rows $j < w$.*

Proof. By Lemma I.4, $w \leq r+1$. left^6 takes 7 units to $-7 = \mathrm{cell}(0,5)$. For $j \leq m$ let I_j be the configuration at $\mathrm{cell}(0,5)$ whose tape is τ with track 5 carrying $m-j$ and track 8 carrying $w-j$, the cells between the old and the new holes set to 0. For $j < m$ the scanned cell $(0,5)$ holds 1, since $m-j \geq 1$. The pass returns to the root (9 units, Lemma I.11(b)), moves the hole of track 8 from row $w-j$ to row $w-j-1$ ($36(w-j-1)+59$) and that of track 5 from row $m-j$ to row $m-j-1$ ($36(m-j-1)+53$), by Lemma I.66(b), whose origin marks are there because $w-j-1 \leq r$; and left^6 returns to $(0,5)$ (7). With its test the pass takes $36(w+m-2j)+57$ units and ends in I_{j+1}. At I_m the cell $(0,5)$ is the hole of track 5, so it holds 0. By Lemma I.6 the loop ends in I_m, and its units are

$$\sum_{j<m} \big(36(w+m-2j)+57\big) + 1 = 36mw + 93m + 1.$$

home_5 from row 0 adds 9. Each of $36mw$, $93m$ and 17 is at most its coefficient times $(m+w+1)^2$. □

Lemma I.192 (Dropping digits)**.** *Let*

$$\begin{aligned}
\mathsf{flag} &= \mathsf{left}^{11}; \mathsf{branch}\big(\mathsf{right}^{11}; \mathsf{cap},\ \mathsf{right}^7; \mathsf{write}\,1; \mathsf{right}^4; \mathsf{cap}\big),\\
\mathsf{cap} &= \mathsf{left}^{24}; \mathsf{branch}(G_0, G_1), \qquad G_b = \mathsf{right}^{13}; \mathsf{write}\, b; \mathsf{right}^{11},\\
S_b &= \mathsf{right}^{16}; \mathsf{write}\, b; \mathsf{left}^{31},\\
\mathsf{shr} &= \mathsf{left}^{16}; \mathsf{loop}\big(\mathsf{left}; \mathsf{branch}(S_0, S_1)\big); \mathsf{right}^{15}; \mathsf{write}\,0; \mathsf{left}^{15},\\
\mathsf{drop}_1 &= \mathsf{flag}; \mathsf{left}^{23}; \mathsf{shr}; \mathsf{right}^{16}; \mathsf{scanR}; \mathsf{right}^7,\\
\mathsf{drop} &= \mathsf{left}^9; \mathsf{loop}(\mathsf{home}_8; \mathsf{drop}_1; \mathsf{retr}_8; \mathsf{left}^9); \mathsf{home}_8.
\end{aligned}$$

Let a word x of length $\ell \geq 1$ lie on tracks 6 and 7 from row 1, let $\tau(0,6) = 0$, $\tau(0,10) = g$ and $\tau(0,3) = s$.

(a) drop_1 *runs from $(-1,\tau)$ to $(-1,\tau')$ within $71\ell + 128 \leq 200(\ell+1)$ units, where, with $(x', g', s') = \theta(x,g,s)$, the word x' lies on tracks 6 and 7 from row 1, $\tau'(0,10) = g'$, $\tau'(0,3) = s'$, $\tau'(0,6) = 0$, and τ' differs from τ only in the cells $(0,3)$, $(0,10)$ and $(j,7)$, $1 \leq j \leq \ell$.*

(b) Let moreover τ be a workspace of r rows whose output buffer holds some word with capacity r, let $\ell \le r$, and let track 8 carry k. Then drop *runs from $(-1,\tau)$ to $(-1,\tau')$ within $500k(r+1)+23$ units, where, with $(x_k, g'_k, s'_k) = \theta^k(x,g,s)$, the word x_k lies on tracks 6 and 7 from row 1, $\tau'(0,10) = g'_k$, $\tau'(0,3) = s'_k$, $\tau'(0,6) = 0$, track 8 carries 0, and τ' differs from τ only on tracks 3, 7, 8 and 10, in rows $j \le r$.*

Proof. (a) left^{11} (12 units) reaches $(0,10)$, which holds g, and the branch tests it (1). If $g=0$, right^{11} (12) returns to the root; if $g=1$, right^{7} (8) reaches $(0,3)$, the write sets it to 1 and right^{4} (5) returns. Either way $(0,3)$ now holds $s \vee g$. Then cap walks to $(1,7) = -25$ (25), tests x_0 (1), returns to $(0,10)$ (14), writes x_0 there (1) and returns to the root (12): 53 units. So flag takes 78 or 80 units. left^{23} (24) reaches $(1,6)$. In shr, left^{16} (17) reaches $(2,6)$; for $1 \le j \le \ell$ let I_j be the configuration at $(j+1,6)$ whose tape has x_i in $(i,7)$ for $1 \le i < j$, the other cells as before. From I_j with $j < \ell$ the presence bit $(j+1,6)$ is 1; the pass reads x_j in $(j+1,7)$, writes it into $(j,7)$ and walks to $(j+2,6)$: $1+1+1+17+1+32 = 53$ units, ending in I_{j+1}. At I_ℓ the cell $(\ell+1,6)$ is the end of the word, and the loop stops (1). right^{15} (16) reaches $(\ell,7)$, which is set to 0, and left^{15} (16) returns: shr takes $53(\ell-1)+51$ units and leaves $x_1 \cdots x_{\ell-1}0$ on tracks 6 and 7 from row 1. right^{16} (17) reaches $(\ell,6)$, and scanR passes the presence marks of the rows $\ell, \dots, 1$ and stops at $(0,6) = 0$ ($18\ell+1$, Lemma I.10); right^{7} (8) returns to the root. In all $71\ell+126$ or $71\ell+128$ units.

(b) left^{9} (10) reaches $(0,8)$. By Lemma I.4, $k \le r+1$. For $j \le k$ let J_j be the configuration at $(0,8)$ whose tape holds $\theta^j(x,g,s)$ as in (a) and carries $k-j$ on track 8. From J_j with $j<k$ the scanned cell holds 1; the pass returns to the root (12), runs drop_1 (at most $200(\ell+1)$ units, by (a), which writes no cell of track 8), moves the hole of track 8 from row $k-j$ to row $k-j-1$ ($36(k-j-1)+59$, Lemma I.66(b)) and returns to $(0,8)$ (10). With its test it takes at most $1+12+200(\ell+1)+36r+59+10 \le 500(r+1)$ units, since $\ell \le r$, and ends in J_{j+1}. At J_k the cell $(0,8)$ holds 0. By Lemma I.6 the loop takes at most $500k(r+1)+1$ units, and home_8 from row 0 adds 12. □

Lemma I.193 (Choosing and adding the carry)**.** *Let*

$$\begin{aligned}
\mathsf{cflag}_0 &= \mathsf{skip}, \qquad C_b = \mathsf{left}^7;\mathsf{write}\,b;\mathsf{right}^{11},\\
\mathsf{cflag}_1 &= \mathsf{left}^{11};\mathsf{branch}\big(\mathsf{right}^{11},\ \mathsf{right}^7;\mathsf{branch}(C_0,C_1)\big),\\
\mathsf{incm} &= \mathsf{left}^{11};\mathsf{branch}\big(\mathsf{right}^{11},\ \mathsf{left}^{13};\mathsf{carry}^{\downarrow 16};\mathsf{home}_7\big).
\end{aligned}$$

(a) For a bit e, cflag_e *runs from $(-1,\tau)$ within 50 units to $(-1,\tau')$, where τ' is τ with $(0,10)$ set to $g \wedge (\neg e \vee s)$, $g = \tau(0,10)$ and $s = \tau(0,3)$.*

(b) Let a word x of length ℓ lie on tracks 6 and 7 from row 1, let $c = \tau(0,10)$, let τ have origin marks up to ℓ, and let $\mathrm{val}(x) + c < 2^\ell$. Then incm *runs from $(-1,\tau)$ to $(-1,\tau')$ within $100(\ell+1)$ units, where a word x' of length ℓ with $\mathrm{val}(x') = \mathrm{val}(x)+c$ lies on tracks 6 and 7 from row 1, and τ' differs from τ only in the cells $(j,7)$, $1 \le j \le \ell$.*

Proof. (a) For $e=0$ the skip takes one unit and $g \wedge (\neg 0 \vee s) = g$. For $e=1$, left^{11} (12) reaches $(0,10)$. If $g=0$, right^{11} returns (25 units in all). If $g=1$, then right^{7} (8) reaches $(0,3)$, the

branch reads s, and C_s walks back to $(0,10)$ (8), writes s and returns (12): 43 units. The new guard is $s = g \wedge s$.

(b) left^{11} (12) reaches $(0,10)$. If $c = 0$, right^{11} returns: 25 units. If $c = 1$, then $\mathrm{val}(x) + 1 < 2^\ell$, so x has a 0; write $x = 1^q 0 u$ with $q < \ell$. left^{13} (14) reaches $(1,7) = -25$. Track 7 from row 1 is the descending tape with origin cell$(1,7)$ and stride 16, on which carry runs in $3q+2$ units and turns 1^q0 into 0^q1 (Lemma I.23(a)); by Lemma I.9, $\mathsf{carry}^{\downarrow 16}$ does so on the track within $17(3q+2)$ units, ends at $(1+q,7)$ and writes no other cell. home_7 from row $1+q \le \ell$ takes $18(1+q)+11$. In all at most $69q + 90 \le 100(\ell+1)$ units, and $\mathrm{val}(0^q1u) = \mathrm{val}(1^q0u) + 1$. □

The three routines combine into the arithmetic part of the stage. It reads the sign of z once, in the cell $(0,7)$, and branches into two copies of the same program that differ only in the carry rule.

Lemma I.194 (Rounding the stored magnitude)**.** *Let*

$$\mathsf{rnd}_e = \mathsf{set}_6(0); \mathsf{set}_{10}(0); \mathsf{set}_3(0); \mathsf{drop}; \mathsf{cflag}_e; \mathsf{incm}; \mathsf{set}_{10}(e),$$
$$\mathsf{DropCarry} = \mathsf{diff}; \mathsf{left}^8; \mathsf{branch}(\mathsf{right}^8; \mathsf{rnd}_0,\ \mathsf{right}^8; \mathsf{rnd}_1).$$

Let $m \le w$, W and r be natural numbers and z an integer such that z and $\mathrm{near}(2^m z/2^w)$ fit W, and let τ be a workspace of r rows on which the binary stream holds $\mathrm{sw}_W(z)$, tracks 5, 8 and 9 carry m, w and W, and the output buffer holds the empty word with capacity r. Put $e = [z < 0]$ and $n = |z|$. Then $\mathsf{DropCarry}$ *runs from $(-1,\tau)$ to $(-1,\tau_1)$ in t_1 units, where a word x' of length $W+1$ with $\mathrm{val}(x') = \mu_e(n, w-m)$ lies on tracks 6 and 7 from row 1, $\tau_1(0,10) = e$, track 5 carries 0, track 9 carries W, the output buffer holds the empty word with capacity r, τ_1 is a workspace of r rows, and*

$$t_1 + r + 1 \le 2000(m + w + W + r + 1)^2.$$

Proof. The stream $\mathrm{sw}_W(z)$ has $W+2$ bits; its last presence mark is in row $W+1$, and the rows beyond r are blank, so $W+1 \le r$; also $w \le r+1$ (Lemma I.4). Put $k = w - m$ and let x be the $W+1$ digits of n. By Lemma I.191, diff takes $36mw + 93m + 17$ units and leaves 0 on track 5 and k on track 8. left^8 (9) reaches $(0,7)$, which holds e; the branch (1) and right^8 (9) lead to rnd_e at the root. The three writes (53 units) clear $(0,6)$, $(0,10)$ and $(0,3)$: now x lies on tracks 6 and 7 from row 1 with guard 0 and sticky bit 0. By Lemma I.192(b), with $\ell = W+1 \le r$, drop takes at most $500k(r+1)+23$ units and leaves $x^{(k)}$, g_k and s_k. By Lemma I.193(a), cflag_e (at most 50) sets $(0,10)$ to $c = g_k \wedge (\neg e \vee s_k)$. By Lemma I.190, $\mathrm{val}(x^{(k)}) + c = \mu_e(n,k)$, and by Lemma I.189, $\mu_e(n,k) = |\mathrm{near}(2^m z/2^w)| < 2^{W+1}$. So Lemma I.193(b) applies: incm (at most $100(W+2)$) leaves a word x' of length $W+1$ with value $\mu_e(n,k)$. The last write (25) puts e into $(0,10)$.

The cells written are $(0,3)$, $(0,6)$, $(0,10)$, cells of track 7 in rows $1,\dots,W+1$, of track 8 in rows $j < w$ and of track 5 in rows $j < m$, all in rows $j \le r$. Tracks 0, 1, 9, 11, 12 and 13 are not written, so τ_1 is a workspace of r rows, track 9 still carries W, and the output buffer still holds the empty word with capacity r. In all

$$t_1 \le 36mw + 93m + 187 + 500k(r+1) + 100(W+2),$$

and with $Y = m + w + W + r + 1$, $k \le w$ and $Y \ge 1$, $t_1 + r + 1 \le (36 + 93 + 500 + 101 + 388)Y^2 \le 2000Y^2$. □

Before the answer can be framed, the workspace must have room for the frame in the output buffer: the frame of a word of $W+2$ bits has $2(W+2)+1$ bits. The machine does not compute this length. It counts its own rows, which are more than $W+1$, and doubles them.

Lemma I.195 (Room for the frame)**.** *Let* $\mathsf{Room} = \mathsf{grow};\mathsf{hc}_{13};\mathsf{zero}_2;\mathsf{alloc};\mathsf{clear}_{12}$*, with* hc_{13} *the header count of Lemma I.91(a), which copies the row headers of track* 0 *into track* 12*, and* alloc *the growth routine of Lemma I.67. Let* τ *be a workspace of* r *rows on which the output buffer holds the empty word with capacity* r*, tracks* 5 *and* 9 *carry* 0 *and* W*, and a word* x *of length* $W+1$ *lies on tracks* 6 *and* 7 *from row* 1*. Put* $r' = 2r+3$*. Then* Room *runs from* $(-1,\tau)$ *to* $(-1,\tau')$ *in* t *units, where* τ' *is a workspace of* r' *rows, its output buffer holds the empty word with capacity* r'*,* x *still lies on tracks* 6 *and* 7 *from row* 1*,* $\tau'(0,10) = \tau(0,10)$*, tracks* 5 *and* 9 *still carry* 0 *and* W*,* $2(W+2)+1 \le r'$*, and*

$$t + r' + 1 \le 12000(r+2)^2.$$

Proof. As in Lemma I.194, $W+1 \le r$, which gives $2(W+2)+1 \le r'$. grow ($72r+99$, Lemma I.17) gives a workspace of $r+1$ rows with origin marks up to $r+1$. hc_{13} (within $300(r+2)$ units, Lemma I.91(a)) puts the value $r+1$ on track 12 and changes only tracks 12 and 13, and zero_2 (11) puts the hole of track 2 in row 0. Lemma I.67 with $R = c = r+1$ then gives, within $2000(2r+4)^2 = 8000(r+2)^2$ units, a workspace of $R+1+c = r'$ rows with origin marks up to r'; it writes only track 2, the headers and the new rows. clear_{12} ($63r'+33$) blanks track 12 in every row.

No routine writes track 11 in row 0, so track 11 still carries 0; track 12 is blank in rows $j \le r'$; so the buffer holds the empty word with capacity r'. The new rows carry the pattern β of Lemma I.17, which is 0 on tracks 6, 7, 9 and 10; the rows $j \le r$ are written only on tracks 0, 2, 12 and 13. So the word x, whose end mark lies in row $W+2 \le r+1$, the cell $(0,10)$, and the counters 0 and $W \le r$ on tracks 5 and 9 are kept. The count is $t + r' + 1 \le 8000(r+2)^2 + 500r + 936 \le 12000(r+2)^2$. □

The last routine writes the sign. It must not write the sign 1 in front of a zero magnitude: the answer word of 0 has the sign bit 0. The OR of the digits decides, and the program computes it with the following scan.

Lemma I.196 (The OR of a word)**.** *Let*

$$\mathsf{or} = \mathsf{loop}(\mathsf{left};\mathsf{branch}(\mathsf{left}^3;\mathsf{branch}(O_0,O_1),\ \mathsf{left}^3;O_1)), \qquad O_b = \mathsf{left}^{16};\mathsf{write}\,b;\mathsf{right}^4.$$

Let a word x *of length* ℓ *lie on tracks* 6 *and* 7 *from row* q*, and let* $a = \tau(q,10)$*. Then* or *runs from* $(\mathrm{cell}(q,6),\tau)$ *to* $(\mathrm{cell}(q+\ell,6),\tau')$ *in at most* $31\ell+1$ *units, where* τ' *is* τ *with* $(q+j+1,10)$ *set to* $a \vee x_0 \vee \cdots \vee x_j$ *for every* $j < \ell$*.*

Proof. Let I_j be the configuration at $(q+j,6)$ in which $(q+i+1,10)$ holds $a_{i+1} = a \vee x_0 \vee \cdots \vee x_i$ for $i < j$; put $a_0 = a$, so $(q+j,10)$ holds a_j. From I_j, $j < \ell$, the presence bit is 1; the pass reads x_j in $(q+j,7)$ and moves to $(q+j,10)$ ($1+1+1+4$ units). If $x_j = 0$ it reads a_j (1) and runs O_{a_j}; if $x_j = 1$ it runs O_1. O_b writes b into $(q+j+1,10)$ and moves to $(q+j+1,6)$ (23). In both cases the bit written is $a_j \vee x_j = a_{j+1}$, and the pass takes 31 or 30 units. At I_ℓ the presence bit is 0 (1 unit). □

Lemma I.197 (Writing the sign). *Let* $\mathsf{sign}(b) = \mathsf{write}\,1;\mathsf{left};\mathsf{write}\,b;\mathsf{right}$ *and*

$$\begin{aligned}\mathsf{Emit} &= \mathsf{left}^{27};\mathsf{write}\,0;\mathsf{right}^{4};\mathsf{or};\mathsf{left}^{4};\mathsf{branch}(E_{\mathrm{z}},E_{\mathrm{n}}),\\ E_{\mathrm{z}} &= \mathsf{home}_{10};\mathsf{left}^{7};\mathsf{sign}(0);\mathsf{right}^{7},\\ E_{\mathrm{n}} &= \mathsf{home}_{10};\mathsf{left}^{11};\mathsf{branch}\big(\mathsf{right}^{4};\mathsf{sign}(0);\mathsf{right}^{7},\ \mathsf{right}^{4};\mathsf{sign}(1);\mathsf{right}^{7}\big).\end{aligned}$$

Let a word x of length $W+1$ lie on tracks 6 and 7 from row 1, let $\tau(0,10)=e$, let τ have origin marks up to $W+2$, and put $y=(-1)^{e}\,\mathrm{val}(x)$. Then Emit *runs from $(-1,\tau)$ to $(-1,\tau')$ within $200(W+2)$ units, where the binary stream of τ' holds $\mathrm{sw}_W(y)$, and τ' differs from τ only in the cells $(0,6)$, $(0,7)$ and $(j,10)$, $1\le j\le W+2$.*

Proof. left^{27} (28) reaches $(1,10)=-28$, the write sets it to 0 (1), and right^{4} (5) reaches $(1,6)$. By Lemma I.196 with $q=1$ and $a=0$, or takes at most $31(W+1)+1$ units, ends at $(W+2,6)$, and leaves $o=[\mathrm{val}(x)>0]$ in $(W+2,10)$; the cell $(0,10)$ is not written. left^{4} (5) reaches $(W+2,10)$ and the branch reads o (1). If $o=0$, E_{z} returns to the root ($18(W+2)+14$), reaches $(0,6)=-8$ (8), writes 1 there and 0 into $(0,7)$ (4) and returns (8). If $o=1$, E_{n} returns to the root, reaches $(0,10)$ (12), reads e (1), moves to $(0,6)$ (5), writes 1 there and e into $(0,7)$ (4) and returns (8). In all at most $49(W+1)+103\le 200(W+2)$ units. Now $(0,6)=1$, $(0,7)=e\wedge o=[y<0]$, and x follows from row 1. Since $|y|=\mathrm{val}(x)<2^{W+1}$ and x has $W+1$ digits, x is the word of the $W+1$ lowest digits of $|y|$, so the binary stream holds $[y<0]\,x=\mathrm{sw}_W(y)$. □

I.14.3 The stage

Theorem I.198 (Round and emit). *Let* $\mathsf{RoundEmit}=(\mathsf{DropCarry};\mathsf{Room});\mathsf{Emit}$ *and $K_{\mathrm{e}}=2000+12000\cdot 2001^{2}+400=48\,048\,014\,400$. Let $m\le w$, W and r be natural numbers and z an integer such that z and $\mathrm{near}(2^{m}z/2^{w})$ fit W. Let τ be a workspace of r rows on which the binary stream holds $\mathrm{sw}_W(z)$, tracks 5, 8 and 9 carry the counters m, w and W, and the output buffer holds the empty word with capacity r. Then* RoundEmit *runs from $(-1,\tau)$ to $(-1,\tau')$ in some number t of units, and there is r' such that:*

(a) the binary stream of τ' holds $\mathrm{sw}_W\big(\mathrm{near}(2^{m}z/2^{w})\big)$;

(b) track 5 of τ' carries 0, the output buffer holds the empty word with capacity r', and τ' is a workspace of r' rows;

(c) $2(W+2)+1\le r'$;

(d) $t+r'+1\le K_{\mathrm{e}}(m+w+W+r+1)^{7}$.

The stage reads tracks 5, 6, 7, 8 and 9, uses the cells $(0,3)$ and $(0,10)$ and tracks 10, 12 and 13 as scratch, and adds rows with track 2 as their counter. The saved request and the step count are not needed any more, and the contract says nothing about them. The exponent is $7=2\cdot(2+1)+1$: the arithmetic part has exponent 2, the growth of the workspace is quadratic in the rows and so turns it into $2(2+1)$, and the final scan adds one more.

Proof. Put $Y=m+w+W+r+1$, $e=[z<0]$ and $n=|z|$. By Lemma I.194, DropCarry runs to a tape τ_1, a workspace of r rows, in t_1 units with $t_1+r+1\le 2000Y^{2}$; on τ_1 a word

x' of length $W+1$ and value $\mu_e(n, w-m)$ lies on tracks 6 and 7 from row 1, $(0,10)$ holds e, tracks 5 and 9 carry 0 and W, and the output buffer holds the empty word with capacity r. By Lemma I.195, Room runs to a tape τ_2, a workspace of $r' = 2r+3$ rows with the empty buffer of capacity r', in t_2 units with $t_2 + r' + 1 \le 12000(r+2)^2$, and keeps x', the cell $(0,10)$ and the counter 0 on track 5. By Lemma I.197, whose origin marks up to $W+2 \le r'$ are those of the buffer, Emit runs in $t_3 \le 200(W+2)$ units to τ', whose binary stream holds $\mathrm{sw}_W(y)$ with $y = (-1)^e \mu_e(n, w-m)$; by Lemma I.189, $y = \mathrm{near}(2^m z/2^w)$. Emit writes only row 0 and track 10 in rows $j \le W+2$, so (b) still holds, and (c) is part of Lemma I.195.

For (d): since $Y \ge 1$, $r+2 \le t_1 + r + 1 + 1 \le 2000Y^2 + Y^3 \le 2001Y^3$, so $t_2 + r' + 1 \le 12000 \cdot 2001^2 Y^6$; also $t_1 \le 2000Y^6$ and $t_3 \le 200(W+2) \le 400Y \le 400Y^7$. Hence

$$t + r' + 1 = t_1 + t_2 + t_3 + r' + 1 \le (2000 + 12000 \cdot 2001^2)Y^6 + 400Y^7 \le K_{\mathrm{e}} Y^7. \qquad \square$$

I.15 Framing the answer and erasing the workspace

At the end of the arithmetic the answer is a signed word in the scratch area, and the input cells are blank. Definition 15.31 asks for more: the machine must halt with the frame of that word written from cell 0 and 0 in every other cell of the tape. Take $W = 1$ and $a_m = -1$. The signed word is $\mathrm{sw}_1(-1) = 110$: the sign 1, then the two digits $1, 0$ of 1. Its frame is 1110 110. The last stage copies the three bits of the stream into the output buffer twice, first as three marks and then as themselves, with a 0 between the two passes; writes the seven bits of the buffer into the cells $0, \dots, 6$; and erases every row of the workspace and the root. For a word u we write $\iota(u)$ for the tape that holds u from cell 0 and 0 in every other cell; the stage ends with the head on cell 0 and the tape $\iota(1110110)$.

The erasure is where the layout of Section I.1 pays off. By Lemma I.4 every cell that holds 1 on a workspace of r rows is the root or lies in a row $j \le r$, and the rows $j < r$ are exactly those whose header on track 0 holds 1. So the machine can find every written cell without knowing r: it walks down the headers, clears each row, clears the sentinel row, and walks back to the root.

I.15.1 The frame in the buffer

The output buffer receives bits through the routine $\mathsf{put}_{11}(b)$ of Lemma I.124. Track 5 serves as the read counter of the stream. For bits h and b put

$$T_b = \mathsf{home}_7; \mathsf{put}_{11}(b); \mathsf{seek}_5; \mathsf{next}; \mathsf{left}, \qquad \mathsf{tx}(h) = \mathsf{left}; \mathsf{branch}(T_{h\vee 0}, T_{h\vee 1}).$$

Lemma I.199 (Copying the stream into the buffer)**.** *Let the binary stream of τ hold a word v of length n, let track 5 carry 0, and let the output buffer hold a word w with capacity R, where $|w| + n \le R$. Let h be a bit and let v^h be 1^n if $h = 1$ and v if $h = 0$. Then $\mathsf{loop}(\mathsf{tx}(h))$ runs from $(\mathrm{cell}(0,6), \tau)$ to $(\mathrm{cell}(n,6), \tau')$ in exactly $36n^2 + (36|w| + 82)n + 1$ units, where the output buffer of τ' holds wv^h with capacity R, track 5 carries n, and τ' differs from τ only on tracks 5, 11, 12 and 13, in rows $j \le R$.*

Proof. For $k \le n$ let I_k be the configuration at $(k,6)$ whose tape is τ with track 5 carrying k and the buffer holding w followed by the first k bits of v^h. From I_k, $k < n$, the presence bit

$(k, 6)$ is 1. The pass moves to $(k, 7)$ (1), reads v_k (1) and runs T_b with $b = h \vee v_k$, the kth bit of v^h: home_7 from row k ($18k + 11$), $\mathsf{put}_{11}(b)$ at buffer length $|w| + k$ ($36(|w| + k) + 75$; the origin marks up to $|w| + k + 1 \le R$ are those of the buffer), seek_5 to the hole $(k, 5)$ ($18k + 8$), next (20), which advances the counter, and left (1) to $(k + 1, 6)$. With its test the pass takes $72k + 36|w| + 118$ units and ends in I_{k+1}. At I_n the presence bit is 0. By Lemma I.6 the loop takes $\sum_{k<n}(72k + 36|w| + 118) + 1 = 36n^2 + (36|w| + 82)n + 1$ units. □

Lemma I.200 (Framing the stream)**.** *Let*

$$\begin{aligned}\mathsf{Frame} = {} & \mathsf{pres}_5; \mathsf{loop}(\mathsf{tx}(1)); \mathsf{home}_6; \mathsf{put}_{11}(0); \mathsf{reset}_5; \\ & \mathsf{pres}_5; \mathsf{loop}(\mathsf{tx}(0)); \mathsf{home}_6; \mathsf{reset}_5.\end{aligned}$$

Let the binary stream of τ hold a word v of length n, let track 5 carry 0, and let the output buffer hold w with capacity R, where $|w| + 2n + 1 \le R$. Then Frame *runs from $(-1, \tau)$ to $(-1, \tau')$ in exactly*

$$F(|w|, n) = 108n^2 + 72|w|n + 344n + 36|w| + 189 \le 700(|w| + n + 1)^2$$

units, where the output buffer of τ' holds $w1^n0v$, the word w followed by the frame of v, with capacity R, track 5 carries 0, and τ' differs from τ only on tracks 5, 11, 12 and 13, in rows $j \le R$.

Proof. pres_5 (10 units, Lemma I.11) reaches $(0, 6)$. By Lemma I.199 with $h = 1$ the first loop takes $36n^2 + (36|w| + 82)n + 1$ units, leaves $w1^n$ in the buffer and n on track 5, and stops at $(n, 6)$. home_6 ($18n + 10$) returns to the root, $\mathsf{put}_{11}(0)$ ($36(|w| + n) + 75$) appends the separator, and reset_5 ($36n + 36$, Lemma I.14(c)) puts the hole of track 5 back in row 0. pres_5 (10) and the second loop, with $h = 0$ and the buffer word of length $|w| + n + 1$, take $10 + 36n^2 + (36(|w| + n + 1) + 82)n + 1$ units and append v; the hypothesis of Lemma I.199 holds because $|w| + n + 1 + n \le R$. home_6 ($18n + 10$) and reset_5 ($36n + 36$) end at the root with track 5 carrying 0. The sum is $F(|w|, n)$, and each of its five coefficients is at most 700 times the corresponding coefficient of $(|w| + n + 1)^2$. □

I.15.2 Writing the output

The buffer is turned into the output by a single raw marker in the input cells, in the sense of Definition I.33, and the routines raw and unraw of Lemma I.34, which walk from the root to the marker and back. The machine has no addresses, so it cannot write the bit of row j into cell j directly. Instead it first pushes the marker from cell 0 to cell $n - 1$, one cell per row of the buffer; then it walks the buffer backwards, and at each row replaces the marker by the

row's bit and puts a new marker one cell to the left. Let

$$\begin{aligned}
\mathsf{oprep} &= [\mathsf{reset}_{11}, \mathsf{right}, \mathsf{write}\,1, \mathsf{left}, \mathsf{pres}_{11}], \qquad \mathsf{onil} = [\mathsf{home}_{12}, \mathsf{raw}, \mathsf{write}\,0, \mathsf{left}],\\
\mathsf{osp} &= [\mathsf{home}_{12}, \mathsf{raw}, \mathsf{write}\,0, \mathsf{right}, \mathsf{write}\,1, \mathsf{unraw}, \mathsf{inc}_{11}, \mathsf{pres}_{11}, \mathsf{left}^{16}],\\
\mathsf{ospread} &= [\mathsf{left}^{16}, \mathsf{loop}(\mathsf{osp}), \mathsf{home}_{12}],\\
U_b &= [\mathsf{home}_{13}, \mathsf{raw}, \mathsf{write}\,b, \mathsf{left}, \mathsf{write}\,1, \mathsf{unraw}, \mathsf{seek}_{11}, \mathsf{right}^{10}],\\
\mathsf{oul} &= [\mathsf{left}^{12}, \mathsf{branch}(U_0, U_1), \mathsf{left}^{10}, \mathsf{write}\,1, \mathsf{right}^{16}, \mathsf{write}\,0, \mathsf{right}^{10}],\\
L_b &= [\mathsf{home}_{13}, \mathsf{raw}, \mathsf{write}\,b, \mathsf{left}], \qquad \mathsf{olast} = [\mathsf{left}^{12}, \mathsf{branch}(L_0, L_1)],\\
\mathsf{ounload} &= [\mathsf{seek}_{11}, \mathsf{right}^{10}, \mathsf{loop}(\mathsf{oul}), \mathsf{olast}],\\
\mathsf{Out} &= \big[\mathsf{oprep}, \mathsf{branch}\big(\mathsf{onil}, [\mathsf{ospread}, \mathsf{ounload}]\big)\big].
\end{aligned}$$

Lemma I.201 (Writing the buffer to the input cells)**.** *Let u be a word of length n and let τ be a tape with $\tau(-1) = 1$ and $\tau(x) = 0$ for every $x \geq 0$, with origin marks up to n, on which track 11 carries n, $\tau(j, 12) = [j < n]$ for $j \leq n$, and $\tau(j, 13) = u_j$ for $j < n$. Then* Out *runs from $(-1, \tau)$ to $(-1, \tau')$ in exactly $O(n)$ units, where $O(0) = 108$ and $O(n) = 66n^2 + 209n - 34$ for $n \geq 1$, so that $O(n) \leq 500(n+1)^2$. In τ' the cell x holds u_x for $0 \leq x < n$ and 0 for $x \geq n$, track 11 carries 0, and every cell $x < 0$ other than the cells $(j, 11)$, $j \leq n$, holds what it held in τ.*

Proof. oprep resets track 11 ($36n + 60$, Lemma I.14(c)), writes a raw marker at cell 0 (3) and moves to $(0, 12)$ (16); with the list, $36n + 80$ units. The branch reads $(0, 12) = [n > 0]$ (1). If $n = 0$, onil returns to the root (16), finds the marker (7, Lemma I.34(b)), erases it and steps back to the root (3, with the list): in all $36 \cdot 0 + 80 + 1 + 26 + 1 = 108$ units.

Let $n \geq 1$. In ospread, left^{16} (17) reaches $(1, 12)$. For $k \leq n - 1$ let A_k be the configuration at $(k + 1, 12)$ whose tape has the raw marker at k, the counter k on track 11, and otherwise agrees with τ off track 11. From A_k, $k < n - 1$, the cell $(k + 1, 12)$ holds 1; the pass returns to the root ($18(k + 1) + 16$), finds the marker ($6k + 7$), moves it to $k + 1$ (3), returns ($6(k + 1) + 7$), advances track 11 ($36k + 66$) and moves to $(k + 2, 12)$ ($18(k + 1) + 16 + 17$): with the list and the test, $84k + 176$ units, ending in A_{k+1}. At A_{n-1} the cell $(n, 12)$ holds 0, and home_{12} from row n costs $18n + 16$. So ospread takes $18n + 34 + \big(42(n-1)^2 + 134(n-1) + 1\big)$ units.

In ounload, seek_{11} ($18(n - 1) + 14$) and right^{10} (11) reach $(n - 1, 1)$. For $k \leq n - 1$ let B_k be the configuration at $(k, 1)$ whose tape has the raw marker at k, the cells x with $k < x < n$ holding u_x, the counter k on track 11, and otherwise agrees with τ off track 11. From B_{k+1} the origin mark $(k + 1, 1)$ is 1. The pass reads u_{k+1} in $(k + 1, 13)$ (13, and the branch 1); $U_{u_{k+1}}$ returns to the root ($18(k + 1) + 17$), finds the marker at $k + 1$ ($6(k + 1) + 7$), writes u_{k+1} there and a new marker at k (3), returns ($6k + 7$) and goes back to $(k + 1, 1)$ ($18(k + 1) + 14 + 11$, and the list 1); the rest of oul moves the hole of track 11 from row $k + 1$ to row k and reaches $(k, 1)$ (42 units). With the test the pass takes $48k + 159$ units and ends in B_k. At B_0 the origin mark $(0, 1)$ is 0, and olast reads u_0, finds the marker at 0 and replaces it by u_0 (42 units). So ounload takes $18(n - 1) + 68 + \big(24(n-1)^2 + 135(n-1) + 1\big)$ units. With oprep, the branch and the two lists the total is $O(n)$. No cell $x \geq n$ is written, and among the cells $x < 0$ only track 11 changes. □

Lemma I.202 (Erasing the workspace)**.** *Let*

$$\mathsf{Erase} = [\mathsf{left}, \mathsf{loop}(\mathsf{put}(0^{16})), \mathsf{put}(0^{16}), \mathsf{findR}, \mathsf{write}\,0],$$

with findR *the mirror of* findL *(Lemma I.34(a)). Let $\tau(j,0) = 1$ for $j < r$, $\tau(r,0) = 0$, $\tau(-1) = 1$, and let every cell of every row $j > r$ hold 0. Then* Erase *runs from $(-1,\tau)$ to $(-1,\tau')$ in $130r+144$ units, where $\tau'(x) = 0$ for every $x < 0$ and $\tau'(x) = \tau(x)$ for every $x \geq 0$.*

Proof. left (1) reaches the header $(0,0)$. For $j \leq r$ let I_j be the configuration at $(j,0)$ in which the rows before j hold 0. From I_j, $j < r$, the header holds 1; $\mathsf{put}(0^{16})$ (33 units, Lemma I.12(c)) writes 0 in the sixteen cells of row j and stops at $(j+1,0)$, so the pass takes 34 units and ends in I_{j+1}. At the sentinel header the loop stops (1), and $\mathsf{put}(0^{16})$ (33) clears row r and stops at $(r+1,0) = -18-16r$. Now every cell from there up to -2 holds 0 and the root holds 1, so findR passes $16r+17$ zeros to the root in $6(16r+17)+5$ units, and the write clears the root. Every negative cell is now 0: the rows $j \leq r$ were cleared and the rows $j > r$ were blank. With the list the count is $1 + (34r+1) + 33 + (96r+107) + 1 + 1 = 130r+144$. □

Lemma I.203 (Finalizing)**.** *Let* $\mathsf{Finalize} = \mathsf{Out}; \mathsf{Erase}; \mathsf{right}$*. Let τ be a workspace of r rows whose output buffer holds a word u with capacity r. Then* Finalize *runs from $(-1,\tau)$ to $(0,\iota(u))$ in exactly $O(|u|) + 130r + 145$ units.*

Proof. The buffer gives the hypotheses of Lemma I.201 with $n = |u|$: the counter n on track 11, the presence cells $(j,12) = [j < n]$ for $j \leq n \leq r$, the bits on track 13 and the origin marks; the workspace gives $\tau(-1) = 1$ and the blank input cells. After Out the input cells hold u followed by 0s, and only track 11 in rows $j \leq n \leq r$ has changed among the negative cells, so the headers, the sentinel, the blank rows beyond r and the root are as in τ. Lemma I.202 clears every negative cell and keeps the input cells, and right moves the head to cell 0: the tape is $\iota(u)$, after $O(n) + (130r+144) + 1$ units. □

I.15.3 The stage

Theorem I.204 (Frame and erase)**.** *Let* $\mathsf{FrameErase} = \mathsf{Frame}; \mathsf{Finalize}$*. Let τ be a workspace of r rows on which the binary stream holds a word v, track 5 carries 0, and the output buffer holds a word w with capacity r, where $|w| + 2|v| + 1 \leq r$. Then* FrameErase *runs from $(-1,\tau)$ to $(0,\iota(w\,1^{|v|}0v))$ in exactly*

$$D(w,v,r) = F(|w|,|v|) + O(|w| + 2|v| + 1) + 130r + 145$$

units. If w is empty, $D(w,v,r) + 1 \leq 2000(r+1)^2$.

Proof. By Lemma I.200 with $R = r$, Frame takes $F(|w|,|v|)$ units, and afterwards the buffer holds $u = w\,1^{|v|}0v$ with capacity r. It writes only tracks 5, 11, 12 and 13 in rows $j \leq r$, so the tape is still a workspace of r rows, and $|u| = |w| + 2|v| + 1 \leq r$. Lemma I.203 finishes. If w is empty, then $|v| + 1 \leq r + 1$ and $2|v| + 2 \leq r + 1$, so $F(0,|v|) \leq 700(r+1)^2$ and $O(2|v|+1) \leq 500(r+1)^2$, and $1200(r+1)^2 + 130r + 146 \leq 2000(r+1)^2$. □

The stage is quadratic in the rows, so a bound of the form of Definition I.7 for everything before it doubles its exponent.

Lemma I.205 (Paying for the frame). *Let $Y \geq 1$, and let t, r, K and g be natural numbers with $t + r + 1 \leq KY^g$. If v is a word with $2|v| + 1 \leq r$ and ε is the empty word, then*

$$t + D(\varepsilon, v, r) + 1 \leq 2001(K+1)^2 Y^{2g}.$$

Proof. By Theorem I.204, $D(\varepsilon, v, r) + 1 \leq 2000(r+1)^2 \leq 2000K^2Y^{2g}$, since $r + 1 \leq KY^g$. Also $t \leq KY^g \leq (K+1)^2Y^{2g}$. Add. □

I.16 The whole count

The stages are now all in place. This section puts them together into the one program P of Theorem 15.33, computes its exponent λ from the exponents of the stages, writes down its constant K, and proves the theorem. The program is

$$\begin{aligned} \mathsf{P} &= \mathsf{Parse}; (\mathsf{Decode}; \mathsf{Solve}), \\ \mathsf{Solve} &= \big(\mathsf{Schedule}; (\mathsf{Traj}; \mathsf{RoundEmit})\big); \mathsf{FrameErase}, \\ \mathsf{Traj} &= \mathsf{Init}; \big(\mathsf{repeat}_2(\mathsf{Step}); (\mathsf{Select}; \mathsf{empty})\big), \qquad \mathsf{Step} = \mathsf{Build}; \mathsf{Sum}, \end{aligned}$$

with the stage programs of Sections I.7 to I.15, the program $\mathsf{empty} = \mathsf{clear}_{11,12}$ of Lemma I.125, and the loop repeat_2 of Lemma I.18, which runs the step once for each of the S marks of the step count on track 14. Table I.5 lists every stage with the result that proves its contract and the exponent of its bound. Throughout, a problem with a coefficient field of dimension d is fixed, with initial point y_0, radius $M \geq 1$ and natural end time T, satisfying the region certificate at level n_0; $X = m + n_0 + \kappa + 1$, $W = W_m$, and $g = 118$ is the exponent of one entry of the coefficient table (Theorem I.119).

The exponent is assembled in four steps. The trajectory and the rounding compose to g_{tr}; the schedule in front of them gives g_{f} by Lemma 15.36(c); the frame doubles it; and the reading of the request, with exponent 11 in X alone, multiplies by 11 after one more step of Lemma 15.36(a):

$$\lambda = 11\Big(2(g_q+1)\big(8(762700(g+20)+2)+1\big)+1\Big) = 11\big(2 \cdot 842\,020\,817\,(g_q+1)+1\big),$$

with $g = 118$ and

$$g_q + 1 = 111\,736\,803\,796\,728\,983\,743\,307\,618\,367\,111\,577\,196\,298\,253.$$

As a numeral,

$$\lambda = 2\,069\,863\,726\,081\,589\,698\,020\,631\,180\,195\,588\,915\,541\,683\,736\,068\,119\,433.$$

It is a fixed number: nothing in it depends on the problem, the coordinate or the precision.

Stage	Result	Exponent of the bound
Parse	Theorem I.42	3, in X
Decode	Theorem I.58	8, in X
Parse; Decode	Corollary I.59	$g_d + 3 = 11$, in X
Schedule	Theorem I.88	$g_q = 23040 \cdot 6^{51} + 12$
Init	Theorem I.102	190674
one coefficient, one row	Theorems I.119, I.120	$g = 118$
Build	Theorem I.121	$g + 2 = 120$
Sum	Theorem I.161	8
Step	Theorem I.162	$g + 18 = 136$
repeat_2(Step)	Theorem I.184	$g + 19 = 137$
Select; empty	Theorem I.187	3
Traj	Theorem I.188	$g_{\mathrm{t}} = 762700(g + 20) = 105\,252\,600$
RoundEmit	Theorem I.198	7, in $m + w + W + r + 1$
Traj; RoundEmit	Lemma I.206	$g_{\mathrm{tr}} = 8(g_{\mathrm{t}} + 2) = 842\,020\,816$
Schedule; (Traj; RoundEmit)	Lemma I.207	$g_{\mathrm{f}} = (g_q + 1)(g_{\mathrm{tr}} + 1)$
FrameErase	Theorem I.204	2, in the rows
Solve	Lemma I.208	$g_{\mathrm{sol}} = 2g_{\mathrm{f}}$
P	Theorem I.209	$\lambda = (g_d + 3)(g_{\mathrm{sol}} + 1)$, in $m + 1$

Table I.5. The stages of the program P in the order in which they run, the result that proves the contract of each, and the exponent of its bound. Unless stated otherwise the bound is $K(X + r + 1)^g$ of Definition I.7, with r the rows in use when the stage starts.

Lemma I.206 (The arithmetic after the schedule)**.** *Let K_{t} be the constant of Theorem I.188 for the problem, $g_{\mathrm{t}} = 105\,252\,600$ its exponent, and $c_W = 4(q_P + l_P + c_P) + 4$ the constant of Proposition I.32(e). Put $A_{\mathrm{tr}} = K_{\mathrm{t}} + c_W + 3$, $K_{\mathrm{tr}} = K_{\mathrm{t}} + K_{\mathrm{e}} A_{\mathrm{tr}}^7$ and $g_{\mathrm{tr}} = 8(g_{\mathrm{t}} + 2)$. Let $i < d$, m and r be given, and let τ be a scheduled request of the request for coordinate i at precision m with r rows (Definition I.85). Then* Traj; RoundEmit *runs from $(-1, \tau)$ to $(-1, \tau')$ in t units, and there is r' such that the binary stream of τ' holds $\mathrm{sw}_W(a_m)$, track 5 carries 0, the output buffer holds the empty word with capacity r', τ' is a workspace of r' rows, $2(W + 2) + 1 \le r'$, and*

$$t + r' + 1 \le K_{\mathrm{tr}}(X + r + 1)^{g_{\mathrm{tr}}}.$$

Proof. Put $Y = X + r + 1$, $n = \max(m, n_0)$ and $w = N(n)$. By Theorem I.188, Traj runs from $(-1, \tau)$ to $(-1, \tau_1)$ in t_1 units, where the binary stream of τ_1 holds $\mathrm{sw}_W(z)$ with $z = (Z_S)_i$, the integer vector of Lemma 15.35 after S steps at level n; tracks 5, 8 and 9 carry m, w and W; the output buffer holds the empty word with capacity r_1; τ_1 is a workspace of r_1 rows; and $t_1 + r_1 + 1 \le K_{\mathrm{t}} Y^{g_{\mathrm{t}}}$. The integer z fits W by Proposition I.32(a), since the region certificate keeps the stored vectors in the ball. By Lemma 15.35, $a_m = \mathrm{near}(2^m z / 2^w)$, and a_m fits W by Lemma H.5. Also $m \le w$, because $w = n + \kappa$. So Theorem I.198 applies to τ_1: RoundEmit runs to a tape τ' with all the stated properties in t_2 units, with $t_2 + r' + 1 \le K_{\mathrm{e}}(m + w + W + r_1 + 1)^7$.

Now $m \le X \le Y$, $w \le X \le Y$, $W \le c_W X^2 \le c_W Y^2$ by Proposition I.32(e), and $r_1 + 1 \le K_{\mathrm{t}} Y^{g_{\mathrm{t}}}$. Each of these powers of Y is at most $Y^{g_{\mathrm{t}}+2}$, so $m + w + W + r_1 + 1 \le A_{\mathrm{tr}} Y^{g_{\mathrm{t}}+2}$ and $t_2 + r' + 1 \le K_{\mathrm{e}} A_{\mathrm{tr}}^7 Y^{7(g_{\mathrm{t}}+2)}$. With $t_1 \le K_{\mathrm{t}} Y^{g_{\mathrm{t}}}$ and $Y \ge 1$, both bounds are at most their constants times $Y^{8(g_{\mathrm{t}}+2)}$; add them. □

Lemma I.207 (After the decoded request)**.** *Let K_q be the constant of Theorem I.88 for the problem and g_q its exponent, and put $K_{\mathrm{f}} = K_q + K_{\mathrm{tr}}(K_q+1)^{g_{\mathrm{tr}}}$ and $g_{\mathrm{f}} = (g_q+1)(g_{\mathrm{tr}}+1)$. Let τ be a decoded request of the request for coordinate i at precision m with r rows (Definition I.56). Then* Schedule; (Traj; RoundEmit) *runs from $(-1,\tau)$ to $(-1,\tau')$ in t units, where τ' has the properties of Lemma I.206 for some r', and*

$$t + r' + 1 \le K_{\mathrm{f}}(X + r + 1)^{g_{\mathrm{f}}}.$$

Proof. By Theorem I.88, Schedule runs from $(-1,\tau)$ to a scheduled request with r_1 rows in t_1 units, with $t_1 + r_1 + 1 \le K_q(X+r+1)^{g_q}$. By Lemma I.206 the rest runs from there in t_2 units, with $t_2 + r' + 1 \le K_{\mathrm{tr}}(X + r_1 + 1)^{g_{\mathrm{tr}}}$. Lemma 15.36(c), with $K_1 = K_q$, $g_1 = g_q$, $K_2 = K_{\mathrm{tr}}$ and $g_2 = g_{\mathrm{tr}}$, gives the bound. □

Lemma I.208 (Solving from the decoded request)**.** *Put $K_{\mathrm{sol}} = 2001(K_{\mathrm{f}}+1)^2$ and $g_{\mathrm{sol}} = 2g_{\mathrm{f}}$, and let τ be as in Lemma I.207. Then* Solve *runs from $(-1,\tau)$ to $(0,\iota(\hat a_m))$ in t units, where $\hat a_m$ is the answer word of a_m at width W (Definition 15.32), and*

$$t + 1 \le K_{\mathrm{sol}}(X + r + 1)^{g_{\mathrm{sol}}}.$$

Proof. By Lemma I.207, the first part of Solve runs in t_1 units to a tape τ', with $t_1 + r' + 1 \le K_{\mathrm{f}} Y^{g_{\mathrm{f}}}$, $Y = X + r + 1$. On τ' the binary stream holds $v = \mathrm{sw}_W(a_m)$, track 5 carries 0, the output buffer holds the empty word with capacity r', and $2|v| + 1 = 2(W+2) + 1 \le r'$. By Theorem I.204 with w empty, FrameErase runs from $(-1,\tau')$ to $(0,\iota(1^{|v|}0v))$ in $D(\varepsilon, v, r')$ units, and $1^{|v|}0v$ is the frame of v, which is $\hat a_m$. Lemma I.205 with $K = K_{\mathrm{f}}$ and $g = g_{\mathrm{f}}$ gives $t_1 + D(\varepsilon, v, r') + 1 \le 2001(K_{\mathrm{f}}+1)^2 Y^{2g_{\mathrm{f}}}$. □

Theorem I.209 (The whole program and its count)**.** *Let*

$$\lambda = 11\big(2(g_q+1)(8(762700\cdot 138 + 2) + 1) + 1\big),$$

the numeral displayed above. Let a problem have a coefficient field, with rational initial point y_0, radius $M \ge 1$ and natural end time T, and let it satisfy the region certificate at level n_0. Put $c = n_0 + \kappa + 1$, let $A = 100000(c_{\mathrm{req}}+2)^3 + K_d c_{\mathrm{num}}^8$ be the constant of Corollary I.59, and

$$K = \big(A + K_{\mathrm{sol}}(A+1)^{g_{\mathrm{sol}}}\big)c^{\lambda},$$

with K_{sol} and g_{sol} as in Lemma I.208. Then for every coordinate $i < d$ and every precision m:

(a) P *has a counted execution in some $t \le K(m+1)^{\lambda}$ units from the head on cell 0 and the input tape of the request for coordinate i at precision m (Definition 15.32) to the head on cell 0 and the tape on which the answer word of a_m is written from cell 0, with 0 in every other cell;*

(b) $|a_m| < 2^{W_m+1}$, and the machine $H = \mathrm{machine}(\mathsf{P})$ runs from that request to the answer word of a_m within $K(m+1)^{\lambda}$ steps.

The same holds for every polynomial field whose quadratized problem satisfies the region certificate at level n_0: with the constants of the quadratized problem, H runs from the request of the quadratized problem for the variable w_{e_i} at precision m to the answer word of the answer at precision m of the polynomial solution x_i within $K(m+1)^{\lambda}$ steps, and that answer is less than 2^{W_m+1} in absolute value.

The program P, and hence H, and the number λ are fixed before the problem is given; only K depends on it. That is the order of quantifiers of Theorem 15.33.

Proof. (a) Let u be the request. By Corollary I.59, $\mathsf{Parse};\mathsf{Decode}$ runs from the head on cell 0 of the input tape of u to $(-1,\tau_D)$, a decoded request with r_D rows, in t_P+t_D units with $t_P+t_D+r_D+1\le AX^{11}$. By Lemma I.208, Solve runs from $(-1,\tau_D)$ to $(0,\iota(\hat{a}_m))$ in t_S units with $t_S+1\le K_{\mathrm{sol}}(X+r_D+1)^{g_{\mathrm{sol}}}$. The counted executions compose in sequence. Lemma 15.36(a), with $N=X\ge 1$, $e=11$, $s=t_P+t_D$, $r=r_D$, $t=t_S$, $r'=0$, $K=K_{\mathrm{sol}}$ and $g=g_{\mathrm{sol}}$, gives

$$t_P+t_D+t_S+1\le \big(A+K_{\mathrm{sol}}(A+1)^{g_{\mathrm{sol}}}\big)X^{11(g_{\mathrm{sol}}+1)},$$

and $11(g_{\mathrm{sol}}+1)=\lambda$. Since $X=m+c$ with $c\ge 1$, Lemma 15.36(b) gives $X^{\lambda}\le c^{\lambda}(m+1)^{\lambda}$, so $t\le K(m+1)^{\lambda}$.

(b) $|a_m|<2^{W_m+1}$ is Lemma H.5. The final tape of (a) is the answer word written from cell 0 with 0 everywhere else, so by Lemma 15.34(b) the machine machine(P) runs from u to the answer word within t steps.

For a polynomial field, the polynomial solution x_i is by Definition 15.27 the certified real of Theorem 15.22 for the variable w_{e_i} of the quadratized problem, which has a coefficient field and satisfies the region certificate at level n_0. Parts (a) and (b), applied to the quadratized problem and the coordinate of w_{e_i}, give the claim with the same H and λ. □

Appendix J

Proofs for the calculus of δ's objects

This appendix carries the verifications deferred in Chapter 16: the equivalence tools behind the compiler, the remaining compiler cases, the derivation of every true equational sentence in the full calculus, the programs that track the maps used, the readings of the sentences in the model of named objects, the computability of addresses, the names of the cover series, and the tent function. Notation is that of the chapter. Arguments about the model of named objects are carried out in ordinary classical mathematics, as there.

J.1 Equivalence and reindexing

Recall that $A \equiv B$ means that the calculus derives $A \vdash B$ and $B \vdash A$. The same constructions work in the computable calculus.

Lemma J.1 (Equivalence is a congruence)**.** *The relation $\equiv$ is reflexive, symmetric and transitive. If $A \equiv A'$ and $B \equiv B'$, then $A \wedge B \equiv A' \wedge B'$, $A \vee B \equiv A' \vee B'$, $A \to B \equiv A' \to B'$ and $f^*A \equiv f^*A'$ for every map f; and for formulas in a context $G \times X$, $A \equiv A'$ gives $\forall_X A \equiv \forall_X A'$ and $\exists_X A \equiv \exists_X A'$. Moreover $f^*\top \equiv \top$, $f^*\bot \equiv \bot$ and $g^*f^*A \equiv (f \circ g)^*A$.*

Proof. Reflexivity is construction (2), and transitivity is cut. For conjunction, eliminate each conjunct of $A \wedge B$, cut with $A \vdash A'$ and $B \vdash B'$, and introduce. For disjunction, apply (11) to $A \vee B$, with branches that project A from $(A \vee B) \wedge A$, cut with $A \vdash A'$ and introduce $A' \vee B'$. For implication, from $(A \to B) \wedge A'$ derive A by cutting $A' \vdash A$, apply (13) to obtain B, cut with $B \vdash B'$, and close with (12). Reindexing is construction (18) applied to both derivations.

For the universal quantifier we use the *generic instance*. Let $\pi = \pi_1 : G \times X \to G$ and let $v = \mathrm{map}(\pi_2)$ be the newest variable. The composite of $\mathrm{gr}(v) : G \times X \to (G \times X) \times X$ and $\pi \times \mathrm{id} : (G \times X) \times X \to G \times X$ sends (g, x) to $((g, x), x)$ and then to (g, x): it is the identity. By rules 12–15, $\mathrm{gr}(v)^*(\pi \times \mathrm{id})^*A \equiv A$. Now $\pi^*\forall_X A \equiv \forall_X(\pi \times \mathrm{id})^*A$ by rules 26 and 27; universal elimination at v and the generic instance give $\pi^*\forall_X A \vdash A$, a cut with $A \vdash A'$ gives $\pi^*\forall_X A \vdash A'$, and (14) gives $\forall_X A \vdash \forall_X A'$. For the existential quantifier, $A \vdash A'$ and the generic instance give $A \vdash \mathrm{gr}(v)^*(\pi \times \mathrm{id})^*A'$, so existential introduction at v gives $A \vdash \exists_X(\pi \times \mathrm{id})^*A'$, and this formula is $\equiv \pi^*\exists_X A'$ by rules 28 and 29. Hence $\pi^*\exists_X A \wedge A \vdash \pi^*\exists_X A'$, and (17), applied to $\exists_X A \vdash \exists_X A$, gives $\exists_X A \vdash \exists_X A'$. The statements about $\top$ and $\bot$ are the derivations given after Table 16.2, and the last statement is rules 14 and 15. □

The coherence lemma (Lemma 16.12) uses these facts at every connective and quantifier. Its atomic cases rest on the substitution lemma for number terms, which says, for the three operations: evaluating t with s substituted at position k, at a valuation ρ, equals evaluating t at ρ with the value of s inserted at k; evaluating t lifted by d at cutoff c at ρ equals evaluating t at ρ with the d positions from c deleted; and a renaming of set positions does not touch number

terms. Each is proved by induction on t, the case of a variable being the definition of the operation (Definition 5.2).

J.2 The remaining cases of the compiler

We complete the proof of Theorem 16.11. In each case the induction hypothesis gives the translated premises.

Falsehood, conjunction and disjunction introductions, conjunction eliminations and implication elimination translate to the constructions of the same name, with the same premises: the translation commutes with these connectives and leaves the context unchanged.

Disjunction elimination. The HAS premises are $\Gamma \vdash A \vee B$, $A, \Gamma \vdash C$ and $B, \Gamma \vdash C$. Cut the translated second and third premises with the swaps $\Gamma^\circ \wedge A^\circ \vdash A^\circ \wedge \Gamma^\circ$ and $\Gamma^\circ \wedge B^\circ \vdash B^\circ \wedge \Gamma^\circ$, and apply (11).

Existential introduction at a term t. The premise is $\Gamma^\circ \vdash A(t)^\circ$, which is $\equiv$ to

$$\Gamma^\circ \vdash \mathrm{gr}(\ulcorner t \urcorner)^* \mathrm{push}_{\mathbb{N}}^* A^\circ$$

by (16.1). Construction (16) at $\ulcorner t \urcorner$ gives $\Gamma^\circ \vdash \exists_{\mathbb{N}} \mathrm{push}_{\mathbb{N}}^* A^\circ$.

Existential elimination. The premises are $\Gamma \vdash \exists x\, A$ and $A, \Gamma\!\uparrow \vdash C\!\uparrow$. Reindex the second translated premise along $\mathrm{push}_{\mathbb{N}}$ and push the reindexing inside the conjunction:

$$\mathrm{push}_{\mathbb{N}}^* A^\circ \wedge \mathrm{push}_{\mathbb{N}}^* (\Gamma\!\uparrow)^\circ \vdash \mathrm{push}_{\mathbb{N}}^* (C\!\uparrow)^\circ.$$

The lifting instance of the coherence lemma replaces $\mathrm{push}_{\mathbb{N}}^*(\Gamma\!\uparrow)^\circ$ by $\pi_1^* \Gamma^\circ$ and $\mathrm{push}_{\mathbb{N}}^*(C\!\uparrow)^\circ$ by $\pi_1^* C^\circ$; a swap gives $\pi_1^* \Gamma^\circ \wedge \mathrm{push}_{\mathbb{N}}^* A^\circ \vdash \pi_1^* C^\circ$, and (17) with the first premise gives $\Gamma^\circ \vdash C^\circ$.

The set quantifiers. The four rules are the four number-quantifier cases with $\mathrm{push}_{\mathbb{S}}$ in place of $\mathrm{push}_{\mathbb{N}}$, the shift of set positions in place of lifting, and the instance of a set variable in place of (16.1). For universal introduction, the premise $\Gamma^+ \vdash A$ is reindexed along $\mathrm{push}_{\mathbb{S}}$ and $\mathrm{push}_{\mathbb{S}}^*(\Gamma^+)^\circ \equiv \pi_1^* \Gamma^\circ$; for universal elimination at X_j, construction (15) at the term $\ulcorner X_j \urcorner$ gives $\Gamma^\circ \vdash \mathrm{gr}(\ulcorner X_j \urcorner)^* \mathrm{push}_{\mathbb{S}}^* A^\circ \equiv A(X_j)^\circ$; existential introduction and elimination follow the number cases.

Reflexivity is construction (19) at $\ulcorner t \urcorner$.

Together with the cases in the text, every one of the twenty-eight rules of HAS (twenty logical and equality rules, the six arithmetic axioms, induction and comprehension) has been treated.

J.3 The full calculus derives every true equational sentence

We prove Proposition 16.15, reasoning in a classical metatheory. Throughout, G is a context, P a formula in context G, 1 the term $S(0)$, and "equational" means equational over constant carriers. For a term $t : G \vdash X$ and $g \in G_1$ we write $t(g)$ for the later value $[\![t]\!]_1(g)$, and for numbers v we use the numeral v, the term of the constant map with value v. The indicator χ_A of a formula A is the map of Proposition 16.15, and we write χ_A also for its term $\mathrm{map}(\chi_A)$.

(E) Equal later values. A map $f : G \to X$ into a constant carrier is determined by its later component: $f_0(g) = h_X(f_0(g)) = f_1(h(g))$ for every $g \in G_0$. Conversely every function $G_1 \to X_1$, composed with the passage at the earlier stage, is a map of carriers. So if X is constant and $t, u : G \vdash X$ have the same later values, then $[\![t]\!] = [\![u]\!]$, and Remark 16.7 derives $P \vdash t = u$. In the same way a derived equation may have its sides replaced by terms with the same maps: from $P \vdash t = u$, rule 13 and rule 18 at the identity give $P \vdash \mathrm{map}([\![t]\!] \circ \mathrm{id}) = \mathrm{map}([\![u]\!] \circ \mathrm{id})$, which is literally the judgment $P \vdash \mathrm{map}([\![t']\!] \circ \mathrm{id}) = \mathrm{map}([\![u']\!] \circ \mathrm{id})$ when $[\![t']\!] = [\![t]\!]$ and $[\![u']\!] = [\![u]\!]$, and rule 19 and rule 12 return to $P \vdash t' = u'$. In particular two formulas $\chi_B = 1$ whose indicators are the same map are the same formula; for a map $f : D \to G$ the indicator of f^*B is $\chi_B \circ f$, and rules 18 and 19 make $f^*(\chi_B = 1)$ and $\chi_{f^*B} = 1$ interderivable.

(S) Substitution through a map. Let $a, b : G \vdash X$, let Y be constant and let $\varphi, \psi : G \times X \to Y$ be maps; write $\varphi(a)$ for the term $\mathrm{map}(\varphi \circ \mathrm{gr}(a))$. Let B be the equation $\mathrm{map}(\varphi) = \mathrm{map}(\psi)$ in context $G \times X$. From $P \vdash a = b$ and $P \vdash \varphi(a) = \psi(a)$, rule 19 at $\mathrm{gr}(a)$ gives $P \vdash \mathrm{gr}(a)^*B$, construction (20) gives $P \vdash \mathrm{gr}(b)^*B$, and rule 18 gives $P \vdash \varphi(b) = \psi(b)$. When $\varphi(a)$ and $\psi(a)$ have the same later values, the second premise holds by (E); so a derived equation $a = b$ gives $P \vdash s = s'$ for all terms s, s' of Y whose later values are those of $\varphi(b)$ and $\psi(b)$. With $Y = X$ constant, $\varphi(g, z) = z$ and $\psi(g, z) = a(g)$, this gives symmetry: $P \vdash a = b$ yields $P \vdash b = a$.

(Z) Refutation. If $P \vdash s = s'$ is derivable for terms $s, s' : G \vdash \mathbb{N}$ with $s(g) = 0$ and $s'(g) = 1$ at every $g \in G_1$, then symmetry and (E) give $P \vdash S(0) = 0$, rule 1 and construction (13) give $P \vdash \bot$, and (4) gives $P \vdash Q$ for every Q.

(D) Case split. Let $D(x)$ be $x = 0 \vee \exists_{\mathbb{N}} y(x = S(y))$ in context $G \times \mathbb{N}$, with x the newest variable. By rules 12–29 and the composition rules, $\mathrm{gr}(0)^*D$ is equivalent to $0 = 0 \vee \exists_{\mathbb{N}} y(0 = S(y))$, whose left side is (19), and succ^*D to $S(x) = 0 \vee \exists_{\mathbb{N}} y(S(x) = S(y))$, whose right side follows by existential introduction at x and (19). So $\mathrm{Ind}(D)$ is derivable, and rule 7 gives $P \vdash \forall_{\mathbb{N}} D$. Let $t : G \vdash \mathbb{N}$ take only the values 0 and 1. Universal elimination at t gives $P \vdash t = 0 \vee \exists_{\mathbb{N}} y(t = S(y))$. In the second case, under existential elimination, apply (S) to the equation $t = S(y)$ with $\varphi(g, z) = t(g)$ and $\psi(g, z) = e(z)$, where $e(0) = 0$ and $e(z) = 1$ for $z \geq 1$: the two sides agree at t, because t takes only the values 0 and 1, and at $S(y)$ they are t and a term with the value 1. So $P \vdash t = 0 \vee t = 1$.

(T) Transfer. Let $C_1, \ldots, C_m$ and A be formulas in context G and $v_1, \ldots, v_m \in \{0, 1\}$, and suppose that every $g \in G_1$ with $\chi_{C_i}(g) = v_i$ for every i satisfies $g \Vdash_1 A$. Then from $P \vdash \chi_{C_i} = v_i$ for $i = 1, \ldots, m$ the calculus derives $P \vdash \chi_A = 1$. Let $F(g, z_1, \ldots, z_m)$ be 1 when $z_i = v_i$ for every i, and $\chi_A(g)$ otherwise. The equation $\chi_A = F(g, \chi_{C_1}, \ldots, \chi_{C_m})$ holds at every later object, by the hypothesis, so it is derivable by (E). Substituting the equations $\chi_{C_i} = v_i$ one at a time by (S) turns its right side into $F(g, v_1, \ldots, v_m)$, which has the value 1.

The induction. For every equational formula A in every context G, the calculus derives $\chi_A = 1 \vdash A$ and $A \vdash \chi_A = 1$. The proof is by induction on A, for all contexts at once.

$\top$ and $\bot$. $\chi_\top$ has the value 1 and $\chi_\bot$ the value 0 at every later object. So $\chi_\top = 1$ is derivable by (E), and $\top$ by (3); $\chi_\bot = 1$ gives $\bot$ by (Z), and (4) gives the converse.

An equation $t = u$, with X constant. From $\chi_{t=u} = 1$, apply (S) with $\varphi(g, z) = t(g)$ and $\psi(g, z) = u(g)$ when $z = 1$ and $t(g)$ otherwise: the two sides agree at $\chi_{t=u}$, and at 1 they are t and u. From $t = u$, apply (S) with $\varphi(g, z) = 1$ when $z = t(g)$ and 0 otherwise, and $\psi(g, z) = 1$: the two sides agree at t, and at u the left side has the values of $\chi_{t=u}$.

Conjunction. From $\chi_{A\wedge B} = 1$, transfer gives $\chi_A = 1$ and $\chi_B = 1$, the induction hypotheses give A and B, and (6) gives $A \wedge B$. From $A \wedge B$, constructions (7–8) and the induction hypotheses give $\chi_A = 1$ and $\chi_B = 1$, and transfer with $m = 2$ gives $\chi_{A\wedge B} = 1$.

Disjunction. From $\chi_{A\vee B} = 1$, the case split (D) gives $\chi_A = 0 \vee \chi_A = 1$, and (11) treats the cases. If $\chi_A = 1$, the induction hypothesis gives A and (9) gives $A \vee B$. If $\chi_A = 0$, transfer from $\chi_{A\vee B} = 1$ and $\chi_A = 0$ gives $\chi_B = 1$, so B, and (10) gives $A \vee B$. From $A \vee B$, apply (11): in each branch the induction hypothesis and transfer give $\chi_{A\vee B} = 1$.

Implication. By (12), it is enough to derive $\chi_{A\to B} = 1 \wedge A \vdash B$. The induction hypothesis gives $\chi_A = 1$; at a later object an implication is read pointwise, so transfer from $\chi_{A\to B} = 1$ and $\chi_A = 1$ gives $\chi_B = 1$, and the induction hypothesis gives B. Conversely, from $A \to B$ split on χ_A by (D) and (11). If $\chi_A = 0$, transfer gives $\chi_{A\to B} = 1$. If $\chi_A = 1$, the induction hypothesis gives A, construction (13) gives B, the induction hypothesis gives $\chi_B = 1$, and transfer gives $\chi_{A\to B} = 1$.

The universal quantifier, with X constant. The judgment $\chi_{\forall_X A} = 1 \vdash \forall_X A$ follows by construction (14) from

$$\pi^*(\chi_{\forall_X A} = 1) \vdash A \qquad \text{in context } G \times X.$$

By (E) the hypothesis is $\chi_{\pi^*\forall_X A} = 1$; at every later (g, x) where it holds, $g \Vdash_1 \forall_X A$ and so $(g, x) \Vdash_1 A$; transfer gives $\chi_A = 1$, and the induction hypothesis gives A. For the converse, if X_1 is empty, $\chi_{\forall_X A}$ has the value 1 everywhere and (E) suffices. Otherwise let $c_1(g)$ be an $x \in X_1$ with $(g, x) \nVdash_1 A$ when one exists, and a fixed element of X_1 when none does; by (E) this is a map $c : G \to X$, a counterexample chosen in the metatheory. Construction (15) at $\mathrm{map}(c)$ gives $\forall_X A \vdash \mathrm{gr}(c)^* A$, and the induction hypothesis reindexed along $\mathrm{gr}(c)$ by (18), with (E), gives $\chi_{\mathrm{gr}(c)^* A} = 1$. If A holds at $(g, c_1(g))$, it holds at every (g, x), since otherwise $c_1(g)$ would be a counterexample; so transfer gives $\chi_{\forall_X A} = 1$.

The existential quantifier, with X constant. If X_1 is empty, $\chi_{\exists_X A}$ has the value 0 everywhere, and (Z) gives $\chi_{\exists_X A} = 1 \vdash \exists_X A$. Otherwise let $w_1(g)$ be an $x \in X_1$ with $(g, x) \Vdash_1 A$ when one exists, and a fixed element otherwise; by (E) this is a map $w : G \to X$. Where $\chi_{\exists_X A}(g) = 1$, A holds at $(g, w_1(g))$, so transfer gives $\chi_{\mathrm{gr}(w)^* A} = 1$; with (E) and the induction hypothesis reindexed along $\mathrm{gr}(w)$ this gives $\mathrm{gr}(w)^* A$, and (16) at $\mathrm{map}(w)$ gives $\exists_X A$. For the converse, by (17) it is enough to derive $\pi^*\exists_X A \wedge A \vdash \pi^*(\chi_{\exists_X A} = 1)$ in context $G \times X$. The induction hypothesis gives $\chi_A = 1$. At every later (g, x) with $\chi_A(g, x) = 1$, $g \Vdash_1 \exists_X A$; so transfer gives $\chi_{\pi^*\exists_X A} = 1$, which is $\pi^*(\chi_{\exists_X A} = 1)$ by (E).

Reindexing $f^* A$, for $f : D \to G$. By (E), $f^*(\chi_A = 1)$ and $\chi_{f^* A} = 1$ are interderivable, and construction (18) applied to the two induction hypotheses gives $f^*(\chi_A = 1) \vdash f^* A$ and $f^* A \vdash f^*(\chi_A = 1)$.

Conclusion. If $g \Vdash_1 A$ for every $g \in G_1$, then χ_A and 1 have the same later values, (E) gives $P \vdash \chi_A = 1$, and a cut with $\chi_A = 1 \vdash A$ gives $P \vdash A$. No step uses a base rule other than rules 1, 7 and 12–29 and the constructions; the truth of A enters only through the indicator and the maps chosen in the metatheory.

J.4 Programs that track the maps

All programs below are built from given programs by the operations of Definition 10.5, so they exist by Proposition 10.6; from a program for a function of two arguments a program for the function of the second argument is computed from the first [19, Chapter XII].

Structural maps. The identity is tracked by the program that returns its input, a composite by the composite program, the projections by the two unpairing functions, and a pairing $\langle f, g\rangle$ by the program that pairs the outputs of the trackers of f and g. Evaluation $\mathrm{ev} : Y^X \times X \to Y$ is tracked by the universal program that sends e to $\{(e)_0\}((e)_1)$: if $(e)_0$ names f and $(e)_1$ names x, the run halts with a name of $f(x)$, by the definition of a name of a function. If c tracks $k : Z \times X \to Y$, the curried map λk is tracked by the program that sends a name a of z to a program that, on input b, runs c on the number of the pair (a, b); for every name b of every x that program halts with a name of $k_1(z, x)$, so its output names the function $x \mapsto k_1(z, x)$. A map into Ω is tracked by the program that returns 0, because 0 names every truth value; the same program, read as a program on every input, names every later object of Ω^X.

Numbers and finite data. Successor, sum and product on $\mathbb{N}$ are tracked by their programs, since a name of a number is the number. A constant map with a named value is tracked by a program that ignores its input and returns the name. Coding and decoding of strings, bits and ratio records are computable.

The line. The quotient map $\mathscr{R} \to \mathscr{L}$ is tracked by the identity program, since a name of s names its class. The point $\eta(q)$ of a ratio record is named by the program that returns the code of q on every input. Negation, sum, product, maximum and minimum of points are computed stage by stage from the stages of given regular sequences by the formulas of Proposition 12.19: the sum at stage n reads both sequences at $2n + 1$, and the product reads a stage computed from the first stages of both sequences. So one program for each operation computes, from programs for s and t, a program for the result; this tracks the operation on $\mathscr{L}$, because its value on classes does not depend on the representatives. Multiplication by a fixed ratio is a product with a named point. For a fixed N, the reciprocal $n \mapsto 1/s_{4(N+1)^2(n+1)-1}$ of Lemma F.11 is computed from a program for s; the stage N is part of the program, not computed from s.

The translation of HAS. A name of an environment is a pair of a program for the number valuation ρ and a name of the set valuation. Evaluating a number term at ρ composes the program for ρ with the finitely many arithmetic operations of the term, so the term maps $\ulcorner t \urcorner$ are tracked; $\ulcorner X_j \urcorner$ is tracked by reading the j-th entry of the set valuation through the universal program. The map $\mathrm{push}_{\mathbb{N}}$ is tracked by the program that sends a name of $((\rho, \sigma), n)$ to a name of $(n \cdot \rho, \sigma)$, where $n \cdot \rho$ is computed by a program that returns n on input 0 and, on input $k + 1$, runs the program for ρ on input k; the map $\mathrm{push}_{\mathbb{S}}$ is the same for set valuations; successor, projections and graphs of tracked terms are tracked as above.

The maps of the compactness sentences. Section 16.10 writes every relation as an equation, so each map that the equations of Definition 16.30 apply must be tracked; otherwise its term would not belong to the computable calculus. Each map below is tracked by the program described.

- *Strings.* The length $|w|$, the digit w_i (with the value 0 for $i \geq |w|$), the concatenation wt and the extension wb by one digit are computable functions of codes, and a name of a

string or of a bit is its code; so each is tracked by its own program. The empty string and the bit 1 are constants with named values. These maps let a string variable stand for the initial segment $\bar{p}(n)$, which turns $D(\bar{p}(n)) = 1$ into an equation.

- *Applications.* The terms $p(i)$, $D(w)$ and $f(x)$ apply a sequence, a decidable set or a function on the line to its argument; they are instances of evaluation, tracked by the universal program.

- *The line.* The maximum, the sum and the negation are the operations of the previous paragraphs. The map $k \mapsto \eta(1/(k+1))$ is tracked by the program that computes from k a program returning the code of $1/(k+1)$ on every input, and the points 0 and 1 are named in the same way. With these maps the order of the line becomes the equation $\max(a, b) = b$, and strict order and closeness become existential and conjunctive combinations of such equations.

- *Covers.* The endpoints $a_c(k)$ and $b_c(k)$ are tracked by the program that runs the given name of c on k, decodes the two ratios of the k-th interval and returns programs for their constant sequences. They write membership in an interval as two strict inequalities.

- *Nets.* The number $\ell_K(n)$ and the point $y_K(n, i)$ are tracked by the program that runs the given name of K on n, obtains the list of programs of the entries of the n-th net, and returns its length, or its i-th entry, or a program for the constant sequence 0 when $i \geq \ell_K(n)$. The map $r(n) = 2^n - 1$ on numbers is computable. With them membership in a presented set becomes a condition on its nets, as in Definition 16.30.

- *Packages.* A name of a presented uniform map or of a map package lists programs for its parts. The set K_M is tracked by the program that returns the component naming the set, and the value $f_M(x)$ by the program that runs the component naming the function on the given name of x. They let one variable M range over maps together with their sets.

J.5 Reading the sentences in the model

The separations evaluate closed sentences at the point of **1** in the model of named objects. The evaluation unfolds Definition 16.20. Three observations shorten it: every number, string and bit is named, so a quantifier over $\mathbb{N}$, W or B ranges over all of them; a quantifier over Seq, P or D ranges over the computable objects, by Proposition 16.21(a); and an equation is read at the later stage as the equality of the values of its two sides.

Proof of Lemma 16.31. The quantifiers over $\mathbb{N}$ and W in Definition 16.30 range over all numbers and all strings.

Numbers. Some k has $n + k = N$ exactly when $n \leq N$, and some k has $i + (k+1) = n$ exactly when $i < n$.

Hits. If $D(\bar{p}(n)) = 1$, the string $\bar{p}(n)$ is a witness for $\mathrm{Hit}(D, p, n)$. Conversely a witness w has length n and the digits $p(0), \dots, p(n-1)$, so $w = \bar{p}(n)$ and $D(\bar{p}(n)) = 1$. The same argument, with every string t in the last conjunct, gives CHit.

The order of the line. By Proposition 12.19(5), $\max(a,b)$ is the least upper bound of a and b. If $a \le b$, then b is an upper bound, so $\max(a,b) \le b \le \max(a,b)$ and $\max(a,b) = b$ by item 2; conversely $\max(a,b) = b$ gives $a \le b$. If $0 < y$, Lemma 12.52, item 1, gives a ratio q with $0 < \eta(q) < y$; since $q > 0$ is a ratio, $1/(k+1) \le q$ for some k, and $\eta(1/(k+1)) \le \eta(q) \le y$. Conversely, if $\eta(1/(k+1)) \le y$, then cotransitivity of $<$ (Proposition 12.19(1)) applied to $0 < \eta(1/(k+1))$ gives $0 < y$ or $y < \eta(1/(k+1))$, and the second contradicts $\eta(1/(k+1)) \le y$. Since $a < b$ means $b - a > 0$, the strict order follows. Closeness within $1/(j+1)$ is by definition the conjunction of the two inequalities, the k-th interval of c consists of the points strictly between its ends (Definition 13.5), and $[0,1]$ consists of the points x with $0 \le x \le 1$.

Membership. If $x \in K$, the n-th net has an entry within $2^{-n} = 1/(r(n)+1)$ of x (Definition 13.22), at an index below $\ell_K(n)$. Conversely, suppose every net has such an entry. For every k the entry of the k-th net names a member of K within $2^{-k} \le 1/(k+1)$ of x, so $x \in K$ by Proposition 13.23. No quantifier over points occurs in the formula, so the restriction of the model to named points does not enter. □

Lemma J.2 (The sentences in the model)**.** *At the point of* **1** *in the model of named objects:*

(a) LPO *holds exactly when every computable binary sequence vanishes or has an entry* 1*;*

(b) $\mathsf{WLPO}_{\mathsf{seq}}$ *holds exactly when for every computable binary A there is a computable binary f with $f(i) = 0$ exactly when $A\langle i,n\rangle = 0$ for every n;*

(c) $\mathsf{FAN}^{=}_{\Delta}$ *holds exactly when every computable decidable set of strings that every computable sequence meets has one bound N such that every computable sequence meets it at a length $n \le N$, and* $\mathsf{FAN}^{=}_{c}$ *likewise with c-sets;*

(d) $\mathsf{WKL}^{=}$ *holds exactly when every computable decidable tree with a member of every length has a computable path;*

(e) $\mathsf{HB}^{=}$ *holds exactly when for every named presented set K and every named cover, if every named point of K lies in an interval of the cover, then some N has every named point of K in one of the first N intervals; and* $\mathsf{HB}^{=}_{[0,1]}$ *holds exactly when the same is true of* $[0,1]$ *and every named cover;*

(f) $\mathsf{PLB}^{=}$ *and* $\mathsf{PINF}^{=}$ *hold exactly when the statements of Chapter 13 hold for every named presented uniform map, with every quantifier over points restricted to the named points, and* $\mathsf{PLB}^{=}_{[0,1]}$ *and* $\mathsf{PINF}^{=}_{[0,1]}$ *likewise for every named function on the line that is uniformly continuous on the named points of* $[0,1]$*;* $\mathsf{HC}^{=}$ *holds exactly when every named map package that is pointwise continuous on the named points of its set is uniformly continuous on them, both notions read with the named points only, and* $\mathsf{HC}^{=}_{[0,1]}$ *likewise for every named function on the line and the named points of* $[0,1]$.

Proof. Each item unfolds the clauses of Definition 16.20 and uses the three observations. In (a), $\mathrm{Bin}(\alpha)$ reads as "every value of α is 0 or 1", and the disjunction is read classically. In (b), the inner universal quantifiers range over all numbers, and the pairing term $\langle i,n\rangle$ is read as its value. In (c) to (f) the abbreviations of Definition 16.30 read as the relations they abbreviate, by Lemma 16.31. In (c) and (d), the quantifiers over strings and bits range over all of them, so

Tree(T) and Inf(T) read as in Definition 12.71, and the quantifiers over sequences range over the computable ones. In (e) and (f), the quantifiers over points range over the named points, and the quantifiers over packages, covers and maps over the named ones. □

J.6 Addresses and the dead cover

We use the geometry and the facts (a)–(c) of the proof of Theorem 13.6: the number $\kappa(s)$, the Cantor point $c(s)$, the Cantor interval $J(s)$, the cells $C(s)$, and $u_n = 1/(8 \cdot 4^n)$.

(C3) Addresses of computable sequences are computable. In fact (c), the digit given to a string s of length n reads one stage of the regular sequence, a stage computed from n, and compares it with the ratio $(32\kappa(s) + 16)/2^{2n+5}$, computed from the digits already chosen. So the address is defined by a primitive recursion on the depth from finitely many stages of the sequence, and a program for the sequence gives a program for its address. The recursion is defined for every regular sequence, and the inductive step of fact (c), that $x \in C(s)$ gives $x \in C(sd)$ for the chosen digit d, holds for every point x; so $x \in C(\bar{a}_r(n))$ for every n whenever x lies in $C(\emptyset) = (-1, 2)$.

Names. The Cantor point $c(t)$ is a ratio, so it is named. The dead cover of T_K is named: from k a program decodes w_k, decides $w_k \in T_K$ by Lemma 16.29(a), and returns the ends of $(0, 0)$ or of $C(w_k)$, which are computed from $\kappa(w_k)$ and $|w_k|$ by the recursion that defines the cells. The presentation of $[0, 1]$ in Chapter 13 has, at every precision, a finite list of dyadic ratios; a program computes from n the list of the programs of their constant sequences, so the presented set $[0, 1]$ is named.

J.7 The cover series as a named map

Let c be a cover named by a program, and use the notation of Lemma 13.26. The functions θ_i are built from translation by the ratios a_i and b_i, negation, max, min and the constant 1, and the weights ω_i are ratios. By the programs of Section J.4, one program computes from i, from the program of the cover and from a name of x a name of $\omega_i\theta_i(x)$, and so, by finite summation, a name of the partial sum $\sum_{i<n} \omega_i\theta_i(x)$ from n and a name of x. The diagonal of Theorem 12.25 reads the stage $2n + 1$ of the $(2n + 1)$-th of these names, so one program computes a name of $f_c(x)$ from a name of x. This program is the transformer of the presented uniform map, and it names f_c as a function on the line. A modulus is computed by a program: $f_c(x)$ is within $1/(n + 1)$ of the partial sum at n, by Theorem 12.25; the partial sums at n of two points within $1/(j + 1)$ are within $1/(j + 1)$, because each θ_i moves by at most the move of its argument (Lemma 12.52, item 5) and the weights ω_i add up to less than 1; so at $n = 4k + 3$, points within $1/(2k + 2)$ have values within $1/(4k + 4) + 1/(2k + 2) + 1/(4k + 4) = 1/(k + 1)$, and $\mu(k) = 2k + 1$ is a modulus. Pairing these programs with a program naming the presented set gives a name of the presented uniform map.

J.8 The tent function

We use the facts (d) and (e) of the proof of Theorem 13.13, which hold for every decidable binary tree, in particular for T_K: overlaps of different strings are disjoint and contain no Cantor

point; and if $u \notin T_K$, every tent of a member of T_K of length at least $|u|$ vanishes on $C(u)$.

(T1) Values at the special points. Every tent vanishes at every Cantor point, so every head does, and $\tau(c(s)) = 0$. For $s \in T_K$ the point $g(s)$ lies in $O(s)$ and in no other overlap, so $H_N(g(s)) = \tau_s(g(s)) = 1$ for every $N > |s|$, and $\tau(g(s)) = 1$.

(T2) Heads stop changing on a dead cell. If $u \notin T_K$, fact (e) gives $H_M = H_{|u|}$ on $C(u)$ for every $M \geq |u|$, so $\tau = H_{|u|}$ on $C(u)$.

(T3) The tent function is named. Let a program compute a regular sequence r naming a point x. The program for τ computes the address digits of r by (C3) and searches for the first depth n with $\bar{a}_r(n) \notin T_K$. The address is computable, so by Lemma 16.29(d) the search ends. If x lies in $(-1, 2)$, then $x \in C(\bar{a}_r(n))$ by (C3), and (T2) gives $\tau(x) = H_n(x)$. Otherwise every overlap, which lies in $[0, 1]$, misses x, so every tent vanishes at x, every head is 0 there, and again $\tau(x) = H_n(x)$. The program does not decide which case holds; the two cases are distinguished only in this proof. It then computes a name of $H_n(x)$: the members of T_K of length below n are listed by Lemma 16.29(a), and each tent is built from the tracked operations of Section J.4 and the ratios $g(s)$ and $8 \cdot 4^{|s|}$. Since $\tau(x)$ depends only on x, the program tracks τ on all named points.

(T4) Continuity at named points of $[0, 1]$. Let $x \in [0, 1]$ be named, and let u be an initial segment of its address outside T_K, as in the proof of Theorem 16.34; then $x \in C(u)$. The cell is open with ratio ends, so some j_0 puts every point within $1/(j_0 + 1)$ of x inside $C(u)$, by Lemma 12.52, item 3. On $C(u)$, $\tau = H_{|u|}$ by (T2), a sum of fewer than $2^{|u|}$ tents of strings shorter than $|u|$, each of which changes by at most $8 \cdot 4^{|u|}$ times the change of its argument. Given k, choose $j \geq j_0$ with $2^{|u|} \cdot 8 \cdot 4^{|u|}/(j + 1) \leq 1/(k + 1)$. Every named $y \in [0, 1]$ within $1/(j + 1)$ of x has $\tau(y)$ within $1/(k + 1)$ of $\tau(x)$.

(T5) No uniform modulus. For every j, Lemma 16.29(c) gives $s \in T_K$ of length $n = j + 1$. The ratios $c(s)$ and $g(s)$ lie in $[0, 1]$, at distance $4u_n = 2^{-(2n+1)} \leq 1/(j + 1)$, and by (T1) $\tau(c(s)) = 0$ and $\tau(g(s)) = 1$.

Appendix K

The comparison with Peano arithmetic: routine cases

This appendix proves the results of Section 20.3 whose proofs there were deferred. The conventions of that section apply: every derivation is posit-free, $\Delta \vdash \varphi$ is a posit-free derivation of φ from the hypotheses Δ, and a statement with free variables holds for every substitution of terms. Unless stated otherwise, every derivation constructed here inducts on formulas with quantifiers, so its ledger records the induction entry.

The elementary facts of arithmetic are used freely: $0+t = t$ and $S(t)+u = S(t+u)$; associativity and commutativity of addition; cancellation $a + c = b + c \to a = b$ and its left form; $0 \cdot t = 0$ and $S(t) \cdot u = t \cdot u + u$; distributivity, commutativity and associativity of multiplication; and monotonicity, $a \le b \to a \cdot c \le b \cdot c$, which follows from distributivity. Each has a derivation by the rules of Chapter 5 whose inductions are on equations or on implications between equations, so these derivations have the empty ledger; Chapter 3 gives the same inductions for the constructed numbers.

K.1 Order and pairing

Proof of Lemma 20.2. Every number is zero or a successor: induct on x in $x = 0 \vee \exists y(x = S(y))$.

Order split, $a < b \vee b \le a$. Induct on a, keeping the statement universal in b. At $a = 0$: if $b = 0$ then $b \le 0$ with witness 0; if $b = S(t)$ then $0 < b$ with witness t. At $a + 1$: if $b = 0$ then $b \le a + 1$ with witness $a + 1$. Otherwise $b = t + 1$, and the induction hypothesis at a, t gives $t = a + (k + 1)$, so $b = (a + 1) + (k + 1)$ and $a + 1 < b$; or $a = t + l$, so $a + 1 = (t + 1) + l = b + l$ and $b \le a + 1$.

Disjointness. From $a + (k + 1) = b$ and $b + l = a$ follows $a + ((k + 1) + l) = a + 0$, so cancellation gives $(k + 1) + l = 0$, that is $S(k + l) = 0$, which separation refutes.

Totality. By the order split, take $z = b \cdot b + a$ when $a < b$ and $z = (a \cdot a + a) + b$ when $b \le a$; each branch of G holds by reflexivity.

Uniqueness. If both graphs use the same branch, both values equal the same term. A crossed case supplies $a < b$ and $b \le a$, which disjointness refutes.

Injectivity. Let s be the larger of a, b. If $a < b$ then $s = b$ and $s \cdot s \le \langle a, b\rangle < s \cdot s + s$; if $b \le a$ then $s = a$ and $s \cdot s + s \le \langle a, b\rangle \le s \cdot s + 2s < (s + 1) \cdot (s + 1)$. So every pair with larger coordinate s has its value in $[s^2, (s + 1)^2)$, and these intervals are disjoint for different s by monotonicity: if $s < s'$ then $s + 1 \le s'$ and $(s + 1)^2 \le s'^2$. Equal values therefore have the same s. Within one s the first branch lies below $s^2 + s$ and the second at or above it, so equal values use the same branch, and cancellation in $s^2 + a = s^2 + c$, or in $s^2 + s + b = s^2 + s + d$, identifies the remaining coordinate.

Decoding. First derive $G(a,b,z) \to \exists a', b'\, G(a',b',z+1)$ by four cases. If $a+1=b$, take $(b,0)$: the second branch gives $b\cdot b+b=z+1$. If $a+1<b$, take $(a+1,b)$: the first branch gives $b\cdot b+(a+1)$. If $b=a$, take $(0,a+1)$: the first branch gives $(a+1)\cdot(a+1)=a\cdot a+2a+1$, and $z=a\cdot a+a+a$. If $b<a$, take $(a,b+1)$: the second branch gives $a\cdot a+a+(b+1)$. The cases are exhaustive by the order split. Then induct on z in $\exists a,b\, G(a,b,z)$, starting from $G(0,0,0)$. □

K.2 The computation formula

Coding finite lists

Lemma K.1 (Remainders). *There are derivations of:* $\mathrm{Rem}(a,d,r) \to \mathrm{Rem}(a,d,t) \to r=t$; *and* $0<d \to \exists r\, \mathrm{Rem}(a,d,r)$. *Consequently* $\beta(A,B,i,w)$ *and* $\beta(A,B,i,v)$ *imply* $w=v$, *and every* A,B,i *have some* w *with* $\beta(A,B,i,w)$.

Proof. For uniqueness write $a=qd+r=pd+t$ with $r<d$ and $t<d$. If $q<p$, then $q+1\le p$ and $qd+r<(q+1)d\le pd\le pd+t$, a contradiction; $p<q$ is symmetric; so $q=p$ and cancellation gives $r=t$. For existence induct on a. At 0 use quotient 0 and remainder 0, with the predecessor of d as the slack witness. At $a+1$, open $a=qd+r$ and $r+(s+1)=d$: if $s=0$ then $a+1=(q+1)d+0$; if $s=t+1$ then $a+1=qd+(r+1)$ and $(r+1)+(t+1)=d$. The modulus $1+(i+1)B$ of β is positive. □

Lemma K.2 (Combining two residues). *If* M,d *are positive with* $uM=vd+1$, *and* $r<M$, $s<d$, *then some* z *has* $\mathrm{Rem}(z,M,r)$ *and* $\mathrm{Rem}(z,d,s)$.

Proof. Write $d=h+1$ and put $w=s+hr$ and $z=r+uMw$. Then $z=(uw)M+r$, and, using $uM=vd+1$ and $r+hr=dr$, $z=r+(vd+1)(s+hr)=(r+vw)d+s$. □

Three identities supply the coefficients that a coprimality theorem would otherwise provide. If $b=kt$, and $g=ab$, $h=(a+k)b$, $c=a^2t$, then

$$(1+h)(c+1+g)=(1+g)(g+c+h)+1, \tag{K.1}$$

which follows on expanding, using $h=g+kb$ and $g^2=ckb$. If $nu=mv+1$ and $m=a+1$, then $m(va+1)=n(ua)+1$. If $au=nv+1$ and $cp=nq+1$, then $(ac)(up)=n(nvq+v+q)+1$.

Lemma K.3 (Appending an entry). *For all* j,A,B,w *there are* y,b *such that* $\beta(A,B,i,r)\to\beta(y,b,i,r)$ *for every* $i<j$ *and every* r, *and* $\beta(y,b,j,w)$.

Proof. Some positive L is divisible by every number from 1 to $j+1$: induct, with $L(n+2)$ covering $n+2$. Put $b=L(A+w)$ and $m_i=1+(i+1)b$ for $i\le j$. Every old entry is at most A and $w\le b$, so all are below every m_i. For $i<l\le j$ the difference $l-i$ divides b; writing $b=kt$ with $k=l-i$, identity (K.1) with $a=i+1$ gives an equation $m_lu=m_iv+1$, and the second identity reverses it. Now process positions $0,\dots,j-1$ in order, by induction, keeping a positive M and $x<M$ such that every processed modulus divides M, x has the required residue at each processed position, and for every remaining $l\le j$ some μ,ν satisfy $M\mu=m_l\nu+1$. Start with $M=1$, $x=0$. At position q, take the old entry r there (Lemma K.1), combine x modulo

M with r modulo m_q by Lemma K.2, and reduce modulo $M' = Mm_q$. Reduction keeps both residues, by remainder uniqueness applied to the quotients, and every earlier divisor of M divides M'. The third identity gives the coefficients of M' against each remaining m_l. At the end, combine x modulo M with w modulo m_j; the result y keeps every earlier residue and has residue w at j. □

Proof of Theorem 20.3(4), finite choice. Assume $\forall i\, \exists r\, \varphi(i, r)$ and induct on j. At 0 take $A = B = 0$. At $j + 1$ the induction hypothesis gives A, B for every $i < j$, the premise gives r for $i = j$, and Lemma K.3 gives parameters that keep the earlier entries and add r at j. □

Rows and the exact graph

Write

$$\mathrm{Packed}(K, e, n, y, w) \;:=\; \exists p, q \big(G(n, y, p) \wedge G(e, p, q) \wedge G(K, q, w)\big),$$

so that $\mathrm{Row}(A, B, i; K, e, n, y)$ is $\exists w(\beta(A, B, i, w) \wedge \mathrm{Packed}(K, e, n, y, w))$. By remainder uniqueness and three uses of pairing injectivity, *a row determines its four coordinates*: two rows at the same position of the same table have equal coordinates.

Proof of Theorem 20.3(1). Say that c returns y on n with allowance K when a finite computation tree built by the evaluation clauses has c, n, y at its root, where every node with allowance $s + 1$ has input at most s, calls of a proper part keep the allowance of the node, and the two calls that repeat the program use s. Every computation of Definition 10.5 has such a tree: choose the allowance one more than the sum of the largest input and the number of nested repetitions. Given a tree, list its nodes with each parent before its children, the root first and then the lists of its children; each node's children stand later in the list, so every row satisfies its local condition with later rows. Finite-list coding stores the packed rows, and the root is row 0. Conversely, from an indexed table and a selected row whose program field is the code of an actual program c, build a tree by strong induction on the number of rows after the selected one: every dependency is later, so has fewer rows after it. The constructor of c eliminates every incompatible alternative, because the four basic codes are below 4, compound codes are $4u + b$ with $u \geq 1$ and $b < 4$, and $4u + b = 4v + c$ with $b, c < 4$ forces $b = c$ and $u = v$; pairing injectivity then identifies the part codes. Each alternative rebuilds the corresponding clause. □

Proof of Theorem 20.3(2), functionality. Prove the following *comparison*: if $\mathrm{Indexed}(N, A, B)$ and $\mathrm{Indexed}(M, C, D)$, $i < N$, $j < M$, and row i of the first table is (K, e, n, y) and row j of the second is (L, e, n, t), then $y = t$; the allowances need not agree. Induct strongly on the rank d with $i + d = N$, keeping everything else universal; strong induction is ordinary induction on u in $\forall d \leq u\, F(d)$. Indexedness and row uniqueness supply the local condition of each selected row. A dependency of the first row has a position p with $i < p < N$, and its rank is smaller. Constructor separation, as in the proof of part (1), leaves only matching alternatives.

- *Zero, successor*: both outputs are 0, or both $n + 1$.
- *Projections*: pairing injectivity on the two decompositions of n.

- *Pairing*: compare the dependencies for f on n and for g on n, by the induction hypothesis; pairing uniqueness gives $y = t$.
- *Composition*: compare the dependencies for g on n, giving equal intermediate values; then those for f.
- *Recursion start*: injectivity on $\langle a, 0\rangle$; compare the dependencies for f. A crossed case would give $0 = m + 1$.
- *Recursion step*: injectivity identifies a and m; compare the dependencies for the previous call, which have program e and input $\langle a, m\rangle$ although their allowances may differ, to get equal previous values; the step inputs are then equal; compare the dependencies for g.
- *Search stop*: both outputs are m. A crossed case compares the two test dependencies and equates 0 with a positive value.
- *Search continue*: injectivity identifies a, m; compare the two dependencies for the rest of the search.

Functionality follows by opening the two applications. Every case split eliminates a derived disjunction. □

Closure of the computation formula

For a table (N, A, B), say that (A', B') *raises it by* d when every row (K, e, n, y) of (A, B) at a position $i < N$ is the row $(K + d, e, n, y)$ of (A', B') at i.

Lemma K.4 (Table operations)**.** *There are derivations of the following.*

1. *Every table has a raise by every d, and a raise of an indexed table is indexed. Consequently, if $K \le L$ and some indexed table has a row (K, e, n, y), some indexed table has a row (L, e, n, y).*
2. *Two indexed tables with selected rows can be placed in one indexed table that contains both rows.*
3. *If (T, A, B) is indexed and every table (C, D) that lists the rows of (A, B) one position later satisfies $\mathrm{Local}_{T+1,C,D,0}(K, e, n, y)$, then some indexed table has the row (K, e, n, y).*

Proof. (1) At each position, remainder existence gives the entry, three decodings write it as $\langle K, \langle e, \langle n, y\rangle\rangle\rangle$, and pairing totality packs $(K + d, e, n, y)$; finite choice stores these. Every local condition transfers: the guard $K = k + 1$, $n \le k$ becomes $K + d = (k + d) + 1$, $n \le k + d$; each dependency keeps its position and moves its allowance by d. For the consequence, write $L = K + d$.

(2) Every position below $N + M$ is below N or equal to $N + j'$ with $j' < M$; finite choice stores the entries of the first table in place and those of the second shifted by N. Rows, local conditions and dependencies move with them, and the strict order of dependency positions is kept.

(3) Pairing totality packs the new row; finite choice stores it at position 0 and the old entries one place later. The hypothesis gives the local condition at 0; each old row keeps its local condition shifted by one. □

Proof of Theorem 20.3(3), closure. For a basic clause, the one-row table whose two parameters are the packed row z lists it, since $\beta(z, z, 0, z)$ holds with quotient 0 because $z < 1 + z$; its local condition holds with allowance $n + 1$ and the clause's alternative, using the supplied graph for a projection. For a compound clause, open the premise computations, with allowances K and L, and put $k := K + (L + n)$, so that $n, K, L \le k$. Raise each premise table that computes a proper part to allowance $k + 1$, and a premise that repeats the program (the previous step of a recursion or the rest of a search) to allowance k, by Lemma K.4(1). Place the two tables in one by part (2); a clause with one premise uses its table twice. Add the parent row $(k + 1, e, n, y)$ in front by part (3): the premise rows sit at positions after 0, so they are dependencies of the new row, the guard is $k + 1 = k + 1$ with $n \le k$, and the clause's alternative holds with the supplied code relation, graphs and test value. The identity law follows from the decoding of n, the two projection laws and the pair law. □

K.3 The record test

The checking programs are built from the eight constructors without the search constructor. Write one $:=$ succ $\circ$ zero, $\mathrm{const}_0 :=$ zero, $\mathrm{const}_{k+1} :=$ succ $\circ\, \mathrm{const}_k$, prev $:=$ pair(left, left $\circ$ right), and $f \oplus g$, $f \otimes g$, $f \ominus g$ for add, mul, $\dot{-}$ composed with pair(f, g). A recursion rec(f, g) receives $\langle a, m\rangle$ and recurses on m; at a successor its step receives $\langle a, \langle m, x\rangle\rangle$, with x the previous value.

Program	Constructor expression
ζ	rec(one, zero $\circ$ prev) $\circ$ pair(zero, id)
pred	rec(zero, right $\circ$ prev) $\circ$ pair(zero, id)
$\dot{-}$	rec(id, pred $\circ$ (right $\circ$ right))
add	rec(id, succ $\circ$ (right $\circ$ right))
mul	rec(zero, add $\circ$ pair(left, right $\circ$ right))
cond(t, f, g)	rec$(f, (g \circ \mathrm{left}) \circ \mathrm{prev}) \circ \mathrm{pair}(\mathrm{id}, t)$: the value of f where t returns 0, of g elsewhere
rem	rec(zero, cond(eq(inc, left), inc, zero)) $\circ$ pair(right, left), with inc $:=$ (right $\circ$ right) $\oplus$ one; on $\langle m, d\rangle$ it recurses on the dividend m
entry(A, B, I)	rem $\circ$ pair$(A, ((I \oplus \mathrm{one}) \otimes B) \oplus \mathrm{one})$
eq(f, g)	$\zeta \circ ((f \ominus g) \oplus (g \ominus f))$
le(f, g), lt(f, g)	$\zeta \circ (f \ominus g)$ and $\zeta \circ ((\mathrm{succ} \circ f) \ominus g)$
and, or	$f \otimes g$ and $f \oplus g$
bnd(t, b)	norm(agg(t) $\circ$ pair(id, b)), with agg$(t) :=$ rec(one, mul $\circ$ pair(right $\circ$ right, $t \circ$ prev)) and norm$(f) :=$ rec(zero, (one $\circ$ left) $\circ$ prev) $\circ$ pair(id, f)
Ver	and$\big($bnd(row, slot_5), and(lt(slot_2, slot_5), rowEq(slot_4, slot_3, slot_2; slot_1, e, n, slot_0))$\big)$
T, $\bigcup$	$\zeta \circ$ Ver; right $\circ$ (search(T) $\circ$ pair(id, zero))

The roman add, mul and pred are programs, not the terms of (20.2) or the predecessor of Chapter 9. Here $\mathrm{slot}_k := (\mathrm{right} \circ d_k) \circ \mathrm{right}$ reads the field of c at depth k from the right on the input $z = \langle\langle e, n\rangle, c\rangle$, while $\mathrm{e} := \mathrm{left} \circ \mathrm{left}$ and $\mathrm{n} := \mathrm{right} \circ \mathrm{left}$ read e and n; $\mathrm{rowEq}(A, B, I; K, E, M, Y)$ is $\mathrm{eq}(\mathrm{entry}(A, B, I), \mathrm{pair}(K, \mathrm{pair}(E, \mathrm{pair}(M, Y))))$. The row check row runs at a cell $\langle z, j\rangle$. It reads the justification $\mathrm{entry}(\mathrm{slot}_7 \circ \mathrm{left}, \mathrm{slot}_6 \circ \mathrm{left}, \mathrm{right})$ and the row $\mathrm{entry}(\mathrm{slot}_4 \circ \mathrm{left}, \mathrm{slot}_3 \circ \mathrm{left}, \mathrm{right})$, takes their fields by projections, and returns the conjunction of the guard tests $K = s + 1, n \leq s$ with the disjunction of the ten case tests of Table 20.2, each the conjunction of the equality, order and named-row tests of its line. A named row $j' \mapsto (k', e', n', y')$ is tested by $\mathrm{lt}(\mathrm{right}, j')$, $\mathrm{lt}(j', \mathrm{slot}_5 \circ \mathrm{left})$ and the row comparison at j'.

Lemma K.5 (The test programs)**.** *There are derivations of: ζ returns* 1 *on* 0, 0 *on every successor, and* 0 *or* 1 *on every input; $\dot{-}$ on $\langle a, m\rangle$ returns y when $m + y = a$ and* 0 *when $a \leq m$;* add *and* mul *on $\langle a, m\rangle$ return $a + m$ and $a \cdot m$; if two programs return a and b on n, their equality, order and strict order tests return values on n that are positive exactly when $a = b$, $a \leq b$ and $a < b$, and their conjunction and disjunction return $a \cdot b$ and $a + b$;* rem *on $\langle m, b + 1\rangle$ returns the remainder of m modulo $b + 1$; and the entry program returns the value w with $\beta(A, B, i, w)$. Every search-free program returns a value on every input.*

Proof. Each recursion program is correct by induction along its count, using the recursion laws. Truncated difference returns y on $\langle a, m\rangle$ when $m + y = a$, by induction on m, and 0 on $\langle a, a + d\rangle$, by induction on d. Addition and multiplication are inductions on m. The remainder program keeps the invariant that after m steps its value r satisfies $m = q(b + 1) + r$ and $r + (s + 1) = b + 1$ for some q, s; each step resets to 0 or adds one. The predecessor and ζ are recursions whose step ignores the recursive value. The "only this value" halves follow from functionality: for example, if ζ returned 1 on $k + 1$ it would also return 0 there, giving $1 = 0$. For the tests: if $a = b$ both differences are 0 and ζ returns 1; if $a \neq b$ the order split makes one difference a successor, so the sum is a successor and ζ returns 0. Totality of search-free programs is an induction on the program: the basic laws, the pair and composition laws with pairing totality, and for recursion an induction on the count. □

Proof of Lemma 20.6. The bounded test multiplies the values of the row check at the cells $\langle z, 0\rangle, \ldots, \langle z, N - 1\rangle$ by a recursion on the number of cells and normalizes the product to 0 or 1. By induction on N, it returns a value that is positive exactly when every cell value is positive, since a product of natural numbers is positive exactly when its factors are. For the row check, the graph $G(X, j, p)$ and the projection laws give the field values at p, the entry program gives the justification w with $\beta(C, D, j, w)$ and the row, and the decoder reads their fields. Each atomic test returns a value that is positive exactly when its relation holds of the fields read, by Lemma K.5; a named-row test computes the entry at j' and compares it with the packed row, and pairing injectivity identifies the four coordinates. Products and sums turn these into the case tests, their disjunction and the guard, and the resulting equivalence has the explicit check of Definition 20.4 on its right side. Every program here is search-free, so every value exists, and no value is found by a search: every number the check compares is a field of w or of a named row. Finally T returns ζ of the value of Ver, which is 0 or 1. □

K.4 Closures

For a number a, let κ_a be the program $\mathrm{succ} \circ (\cdots \circ (\mathrm{succ} \circ \mathrm{zero}))$ with a successors; it returns a on every input. Its code $\mathrm{quote}(a)$ satisfies $\mathrm{quote}(0) = 0$ and $\mathrm{quote}(a+1) = 4\langle 1, \mathrm{quote}(a)\rangle + 6$. The *quotation program* qt computes quote by primitive recursion, with pairing, multiplication by 4 and addition of a constant in its step. The *assembler* asm, on input $\langle e, a\rangle$, returns the code

$$4\big\langle e,\ 4\langle \mathrm{quote}(a), \ulcorner \mathrm{id} \urcorner\rangle + 4\big\rangle + 6$$

of the program $e \circ \mathrm{pair}(\kappa_a, \mathrm{id})$, where $\ulcorner \mathrm{id} \urcorner$ is the code of id. The *closure generator* of a program c is

$$\Lambda_c \ := \mathrm{asm} \circ \mathrm{pair}(\kappa_{\ulcorner c \urcorner}, \mathrm{id}).$$

On an environment R it returns the code of $c \circ \mathrm{pair}(\kappa_R, \mathrm{id})$. The body c occurs only inside the constant $\ulcorner c \urcorner$, so constructing the closure never runs c.

Proof of Lemma 20.13. The generator returns. The programs qt, asm, $\kappa_{\ulcorner c \urcorner}$ and id contain no search, so by Lemma K.5 Λ_c returns a value f on every R, and the closure laws show that f is the code displayed above, with every pairing witness retained.

Quoted constants. By induction on a: the program with code $\mathrm{quote}(a)$ returns a on every input. At 0 it is zero. At $a+1$ its code is $4\langle 1, \mathrm{quote}(a)\rangle + 6$, the composition of the successor with the program of code $\mathrm{quote}(a)$, which returns a on every input by the induction hypothesis; the composition and successor laws give $a+1$ on every input. The induction formula quantifies over all inputs.

Behavior. Take x, p, y with $G(R, x, p)$ and $\mathrm{App}(c, p, y)$. The quoted constant returns R on x and id returns x, so the pair law with $G(R, x, p)$ gives the paired value p, and the composition law with $\mathrm{App}(c, p, y)$ gives $\mathrm{App}(f, x, y)$. □

K.5 Reading a variable

Lemma K.6 (Dropping entries)**.** *Let k be at most the length of Θ, and let Θ_k be Θ with its first k entries removed. A context that derives $R \approx_\Theta R'$ derives that some z, z' satisfy $\mathrm{App}(d_k, R, z)$, $\mathrm{App}(d_k, R', z')$ and $z \approx_{\Theta_k} z'$.*

Proof. Recurse on k. At 0 take $z = R$, $z' = R'$, with the identity law. At $k+1$, $\Theta = A, \Theta'$; open $R \approx_{A,\Theta'} R'$ to witnesses r, r', x, x' with $G(r, x, R)$, $G(r', x', R')$ and $r \approx_{\Theta'} r'$. The recursive construction at k gives z, z' for r, r'. The left-projection laws give $\mathrm{App}(\mathrm{left}, R, r)$ and its counterpart, and the composition law gives $\mathrm{App}(d_{k+1}, R, z)$, since $d_{k+1} = d_k \circ \mathrm{left}$. The conclusion contains none of the opened witnesses, so the eliminations close. □

For the variable case of Theorem 20.15, let x_k have type A, so $\Theta_k = A, \Theta''$. The lemma gives $z \approx_{A,\Theta''} z'$; open it to $G(s, v, z)$, $G(s', v', z')$ and $v \simeq_A v'$. The right-projection law gives $\mathrm{App}(\mathrm{right}, z, v)$, and the composition law gives $E_{x_k}(R, v)$, since $c(x_k) = \mathrm{right} \circ d_k$; likewise $E_{x_k}(R', v')$.

K.6 The conversion cases

The remaining contractions in the proof of Theorem 20.22 follow the two compiled programs in the same way. In each, the parts b, g, n of the redex step to b', g', n', and the induction hypotheses relate their values at R and R'.

Iteration at zero, $\mathsf{I}(B, b, g, 0) \Rrightarrow b'$. At R the compiled program evaluates the step to γ, the seed to v and the count to 0, packs a with $G(\gamma, v, a)$ and $\langle a, 0\rangle$, and the recursion-start law with $\mathrm{App}(\mathrm{right}, a, v)$ returns v. The induction hypothesis relates v to the value of b' at R'.

Indexed recursion at zero, $\mathsf{R}(B, b, g, 0) \Rrightarrow b'$: the same, with $\mathsf{It}^{\mathrm{ind}}$.

Indexed recursion at a successor, $\mathsf{R}(B, b, g, S(n)) \Rrightarrow g'\, n'\, \mathsf{R}(B, b', g', n')$. At R the recursion-step law shows that $\mathsf{It}^{\mathrm{ind}}$ on $\langle a, m+1\rangle$ returns s through a previous value x at count m, an output h of γ on m, and an output s of h on x, each application run through U and identified by universal application and functionality. At R' the right side applies γ', related to γ, to the value m of n', and the resulting code to the value x' of $\mathsf{R}(B, b', g', n')$, which the congruence case relates to x. The arrow clause of $\gamma \simeq_{O\to(B\to B)} \gamma'$ at $m = m$ relates h to the code obtained on the right, and its arrow clause at $x \simeq x'$ relates s to the output on the right.

Congruences. A step that rebuilds a constructor around stepped parts is related by the case of Theorem 20.15 for that constructor, applied to the related values of the parts at R and R'. For an abstraction $\lambda A.\, b \Rrightarrow \lambda A.\, b'$ the arrow clause between the two closure codes is derived as in that theorem, with the induction hypothesis for $b \Rrightarrow b'$ at the related extended environments.

K.7 The type conversion case

In Lemma 20.34, a term of type A may be given type B when $A \equiv B$. The reading Adm_T depends on T only through its *skeleton* $\|T\|$: the type obtained by keeping O, $\mathbf{0}$, products and D, and replacing $\mathrm{Id}_A(t, u)$ by $\mathrm{Id}_{|A|}(|t|, |u|)$ and $D(t)$ by $D(|t|)$. Erasure commutes with substitution and sends each generating equation of conversion to an equation between skeletons: the function, iteration, indexed recursion and dependent eliminator equations to computation equations (20.1) of the language of functions; the identity-elimination equation to $|d| = |d|$; and the two discriminator equations to themselves. So conversion of skeletons is generated by the five contractions inside terms and the two discriminator contractions $D(0) \Rrightarrow \mathrm{Id}_O(0, 0)$ and $D(S(t)) \Rrightarrow \mathbf{0}$. These seven contractions are left-linear and do not overlap; erasure has removed the identity-elimination equation, whose left side repeats a variable, so the complete-development argument of Lemma 20.1 gives $\|A\|$ and $\|B\|$ a common reduct, and every erased term on the way stays typed.

The reading moves along each parallel step in both directions. Inside an identity or discriminator type the step changes a term s to s' with $s \equiv s'$; the computation rule proves $s = s'$, and replacement moves each conjunct. At $D(0) \Rrightarrow \mathrm{Id}_O(0, 0)$ the reading $(0 =_O 0) \wedge (v =_O 0)$ becomes $(v =_O 0) \wedge (0 =_O 0)$, which exchanges the conjuncts. At $D(S(t)) \Rrightarrow \mathbf{0}$ the reading $(S(|t|) =_O 0) \wedge (v =_O 0)$ gives $\bot$ by the rule *successor is not zero*, and $\bot$ gives it back by the rule for falsehood. At a product, the step acts on the domain and the codomain, and the induction hypotheses move the antecedent backward and the consequent forward. No step depends on the fragment, so the case holds in each fragment, with or without the discriminator.

Appendix L

One universe and bounded arithmetic: the proofs

This appendix proves Theorems 20.44 and 20.46. Sections L.1–L.8 prove the first: the calculus $\mathscr{F}_U$ with one universe derives the negative translation of Con(PA), extends $\mathscr{E}^+$, and is sound for a set model. Sections L.9–L.12 prove the second: over bounded arithmetic, pair formation and exponentiation derive each other. The two parts are independent.

The first part is long because Con(PA) is long. The sentence says that no finite tables describe an accepted proof of $\bot$, and an accepted proof involves every schema of the checker, every syntactic operation the schemas use, and the bookkeeping that keeps each code in range. The idea is short: inside $\mathscr{F}_U$ every formula code receives a type, built by recursion into the universe; every line the tables certify is shown to have an inhabited type; and the type of $\bot$ is empty. The rest of the part makes that idea precise.

L.1 The checker and the sentence Con(PA)

Codes. Pairs are coded by $P(a,b) = 2^a(2b+1) - 1$. Every number n has exactly one representation $n+1 = 2^a(2b+1)$, so P is a bijection from pairs of numbers onto numbers, and $a \le P(a,b)$ and $b \le P(a,b)$. Lists are coded by $L([\,]) = 0$ and $L(x :: xs) = 1 + P(x, L(xs))$, and a *node* with tag j and children $t_1, \dots, t_m$ by

$$C_j(t_1, \dots, t_m) = 1 + P\big(j, L([t_1, \dots, t_m])\big).$$

Every positive number is exactly one node: decoding $n-1 = P(j,\ell)$ and the list ℓ gives its tag $\mathrm{tag}(n) = j$ and its children, and each child is smaller than n. We put $\mathrm{tag}(0) = 0$ and give 0 no children. The ith child of n, or 0 when there is none, is written $a_i(n)$ for $i = 1, 2, 3$.

The tags have fixed meanings. Natural data are $C_0()$ and $C_1(n)$, with value nv: $\mathrm{nv}(C_1(n)) = \mathrm{nv}(n) + 1$ and $\mathrm{nv} = 0$ otherwise. Terms are variables $C_2(k)$, with k natural data giving a de Bruijn index, zero $C_3()$, successor $C_4(t)$, sum $C_5(s,t)$ and product $C_6(s,t)$. Formulas are equations $C_7(s,t)$, falsehood $C_8()$, implications $C_9(A,B)$ and universal quantifications $C_{10}(A)$, whose body has the bound variable at index 0. Lists of parameters are $C_{15}()$ and $C_{16}(h,\ell)$. Proof nodes are axiom nodes $C_{12}(c,j,\ell)$, modus ponens nodes $C_{13}(c,p,q)$ and generalization nodes $C_{14}(c,p)$; their first child c is the stored conclusion. The code of the zero term is $C_3() = 1 + P(3,0) = 8$. A code *parses* as a term or a formula when its tags and arities are those listed, recursively.

Operations on codes. Four operations act on codes by recursion on the code, leaving a code with any other tag unchanged:

- the *term shift* $\mathrm{tsh}(t,d)$ adds d to the index of every variable of t;

- the *substitution* $\mathrm{sub}(x, d, t)$ replaces the variable with index d by $\mathrm{tsh}(t, d)$, lowers every larger index by one, keeps every smaller index, and passes under C_{10} with $d + 1$ in place of d;
- the *lifting* $\mathrm{lift}(x, c)$ raises every index at least c by one, passing under C_{10} with $c + 1$;
- the *successor shift* $\mathrm{plus}(x, d)$ replaces the variable with index d by its successor, passing under C_{10} with $d + 1$.

Write $A[t]$ for $\mathrm{sub}(A, 0, t)$, $\uparrow A$ for $\mathrm{lift}(A, 0)$ and A^{+} for $\mathrm{plus}(A, 0)$, and $\neg A$ for $C_9(A, C_8())$.

The checker. The fourteen schemas take a list of parameters, each of which must parse as stated (A, B, C formulas, s, t terms):

0	$A \to (B \to A)$	7	$\neg(S(t) = 0)$
1	$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$	8	$S(s) = S(t) \to s = t$
2	$(\neg B \to \neg A) \to (A \to B)$	9	$s + 0 = s$
3	$\forall A \to A[t]$	10	$s + S(t) = S(s + t)$
4	$\forall(\uparrow A \to B) \to (A \to \forall B)$	11	$s \cdot 0 = 0$
5	$t = t$	12	$s \cdot S(t) = s \cdot t + s$
6	$s = t \to (A[s] \to A[t])$	13	$A[0] \to (\forall(A \to A^{+}) \to \forall A)$

The checker with fuel k is K_k: $K_0(p) = 0$, and $K_{k+1}(p) = 1$ exactly when $p \neq 0$ and one of the following holds.

- $p = C_{12}(c, j, \ell)$, c parses as a formula, j is natural data, ℓ is a list of parameters, and c is the instance of schema $\mathrm{nv}(j)$ at ℓ.
- $p = C_{13}(c, p_1, q_1)$, $K_k(p_1) = K_k(q_1) = 1$, c parses as a formula, and the stored conclusion of q_1 is C_9(stored conclusion of p_1, c).
- $p = C_{14}(c, p_1)$, $K_k(p_1) = 1$, c parses as a formula, and $c = C_{10}$(stored conclusion of p_1).

Otherwise $K_{k+1}(p) = 0$. The checker accepts p when $K_{p+1}(p) = 1$; since children are smaller than their parents, the fuel $p + 1$ is enough. An *accepted proof of* $\bot$ is an accepted p whose stored conclusion is $C_8()$. The checker is primitive recursive: each of its parts is, and every recursion it uses is on a smaller code or a smaller fuel.

Certificate tables. A proof is described by thirteen finite tables: tag, a_1, a_2, a_3, the *slacks* w_1, w_2, w_3 with $a_i(x) + S(w_i(x)) = x$ for a node x, nv, the fuel table $K(f, q)$, and the tables $\mathrm{sub}(x, d, t)$, $\mathrm{tsh}(t, d)$, $\mathrm{lift}(x, c)$ and $\mathrm{plus}(x, d)$. A table is a function of numbers; only its values at arguments at most a bound B are constrained. The *clauses* are universal implications between equations and negated equations of table values, with an equation or $\bot$ as conclusion. There are 247 of them, in eleven families, summarized here and listed in full below. In the summary, $x \leq B$ abbreviates $x + k = B$ with a further variable k of the clause.

(a) *Children* (3): $x \leq B \to \mathrm{tag}(x) = S(t) \to a_i(x) + S(w_i(x)) = x$, for $i = 1, 2, 3$.

(b) *Zero* (2): $8 \le B$ and $\mathrm{tag}(8) = 3$.

(c) *Fuel* (2): $K(0,x) = 0$ for $x \le B$; and if $f, q \le B$ and $K(S(f), q) = 1$, then $\mathrm{tag}(q)$ is 12, 13 or 14, in the form $\mathrm{tag}(q) \neq 12 \to \mathrm{tag}(q) \neq 13 \to \mathrm{tag}(q) \neq 14 \to \bot$.

(d) *Modus ponens* (6): if $f, q \le B$, $K(S(f), q) = 1$ and $\mathrm{tag}(q) = 13$, then $K(f, a_2(q)) = K(f, a_3(q)) = 1$, $\mathrm{tag}(a_1(a_3(q))) = 9$, $a_1(a_1(a_3(q))) = a_1(a_2(q))$, $a_2(a_1(a_3(q))) = a_1(q)$, and $\mathrm{tag}(a_3(q)) \neq 0$.

(e) *Generalization* (3): under the same premises with tag 14, $K(f, a_2(q)) = 1$, $\mathrm{tag}(a_1(q)) = 10$ and $a_1(a_1(q)) = a_1(a_2(q))$.

(f) *Axioms* (143): under the premises with tag 12, the schema number $\mathrm{nv}(a_2(q))$ is one of $0, \dots, 13$, in the form $\mathrm{nv}(a_2(q)) \neq 0 \to \cdots \to \mathrm{nv}(a_2(q)) \neq 13 \to \bot$; and, for each schema, the list $a_3(q)$ has as many cells (tag 16) as the schema has parameters, and the stored conclusion $a_1(q)$ has the decoded shape of the instance, read through the tags and children of its nodes. For schema 0, for example, with parameters $P_1 = a_1(a_3(q))$ and $P_2 = a_1(a_2(a_3(q)))$, the seven clauses say that $a_3(q)$ and $a_2(a_3(q))$ have tag 16, that $a_1(q)$ and $a_2(a_1(q))$ have tag 9, and that $a_1(a_1(q)) = P_1$, $a_1(a_2(a_1(q))) = P_2$ and $a_2(a_2(a_1(q))) = P_1$. Where a schema uses sub, lift or plus, the clause names the table value; and where it substitutes a term parameter, a clause refutes each of the formula tags 7, 8, 9, 10 for that parameter.

(g) *Substitution* (23), *term shift* (16), *lifting* (22) and *successor shift* (23): if the table gives the result y of the operation at its arguments, under the guards

$$\begin{aligned} \mathrm{sub}(x,d,t) = y &: \quad x + d \le B,\ t \le B,\ y \le B; \\ \mathrm{tsh}(t,d) = y &: \quad t \le B,\ d \le B,\ y \le B; \\ \mathrm{lift}(x,c) = y &: \quad x + c \le B,\ y \le B; \\ \mathrm{plus}(x,d) = y &: \quad x + d \le B,\ y \le B, \end{aligned}$$

then y has the tag and the children that the recursion of the operation prescribes, case by case on the tag of the argument; for a variable the cases compare its index with the depth; and a code with any other tag is returned unchanged. For the term shift, a result with a formula tag comes from an argument with the same tag. The depth is bounded together with the code: at a node x with tag 10 a clause reads the body $a_1(x)$ at depth $d+1$, and $a_1(x) + d + 1 \le x + d \le B$, so every depth or level that a clause reads is at most B.

(h) *Root* (4): $p \le B$, $K(S(p), p) = 1$, $\mathrm{tag}(a_1(p)) = 8$, and $\mathrm{tag}(p) \neq 0$.

The clauses in (f) are 1 + 142 in number: 7, 16, 13, 10, 11, 4, 18, 7, 11, 6, 11, 6, 12 and 10 for schemas 0 to 13.

The clauses. The clause C_i, for $i = 0, \dots, 246$, is the universal closure, over the variables named below in the order given, of the displayed implication. For a group whose clauses share hypotheses, the shared ones are stated once and come first, in the order shown; each line then lists the remaining hypotheses and the conclusion. We write $a_1 a_3\, q$ for $a_1(a_3(q))$, and

$h \neq m \to \cdots \to h \neq n$ for the hypotheses $h \neq m$, $h \neq m+1$, ..., $h \neq n$ in this order. A hypothesis $u \neq v$ is $\neg(u = v)$, a numeral n is $S^n(0)$, and the variables k, k', k'', k_f, k_q are the slacks of the bounds.

- C_0, C_1, C_2 (children), over x, k, t: $x + k = B \to \mathrm{tag}(x) = S(t) \to a_i(x) + S(w_i(x)) = x$, for $i = 1, 2, 3$ in this order.
- C_3, C_4 (zero): $8 + k_8 = B$; $\mathrm{tag}(8) = 3$.
- C_5 (fuel), over x, k: $x + k = B \to K(0, x) = 0$.

The clauses C_6 to C_{158} are over f, k_f, q, k_q and share the hypotheses $f + k_f = B$, $q + k_q = B$ and $K(S(f), q) = 1$.

- C_6 (fuel): $\mathrm{tag}(q) \neq 12 \to \mathrm{tag}(q) \neq 13 \to \mathrm{tag}(q) \neq 14 \to \bot$.
- C_7 to C_{12} (modus ponens), each after $\mathrm{tag}(q) = 13$: $K(f, a_2\, q) = 1$; $K(f, a_3\, q) = 1$; $\mathrm{tag}(a_1 a_3\, q) = 9$; $a_1 a_1 a_3\, q = a_1 a_2\, q$; $a_2 a_1 a_3\, q = a_1\, q$; $\mathrm{tag}(a_3\, q) = 0 \to \bot$.
- C_{13} to C_{15} (generalization), each after $\mathrm{tag}(q) = 14$: $K(f, a_2\, q) = 1$; $\mathrm{tag}(a_1\, q) = 10$; $a_1 a_1\, q = a_1 a_2\, q$.
- C_{16} (axiom range): $\mathrm{tag}(q) = 12 \to \mathrm{nv}(a_2\, q) \neq 0 \to \cdots \to \mathrm{nv}(a_2\, q) \neq 13 \to \bot$.

The axiom clauses C_{17} to C_{158} have, after the shared hypotheses, the two hypotheses $\mathrm{tag}(q) = 12$ and $\mathrm{nv}(a_2\, q) = s$ for the schema number s; each entry below gives what follows them.

- Schema 0, C_{17} to C_{23}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_2 a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 9$; $a_1 a_1\, q = a_1 a_3\, q$; $\mathrm{tag}(a_2 a_1\, q) = 9$; $a_1 a_2 a_1\, q = a_1 a_2 a_3\, q$; $a_2 a_2 a_1\, q = a_1 a_3\, q$.
- Schema 1, C_{24} to C_{39}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_2 a_3\, q) = 16$; $\mathrm{tag}(a_2 a_2 a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 9$; $\mathrm{tag}(a_1 a_1\, q) = 9$; $a_1 a_1 a_1\, q = a_1 a_3\, q$; $\mathrm{tag}(a_2 a_1 a_1\, q) = 9$; $a_1 a_2 a_1 a_1\, q = a_1 a_2 a_3\, q$; $a_2 a_2 a_1 a_1\, q = a_1 a_2 a_2 a_3\, q$; $\mathrm{tag}(a_2 a_1\, q) = 9$; $\mathrm{tag}(a_1 a_2 a_1\, q) = 9$; $a_1 a_1 a_2 a_1\, q = a_1 a_3\, q$; $a_2 a_1 a_2 a_1\, q = a_1 a_2 a_3\, q$; $\mathrm{tag}(a_2 a_2 a_1\, q) = 9$; $a_1 a_2 a_2 a_1\, q = a_1 a_3\, q$; $a_2 a_2 a_2 a_1\, q = a_1 a_2 a_2 a_3\, q$.
- Schema 2, C_{40} to C_{52}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_2 a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 9$; $\mathrm{tag}(a_1 a_1\, q) = 9$; $\mathrm{tag}(a_1 a_1 a_1\, q) = 9$; $a_1 a_1 a_1 a_1\, q = a_1 a_2 a_3\, q$; $\mathrm{tag}(a_2 a_1 a_1 a_1\, q) = 8$; $\mathrm{tag}(a_2 a_1 a_1\, q) = 9$; $a_1 a_2 a_1 a_1\, q = a_1 a_3\, q$; $\mathrm{tag}(a_2 a_2 a_1 a_1\, q) = 8$; $\mathrm{tag}(a_2 a_1\, q) = 9$; $a_1 a_2 a_1\, q = a_1 a_3\, q$; $a_2 a_2 a_1\, q = a_1 a_2 a_3\, q$.
- Schema 3, C_{53} to C_{62}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_2 a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 9$; $\mathrm{tag}(a_1 a_1\, q) = 10$; $a_1 a_1 a_1\, q = a_1 a_3\, q$; $a_2 a_1\, q = \mathrm{sub}(a_1 a_3\, q, 0, a_1 a_2 a_3\, q)$; $\mathrm{tag}(a_1 a_2 a_3\, q) = 7 \to \bot$; $\mathrm{tag}(a_1 a_2 a_3\, q) = 8 \to \bot$; $\mathrm{tag}(a_1 a_2 a_3\, q) = 9 \to \bot$; $\mathrm{tag}(a_1 a_2 a_3\, q) = 10 \to \bot$.
- Schema 4, C_{63} to C_{73}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_2 a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 9$; $\mathrm{tag}(a_1 a_1\, q) = 10$; $\mathrm{tag}(a_1 a_1 a_1\, q) = 9$; $a_1 a_1 a_1 a_1\, q = \mathrm{lift}(a_1 a_3\, q, 0)$; $a_2 a_1 a_1 a_1\, q = a_1 a_2 a_3\, q$; $\mathrm{tag}(a_2 a_1\, q) = 9$; $a_1 a_2 a_1\, q = a_1 a_3\, q$; $\mathrm{tag}(a_2 a_2 a_1\, q) = 10$; $a_1 a_2 a_2 a_1\, q = a_1 a_2 a_3\, q$.
- Schema 5, C_{74} to C_{77}: $\mathrm{tag}(a_3\, q) = 16$; $\mathrm{tag}(a_1\, q) = 7$; $a_1 a_1\, q = a_1 a_3\, q$; $a_2 a_1\, q = a_1 a_3\, q$.

- Schema 6, C_{78} to C_{95}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_2a_3\,q) = 16$; $\mathrm{tag}(a_2a_2a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 9$; $\mathrm{tag}(a_1a_1\,q) = 7$; $a_1a_1a_1\,q = a_1a_2a_3\,q$; $a_2a_1a_1\,q = a_1a_2a_2a_3\,q$; $\mathrm{tag}(a_2a_1\,q) = 9$; $a_1a_2a_1\,q = \mathrm{sub}(a_1a_3\,q, 0, a_1a_2a_3\,q)$; $a_2a_2a_1\,q = \mathrm{sub}(a_1a_3\,q, 0, a_1a_2a_2a_3\,q)$; $\mathrm{tag}(a_1a_2a_3\,q) = 7 \to \bot$; $\mathrm{tag}(a_1a_2a_3\,q) = 8 \to \bot$; $\mathrm{tag}(a_1a_2a_3\,q) = 9 \to \bot$; $\mathrm{tag}(a_1a_2a_3\,q) = 10 \to \bot$; $\mathrm{tag}(a_1a_2a_2a_3\,q) = 7 \to \bot$; $\mathrm{tag}(a_1a_2a_2a_3\,q) = 8 \to \bot$; $\mathrm{tag}(a_1a_2a_2a_3\,q) = 9 \to \bot$; $\mathrm{tag}(a_1a_2a_2a_3\,q) = 10 \to \bot$.
- Schema 7, C_{96} to C_{102}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 9$; $\mathrm{tag}(a_1a_1\,q) = 7$; $\mathrm{tag}(a_1a_1a_1\,q) = 4$; $a_1a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 3$; $\mathrm{tag}(a_2a_1\,q) = 8$.
- Schema 8, C_{103} to C_{113}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_2a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 9$; $\mathrm{tag}(a_1a_1\,q) = 7$; $\mathrm{tag}(a_1a_1a_1\,q) = 4$; $a_1a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 4$; $a_1a_2a_1a_1\,q = a_1a_2a_3\,q$; $\mathrm{tag}(a_2a_1\,q) = 7$; $a_1a_2a_1\,q = a_1a_3\,q$; $a_2a_2a_1\,q = a_1a_2a_3\,q$.
- Schema 9, C_{114} to C_{119}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 7$; $\mathrm{tag}(a_1a_1\,q) = 5$; $a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 3$; $a_2a_1\,q = a_1a_3\,q$.
- Schema 10, C_{120} to C_{130}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_2a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 7$; $\mathrm{tag}(a_1a_1\,q) = 5$; $a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 4$; $a_1a_2a_1a_1\,q = a_1a_2a_3\,q$; $\mathrm{tag}(a_2a_1\,q) = 4$; $\mathrm{tag}(a_1a_2a_1\,q) = 5$; $a_1a_1a_2a_1\,q = a_1a_3\,q$; $a_2a_1a_2a_1\,q = a_1a_2a_3\,q$.
- Schema 11, C_{131} to C_{136}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 7$; $\mathrm{tag}(a_1a_1\,q) = 6$; $a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 3$; $\mathrm{tag}(a_2a_1\,q) = 3$.
- Schema 12, C_{137} to C_{148}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_2a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 7$; $\mathrm{tag}(a_1a_1\,q) = 6$; $a_1a_1a_1\,q = a_1a_3\,q$; $\mathrm{tag}(a_2a_1a_1\,q) = 4$; $a_1a_2a_1a_1\,q = a_1a_2a_3\,q$; $\mathrm{tag}(a_2a_1\,q) = 5$; $\mathrm{tag}(a_1a_2a_1\,q) = 6$; $a_1a_1a_2a_1\,q = a_1a_3\,q$; $a_2a_1a_2a_1\,q = a_1a_2a_3\,q$; $a_2a_2a_1\,q = a_1a_3\,q$.
- Schema 13, C_{149} to C_{158}: $\mathrm{tag}(a_3\,q) = 16$; $\mathrm{tag}(a_1\,q) = 9$; $a_1a_1\,q = \mathrm{sub}(a_1a_3\,q, 0, 8)$; $\mathrm{tag}(a_2a_1\,q) = 9$; $\mathrm{tag}(a_1a_2a_1\,q) = 10$; $\mathrm{tag}(a_1a_1a_2a_1\,q) = 9$; $a_1a_1a_1a_2a_1\,q = a_1a_3\,q$; $a_2a_1a_1a_2a_1\,q = \mathrm{plus}(a_1a_3\,q, 0)$; $\mathrm{tag}(a_2a_2a_1\,q) = 10$; $a_1a_2a_2a_1\,q = a_1a_3\,q$.

The substitution clauses C_{159} to C_{181} are over x, d, k, t, k', y, k'' (and a further variable j where it occurs, last), with the shared hypotheses $x + d + k = B$, $t + k' = B$, $\mathrm{sub}(x, d, t) = y$ and $y + k'' = B$. After them:

- C_{159}: $\mathrm{tag}(x) = 3 \to \mathrm{tag}(y) = 3$.
- C_{160}: $\mathrm{tag}(x) = 4 \to \mathrm{tag}(y) = 4$.
- C_{161}: $\mathrm{tag}(x) = 5 \to \mathrm{tag}(y) = 5$.
- C_{162}: $\mathrm{tag}(x) = 6 \to \mathrm{tag}(y) = 6$.
- C_{163}: $\mathrm{tag}(x) = 7 \to \mathrm{tag}(y) = 7$.
- C_{164}: $\mathrm{tag}(x) = 8 \to \mathrm{tag}(y) = 8$.
- C_{165}: $\mathrm{tag}(x) = 9 \to \mathrm{tag}(y) = 9$.
- C_{166}: $\mathrm{tag}(x) = 10 \to \mathrm{tag}(y) = 10$.
- C_{167}: $\mathrm{tag}(x) = 4 \to a_1\,y = \mathrm{sub}(a_1\,x, d, t)$.
- C_{168}: $\mathrm{tag}(x) = 5 \to a_1\,y = \mathrm{sub}(a_1\,x, d, t)$.
- C_{169}: $\mathrm{tag}(x) = 6 \to a_1\,y = \mathrm{sub}(a_1\,x, d, t)$.
- C_{170}: $\mathrm{tag}(x) = 7 \to a_1\,y = \mathrm{sub}(a_1\,x, d, t)$.
- C_{171}: $\mathrm{tag}(x) = 9 \to a_1\,y = \mathrm{sub}(a_1\,x, d, t)$.
- C_{172}: $\mathrm{tag}(x) = 5 \to a_2\,y = \mathrm{sub}(a_2\,x, d, t)$.
- C_{173}: $\mathrm{tag}(x) = 6 \to a_2\,y = \mathrm{sub}(a_2\,x, d, t)$.
- C_{174}: $\mathrm{tag}(x) = 7 \to a_2\,y = \mathrm{sub}(a_2\,x, d, t)$.
- C_{175}: $\mathrm{tag}(x) = 9 \to a_2\,y = \mathrm{sub}(a_2\,x, d, t)$.
- C_{176}: $\mathrm{tag}(x) = 10 \to a_1\,y = \mathrm{sub}(a_1\,x, S(d), t)$.
- C_{177}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\,x) + S(j) = d \to y = x$.
- C_{178}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\,x) = d \to y = \mathrm{tsh}(t, d)$.

- C_{179}: $\mathrm{tag}(x) = 2 \to d + S(j) = \mathrm{nv}(a_1\, x) \to \mathrm{tag}(y) = 2$.
- C_{180}: $\mathrm{tag}(x) = 2 \to d + S(j) = \mathrm{nv}(a_1\, x) \to S(\mathrm{nv}(a_1\, y)) = \mathrm{nv}(a_1\, x)$.
- C_{181}: $\mathrm{tag}(x) \neq 2 \to \cdots \to \mathrm{tag}(x) \neq 10 \to y = x$.

The term-shift clauses C_{182} to C_{197} are over t, k, d, k', y, k'' (and a further variable j where it occurs, last), with the shared hypotheses $t + k = B$, $d + k' = B$, $\mathrm{tsh}(t, d) = y$ and $y + k'' = B$. After them:

- C_{182}: $\mathrm{tag}(t) = 3 \to \mathrm{tag}(y) = 3$.
- C_{183}: $\mathrm{tag}(t) = 4 \to \mathrm{tag}(y) = 4$.
- C_{184}: $\mathrm{tag}(t) = 5 \to \mathrm{tag}(y) = 5$.
- C_{185}: $\mathrm{tag}(t) = 6 \to \mathrm{tag}(y) = 6$.
- C_{186}: $\mathrm{tag}(t) = 4 \to a_1\, y = \mathrm{tsh}(a_1\, t, d)$.
- C_{187}: $\mathrm{tag}(t) = 5 \to a_1\, y = \mathrm{tsh}(a_1\, t, d)$.
- C_{188}: $\mathrm{tag}(t) = 6 \to a_1\, y = \mathrm{tsh}(a_1\, t, d)$.
- C_{189}: $\mathrm{tag}(t) = 5 \to a_2\, y = \mathrm{tsh}(a_2\, t, d)$.
- C_{190}: $\mathrm{tag}(t) = 6 \to a_2\, y = \mathrm{tsh}(a_2\, t, d)$.
- C_{191}: $\mathrm{tag}(t) = 2 \to \mathrm{tag}(y) = 2$.
- C_{192}: $\mathrm{tag}(t) = 2 \to \mathrm{nv}(a_1\, y) = \mathrm{nv}(a_1\, t) + d$.
- C_{193}: $\mathrm{tag}(t) \neq 2 \to \cdots \to \mathrm{tag}(t) \neq 6 \to y = t$.
- C_{194}: $\mathrm{tag}(y) = 7 \to \mathrm{tag}(t) = 7$.
- C_{195}: $\mathrm{tag}(y) = 8 \to \mathrm{tag}(t) = 8$.
- C_{196}: $\mathrm{tag}(y) = 9 \to \mathrm{tag}(t) = 9$.
- C_{197}: $\mathrm{tag}(y) = 10 \to \mathrm{tag}(t) = 10$.

The lifting clauses C_{198} to C_{219} are over x, c, k, y, k' (and a further variable j where it occurs, last), with the shared hypotheses $x + c + k = B$, $\mathrm{lift}(x, c) = y$ and $y + k' = B$. After them:

- C_{198}: $\mathrm{tag}(x) = 3 \to \mathrm{tag}(y) = 3$.
- C_{199}: $\mathrm{tag}(x) = 4 \to \mathrm{tag}(y) = 4$.
- C_{200}: $\mathrm{tag}(x) = 5 \to \mathrm{tag}(y) = 5$.
- C_{201}: $\mathrm{tag}(x) = 6 \to \mathrm{tag}(y) = 6$.
- C_{202}: $\mathrm{tag}(x) = 7 \to \mathrm{tag}(y) = 7$.
- C_{203}: $\mathrm{tag}(x) = 8 \to \mathrm{tag}(y) = 8$.
- C_{204}: $\mathrm{tag}(x) = 9 \to \mathrm{tag}(y) = 9$.
- C_{205}: $\mathrm{tag}(x) = 10 \to \mathrm{tag}(y) = 10$.
- C_{206}: $\mathrm{tag}(x) = 4 \to a_1\, y = \mathrm{lift}(a_1\, x, c)$.
- C_{207}: $\mathrm{tag}(x) = 5 \to a_1\, y = \mathrm{lift}(a_1\, x, c)$.
- C_{208}: $\mathrm{tag}(x) = 6 \to a_1\, y = \mathrm{lift}(a_1\, x, c)$.
- C_{209}: $\mathrm{tag}(x) = 7 \to a_1\, y = \mathrm{lift}(a_1\, x, c)$.
- C_{210}: $\mathrm{tag}(x) = 9 \to a_1\, y = \mathrm{lift}(a_1\, x, c)$.
- C_{211}: $\mathrm{tag}(x) = 5 \to a_2\, y = \mathrm{lift}(a_2\, x, c)$.
- C_{212}: $\mathrm{tag}(x) = 6 \to a_2\, y = \mathrm{lift}(a_2\, x, c)$.
- C_{213}: $\mathrm{tag}(x) = 7 \to a_2\, y = \mathrm{lift}(a_2\, x, c)$.
- C_{214}: $\mathrm{tag}(x) = 9 \to a_2\, y = \mathrm{lift}(a_2\, x, c)$.
- C_{215}: $\mathrm{tag}(x) = 10 \to a_1\, y = \mathrm{lift}(a_1\, x, S(c))$.
- C_{216}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\, x) + S(j) = c \to y = x$.
- C_{217}: $\mathrm{tag}(x) = 2 \to c + j = \mathrm{nv}(a_1\, x) \to \mathrm{tag}(y) = 2$.
- C_{218}: $\mathrm{tag}(x) = 2 \to c + j = \mathrm{nv}(a_1\, x) \to \mathrm{nv}(a_1\, y) = S(\mathrm{nv}(a_1\, x))$.
- C_{219}: $\mathrm{tag}(x) \neq 2 \to \cdots \to \mathrm{tag}(x) \neq 10 \to y = x$.

The successor-shift clauses C_{220} to C_{242} are over x, d, k, y, k' (and a further variable j where it occurs, last), with the shared hypotheses $x + d + k = B$, $\mathrm{plus}(x, d) = y$ and $y + k' = B$. After them:

- C_{220}: $\mathrm{tag}(x) = 3 \to \mathrm{tag}(y) = 3$.
- C_{221}: $\mathrm{tag}(x) = 4 \to \mathrm{tag}(y) = 4$.
- C_{222}: $\mathrm{tag}(x) = 5 \to \mathrm{tag}(y) = 5$.
- C_{223}: $\mathrm{tag}(x) = 6 \to \mathrm{tag}(y) = 6$.
- C_{224}: $\mathrm{tag}(x) = 7 \to \mathrm{tag}(y) = 7$.
- C_{225}: $\mathrm{tag}(x) = 8 \to \mathrm{tag}(y) = 8$.
- C_{226}: $\mathrm{tag}(x) = 9 \to \mathrm{tag}(y) = 9$.
- C_{227}: $\mathrm{tag}(x) = 10 \to \mathrm{tag}(y) = 10$.

- C_{228}: $\mathrm{tag}(x) = 4 \to a_1\, y = \mathrm{plus}(a_1\, x, d)$.
- C_{229}: $\mathrm{tag}(x) = 5 \to a_1\, y = \mathrm{plus}(a_1\, x, d)$.
- C_{230}: $\mathrm{tag}(x) = 6 \to a_1\, y = \mathrm{plus}(a_1\, x, d)$.
- C_{231}: $\mathrm{tag}(x) = 7 \to a_1\, y = \mathrm{plus}(a_1\, x, d)$.
- C_{232}: $\mathrm{tag}(x) = 9 \to a_1\, y = \mathrm{plus}(a_1\, x, d)$.
- C_{233}: $\mathrm{tag}(x) = 5 \to a_2\, y = \mathrm{plus}(a_2\, x, d)$.
- C_{234}: $\mathrm{tag}(x) = 6 \to a_2\, y = \mathrm{plus}(a_2\, x, d)$.
- C_{235}: $\mathrm{tag}(x) = 7 \to a_2\, y = \mathrm{plus}(a_2\, x, d)$.
- C_{236}: $\mathrm{tag}(x) = 9 \to a_2\, y = \mathrm{plus}(a_2\, x, d)$.
- C_{237}: $\mathrm{tag}(x) = 10 \to a_1\, y = \mathrm{plus}(a_1\, x, S(d))$.
- C_{238}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\, x) = d \to \mathrm{tag}(y) = 4$.
- C_{239}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\, x) = d \to \mathrm{tag}(a_1\, y) = 2$.
- C_{240}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\, x) = d \to \mathrm{nv}(a_1 a_1\, y) = d$.
- C_{241}: $\mathrm{tag}(x) = 2 \to \mathrm{nv}(a_1\, x) \neq d \to y = x$.
- C_{242}: $\mathrm{tag}(x) \neq 2 \to \cdots \to \mathrm{tag}(x) \neq 10 \to y = x$.

The root clauses C_{243} to C_{246} are $p + k_p = B$, $K(S(p), p) = 1$, $\mathrm{tag}(a_1\, p) = 8$ and $\mathrm{tag}(p) = 0 \to \bot$.

Each table is coded by two numbers through Gödel's β: the value of a table with codes c, d at the index i is the remainder of c modulo $1 + (i+1)d$. The index of one argument a is a; of two, $a + (B+1)b$; of three, $a + (B+1)(b + (B+1)c)$; and the fuel table $K(f, q)$ uses $q + (B+1)f$, with $B + 1$ written $S(B)$. In a clause, a lookup $T(\vec{a}) = v$ is written

$$\forall v\, \forall q\, \forall j \left(c_T = v + S(m) \cdot q \to v + S(j) = S(m) \to \cdots \right), \qquad m = S(i) \cdot d_T,$$

with i the index of $\vec{a}$: it says that v is the remainder of c_T modulo $S(m)$. Each occurrence of a table value in a clause is replaced by its own lookup, as follows. The hypotheses are taken in order and then the conclusion; in an equation, the lookups of its left side come before those of its right side; in a term, the lookups of the arguments of a table come before the lookup of that table. The three quantifiers and two equations of a lookup are placed immediately before the hypothesis or conclusion in which its value occurs, and that value v replaces the table value there. A negated hypothesis $\neg(u = v)$ has its lookups placed before it in the same way.

Definition L.1 (The consistency sentence)**.** Con(PA) is the sentence

$$\forall B\, \forall p\, \forall k_p\, \forall k_8\, \forall c_1\, \forall d_1 \cdots \forall c_{13}\, \forall d_{13}\, \left(C_0 \to C_1 \to \cdots \to C_{246} \to \bot \right)$$

over thirty variables, where $C_0, \ldots, C_{246}$ are the clauses listed above, written with β-lookups, k_8 is the slack of $8 \leq B$ and k_p that of $p \leq B$, and $(c_1, d_1), \ldots, (c_{13}, d_{13})$ code the tables tag, a_1, a_2, a_3, w_1, w_2, w_3, nv, K, sub, tsh, lift and plus, in this order. It is built from $=$, $\bot$, $\to$ and $\forall$ only, and it is closed.

In $\mathbb{N}$, a lookup clause holds exactly when v is the remainder, so Con(PA) is true in $\mathbb{N}$ exactly when no values of the thirty variables satisfy all 247 clauses.

Lemma L.2 (An accepted proof of $\bot$ refutes the sentence)**.** *If the checker accepts a proof p of $\bot$, then* Con(PA) *is false in* $\mathbb{N}$.

Proof. Put $B = \max(p, 8)$ and take the actual tables of the checker: tag, the children and their slacks, nv, K and the four operations, each listed at every index below $(B+2)^3$ and coded by β through the Chinese remainder theorem. Every clause is then a true statement about the checker. The guards of the operation clauses keep every argument of a lookup at most B: at a node with tag 10 the depth $d + 1$ satisfies $a_1(x) + d + 1 \leq x + d \leq B$, and the other recursive

calls keep or lower the sum of code and depth. So every index a clause reads is the index of the arguments it names, and the value found there is the actual value of the operation. The child and zero clauses are the decoding facts above. The fuel, modus ponens, generalization and axiom clauses restate the definition of K_{k+1} at an accepted node. The operation clauses restate the recursions of sub, tsh, lift and plus. The root clauses hold because p is accepted, $p \neq 0$, and its stored conclusion is $C_8()$. So these values satisfy every clause, and the sentence fails at them. □

L.2 The calculus with one universe and its arithmetic

$\mathscr{F}_U$ has the types and terms of $\mathscr{E}^+$ except the primitive discriminator, and adds a type U of codes, a decoding $\mathrm{El}(a)$ of each $a : U$, and four code constructors: $\hat{O}$, $\hat{\mathbf{0}}$, $\widehat{\mathrm{Id}}(a,t,u)$ for $t,u : \mathrm{El}(a)$, and $\hat{\Pi}(a,b)$ for $a : U$ and a family b of codes over $\mathrm{El}(a)$. The decoding equations are conversions:

$$\mathrm{El}(\hat{O}) \equiv O, \quad \mathrm{El}(\hat{\mathbf{0}}) \equiv \mathbf{0}, \quad \mathrm{El}(\widehat{\mathrm{Id}}(a,t,u)) \equiv \mathrm{Id}_{\mathrm{El}(a)}(t,u), \quad \mathrm{El}(\hat{\Pi}(a,b)) \equiv \textstyle\prod_{x:\mathrm{El}(a)} \mathrm{El}(b(x)).$$

The iterator, the indexed recursor and the dependent eliminator of the orbit are kept, and the motive of the dependent eliminator may be any type, including U and products containing U. There is no code for U itself. We write $\hat{\to}$, $\hat{\neg}$ and $\hat{\mathbf{1}} = \widehat{\mathrm{Id}}(\hat{O},0,0)$ for the codes of function types, of negations and of a type with one element, and $\hat{\leftrightarrow}(a,b)$ for the code of $(\mathrm{El}\,a \to \mathrm{El}\,b) \times (\mathrm{El}\,b \to \mathrm{El}\,a)$, built as a product over a selector in O.

Lemma L.3 (Arithmetic in $\mathscr{F}_U$)**.** *$\mathscr{F}_U$ has terms for addition and multiplication (by iteration on the second argument), the predecessor, and an equality test* $\mathrm{eq} : O \to O \to O$*, and derives:*

(a) start separation $\prod_{n:O}(\mathrm{Id}_O(0,Sn) \to \mathbf{0})$ *and injectivity of S;*

(b) $\mathrm{eq}(x,x) = 1$*;* $\mathrm{eq}(x,y) = S(k)$ *gives* $\mathrm{Id}_O(x,y)$*;* $\mathrm{eq}(x,y) = 0$ *refutes it, and a refuted identity has* $\mathrm{eq}(x,y) = 0$*; hence identity on O is decidable and stable:* $\neg\neg\,\mathrm{Id}_O(x,y) \to \mathrm{Id}_O(x,y)$*;*

(c) the laws of addition: associativity, commutativity, $S(x)+y = S(x+y)$*, and cancellation;*

(d) division with remainder: there are terms $\mathrm{rem}(a,d)$*,* $\mathrm{quot}(a,d)$ *and* $\mathrm{slack}(a,d)$ *with* $a = \mathrm{rem}(a,d) + S(d)\cdot\mathrm{quot}(a,d)$ *and* $\mathrm{rem}(a,d) + S(\mathrm{slack}(a,d)) = S(d)$*;*

(e) strong induction: from $\prod_n\big((\prod_{y,j}(y + Sj = n \to \Phi(y))) \to \Phi(n)\big)$ *it derives* $\prod_x \Phi(x)$*, for every family* Φ*.*

Proof. (a) Let $\delta(t)$ be the code defined by the dependent eliminator into U with $\delta(0) = \hat{\mathbf{1}}$ and $\delta(Sn) = \hat{\mathbf{0}}$. Then $\mathrm{El}\,\delta(0) \equiv \mathrm{Id}_O(0,0)$ has the element refl_0, and transporting it along a path from 0 to Sn in the family $\mathrm{El}\,\delta(y)$ gives an element of $\mathrm{El}\,\delta(Sn) \equiv \mathbf{0}$. Injectivity is Theorem 9.2, whose proof uses only the indexed recursor. So δ plays the part of the discriminator D of $\mathscr{E}^+$: its two equations $\mathrm{El}\,\delta(0) \equiv \mathrm{Id}_O(0,0)$ and $\mathrm{El}\,\delta(Sn) \equiv \mathbf{0}$ are conversions.

(b) Define eq by iteration on the first argument, into functions $O \to O$: $\mathrm{eq}(0,y)$ is 1 at $y = 0$ and 0 at a successor, $\mathrm{eq}(Sx,0) = 0$ and $\mathrm{eq}(Sx,Sy) = \mathrm{eq}(x,y)$. The four facts are inductions on x with y general, using (a). Given $\neg\neg\,\mathrm{Id}_O(x,y)$, the value $\mathrm{eq}(x,y)$ is 0 or a successor; 0 would

refute $\mathrm{Id}_O(x, y)$ and contradict the hypothesis, so it is a successor, and the path follows. A test of a tag against a numeral j is $\mathrm{eq}(\mathrm{tag}\, x, j)$, read as true at a successor.

(c) Each law is an induction on one argument with the dependent eliminator, from the recursion equations of addition.

(d) Define the remainder by iteration on a: $\mathrm{rem}(0, d) = 0$, and $\mathrm{rem}(Sa, d)$ is 0 when $\mathrm{rem}(a, d) = d$ and $S(\mathrm{rem}(a, d))$ otherwise; the quotient rises by one exactly when the remainder returns to 0, and the slack is d minus the remainder. Both equations are proved by induction on a, splitting on the test $\mathrm{eq}(\mathrm{rem}(a, d), d)$ with the path or refutation it provides.

(e) Apply the dependent eliminator to the family $Q(n) = \prod_{y,j}(y + Sj = n \to \Phi(y))$. $Q(0)$ holds by start separation. From $Q(n)$ and the hypothesis at n follows $\Phi(n)$, and with $Q(n)$ this gives $Q(Sn)$: a y with $y + Sj = Sn$ is n when $j = 0$, by injectivity, and otherwise satisfies $y + S(j') = n$. Then $\Phi(x)$ follows from $Q(Sx)$ at $y = x$, $j = 0$. □

L.3 The interpretation of codes

Fix in $\mathscr{F}_U$ a bound B, the numbers p, k_p and k_8, and thirteen tables, functions of numbers with one, two or three arguments, and assume the 247 clauses as hypotheses, with Id_O for equality. This context is the *core context*. Within it, codes are evaluated and interpreted. An *environment* is a function $\rho : O \to O$; it assigns values to de Bruijn indices.

Definition L.4 (Evaluation and interpretation with fuel)**.** Let $E_0(x, \rho) = 0$ and $T_0(x, \rho) = \hat{\mathbf{1}}$, and define E_{k+1} and T_{k+1} by the iterator, with types $O \to (O \to O) \to O$ and $O \to (O \to O) \to U$:

$$E_{k+1}(x, \rho) = \begin{cases} \rho(\mathrm{nv}(a_1 x)) & \mathrm{tag}\, x = 2, \\ 0 & \mathrm{tag}\, x = 3, \\ S(E_k(a_1 x, \rho)) & \mathrm{tag}\, x = 4, \\ E_k(a_1 x, \rho) + E_k(a_2 x, \rho) & \mathrm{tag}\, x = 5, \\ E_k(a_1 x, \rho) \cdot E_k(a_2 x, \rho) & \mathrm{tag}\, x = 6, \\ 0 & \text{otherwise}, \end{cases}$$

$$T_{k+1}(x, \rho) = \begin{cases} \hat{\neg}\hat{\neg}\, \widehat{\mathrm{Id}}\big(\hat{O}, E(a_1 x, \rho), E(a_2 x, \rho)\big) & \mathrm{tag}\, x = 7, \\ \hat{\mathbf{0}} & \mathrm{tag}\, x = 8, \\ T_k(a_1 x, \rho) \mathbin{\hat{\to}} T_k(a_2 x, \rho) & \mathrm{tag}\, x = 9, \\ \hat{\Pi}\big(\hat{O},\ m \mapsto T_k(a_1 x, m \cdot \rho)\big) & \mathrm{tag}\, x = 10, \\ \hat{\mathbf{1}} & \text{otherwise}, \end{cases}$$

where $m \cdot \rho$ is the environment with $0 \mapsto m$ and $i + 1 \mapsto \rho(i)$. The *canonical* evaluation and interpretation are $E = E_{SB}$ and $T = T_{SB}$; the equation case uses E. The case distinctions test the tag against the numerals with eq.

So an equation is read as the double negation of an identity, falsehood as the empty type, implication as the function type and the universal quantifier as the product over O; a code that is not a formula is read as a type with one element. The recursion into U is what the universe adds: $\mathscr{E}^+$ has one type depending on a number, the discriminator $D(t)$, but cannot define a type by recursion on a number.

Lemma L.5 (Unfolding). *For every k, x and ρ: if $\operatorname{tag} x = j$, then $E_{k+1}(x, \rho)$ equals the case j of the definition, and $\operatorname{El} T_{k+1}(x, \rho)$ has maps in both directions to the decoding of the case j; if the tag is none of the listed values, $E_{k+1}(x, \rho) = 0$ and $\operatorname{El} T_{k+1}(x, \rho)$ is inhabited.*

Proof. The step function is applied to $\operatorname{tag} x$; transport along the path $\mathrm{Id}_O(\operatorname{tag} x, j)$ in the family of step values moves the claim to the numeral j, where the tests compute. For the default case, each test against a listed numeral returns 0, because the tag differs from it (Lemma L.3(b)). □

Lemma L.6 (Fuel). *If $x \le B$ and $x + S(j) = k$, then $E_k(x, \rho) = E_{k+1}(x, \rho)$ and $\operatorname{El} T_k(x, \rho) \leftrightarrow \operatorname{El} T_{k+1}(x, \rho)$. Consequently, for every $x \le B$ the canonical E and T satisfy the unfolding equations of Lemma L.5 with E and T on both sides: for instance $E(x, \rho) = S(E(a_1 x, \rho))$ when $\operatorname{tag} x = 4$, and $\operatorname{El} T(x, \rho) \leftrightarrow \prod_m \operatorname{El} T(a_1 x, m \cdot \rho)$ when $\operatorname{tag} x = 10$.*

Proof. Induct on k. At $k = 0$ there is no such x, by start separation. At $k + 1$, unfold both sides at the tag of x. The children of a node $x \le B$ are smaller than x by the child clauses (a), so they satisfy the hypothesis at k, and the induction hypothesis identifies the values or the types of the children. The second statement applies the first at $k = B$ to the children of x. □

L.4 Environments and the syntactic lemmas

Environments are compared pointwise: $\rho \approx \rho'$ when $\rho(i) = \rho'(i)$ for every i; the calculus is not asked for function extensionality. Let $\operatorname{tl} \rho(i) = \rho(Si)$ and $\operatorname{sh}_d \rho(i) = \rho(i + d)$, and define by recursion on d and c:

$$\begin{aligned} \operatorname{ins}_0(\rho, w) &= w \cdot \rho, & \operatorname{ins}_{d+1}(\rho, w) &= \rho(0) \cdot \operatorname{ins}_d(\operatorname{tl} \rho, w), \\ \operatorname{del}_0 \rho &= \operatorname{tl} \rho, & \operatorname{del}_{c+1} \rho &= \rho(0) \cdot \operatorname{del}_c(\operatorname{tl} \rho), \\ \operatorname{upd}_0 \rho &= S(\rho(0)) \cdot \operatorname{tl} \rho, & \operatorname{upd}_{d+1} \rho &= \rho(0) \cdot \operatorname{upd}_d(\operatorname{tl} \rho). \end{aligned}$$

They insert w at position d, delete position c, and raise the value at position d by one. By induction on d and c: $\operatorname{ins}_d(\rho, w)$ has $\rho(i)$ at $i < d$, w at d, and $\rho(d + m)$ at $S(d + m)$; $\operatorname{del}_c \rho$ has $\rho(i)$ at $i < c$ and $\rho(S(c + m))$ at $c + m$; $\operatorname{upd}_d \rho$ has $S(\rho(d))$ at d and $\rho(i)$ elsewhere; and each operation preserves $\approx$ and commutes with $m \cdot -$ by raising the position, for instance $m \cdot \operatorname{ins}_d(\rho, w) \approx \operatorname{ins}_{d+1}(m \cdot \rho, w)$.

Each of the following lemmas is proved by strong induction on the code x (Lemma L.3(e)), case by case on its tag. In each case the table clauses of family (g) give the tag and the children of the result, the unfolding equations of Lemma L.6 reduce the claim to the children, which are smaller and in range by the child clauses, and the induction hypothesis finishes. Only the variable cases need more, and they are given.

Lemma L.7 (Agreement). *If $x \le B$ and $\rho \approx \rho'$, then $E(x, \rho) = E(x, \rho')$ and $\operatorname{El} T(x, \rho) \leftrightarrow \operatorname{El} T(x, \rho')$.*

Proof. A variable is evaluated by applying the environment, and the environments agree there. Under C_{10} the environments $m \cdot \rho$ and $m \cdot \rho'$ agree pointwise. □

Lemma L.8 (Term shift). *If $t \le B$, $d \le B$ and $\mathrm{tsh}(t,d) = y \le B$, then $E(y,\rho) = E(t, \mathrm{sh}_d\,\rho)$.*

Proof. For a variable with index i the result is the variable with index $i{+}d$, and $\rho(i{+}d) = \mathrm{sh}_d\,\rho(i)$. The other term tags commute with the shift. □

Lemma L.9 (Substitution). *Let $x + d \le B$, $t \le B$ and $\mathrm{sub}(x,d,t) = y \le B$, and put $\sigma = \mathrm{ins}_d\big(\rho, E(t, \mathrm{sh}_d\,\rho)\big)$. Then $E(y,\rho) = E(x,\sigma)$. If moreover $\mathrm{tag}\,t$ is not 7, 8, 9 or 10, then $\mathrm{El}\,T(y,\rho) \leftrightarrow \mathrm{El}\,T(x,\sigma)$.*

Proof. Let x be a variable with index i, and compare i with d by trichotomy. If $i < d$, then $y = x$, and $\sigma(i) = \rho(i)$. If $i = d$, then $y = \mathrm{tsh}(t,d)$, whose value is $E(t, \mathrm{sh}_d\,\rho)$ by Lemma L.8, which is $\sigma(d)$. If $i = S(d+m)$, then y is the variable with index $d+m$, and $\sigma(S(d+m)) = \rho(d+m)$. Under C_{10} the depth becomes $d+1$ and the environment $m \cdot \rho$, and $m \cdot \sigma \approx \mathrm{ins}_{d+1}(m \cdot \rho, w)$ with the same w, because $\mathrm{sh}_{d+1}(m \cdot \rho) \approx \mathrm{sh}_d\,\rho$; Lemma L.7 absorbs the pointwise equality. For the interpretation, a variable code is read as $\hat{\mathbf{1}}$; its image under substitution is either a variable or $\mathrm{tsh}(t,d)$, which by the term-shift clauses has a formula tag only if t has, so it is read as $\hat{\mathbf{1}}$ too. □

Lemma L.10 (Lifting and successor shift). *If $x + c \le B$ and $\mathrm{lift}(x,c) = y \le B$, then $E(y,\rho) = E(x, \mathrm{del}_c\,\rho)$ and $\mathrm{El}\,T(y,\rho) \leftrightarrow \mathrm{El}\,T(x, \mathrm{del}_c\,\rho)$. If $x + d \le B$ and $\mathrm{plus}(x,d) = y \le B$, then $E(y,\rho) = E(x, \mathrm{upd}_d\,\rho)$ and $\mathrm{El}\,T(y,\rho) \leftrightarrow \mathrm{El}\,T(x, \mathrm{upd}_d\,\rho)$.*

Proof. For lifting, a variable with index $i < c$ is kept and $\mathrm{del}_c\,\rho(i) = \rho(i)$; a variable with index $c+m$ becomes $S(c+m)$, and $\mathrm{del}_c\,\rho(c+m) = \rho(S(c+m))$. For the successor shift, the variable with index d becomes its successor, whose value is $S(\rho(d)) = \mathrm{upd}_d\,\rho(d)$, and every other variable is kept, with $\mathrm{upd}_d\,\rho(i) = \rho(i)$ for $i \ne d$. Under C_{10} the cutoff or depth rises by one and the environment operations commute with $m \cdot -$. □

Lemma L.11 (Stability). *If $x \le B$, then $\neg\neg\,\mathrm{El}\,T(x,\rho) \to \mathrm{El}\,T(x,\rho)$.*

Proof. An equation is read as a negation, and a negation is stable. The empty type is stable. An implication is stable when its conclusion is, and a product when its body is, by the induction hypothesis. The type with one element is inhabited. □

L.5 The checker is sound inside $\mathscr{F}_U$

Lemma L.12 (The fourteen schemas). *Let $f \le B$, $q \le B$, $K(Sf,q) = 1$, $\mathrm{tag}\,q = 12$ and $\mathrm{nv}(a_2 q) = s$ with $s \le 13$. Then $\mathrm{El}\,T(a_1 q,\rho)$ for every ρ.*

Proof. The axiom clauses for schema s decode the stored conclusion $c = a_1 q$ into nodes whose tags and children are the schema's, with the parameters P_1, P_2, P_3 read from the list $a_3 q$. The unfolding equations of Lemma L.6 then turn $\mathrm{El}\,T(c,\rho)$ into the schema's type built from $\mathrm{El}\,T(P_i,\cdot)$ and $E(P_i,\cdot)$.

0, 1. The types $\alpha \to (\beta \to \alpha)$ and $(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$ have the usual combinators as elements.

2. From $\neg\beta \to \neg\alpha$ and α follows $\neg\neg\beta$, and Lemma L.11 gives β.

3. From $\prod_m \mathrm{El}\, T(A, m \cdot \rho)$ take $m = E(t, \rho)$. Lemma L.9 at $d = 0$, whose term t is not a formula code by the refuting clauses, identifies $\mathrm{El}\, T(A[t], \rho)$ with $\mathrm{El}\, T(A, E(t, \rho) \cdot \rho)$, since $\mathrm{ins}_0(\rho, w) = w \cdot \rho$ and $\mathrm{sh}_0\, \rho \approx \rho$.

4. Given $h : \prod_m(\mathrm{El}\, T(\uparrow A, m \cdot \rho) \to \mathrm{El}\, T(B, m \cdot \rho))$ and $a : \mathrm{El}\, T(A, \rho)$, return $m \mapsto h(m)(a')$, where a' is a moved along Lemma L.10: $\uparrow A$ at $m \cdot \rho$ is A at $\mathrm{del}_0(m \cdot \rho) = \rho$.

5. $\lambda k.\, k(\mathrm{refl})$ inhabits $\neg\neg\, \mathrm{Id}(\nu, \nu)$.

6. From $\neg\neg\, \mathrm{Id}(E(s, \rho), E(t, \rho))$ and $\mathrm{El}\, T(A[s], \rho)$, assume $\neg\, \mathrm{El}\, T(A[t], \rho)$. A path from $E(s, \rho)$ to $E(t, \rho)$ would transport the second hypothesis, read through Lemma L.9 as A at $E(s, \rho) \cdot \rho$, to A at $E(t, \rho) \cdot \rho$, which is $A[t]$ at ρ; so there is no path, against the first hypothesis. Hence $\neg\neg\, \mathrm{El}\, T(A[t], \rho)$, and Lemma L.11 removes the double negation.

7. $\neg\neg\, \mathrm{Id}(S(E(t, \rho)), 0) \to \mathbf{0}$ follows from start separation.

8. From $\neg\neg\, \mathrm{Id}(Su, Sv)$ derive $\neg\neg\, \mathrm{Id}(u, v)$ by composing with injectivity of S.

9–12. The evaluation equations of Lemma L.6 give the identities $E(s + 0) = E(s)$, $E(s + St) = S(E(s + t))$, $E(s \cdot 0) = 0$ and $E(s \cdot St) = E(s \cdot t + s)$ from the recursion equations of addition and multiplication; a path gives its double negation.

13. Let $b : \mathrm{El}\, T(A[0], \rho)$ and $h : \prod_m(\mathrm{El}\, T(A, m \cdot \rho) \to \mathrm{El}\, T(A^+, m \cdot \rho))$. Apply the dependent eliminator to the family $n \mapsto \mathrm{El}\, T(A, n \cdot \rho)$, a family of types obtained from codes in U. The base is b, moved by Lemma L.9 at the zero code 8, whose value is 0 by the zero clauses. The step from n to Sn applies $h(n)$ and moves the result by Lemma L.10: A^+ at $n \cdot \rho$ is A at $\mathrm{upd}_0(n \cdot \rho) \approx S(n) \cdot \rho$. The result is $\prod_n \mathrm{El}\, T(A, n \cdot \rho)$, the reading of $\forall A$. This is the one schema whose proof needs the universe: the motive varies with the code A, which is data inside the derivation.

In each case the clauses give the tags that select the unfolding equations, and the hypotheses $f \le B$ and $q \le B$, with the child clauses, keep every code in range. □

Lemma L.13 (Every certified line is true)**.** *For every f with $f \le SB$: if $q \le B$ and $K(f, q) = 1$, then $\mathrm{El}\, T(a_1 q, \rho)$ for every ρ.*

Proof. Induct on f. At $f = 0$ the fuel clause $K(0, q) = 0$ contradicts $K(0, q) = 1$ by start separation. At Sf, decide tag q against 12, 13 and 14. With tag 12, the schema number is one of $0, \ldots, 13$ by the axiom-range clause, and Lemma L.12 applies. With tag 13, the modus ponens clauses give $K(f, a_2 q) = K(f, a_3 q) = 1$ and an implication $a_1(a_3 q)$ from $a_1(a_2 q)$ to $a_1 q$; the induction hypothesis gives the two premises at ρ, and the unfolding equation for tag 9 applies the second to the first. With tag 14, the generalization clauses give $K(f, a_2 q) = 1$ and $a_1 q = C_{10}(a_1(a_2 q))$ in decoded form; the induction hypothesis at $m \cdot \rho$, for every m, gives the product that reads $a_1 q$. Any other tag contradicts the fuel clause. The bound $f \le SB$ keeps the predecessor fuel within B. □

Proposition L.14 (The core derives $\bot$)**.** *In the core context, $\mathscr{F}_U$ derives* $\mathbf{0}$.

Proof. By the root clauses, $K(Sp, p) = 1$ with $p \le B$, so $Sp \le SB$, and Lemma L.13 gives $\mathrm{El}\, T(a_1 p, \rho_0)$ at the environment ρ_0 that is constantly 0. The tag of p is not 0, so its first child is in range, and its tag is 8; the unfolding equation for tag 8 turns the element into one of $\mathbf{0}$. □

L.6 From the core to the sentence

The negative translation of Con(PA) is the product, over thirty variables of type O and one hypothesis for each clause, of $\mathbf{0}$; the hypothesis of a clause is its negative translation, in which every equation $u = v$ becomes $\neg\neg\, \mathrm{Id}_O(u, v)$. We build a term of this type from Proposition L.14 by substituting for every variable of the core context.

Lemma L.15 (The core context from the hypotheses)**.** *Under the thirty variables and the translated clause hypotheses, $\mathscr{F}_U$ derives every entry of the core context at the following values: the bound and the variables p, k_p, k_8 are themselves; each table T is the lookup $\vec{a} \mapsto \mathrm{rem}(c_T, m)$, the remainder modulo $S(m)$, with $m = S(i) \cdot d_T$ and i the index of $\vec{a}$; and the evaluation and interpretation are E_k and T_k of Definition L.4, whose unfolding equations are Lemma L.5.*

Proof. Take a clause with its table form and its translated arithmetic form. Each lookup in the arithmetic form is a universal statement over v, q and j; instantiate it at $v = \mathrm{rem}(c_T, m)$, $q = \mathrm{quot}(c_T, m)$ and $j = \mathrm{slack}(c_T, m)$ of Lemma L.3(d), and supply its two equations, which division with remainder proves, under double negation by $\lambda k.\, k(\cdot)$. A hypothesis $u = v$ of the table form gives the translated hypothesis $\neg\neg\, \mathrm{Id}(u, v)$ in the same way; a negated equation is its own translation, since $\neg\neg\neg\alpha$ and $\neg\alpha$ imply each other. The conclusion of the arithmetic form is $\neg\neg\, \mathrm{Id}(u, v)$ or $\mathbf{0}$, and stability of identity on O, Lemma L.3(b), gives the identity the table form asks for. The three arithmetic facts used here, division with remainder, the bound on the remainder and stability, are closed derivations of $\mathscr{F}_U$. □

Proof of Theorem 20.44(1). Substituting the values of Lemma L.15 into the derivation of Proposition L.14 gives a derivation of $\mathbf{0}$ in the context of the thirty variables and the translated clause hypotheses, with the three arithmetic facts cut against their closed derivations. Abstracting over that context, in order, gives a closed term of the negative translation of Con(PA). The sentence is closed, the checker is primitive recursive, and by Lemma L.2 an accepted proof of $\bot$ would refute the sentence in $\mathbb{N}$. □

L.7 The embedding and the strictness

This section proves parts (2) and (4); part (3) is proved in Section L.8.

Proof of Theorem 20.44(2). Send each type and term of $\mathscr{E}^+$ to itself, except the discriminator, which goes to $\mathrm{El}\, \delta(t)$ with the code δ of Lemma L.3(a). The two discriminator equations become the decoding conversions of δ, and every other rule of $\mathscr{E}^+$ is a rule of $\mathscr{F}_U$, so induction on derivations sends each derivation of $\mathscr{E}^+$ to one of $\mathscr{F}_U$ of the translated judgment. A negative

arithmetic type contains no discriminator, so it is sent to itself. If Peano arithmetic derives σ, then $\mathscr{E}^+$ proves σ^N by Theorem 20.39, and so does $\mathscr{F}_U$. □

Proof of Theorem 20.44(4). Peano arithmetic is consistent, by Theorem 20.41, so by Theorem 20.42 it does not derive Con(PA). Proposition 20.43 then says that $\mathscr{E}^+$ does not prove $\mathrm{Con(PA)}^N$, and part (1) says that $\mathscr{F}_U$ does. □

L.8 The set model

This section proves part (3). We work in set theory with a *universe of sets*: a transitive set $\mathbb{U}$, closed under pairs, unions, power sets and images of its members under maps into $\mathbb{U}$, and containing the set N of numerals defined below. The sets of rank below a strongly inaccessible cardinal form such a universe. In a type-theoretic metatheory with a hierarchy of type universes $\mathscr{U}_0$, $\mathscr{U}_1$, $\mathscr{U}_2$, the well-founded sets built over $\mathscr{U}_1$ can serve as the sets; those built over $\mathscr{U}_0$, each lifted to the higher level, then form a universe in this sense, and their collection is itself one of the sets. The members of $\mathbb{U}$ are called *small*.

Sets for the constructions. The numerals are $\nu_0 = \emptyset$ and $\nu_{n+1} = \{\nu_n\}$; they are distinct, $N = \{\nu_n : n \in \mathbb{N}\}$, and $\mathrm{dec}(\nu)$ is n when $\nu = \nu_n$ and 0 for any other set, with $\mathrm{s}(\nu) = \nu_{\mathrm{dec}(\nu)+1}$. For a set A and a map F from A to sets, $\Pi(A, F)$ is the set of functions $f \subseteq A \times \bigcup_{x \in A} F(x)$ with domain A and $f(x) \in F(x)$ for every $x \in A$; the graph of a map h on A is $\{(x, h(x)) : x \in A\}$; the application $\mathrm{ap}(f, x)$ is a y with $(x, y) \in f$ when there is one, chosen once and for all, and $\emptyset$ otherwise; and $\mathrm{Id}(x, y)$ is $\{\emptyset\}$ when $x = y$ and $\emptyset$ otherwise. Iteration $\mathrm{it}(b, G, \nu)$ applies G $\mathrm{dec}(\nu)$ times to b, and recursion $\mathrm{rec}(b, G, \nu)$ is the value at $\mathrm{dec}(\nu)$ of the sequence that starts at b and passes from its value y at n to $G(\nu_n, y)$.

Lemma L.16 (The small sets)**.** *N, $\emptyset$ and every $\mathrm{Id}(x, y)$ are small. If A is small and $F(x)$ is small for every $x \in A$, then $\Pi(A, F)$ is small.*

Proof. N is small by assumption, and $\emptyset$ and $\{\emptyset\}$ are built from it by power sets and pairs. $\Pi(A, F)$ is a subset of the power set of $A \times \bigcup_{x \in A} F(x)$; the union of the image of A under F is small, the product of two small sets is built by pairs and power sets, and a subset of a small set is a member of its power set, hence small by transitivity. □

Definition L.17 (The interpretation)**.** An *environment* ρ assigns a set to each variable index; $\rho.x$ puts x at index 0 and $\rho(k)$ at $k+1$. Raw types and terms of $\mathscr{F}_U$, typed or not, receive values $[\![A]\!]_\rho$ and $[\![t]\!]_\rho$:

$$
\begin{array}{lllll}
[\![O]\!] & = N & & [\![\mathbf{0}]\!] & = \emptyset \\
[\![U]\!] & = \mathbb{U} & & [\![\mathrm{El}\, a]\!]_\rho & = [\![a]\!]_\rho \\
[\![x_k]\!]_\rho & = \rho(k) & & [\![0]\!] & = \nu_0 \\
[\![St]\!]_\rho & = \mathrm{s}([\![t]\!]_\rho) & & [\![ft]\!]_\rho & = \mathrm{ap}([\![f]\!]_\rho, [\![t]\!]_\rho) \\
[\![\textstyle\prod_A B]\!]_\rho & = \Pi([\![A]\!]_\rho,\ x \mapsto [\![B]\!]_{\rho.x}) & & [\![\mathrm{Id}_A(t,u)]\!]_\rho & = \mathrm{Id}([\![t]\!]_\rho, [\![u]\!]_\rho)
\end{array}
$$

The value of $\lambda A.\, b$ is the graph of $x \mapsto [\![b]\!]_{\rho.x}$ on $[\![A]\!]_\rho$; refl and the elimination of $\mathbf{0}$ have the value $\emptyset$, and path elimination has the value of its base. The iterator with base b, step g

and argument t has value $\mathrm{it}(\llbracket b\rrbracket_\rho, \mathrm{ap}(\llbracket g\rrbracket_\rho, \cdot), \llbracket t\rrbracket_\rho)$; the indexed recursor and the dependent eliminator have value $\mathrm{rec}(\llbracket b\rrbracket_\rho, (n, y) \mapsto \mathrm{ap}(\mathrm{ap}(\llbracket g\rrbracket_\rho, n), y), \llbracket t\rrbracket_\rho)$. The codes have the values of the types they decode to: $\hat{O}$ has N, $\hat{\mathbf{0}}$ has $\varnothing$, $\widehat{\mathrm{Id}}(a, t, u)$ has $\mathrm{Id}(\llbracket t\rrbracket_\rho, \llbracket u\rrbracket_\rho)$, and $\hat{\Pi}(a, b)$ has $\Pi(\llbracket a\rrbracket_\rho, x \mapsto \llbracket b\rrbracket_{\rho.x})$. So the four decoding conversions preserve values.

Lemma L.18 (Values and substitution)**.** *For a substitution σ, write $\sigma\rho$ for the environment $k \mapsto \llbracket \sigma_k\rrbracket_\rho$. Then $\llbracket A\sigma\rrbracket_\rho = \llbracket A\rrbracket_{\sigma\rho}$ and $\llbracket t\sigma\rrbracket_\rho = \llbracket t\rrbracket_{\sigma\rho}$; in particular $\llbracket b[t]\rrbracket_\rho = \llbracket b\rrbracket_{\rho.\llbracket t\rrbracket_\rho}$.*

Proof. Simultaneous induction on A and t, with σ and ρ general. Under a binder the substitution is lifted, and the lifted substitution evaluated at $\rho.x$ is $(\sigma\rho).x$, since a weakened term has at $\rho.x$ the value it has at ρ. Every other clause of Definition L.17 is computed from the values of the parts. □

Raw conversion does not preserve these values as it stands. The application of $\lambda\mathbf{0}.\, S(0)$ to 0 converts by β to $S(0)$, whose value is ν_1; but the value of the abstraction is the empty graph, so the application has the value $\mathrm{ap}(\varnothing, \nu_0) = \varnothing = \nu_0$. A conversion between two typed expressions may pass through untyped ones, so the values are controlled by an external condition.

Definition L.19 (Safety)**.** Safety of a type or a term at an environment ρ is defined by recursion on the expression. The types O, $\mathbf{0}$ and U, the variables, 0, refl, the elimination of $\mathbf{0}$, $\hat{O}$ and $\hat{\mathbf{0}}$ are safe. Otherwise:

- $\prod_A B$ is safe when A is, and B is safe at $\rho.x$ for every $x \in \llbracket A\rrbracket_\rho$; $\lambda A.\, b$ and $\hat{\Pi}(a, b)$ likewise, with b and with $\llbracket a\rrbracket_\rho$ for the domain;
- $\mathrm{Id}_A(t, u)$ and $\widehat{\mathrm{Id}}(a, t, u)$ are safe when t and u are; $\mathrm{El}\, a$ when a is; path elimination when its base is;
- St is safe when t is and $\llbracket t\rrbracket_\rho \in N$;
- ft is safe when f and t are and there are A and F with $\llbracket f\rrbracket_\rho \in \Pi(A, F)$ and $\llbracket t\rrbracket_\rho \in A$;
- an iterator with base b, step g and argument t is safe when b, g, t are, $\llbracket t\rrbracket_\rho \in N$, and there is a set V with $\llbracket b\rrbracket_\rho \in V$ and $\llbracket g\rrbracket_\rho \in \Pi(V, x \mapsto V)$; an indexed recursor likewise, with $\llbracket g\rrbracket_\rho \in \Pi(N, n \mapsto \Pi(V, y \mapsto V))$; a dependent eliminator likewise, with a family Q of sets in place of V: $\llbracket b\rrbracket_\rho \in Q(\nu_0)$ and $\llbracket g\rrbracket_\rho \in \Pi(N, n \mapsto \Pi(Q(n), y \mapsto Q(\mathrm{s}(n))))$.

If every σ_k is safe at ρ and A is safe at $\sigma\rho$, then $A\sigma$ is safe at ρ, and likewise for terms: by induction on the expression, with Lemma L.18 for the conditions on values and the lifted substitution under binders. In particular $b[t]$ is safe at ρ when b is safe at $\rho.\llbracket t\rrbracket_\rho$ and t is safe at ρ.

Definition L.20 (Parallel reduction)**.** Parallel reduction $\Rightarrow$ on raw types and terms is the least relation that is closed under every constructor (each part reduces in parallel, and every expression reduces to itself) and contains:

- $(\lambda A.\, b)\, t \Rightarrow b'[t']$ when $b \Rightarrow b'$ and $t \Rightarrow t'$;

- for each of the three eliminators, the step at 0 to the reduct of the base, and the step at St: the iterator to $g'(\mathrm{it}\,\ldots\,t')$ and the indexed recursor and dependent eliminator to $g'\,t'\,(\mathrm{rec}\,\ldots\,t')$, with all parts reduced in parallel;
- the four decodings $\mathrm{El}\,\hat{O} \Rightarrow O$, $\mathrm{El}\,\hat{\mathbf{0}} \Rightarrow \mathbf{0}$, $\mathrm{El}\,\widehat{\mathrm{Id}}(a,t,u) \Rightarrow \mathrm{Id}_{\mathrm{El}\,a'}(t',u')$ and $\mathrm{El}\,\hat{\Pi}(a,b) \Rightarrow \prod_{\mathrm{El}\,a'} \mathrm{El}\,b'$;
- the erasures: path elimination with base d reduces to the reduct of d, and $\mathrm{refl}(t)$ and the elimination of $\mathbf{0}$ reduce to $\mathrm{refl}(0)$.

The erasures are not conversions of $\mathscr{F}_U$; they are allowed here because the values ignore those fields, and they make the conversion β of path elimination at refl a reduction.

Lemma L.21 (Parallel reduction preserves safety and value)**.** *If $A \Rightarrow B$ and A is safe at ρ, then B is safe at ρ and $[\![A]\!]_\rho = [\![B]\!]_\rho$; likewise for terms.*

Proof. Induction on the reduction, with ρ general. The closure cases follow from the induction hypotheses part by part; under a binder the domain has the same value before and after, so the hypothesis applies at every $\rho.x$ with x in it. For $(\lambda A.\,b)\,t$: safety gives A_0, F with the graph of $x \mapsto [\![b]\!]_{\rho.x}$ on $[\![A]\!]_\rho$ in $\Pi(A_0,F)$, and $[\![t]\!]_\rho \in A_0$; a graph in $\Pi(A_0,F)$ has domain A_0, so $A_0 = [\![A]\!]_\rho$, the application has the value $[\![b]\!]_{\rho.[\![t]\!]_\rho}$, and b is safe there. By the induction hypotheses and Lemma L.18 this is the value of $b'[t']$, which is safe by the remark after Definition L.19. For an eliminator at St: the value of St is a numeral ν_{n+1} with $n = \mathrm{dec}[\![t]\!]_\rho$, so the eliminator's value is the step applied to its value at t; that value lies in V (or in $Q(\nu_n)$), by induction on n, and the step lies in $\Pi(V, x \mapsto V)$ (or the corresponding product), so the reduct is safe with the same value. At 0 the value is that of the base. The decodings have equal values by Definition L.17, and their safety conditions coincide. Path elimination has the value of its base; $\mathrm{refl}(t)$, $\mathrm{refl}(0)$ and the elimination of $\mathbf{0}$ all have the value $\varnothing$ and are safe. □

Lemma L.22 (Triangle property and confluence)**.** *Let A^* be the complete development of A: every redex of A that is present when A is read from the root is contracted, the parts first, and every other constructor is kept with its parts developed. If $A \Rightarrow B$, then $B \Rightarrow A^*$; likewise for terms. Hence if $A \Rightarrow^* B$ and $A \Rightarrow^* C$ there is D with $B \Rightarrow^* D$ and $C \Rightarrow^* D$.*

Proof. Induction on $A \Rightarrow B$. In a closure case where A is not itself a redex, A^* is the same constructor applied to the developments of the parts, and the induction hypotheses give the step. If A is a redex and the step contracted it, the parts of B reduce to the developments of the parts of A, and parallel reduction is closed under substitution of related terms, so the contractum of B's parts reduces to A^*. If A is a redex and the step kept it, the step from B contracts the same redex in B, with the same result. The cases of the decodings and the erasures are direct. The triangle property makes $\Rightarrow$ satisfy the diamond property, with A^* as the common reduct of any two one-step reducts, and the diamond property passes to $\Rightarrow^*$ by the usual tiling. □

Lemma L.23 (Every conversion joins)**.** *If two raw types are convertible in $\mathscr{F}_U$, both reduce by $\Rightarrow^*$ to a common type; likewise for terms.*

Proof. Write $A \downarrow B$ when A and B have a common reduct. By Lemma L.22 this relation is an equivalence; it is a congruence, since parallel reduction is closed under every constructor; and it is closed under substitution, since parallel reduction is. Each basic conversion of $\mathscr{F}_U$, β for products and for path elimination, the six equations of the eliminators and the four decodings, is a single parallel step. Induction on the conversion therefore gives $A \downarrow B$. □

Corollary L.24 (Coherence)**.** *If A and B are convertible and both safe at ρ, then $[\![A]\!]_\rho = [\![B]\!]_\rho$; likewise for terms.*

Proof. Follow both reduction sequences to the common reduct of Lemma L.23; by Lemma L.21, each step keeps safety and value. □

Proposition L.25 (Soundness of the set model)**.** *Say that ρ satisfies a context when each $\rho(k)$ lies in the value of the type of the kth entry, read at the tail of ρ. Then for every derivable judgment and every ρ that satisfies its context: a derivable type is safe; a derivable term $t : A$ is safe and $[\![t]\!]_\rho \in [\![A]\!]_\rho$; derivable equalities of types or of terms relate safe expressions with equal values; and a derivable substitution sends ρ to an environment that satisfies the source context, with safe terms.*

Proof. Induction on derivations. Variables are the satisfaction of the context. An abstraction has as value a graph in the product, and an application of a function in $\Pi(A, F)$ to a member of A is safe with value in the fiber, which is the value of $B[t]$ by Lemma L.18. The rule of conversion uses Corollary L.24: the derivation of the target type makes it safe, and the induction hypothesis makes the source safe. refl lies in $\mathrm{Id}(x, x) = \{\emptyset\}$. For path elimination, a member of $\mathrm{Id}([\![x]\!], [\![y]\!])$ forces the two values to be equal and the path to be $\emptyset$, the value of refl, so the motive at (y, p) has the value it has at (x, refl), where the base lies. The elimination of $\mathbf{0}$ is vacuous, since nothing lies in $\emptyset$. For the three eliminators of O, the argument is a numeral, the motive's values give the set V or the family Q required for safety, and the value lies in the motive at the argument by induction on the numeral. The universe rules use Lemma L.16: the codes of $\mathscr{F}_U$ have small values, and El is the identity on values. The equality rules hold by the same computations, the β- and ι-rules by Lemma L.21. □

Proof of Theorem 20.44(3). By Proposition L.25, a closed term of $\mathbf{0}$ would have a value in $[\![\mathbf{0}]\!] = \emptyset$; so there is none. A negative arithmetic type is inhabited at an environment of numerals exactly when its formula is true at the decoded environment, by induction on the formula: $\neg\neg\,\mathrm{Id}$ is inhabited exactly when the identity is, a product over N is inhabited exactly when every instance is, and functions between inhabited or empty sets exist exactly as implication says. So every arithmetic sentence whose negative translation $\mathscr{F}_U$ derives is true. A Peano derivation of $\bot$ would give, by part (2), a closed term of $\mathbf{0}$; so Peano arithmetic is consistent. With part (1), Con(PA) is true, and by Lemma L.2 the checker accepts no proof of $\bot$. □

L.9 Completeness and the models of bounded arithmetic

The second part works with the classical rules: the rules of Chapter 5 with excluded middle, without the omniscience and Markov rules. Formulas have de Bruijn variables $x_0, x_1, \ldots$, and a *sentence* has none free. A *structure* interprets 0, S, $+$ and $\cdot$ on a set and satisfies the six *rule checks*: $S(x) \neq 0$, injectivity of S, and the recursion equations of addition and multiplication. A formula is *bounded* when it is built from equations and $\bot$ by $\wedge$, $\vee$, $\to$ and the two bounded quantifiers

$$\forall y \forall z (y + z = t \to \psi), \qquad \exists y \exists z (y + z = t \wedge \psi),$$

read $\forall y \leq t \psi$ and $\exists y \leq t \psi$. A *model of* $I\Delta_0$ is a structure in which every induction instance $\varphi(0) \to \forall x (\varphi(x) \to \varphi(Sx)) \to \forall x \varphi(x)$ of a bounded formula φ holds at every assignment to its other variables. $I\Delta_0$ *derives* φ from a list E when a classical derivation from E has only induction nodes on bounded formulas.

For a set T of formulas, $T \vdash \varphi$ means that finitely many members of T derive φ by the classical rules with no induction node; T is *consistent* when $T \nvdash \bot$.

Theorem L.26 (Completeness of the classical rules)**.** *Let T be a set of sentences and φ a sentence. If φ holds in every structure that satisfies every member of T, then $T \vdash \varphi$.*

Proof. This is Gödel's completeness theorem, proved by Henkin's construction [11]. The calculus has no constants besides 0 and none can be added, so free variables serve as witnesses. The argument uses Zorn's lemma, in the classical metatheory.

Witnesses. Give every formula ψ a variable index w_ψ, injective in ψ and larger than the index of every free variable of ψ; for instance the pair of a code of ψ and a bound on its free variables. The *Henkin axiom* of ψ is $H_\psi = \exists \psi \to \psi[x_{w_\psi}/x_0]$; its free variables lie below $w_\psi + 1$.

Adding the Henkin axioms keeps a consistent set of sentences consistent. Suppose $T, H_{\psi_1}, \ldots, H_{\psi_n} \vdash \bot$ with distinct ψ_i, and let $\psi = \psi_n$ have the largest witness w. The variable x_w is free in no member of T, in no other H_{ψ_i} (whose free variables lie below $w_{\psi_i} + 1 \leq w$), and not in $\exists \psi$. From the hypotheses without H_ψ, the derivation gives $\neg H_\psi$, hence $\exists \psi$ and $\neg \psi[x_w/x_0]$ by classical logic. Generalizing over the fresh x_w gives $\forall x_0 \neg \psi$, which contradicts $\exists \psi$. So the hypotheses without H_ψ already derive $\bot$, and induction on n removes all Henkin axioms.

A maximal set. The union of a chain of consistent sets is consistent, because a derivation uses finitely many members. By Zorn's lemma a consistent set extends to a maximal consistent set S. Such an S contains every formula it derives, contains exactly one of χ and $\neg \chi$, contains a conjunction, disjunction or implication exactly when the truth table on its parts says so, and, if it contains every Henkin axiom, contains $\forall a$ exactly when it contains $a[t]$ for every term t: if $\forall a \notin S$, then $\neg \forall a \in S$, so $\exists \neg a \in S$, and the Henkin axiom of $\neg a$ puts $\neg a[x_w]$ in S. The existential quantifier is treated in the same way.

The term model. Call terms t and s equivalent when $t = s$ is in S. The rules for equality make this a congruence for S, $+$ and $\cdot$, so the classes form a structure, and it satisfies the six rule checks because S contains their instances. At the assignment sending x_i to its class, the value of a term is its class, and by induction on the number of connectives and quantifiers, which substitution does not change, a formula is true exactly when it is in S.

Completeness. If $T \nvdash \varphi$, then $T \cup \{\neg\varphi\}$ is consistent; add the Henkin axioms and extend to a maximal S. The term model satisfies every member of T and $\neg\varphi$, all of them sentences, so φ fails in a structure satisfying T. □

Theorem L.27 (Bounded arithmetic is sound and complete for its models)**.** *For sentences E and φ, $I\Delta_0$ derives φ from E exactly when φ holds in every model of $I\Delta_0$ in which every member of E holds.*

Proof. Soundness is induction on the derivation: each rule preserves truth at every assignment in a structure with the rule checks, and an induction node on a bounded formula is an instance that holds in every model of $I\Delta_0$.

For completeness, let T consist of E and the universal closures of the induction instances of all bounded formulas. A structure that satisfies T is a model of $I\Delta_0$, since an instance at a given assignment follows from its closure, and it satisfies E; so φ holds in it. Theorem L.26 gives finitely many members of T that derive φ without induction. Each closed induction axiom used has its own derivation: one induction node on its bounded formula, followed by generalizations. Substituting these derivations gives an $I\Delta_0$ derivation of φ from E. □

So to show that $I\Delta_0$ derives a sentence from E, it is enough to show the sentence true in an arbitrary model of $I\Delta_0$ satisfying E, and that is what the next three sections do. In such a model every predicate below is defined by a bounded formula with the same truth condition, so every induction used is an instance of bounded induction.

L.10 Arithmetic in a model of bounded arithmetic

Fix a model M of $I\Delta_0$.

Lemma L.28 (Induction, order and division)**.** *In M:*

(a) For a bounded φ: induction; strong induction, from $\forall v(\forall w < v\varphi(w) \to \varphi(v))$ to $\forall v\varphi(v)$; and the least number principle.

(b) Addition and multiplication are associative and commutative, multiplication distributes over addition, 0 and 1 are neutral, and addition cancels.

(c) Every element is 0 or a successor. The relation $x \le y$, defined by $\exists z(x + z = y)$, is a total order, $x < y + 1$ exactly when $x \le y$, and the order is compatible with addition and with multiplication by any element; a product is 0 only when a factor is.

(d) For $d \neq 0$, every a is $q \cdot d + r$ with $r < d$, and q and r are unique.

Proof. Induction for bounded formulas is an axiom of $I\Delta_0$; we prove (b) and (c) first, by induction alone, and then the strong forms in (a). (b) Each law is an induction on an equation, in the order of Chapter 3. (c) The first statement is an induction on $x = 0 \vee \exists y \le x(x = y + 1)$; totality is an induction on y, with x fixed, in the bounded formula $\exists z \le y(x + z = y) \vee \exists z \le x(y + z = x)$; the rest follows by cancellation. (a) Strong induction is induction on the bounded

formula $\forall w \le n\, \varphi(w)$, using $w \le 0 \to w = 0$ and $w \le n+1 \to w \le n \vee w = n+1$ from (c), and the least number principle is strong induction applied to $\neg\varphi$. (d) Existence is an induction on a in $\exists q \le a\, \exists r \le d (a = q \cdot d + r \wedge r < d)$: at $a+1$ either $r+1 < d$, or $r+1 = d$ and q rises by one. Uniqueness: if $q < q'$ then $q' \cdot d + r' \ge q \cdot d + d > q \cdot d + r$. □

Write $\mathrm{quo}(a,d)$ and $\mathrm{rem}(a,d)$ for the quotient and remainder. They satisfy $\mathrm{rem}(a + k \cdot m, d) = \mathrm{rem}(a,d)$ and $\mathrm{rem}(\mathrm{rem}(a,m),d) = \mathrm{rem}(a,d)$ whenever d divides m, by uniqueness. The *digit block* of B at a is $\mathrm{val}(B,a) = \mathrm{rem}(\mathrm{quo}(B,a),a)$; it is unchanged by adding a multiple of m, or by passing to the remainder modulo m, when $a \cdot a$ divides m, and $\mathrm{val}(B,a) = \mathrm{quo}(B,a)$ when $B < a \cdot a$. Every element is $2k$ or $2k+1$ and not both, by division by 2.

An element u is a *power of two* when $u \ne 0$ and in every factorization $u = d \cdot e$ with $d, e \le u$ the factor d is 1 or even.

Lemma L.29 (Powers of two)**.** *In M:* 1 *is a power of two; every divisor of a power of two is a power of two; if* w *is a power of two so is* $2w$*; a power of two other than* 1 *is* $2w$ *for a power of two* w*; if* $u \le v$ *are powers of two then* u *divides* v*; if* $u < v$ *then* $2u \le v$*; a product of powers of two is a power of two; and every* $m \ne 0$ *has a power of two* u *with* $u \le m < 2u$*.*

Proof. The first four facts follow from the definition and parity: an odd divisor of $2w$ other than 1 would divide w, or would make $2w$ odd. Divisibility of comparable powers is a strong induction on v in the bounded formula

$$v \text{ a power of two} \to \forall u \le v (u \text{ a power of two} \to \exists k \le v (v = u \cdot k)) \ :$$

if $v = 1$ then $u = 1$; otherwise $v = 2v'$ and either $u = v$, or $u = 1$, or $u = 2u'$ with $u' \le v'$, and the hypothesis at v' gives $u' \mid v'$. If $u < v$, then $v = u \cdot k$ with $k \ne 1$ a power of two, so k is even and $v \ge 2u$. Products are a strong induction on the second factor, through $u \cdot 2w' = 2(u \cdot w')$. The last fact is an induction on m: if $m < 2u$ fails at $m+1$, then $m+1 = 2u$, and $2u$ is the new power. □

Say that a is a *digit* of N, written $\mathrm{Bit}(a,N)$, when $N = (2q+1) \cdot a + r$ with $r < a$; for a power of two this is binary digit $\log_2 a$. For $a \ne 0$, $\mathrm{Bit}(a,N)$ holds exactly when $a \le \mathrm{rem}(N, 2a)$; it is unchanged by adding a multiple of m or passing to the remainder modulo m when $2a$ divides m; it implies $a \le N$; and $a \le N < 2a$ implies it.

L.11 The exponentiation graph

The graph says $y = 2^x$ by exhibiting the computation of 2^x by repeated squaring, read from the top binary digit of x down. With $x_j = \lfloor x / 2^j \rfloor$, the checkpoints $y_j = 2^{x_j}$ satisfy $y_j = y_{j+1}^2 \cdot 2^{b_j}$, where b_j is a binary digit of x. A certificate stores the checkpoints as the binary digits of one number N and the exponents x_j in the digit blocks of one number B at the checkpoints.

Definition L.30 (The graph)**.** A *step at u* for N and B is one of

- $1 < u$, $\mathrm{Bit}(u^2, N)$, $\mathrm{rem}(N, u^2) < 2u$ and $\mathrm{val}(B, u^2) = 2\,\mathrm{val}(B, u)$;
- $\mathrm{Bit}(2u^2, N)$, $\mathrm{rem}(N, 2u^2) < 2u$ and $\mathrm{val}(B, 2u^2) = 2\,\mathrm{val}(B, u) + 1$.

$\mathrm{Cert}(x, y, N, B)$ says: y is a power of two, $N < 2y$, $\mathrm{Bit}(1, N)$, $\mathrm{Bit}(y, N)$, $B < y^2$, $\mathrm{val}(B, y) = x$, and every power of two $u < y$ with $\mathrm{Bit}(u, N)$ has a step at u. The graph is

$$\mathrm{Pow}_2(x, y) \;:=\; \exists N \le 2y\, \exists B \le y \cdot y\ \mathrm{Cert}(x, y, N, B),$$

with remainders, quotients and digit blocks written through their bounded definitions. It is a bounded formula with free variables x and y.

A step at u names the next checkpoint $w = (1 + b)u^2$ with $b \in \{0, 1\}$ and $u < w$: the digit of N at w is one, no digit of N lies strictly between u and w (the condition $\mathrm{rem}(N, w) < 2u$), and the stored exponent doubles, plus b.

Lemma L.31 (First facts)**.** $\mathrm{Cert}(x, y, N, B)$ *gives* $x < y$*, and* $\mathrm{Pow}_2(x, y)$ *makes y a power of two.* $\mathrm{Pow}_2(0, 1)$ *holds, with* $N = 1$ *and* $B = 0$*, and* $\mathrm{Pow}_2(x, 1)$ *gives* $x = 0$.

Proof. $x = \mathrm{val}(B, y) < y$. At $y = 1$ there is no power of two below y, and every remainder modulo 1 is 0. □

Lemma L.32 (Extension)**.** *If* $\mathrm{Pow}_2(x', y')$*,* $b \in \{0, 1\}$ *and* $y' < (1 + b)y'^2$*, then* $\mathrm{Pow}_2(2x' + b, (1 + b)y'^2)$.

Proof. Let (N', B') certify (x', y'), and put $y = (1 + b)y'^2$, $x = 2x' + b$, $N = N' + y$ and $B = B' + x \cdot y$. Then y is a power of two, $2y' \le y$, and $x < y$ because $x' < y'$. So $N < 2y$, $B < y^2$, $\mathrm{Bit}(y, N)$ since $y \le N < 2y$, and $\mathrm{Bit}(1, N)$ because adding y, a multiple of 2, keeps it. The block of B at y is x, since $B' < y$; blocks at powers $a \le y'$ are unchanged, since a^2 divides y. Let $u < y$ be a power of two with $\mathrm{Bit}(u, N)$. Then $\mathrm{Bit}(u, N')$, because $2u$ divides y, so $u \le N' < 2y'$ and $u \le y'$. If $u < y'$, the old step at u leads to some $w \le y'$, and its three conditions survive, since $2w$ divides y and the blocks at u and w are unchanged. If $u = y'$, the new step leads to y: $\mathrm{rem}(N, y) = N' < 2y'$, and the block at y is $x = 2\,\mathrm{val}(B', y') + b$. □

Lemma L.33 (Decomposition)**.** *If* $\mathrm{Pow}_2(x, y)$ *and* $y \ne 1$*, then there are* x'*,* y' *and* $b \in \{0, 1\}$ *with* $\mathrm{Pow}_2(x', y')$*,* $y' < y$*,* $y = (1 + b)y'^2$ *and* $x = 2x' + b$.

Proof. Let (N, B) certify (x, y) and $m = \mathrm{rem}(N, y)$. Since y is even, $\mathrm{Bit}(1, m)$, so $m \ne 0$; let u be the power of two with $u \le m < 2u$. Then $u < y$ and $\mathrm{Bit}(u, N)$, so there is a step at u to some w. If $w < y$, the digit of m at $w \ge 2u$ would put $w \le m < 2u$; if $w > y$, the top digit y of N would be a digit of $\mathrm{rem}(N, w) < 2u \le y$. So $w = y$, and with $x' = \mathrm{val}(B, u)$ the step gives $y = (1 + b)u^2$ and $x = 2x' + b$. The pair $(\mathrm{rem}(N, 2u), \mathrm{rem}(B, u^2))$ certifies (x', u): its digits are those of N below $2u$, its blocks those of B at powers up to u, and every step at a power $a < u$ leads to a $w_a \le u$ (a larger w_a would see the digit u of N inside $\mathrm{rem}(N, w_a) < 2a \le u$), so it survives the remainders. □

Theorem L.34 (The laws of the graph)**.** *In every model of $I\Delta_0$:*

(a) $\mathrm{Pow}_2(0,1)$*, and* $\mathrm{Pow}_2(0,y) \to y = 1$*;*

(b) upward recursion: $\mathrm{Pow}_2(x,y) \to \mathrm{Pow}_2(x+1,2y)$*;*

(c) downward recursion: $\mathrm{Pow}_2(x+1,q) \to \exists p \le q\,(q = p+p \wedge \mathrm{Pow}_2(x,p))$*;*

(d) functionality: $\mathrm{Pow}_2(x,p) \to \mathrm{Pow}_2(x,p') \to p = p'$*;*

(e) positive values: $\mathrm{Pow}_2(x,p) \to p \neq 0$*;*

(f) initial-segment domain: $\mathrm{Pow}_2(x,p) \to \forall y \le x \exists p'\ \mathrm{Pow}_2(y,p')$*.*

Proof. Each is a strong induction on the power v, over a bounded formula with v as the induction variable; Lemma L.33 supplies the smaller power and Lemma L.32 rebuilds. Positive values hold because p is a power of two.

(a) The first is Lemma L.31. For the second, if $\mathrm{Pow}_2(0,v)$ with $v \neq 1$, decomposition gives $0 = 2x' + b$, so $x' = b = 0$, and $\mathrm{Pow}_2(0,y')$ with $y' < v$ and $v = y'^2$; the hypothesis gives $y' = 1$, hence $v = 1$.

(b) Induct on v in $\forall a \le v(\mathrm{Pow}_2(a,v) \to \mathrm{Pow}_2(a+1,2v))$. At $v = 1$, $a = 0$, and extension with $b = 1$ gives $\mathrm{Pow}_2(1,2)$. Otherwise decompose $a = 2x' + b$, $v = (1+b)y'^2$. If $b = 0$, extend (x',y') with $b = 1$: $(2x'+1, 2y'^2) = (a+1, 2v)$. If $b = 1$, the hypothesis at $y' < v$ gives $\mathrm{Pow}_2(x'+1, 2y')$, and extension with $b = 0$ gives $(2x'+2, 4y'^2) = (a+1, 2v)$.

(c) Induct on v in $\forall a \le v(\mathrm{Pow}_2(a+1,v) \to \exists p \le v(v = p+p \wedge \mathrm{Pow}_2(a,p)))$. At $v = 1$ the exponent would be 0. Otherwise decompose $a + 1 = 2x' + b$, $v = (1+b)y'^2$. If $b = 0$, then $x' = x_0 + 1$ and $a = 2x_0 + 1$; the hypothesis at y' gives $y' = 2p_0$ with $\mathrm{Pow}_2(x_0,p_0)$, and extension with $b = 1$ gives $\mathrm{Pow}_2(2x_0+1, 2p_0^2)$, where $v = 4p_0^2 = 2p_0^2 + 2p_0^2$. If $b = 1$, then $a = 2x'$; at $y' = 1$ take $p = 1$, and otherwise extension with $b = 0$ gives $\mathrm{Pow}_2(2x', y'^2)$, where $v = 2y'^2$.

(d) Induct on v in $\forall a \le v \forall c \le p'\,(\mathrm{Pow}_2(a,v) \to \mathrm{Pow}_2(a,c) \to v = c)$. At $v = 1$, $a = 0$ and $c = 1$ by (a). Otherwise $a \neq 0$, so $c \neq 1$; decompose both: $a = 2x_1 + b_1 = 2x_2 + b_2$ gives $x_1 = x_2$ and $b_1 = b_2$ by parity, the hypothesis at $y_1 < v$ gives $y_1 = y_2$, and $v = (1+b_1)y_1^2 = c$.

(f) Induct on v in $\forall a \le v(\mathrm{Pow}_2(a,v) \to \forall c \le a \exists p' \le v\ \mathrm{Pow}_2(c,p'))$. If $c = a$ take v. If $c < a$, then $a = a_0 + 1$, downward recursion gives $v = 2p_0$ with $\mathrm{Pow}_2(a_0,p_0)$, and the hypothesis at $p_0 < v$ applies to $c \le a_0$. □

Lemma L.35 (The graph in $\mathbb{N}$)**.** *In* $\mathbb{N}$*,* $\mathrm{Pow}_2(x,p)$ *holds exactly when* $p = 2^x$*.*

Proof. $\mathbb{N}$ is a model of $I\Delta_0$. $\mathrm{Pow}_2(n,2^n)$ follows from $\mathrm{Pow}_2(0,1)$ by upward recursion, and functionality gives the converse. □

L.12 Pair formation from exponentiation

Recall from Section 20.3.11 that $w\ \varepsilon\ z$ abbreviates $\exists p \le z \exists q \le z \exists r < p(\mathrm{Pow}_2(w,p) \wedge z = (2q+1)p + r)$: the digit of z at 2^w is one.

Lemma L.36 (Multiplicativity)**.** *In a model of $I\Delta_0$: if $\mathrm{Pow}_2(a,p)$ and $\mathrm{Pow}_2(a+k,q)$, then $q = p \cdot r$ for some r with $\mathrm{Pow}_2(k,r)$. Consequently, if $\mathrm{Pow}_2(a,p)$, $\mathrm{Pow}_2(b,q)$ and $a < b$, then $q = 2m \cdot p$ with $m \neq 0$, and $2p \leq q$.*

Proof. Induct on k in $\forall c \leq q\,(\mathrm{Pow}_2(a+k,c) \to \exists r \leq c\,(\mathrm{Pow}_2(k,r) \wedge c = p \cdot r))$. At $k = 0$, $c = p$ by functionality, and $r = 1$. At $k+1$, downward recursion writes $c = c_0 + c_0$ with $\mathrm{Pow}_2(a+k,c_0)$; the hypothesis gives $c_0 = p \cdot r_0$, and $r = 2r_0$ satisfies $\mathrm{Pow}_2(k+1,r)$ by upward recursion. For the consequence write $b = a + (k+1)$; then r is the value at a positive exponent, which downward recursion writes as $2m$ with $m \neq 0$ a value of the graph. □

Division by a nonzero p is unique, so $(2q+1)p + r = 2m \cdot p + s$ with $r, s < p$ is impossible: the quotients $2q+1$ and $2m$ would be equal.

Theorem L.37 (Exponentiation gives pair formation)**.** *Let M be a model of $I\Delta_0$ in which every x has a p with $\mathrm{Pow}_2(x,p)$. Then for all x and y some z satisfies $w \,\varepsilon\, z \leftrightarrow (w = x \vee w = y)$ for every w.*

Proof. Let $\mathrm{Pow}_2(x,X)$ and $\mathrm{Pow}_2(y,Y)$.

If $x = y$, take $z = X$. Then $x \,\varepsilon\, z$ with $q = r = 0$. If $w \,\varepsilon\, z$ through p, q, r, and $w < x$, then $X = 2m_1 \cdot p$ by Lemma L.36, and $(2q+1)p + r = 2m_1 p + 0$ is impossible; if $w > x$, then $2X \leq p \leq z = X$, which is impossible. So $w = x$.

If $x < y$, take $z = X + Y$, where $Y = 2m \cdot X$ with $m \neq 0$ and $2X \leq Y$. Then $x \,\varepsilon\, z$ because $z = (2m+1)X + 0$, and $y \,\varepsilon\, z$ because $z = 1 \cdot Y + X$ with $X < Y$. Conversely, let $w \,\varepsilon\, z$ through p, q, r. If $w < x$, then $X = 2m_1 p$ and $Y = 2m_2 p$, and $z = 2(m_1 + m_2)p + 0$ contradicts the odd quotient. If $x < w < y$, then $Y = 2m_2 p$ and $X < 2X \leq p$, so $z = 2m_2 p + X$ with $X < p$, again a contradiction. If $w > y$, then $X + Y < 2Y \leq p \leq z$, which is impossible. So $w = x$ or $w = y$. The case $y < x$ is symmetric. □

Proof of Theorem 20.46. Take Pow_2 of Definition L.30. It is bounded, its free variables are x and p, and it defines $p = 2^x$ in $\mathbb{N}$ by Lemma L.35. Its six laws hold in every model of $I\Delta_0$ (Theorem L.34), so $I\Delta_0$ derives them by Theorem L.27: it is a standard bounded graph. The translated pairing axiom holds in every model of $I\Delta_0$ that satisfies $\forall x \exists p\ \mathrm{Pow}_2(x,p)$ (Theorem L.37), so $I\Delta_0$ derives it from that sentence, again by Theorem L.27. With Proposition 20.45, each derives the other. □

Appendix M

Incompleteness and finite set theory: the proofs

This appendix gives the full arguments behind Theorems 20.42 and 20.47, whose proofs in Section 20.3.11 name the classical sources and state the steps. Both run through models: each statement that Peano arithmetic must derive is shown true in an arbitrary model of Peano arithmetic and then read back as a derivation.

M.1 Models of Peano arithmetic

A *model of Peano arithmetic* is a structure for 0, S, $+$ and $\cdot$ that satisfies the six rule checks and every induction instance, at every assignment. A derivation of Peano arithmetic is sound in every model, by induction on the derivation.

Theorem M.1 (Completeness of Peano arithmetic for its models)**.** *A formula true in every model of Peano arithmetic at every assignment has a derivation in Peano arithmetic.*

Proof. Its universal closure ψ is a sentence true in every model. Apply Theorem L.26 with T the universal closures of all induction instances: a structure satisfying T is a model, so finitely many closed induction axioms derive ψ by the classical rules, each axiom has a derivation with one induction node, and instantiation returns the formula. □

In a model of Peano arithmetic, induction holds for every property defined by a formula with parameters, so the arithmetic of Section L.10 holds there with induction on all formulas: the semiring laws, cancellation, the discrete linear order, the least number principle, division with remainder, parity, the pairing (10.3) and Gödel's β. The pairing is $\pi(a,b) = b \cdot b + a$ when $a < b$ and $a \cdot a + a + b$ otherwise; Peano arithmetic derives that its graph is total, functional, injective and onto (Lemma 20.2, proved in Appendix K.1), and that β is total and functional and can extend any sequence by one entry (Lemmas K.1–K.4 of Appendix K.2), so in every model these are functions with the same laws.

M.2 Codes and certificates

Codes. A node with tag j and children a, b, c (unused children 0) has code $\pi(\pi(a, \pi(b, c)), j) - 1$. The tags are those of Section L.1, except that a variable has its de Bruijn index as a raw number and an axiom node its schema number as a raw number: variable 2, zero 3, successor 4, sum 5, product 6, equation 7, falsehood 8, implication 9, universal quantifier 10, axiom $12(c, s, \ell)$, modus ponens $13(c, \text{minor}, \text{major})$, generalization $14(c, p)$, list cells 15 and $16(h, t)$. The zero term has code 8 and $\bot$ has code 63. Every child of a code is at most the code, and smaller when the code and its tag are both positive. The code 0 decodes as tag 1 with all children 0; where tables are built from this decoding (Proposition M.7) we put $\text{tag}(0) = 0$, as in Section L.1.

A formula with $\wedge$, $\vee$ or $\exists$ is coded through the classical definitions $A \wedge B = \neg(A \to \neg B)$, $A \vee B = \neg A \to B$ and $\exists A = \neg\forall\neg A$; write $\ulcorner\varphi\urcorner$ for the code of φ and $\underline{n}$ for the numeral $S^n(0)$.

Certificates. A *certificate* is a length n, a root p, and five sequences of length n, each coded by Gödel's β: a kind sequence and four field sequences. Its records are

- kind 0, a proof node q;
- kind 1, a substitution record (x, d, t, y): $y = \mathrm{sub}(x, d, t)$;
- kind 2, a lifting record (x, c, y): $y = \mathrm{lift}(x, c)$;
- kind 3, a term-shift record (t, d, y): $y = \mathrm{tsh}(t, d)$;
- kind 4, a successor-shift record (x, d, y): $y = \mathrm{plus}(x, d)$.

A record is *locally correct* when it agrees with the other records of the certificate as the clauses of Con(PA) say. A proof record has tag 12, 13 or 14, and an axiom node's schema number is at most 13. A proof node of tag 13 has its minor and major premises among the proof records and the major premise's conclusion is the implication from the minor's conclusion to the node's; a node of tag 14 has its premise among the proof records and concludes the universal quantification of its conclusion; a node of tag 12 has as many list cells as its schema has parameters and its conclusion is the schema's instance, read through decoding and through records of kinds 1–4 for the operations the schema uses, with the term parameters of schemas 3 (one) and 6 (two) not tagged as a formula, that is, with no tag among 7, 8, 9, 10; a record of kind 1–4 has the value that the recursion of its operation prescribes from records at the children. Every record relation occurs positively, so local correctness survives adding records.

$\mathrm{Prov}(x)$ says that some certificate has its root among the proof records, with conclusion x, and has every record locally correct. All quantifiers inside are bounded by the certificate's numbers, so Prov is twelve existential quantifiers over a bounded formula. Let $\mathrm{Pr}(\varphi) = \mathrm{Prov}(\underline{\ulcorner\varphi\urcorner})$.

Lemma M.2 (Certificates in a model)**.** *In every model M of Peano arithmetic:*

(a) two certificates concatenate into one, and a record that is correct against the extended store may be added; all records stay correct;

(b) for every x, d, t some certificate contains a substitution record for (x, d, t), and likewise for the other three operations;

(c) two correct certificates give the same value to records with the same arguments, so the four operations are functions of M with their recursion equations;

(d) at numerals of codes of actual terms and formulas, these functions give the numerals of the codes of the actual results.

Proof. (a) Concatenation and extension are β-comprehension: the sequence of the first n entries followed by the second m is defined by a formula, and induction on the length builds its code by extending one entry at a time. Correctness is monotone. (b) Course-of-values

induction on the code x in the formula "some correct certificate has a record for (x, d, t), for every d and t": the records for the children come first, by the induction hypothesis, the certificates are concatenated, and the record for x is added. (c) Merge the two certificates and compare the two values by course-of-values induction on the code, using the local correctness of each record. (d) Induction on the actual term or formula, reading each record at decoded numeral codes, where the decoding of M agrees with that of $\mathbb{N}$. □

M.3 The first, second and fourth conditions

The first condition rests on a syntactic step: a derivation by the rules of Peano arithmetic (Definition 8.10) becomes a Hilbert derivation of the checker's calculus. Write $\Gamma \vdash_H A$ when A has a Hilbert derivation from the hypotheses Γ: its leaves are members of Γ and instances of the fourteen schemas of Section L.1, and its steps are modus ponens and generalization, which passes from $\uparrow\Gamma \vdash_H A$ to $\Gamma \vdash_H \forall A$, where $\uparrow\Gamma$ lifts every hypothesis, as universal introduction does. With Γ empty these are the proofs the checker accepts. Let $A^\star$ replace $\wedge$, $\vee$ and $\exists$ by their classical definitions $\neg(A \to \neg B)$, $\neg A \to B$ and $\neg\forall\neg A$; codes are assigned through these definitions, so A and $A^\star$ have the same code.

Lemma M.3 (Natural deduction into the Hilbert calculus)**.**

(a) $(\cdot)^\star$ *commutes with lifting, with substitution and with the step* $A \mapsto A^+$ *of the induction rule, whose code is* $\mathrm{plus}(\ulcorner A \urcorner, 0)$.

(b) (Deduction theorem.) If $A, \Gamma \vdash_H B$, *then* $\Gamma \vdash_H A \to B$.

(c) $\Gamma \vdash_H A \to A$. *If* $\Gamma \vdash_H \bot$, *then* $\Gamma \vdash_H A$. *If* $\Gamma \vdash_H \neg\neg A$, *then* $\Gamma \vdash_H A$. *If* $\neg A, \Gamma \vdash_H \bot$, *then* $\Gamma \vdash_H A$.

(d) If $\Gamma \vdash_{\mathrm{PA}} \varphi$, *then* $\Gamma^\star \vdash_H \varphi^\star$. *In particular a theorem* φ *of Peano arithmetic has a closed Hilbert derivation whose conclusion has the code* $\ulcorner \varphi \urcorner$.

Proof. (a) Induction on the formula; the definitions are built from $\to$, $\bot$ and $\forall$, on which lifting and substitution act part by part. The step of the induction rule replaces the variable 0 by its successor and keeps the others, which is the successor shift at depth 0.

(c), first part. $A \to A$ comes from the instance $(A \to ((A \to A) \to A)) \to ((A \to (A \to A)) \to (A \to A))$ of schema 1 and two instances of schema 0 by modus ponens twice.

(b) Induction on the derivation of B. If B is A, use $A \to A$. If B is another hypothesis or an axiom, schema 0 gives $B \to (A \to B)$. A modus ponens step from $C \to B$ and C becomes, by the induction hypotheses $A \to (C \to B)$ and $A \to C$, two modus ponens steps with schema 1. A generalization step concludes $\forall C$ from $\uparrow A, \uparrow\Gamma \vdash_H C$; the induction hypothesis gives $\uparrow\Gamma \vdash_H \uparrow A \to C$, generalization gives $\Gamma \vdash_H \forall(\uparrow A \to C)$, and schema 4 with modus ponens gives $A \to \forall C$.

(c), the rest. From $\bot$: $\bot \to \bot$ is $\neg\bot$, schema 0 gives $\neg A \to \neg\bot$, and schema 2 gives $(\neg A \to \neg\bot) \to (\bot \to A)$. From $\neg\neg A$: schema 0 gives $\neg\neg\neg\neg A \to \neg\neg A$, schema 2 turns it into $\neg A \to \neg\neg\neg A$ and that into $\neg\neg A \to A$. The last statement is the deduction theorem followed by double-negation elimination.

(d) Induction on the derivation. Each rule of Definition 8.10 becomes Hilbert steps as listed below, where "by contradiction from $\neg C$" means: derive $\bot$ with the further hypothesis $\neg C$ and apply the last statement of (c). By (a) the translated premises and conclusions are the translations of the rule's formulas. □

Rule	Hilbert derivation of the translation
hypothesis	the hypothesis
arithmetic axiom	an instance of schemas 7–12
reflexivity $t = t$	schema 5
from $s = t$ and $A[s]$, $A[t]$	schema 6 and modus ponens twice
introduction of $\to$	the deduction theorem
elimination of $\to$	modus ponens
from A and B, $A \wedge B$	under the hypothesis $A \to \neg B$, modus ponens with A and then with B gives $\bot$; the deduction theorem gives $\neg(A \to \neg B)$
from $A \wedge B$, A	by contradiction from $\neg A$: $A \to \neg B$ by the deduction theorem, since A and $\neg A$ give $\bot$ and $\bot$ gives $\neg B$; modus ponens with $\neg(A \to \neg B)$
from $A \wedge B$, B	by contradiction from $\neg B$: schema 0 gives $A \to \neg B$; modus ponens with $\neg(A \to \neg B)$
from A, $A \vee B$	under the hypothesis $\neg A$, A gives $\bot$ and so B; the deduction theorem gives $\neg A \to B$
from B, $A \vee B$	schema 0 gives $\neg A \to B$
from $A \vee B$, $A \vdash C$ and $B \vdash C$, C	by contradiction from $\neg C$: the deduction theorem gives $A \to C$ and $B \to C$; with $\neg C$, A gives $\bot$, so $\neg A$; then B from $\neg A \to B$, C, and $\bot$
from $\bot$, any A	ex falso, (c)
introduction of $\forall$	generalization
from $\forall A$, $A[t]$	schema 3 and modus ponens
from $A[t]$, $\exists A$	under the hypothesis $\forall\neg A$, schema 3 gives $\neg A[t]$, and $A[t]$ gives $\bot$; the deduction theorem gives $\neg\forall\neg A$
from $\exists A$ and $A, \uparrow\Gamma \vdash \uparrow C$, C	by contradiction from $\neg C$: under the lifted hypotheses, which include $\uparrow\neg C$, the premise turns A into $\uparrow C$ and so into $\bot$; the deduction theorem gives $\neg A$, generalization gives $\forall\neg A$, and modus ponens with $\neg\forall\neg A$ gives $\bot$
induction, from $A[0]$ and $\forall(A \to A^{+})$, $\forall A$	schema 13 and modus ponens twice
excluded middle $A \vee \neg A$	$\neg A \to \neg A$, by (c)

Proposition M.4 (D1). *If Peano arithmetic derives φ, it derives* $\mathrm{Pr}(\varphi)$.

Proof. By Lemma M.3(d), a derivation of φ gives a closed Hilbert derivation of the checker's calculus whose conclusion has the code $\ulcorner\varphi\urcorner$. Fix a model M. Every axiom instance of the derivation has a certificate in M: one axiom node over records of kinds 1–4, which exist by Lemma M.2(b) and carry the standard values by (d). Modus ponens and generalization extend certificates by (a). So $\mathrm{Pr}(\varphi)$ holds in M; by Theorem M.1, Peano arithmetic derives it. □

Proposition M.5 (D2)**.** *Peano arithmetic derives* $\mathrm{Pr}(\varphi \to \psi) \to \mathrm{Pr}(\varphi) \to \mathrm{Pr}(\psi)$.

Proof. In a model, concatenate the two certificates and add the modus ponens node $13(\ulcorner\psi\urcorner, \text{root for } \varphi, \text{root for } \varphi \to \psi)$; it is correct against the concatenation. Theorem M.1 finishes. □

Proposition M.6 (D4)**.** *There is a sentence* G *such that Peano arithmetic derives* $G \to \neg\,\mathrm{Pr}(G)$ *and* $\neg\,\mathrm{Pr}(G) \to G$.

Proof. Let $\mathrm{Num}(x, t)$ say that some β-sequence starting at the code 8 of 0, each entry the successor node of the previous one, has t at position x; in a model, at a standard x it holds exactly for the code of $\underline{x}$. Let $\mathrm{Dg}(x, y)$ say that some correct certificate has a substitution record putting the numeral code of x for variable 0 in x, with value y. At the numeral of a formula code $\ulcorner A\urcorner$ it holds, in every model, exactly for $y = \ulcorner A[\underline{\ulcorner A\urcorner}/0]\urcorner$, by Lemma M.2(c) and (d). Put $\theta(x) = \forall y(\mathrm{Dg}(x, y) \to \neg\,\mathrm{Prov}(y))$ and $G = \theta(\underline{\ulcorner\theta\urcorner})$. In every model, G holds exactly when Prov fails at the one y with $\mathrm{Dg}(\ulcorner\theta\urcorner, y)$, which is $\ulcorner G\urcorner$; that is, exactly when $\mathrm{Pr}(G)$ fails. Theorem M.1 gives both implications. □

The proof uses nothing about $\neg\,\mathrm{Prov}$ beyond its being a formula in y; Theorem 20.48 states the general form.

M.4 The bridge to the consistency sentence

Proposition M.7 (D5)**.** *Peano arithmetic derives* $\mathrm{Con}(\mathrm{PA}) \to \neg\,\mathrm{Pr}(\bot)$.

Proof. Let M be a model with a certificate C for the code 63 of $\bot$, with root p. Put $B = p + 8$ and define tables on numbers at most B: the tag and children by decoding, with $\mathrm{tag}(0) = 0$; the slacks from the child sizes; nv as the identity, since natural data are raw numbers in this coding; $K(f, q) = 1$ exactly for the proof records $q < f$ of C; and the four operations by the functions of Lemma M.2(c), indices read in base $B + 1$ as in Section L.1. Each table is defined by a formula, so β-comprehension gives codes for it on any prefix. Now check the 247 clauses. The child clauses hold because children are smaller than nodes; the zero clauses because the zero term has code $8 \le B$ and tag 3; the fuel, modus ponens, generalization and axiom clauses because $K(f, q) = 1$ only at proof records below f, whose local correctness says what each clause asks; the operation clauses because the operations satisfy their recursion equations; and the root clauses because p is a proof record with conclusion 63, whose tag is 8. Under the guards of the operation clauses (Section L.1) every argument of a lookup is at most B, so each lookup reads the entry of the arguments it names. So these values satisfy every clause, and M refutes $\mathrm{Con}(\mathrm{PA})$. In every model, therefore, $\mathrm{Con}(\mathrm{PA})$ implies $\neg\,\mathrm{Pr}(\bot)$, and Theorem M.1 gives the derivation. □

The clauses constrain only the values of the tables, not the coding behind them, so a certificate in this coding gives tables that satisfy the clauses of Definition L.1; the coding was chosen so that the zero term has code 8 and the list cells and schema tags agree.

M.5 The third condition

Fix a model M. Its elements, standard or not, have numerals inside M: by induction in M, every m has a unique *numeral code* $\mathrm{nc}(m)$, with $\mathrm{nc}(0) = 8$ and $\mathrm{nc}(m+1)$ the successor node of $\mathrm{nc}(m)$; at standard m it is the code of $\underline{m}$. Numeral codes are closed: substitution and term shifts leave them unchanged. For a standard formula A and a list c of closed codes, let $\mathrm{cf}_k(c, A)$ be the code of A with the free variables from k on replaced by the codes in c.

Lemma M.8 (Tools for the third condition)**.** *In M:*

(a) Code substitution*: for closed t and closed codes c, substituting t for variable k in $\mathrm{cf}_{k+1}(c, A)$ gives $\mathrm{cf}_k(t :: c, A)$.*

(b) Internal validity*: if a standard formula F with free variables below n holds in every model of Peano arithmetic, then for any n closed term codes c the code $\mathrm{cf}_0(c, F)$ has a certificate.*

(c) Numeral arithmetic*: for all a, b in M the codes of $\underline{a}+\underline{b} = \underline{a+b}$ and $\underline{a}\cdot\underline{b} = \underline{ab}$ have certificates, and so does that of $\neg(\underline{a} = \underline{b})$ when $a \neq b$.*

(d) Bounded completeness*: for a bounded formula A, an assignment e and the numeral codes c of its values, if A holds at e then $\mathrm{cf}_0(c, A)$ has a certificate, and if it fails then the negation of that code has one.*

Proof. (a) Induction on the actual formula A, using the recursion equations of the substitution function and that the plugged codes are closed.

(b) The universal closure of F is a theorem of Peano arithmetic by Theorem M.1, so it has a certificate in M by the construction in the proof of Proposition M.4. Instantiate it n times with schema 3 and modus ponens; by (a) the instances have the codes $\mathrm{cf}_0(c, F)$. The plugged terms are closed and not formula codes, as schema 3 requires.

(c) Induction in M on b. At $b = 0$ the code is an instance of schema 9 or 11. From b to $b+1$, schemas 10 and 12, the induction hypothesis and equality substitution (schema 6) combine by modus ponens; each combination is an instance of (b). For $a \neq b$, induct on the smaller with schemas 7 and 8.

(d) Induction on the actual formula A. For terms, the code of $t[c] = \underline{t(e)}$ has a certificate, by induction on t with (c). An equation follows by symmetry and transitivity, and a false one by (c). The connectives use instances of propositional tautologies, from (b). A true bounded existential quantifier gives its witness's numeral and existential introduction; a false one is the bounded universal quantifier of the negation. A bounded universal quantifier $\forall y \le t\, A(y)$ is an induction in M on the value T of the bound: the induction predicate is "the code of $\forall y \le \underline{T}\, A(y)$ has a certificate", defined by a formula, and each step combines the induction hypothesis, the certificate for $A(\underline{T}+1)$ from the outer induction, and the theorem $\forall n\,\big((\forall y \le n\, A(y)) \to A(Sn) \to \forall y \le Sn\, A(y)\big)$ of Peano arithmetic, through (b); the base is the theorem $(\forall y \le 0\, A(y)) \leftrightarrow A(0)$. □

Proposition M.9 (D3)**.** *Peano arithmetic derives* $\mathrm{Pr}(\varphi) \to \mathrm{Pr}(\mathrm{Pr}(\varphi))$.

Proof. In a model M, let a certificate with twelve components show $\mathrm{Prov}(x)$. The bounded matrix of Prov holds at x and these components, so by Lemma M.8(d) its code with the twelve numeral codes plugged in has a certificate; twelve existential introductions give a certificate for the code of $\mathrm{Prov}(\underline{x})$. At $x = \ulcorner\varphi\urcorner$ this is the code of $\mathrm{Pr}(\varphi)$. So $\mathrm{Pr}(\varphi) \to \mathrm{Pr}(\mathrm{Pr}(\varphi))$ holds in every model, and Theorem M.1 gives the derivation. □

With (D1)–(D5), the propositional argument in the proof of Theorem 20.42 completes the proof of that theorem.

M.6 Finite set theory and binary membership

The calculus of $\mathrm{ZF}^*_{\mathrm{fin}}$ has the natural-deduction rules of the arithmetic calculus with excluded middle, over the language of $\in$ and $=$, with variables as the only terms. Its axioms are extensionality, the empty set, pairing, union, power set, separation and replacement (in the form: if every $u \in x$ has exactly one v with $\psi(u, v)$, some y contains such a v for every $u \in x$) for every formula, foundation (every inhabited set has an $\in$-minimal element), the negation of infinity (no set contains an empty set and, with each of its members y, the set $y \cup \{y\}$), and transitive containment. The completeness argument of Theorem L.26, with Henkin witnesses among the variables, proves: a sentence true in every structure satisfying the axioms has a derivation.

Lemma M.10 (Binary membership in a model of arithmetic)**.** *Let* Pow_2 *be as in Theorem 20.47 and M a model of Peano arithmetic. With $x \in y$ read as $x\ \varepsilon\ y$, M satisfies every axiom of* $\mathrm{ZF}^*_{\mathrm{fin}}$.

Proof. In M, Pow_2 is total, by induction from $\mathrm{Pow}_2(0, 1)$ and upward recursion, and functional, so $x\ \varepsilon\ y$ says that digit x of y is one. A number is determined by its digits, by induction on the larger of two numbers with the same digits: this is extensionality. Every number below 2^b has its digits below b, and for every property defined by a formula and every bound b some number has exactly the digits $x < b$ with the property, by induction on b, adding 2^b when b has it. This *bounded comprehension* gives the empty set (0), pairing ($2^x + 2^y$, or 2^x), union (digits of digits of x lie below x), power set (a number whose digits are digits of x is at most x, so comprehension below $x + 1$ collects them), and separation. Transitive containment: the number with every digit below x contains x, and a digit of one of its digits is again below x. Replacement: for each digit u of x the unique v exists, and by induction on the digits of x there is a bound for all of them; comprehension below that bound collects them. Foundation: the least digit y of an inhabited x has no digit in x, because the digits of y are smaller than y. Negation of infinity: a nonzero number has a largest digit y, and $y \cup \{y\}$, the number $y + 2^y$, is larger, so it is not a digit. □

Lemma M.11 (Arithmetic in a model of finite set theory)**.** *Let V be a nonempty model of $\mathrm{ZF}^*_{\mathrm{fin}}$. Its natural numbers ω^V (the von Neumann numerals), with addition and multiplication defined by recursion, form a model of Peano arithmetic, and the decoder $\mathrm{dec}(n) = \{\mathrm{dec}(k) : \text{digit } k \text{ of } n \text{ is one}\}$, defined by course-of-values recursion on ω^V, is a bijection from ω^V onto V with $\mathrm{dec}(k) \in \mathrm{dec}(n)$ exactly when $k \in n$.*

Proof. Transitive containment and foundation give $\in$-induction for every formula. Recursion on ω^V is justified by sets of finite approximations, which exist by separation and replacement and are unique by induction; it defines addition, multiplication and dec. Satisfaction of an arithmetic formula in ω^V is defined by a formula of set theory, so separation gives induction for every arithmetic formula: ω^V is a model of Peano arithmetic. The decoder is injective, by course-of-values induction, since two numbers with the same decoded digits have the same digits. It is onto, by $\in$-induction: the members of a set x are decoded numbers, by the induction hypothesis; by the negation of infinity a set of natural numbers is bounded, so the indices of those members are the digits of one number n, built by induction on a bound, and $\mathrm{dec}(n) = x$ by extensionality. □

Proof of Theorem 20.47. If $\mathrm{ZF}^*_{\mathrm{fin}}$ derives φ and M is a model of Peano arithmetic, then M read through ε is a model of $\mathrm{ZF}^*_{\mathrm{fin}}$ by Lemma M.10, so φ holds there, which says that φ^a holds in M; by Theorem M.1 Peano arithmetic derives φ^a. Conversely, if Peano arithmetic derives φ^a and V is a nonempty model of $\mathrm{ZF}^*_{\mathrm{fin}}$, then φ^a holds in ω^V, and the decoder of Lemma M.11 carries binary membership on ω^V to membership in V, so φ holds in V; by completeness of the set-theoretic calculus, $\mathrm{ZF}^*_{\mathrm{fin}}$ derives φ. The last assertion is Theorem 20.39. □

Appendix N

Notation and glossary

Symbols are local to the mathematical object under discussion. In particular, O denotes a proof ledger in Chapter 6 and an orbit type in Chapter 9; and $\approx$ is a congruence of signed records in Chapter 7 but the relation of naming the same point in Chapter 12; and $\beta_{\sim}$ is the collision search of a congruence, while $\beta\mathbb{N}$ is the space of ultrafilters on the natural numbers, both in Chapter 7. The accompanying type or definition decides which reading is intended. Two clashes are avoided by a change of type face: the italic I and U are an identity matrix and a unitary in Part V, while the sans-serif I is the iterator of Chapter 9 and U the universal program of Section 20.3. Chapters 18, 19 and 20 recall this at first use. This guide groups the notation by its use; Section N.6 collects the notation of Parts VI to VIII.

Dropped subscripts. The subscript δ marks a constructed object where it is introduced: the start 0_δ, the orbit $\mathbb{N}_\delta$, the signed records $\mathbb{Z}_\delta$, the ratios $\mathbb{Q}_\delta$ and the line $\mathbb{R}_\delta$. Once a construction has been read faithfully into familiar notation, the book drops the subscript on its elements and operations, with the following fixed meaning. From Chapter 3 on, 0 is the start 0_δ, and $1, 2, 3, \ldots$ are $S(0_\delta)$, $S(S(0_\delta))$, $S(S(S(0_\delta)))$, and so on; a numeral n names the orbit expression with n steps, which is legitimate because reading and expansion are inverse, Theorem 1.4. On $\mathbb{Z}_\delta$ and $\mathbb{Q}_\delta$, the symbols 0, 1, $+$, $\cdot$ and $<$ are the constructed operations on classes, and an orbit record n written there stands for its image $\iota(n) = [n, 0]$, or for the ratio $\langle \iota(n); 1 \rangle$; these inclusions preserve the operations, as the formulas (4.2) and (4.5) show, so no equation changes its meaning. In Part II, 0, S, $+$ and $\cdot$ are also symbols of the arithmetic language, and Definition 5.3 reads them as the constructed operations; where it matters, the text says whether a symbol or its value is meant. In Parts III and IV a ratio such as $1/(k+1)$ is a ratio record of $\mathbb{Q}_\delta$, and $\eta(q)$ is the point of $\mathbb{R}_\delta$ it names; where Chapters 13 and 14 use a ratio as a point, it stands for that point. In Part V, 0 and I are the zero and identity matrices of the stated size.

N.1 Arithmetic and maps

$\mathbb{N}_\delta, 0, S$ The carrier of finite orbit expressions, its start, and its step. An expression is 0 or $S(a)$ for an earlier expression a. Chapters 1–3.

r, e The reading of an orbit expression as a natural number and the expansion of a natural number as an orbit expression. Their two round trips are Theorem 1.4.

(M, b, s) A pointed step system: a carrier M, a start b and a map $s : M \to M$. The canonical reading into it is R_M. Chapter 2.

$f \circ g$ Composition, with g acting first. Thus $(f \circ g)(x) = f(g(x))$.

$f : A \to B$ A map with domain A and specified target B. Its range is the set of outputs it actually takes.

$[a, b]$ A class of signed pairs in Chapter 4, when used as the representative of a difference. For rational endpoints in Parts III and IV the same bracket shape denotes a closed interval; the chapter and the type distinguish the two uses.

$\sim$ An equivalence relation specified at its use. Balance of signed pairs, equality of ratios by cross-multiplication, and recognition congruences have different definitions.

$\sim_{i,p}$ The index-period congruence: points before i remain distinct, and the tail from i repeats with positive period p. Definition 7.2 gives the exact relation.

ι The inclusion $n \mapsto [n, 0]$ of the orbit into the constructed integers. Theorem 4.13.

$\langle u; d \rangle$ The class of a ratio record with signed numerator u and nonzero orbit denominator d. Definition 4.9.

N.2 Formulas, judgments and resources

φ, ψ, χ Formulas of the language under discussion. Letters t, u ordinarily denote terms.

$\bot$ The false formula. Negation $\neg\varphi$ abbreviates $\varphi \to \bot$.

$\wedge, \vee, \to$ Conjunction, disjunction and implication. Their introduction and elimination rules appear in Chapter 5 and Appendix D.

$\forall, \exists$ Universal and existential quantification. An existential witness supplies an object and its required property.

$\Gamma \vdash \varphi$ A derivability judgment: there is a finite argument for φ from context Γ using the specified rules. A context is a finite list of formulas.

$\rho \vDash \varphi$ Truth of a formula under an assignment ρ in the stated interpretation. Derivability and truth are different notions, connected in one direction by soundness.

$x_0, x_1, \ldots$ Variable positions. Immediately inside a quantifier, x_0 is its new bound variable. Earlier positions move up.

$\varphi\uparrow, \Gamma\uparrow$ Lift free variable positions past a new binder, in a formula or throughout a context.

$\varphi(t)$ Substitute t at the designated variable position, avoiding capture. It does not mean textual replacement without regard to binders.

$(e, \ell, m; i)$ Four Boolean ledger entries: EM, LPO, MP and induction on a quantified formula. Combining ledgers with $\sqcup$ takes inclusive-or in each coordinate.

EM Excluded middle for every proposition, or its explicitly specified object-language schema.

LPO The limited principle of omniscience. For a decidable predicate, it gives a witness or says that all instances fail. The arithmetic schema applies to every formula and, with induction on every formula, proves every instance of the EM schema. Chapter 6.

MP Markov's principle. For a decidable predicate, double-negated existence yields existence. The arithmetic schema in this book is restricted to quantifier-free formulas, and so restricted, with induction on every formula, it is derivable from the base rules. With induction on quantifier-free formulas only, some instance is not, Proposition 6.14. Chapter 6.

WLPO, LLPO Weak and lesser limited principles of omniscience. Their forms for binary sequences, and their different conclusions, are specified in Definition 7.8 and used again in Chapter 12. The abbreviations do not identify them with LPO or with each other: no two of MP, LLPO, WLPO, LPO and EM are equivalent over intuitionistic second-order arithmetic, Theorem 7.38.

$\mathrm{MP}_{\mathrm{bin}}$ Markov's principle for binary sequences, Definition 7.8. With supplied decisions it is the same principle as MP. Intuitionistic second-order arithmetic does not prove it, Theorem 7.38. In Chapter 12 it is the exact price of a sign for a nonzero quantity, Theorem 12.11.

AC_{ω} Countable choice, Definition 12.21. It selects a sequence of witnesses and decides nothing: intuitionistic arithmetic with variables for sequences and countable choice for numbers and for sequences proves none of LPO, WLPO, LLPO and $\mathrm{MP}_{\mathrm{bin}}$ and refutes none of them, Theorem 12.22. Without unique choice, excluded middle does not replace it in four completeness assertions, Theorem 12.63.

WEM Weak excluded middle: $\neg P \vee \neg\neg P$ for every proposition P. Chapter 12; as a sentence of intuitionistic second-order arithmetic, Definition 7.39.

WMP, $\mathrm{W}(X)$ Ishihara's weak Markov principle and its hypothesis, Definition 7.39; with $\mathrm{MP}^{\vee}$ it splits $\mathrm{MP}_{\mathrm{bin}}$, Remark 7.40.

LLPO_n**,** $\mathrm{C}_i^n(X)$ A decidable set with at most one member misses one of the n residue classes modulo n; $\mathrm{C}_i^n(X)$ says it misses the class of i, Definition 7.39.

FAN_{Δ}**,** FAN_c**,** FAN The decidable, c-set and full fan theorems, Definition 13.2. Heine–Borel for $[0,1]$ is equivalent to FAN_{Δ} with no choice principle, Theorem 13.6; without unique choice, excluded middle with the countable choice that a search can witness does not prove it, Theorem 13.11.

$\mathrm{WLPO}_{\mathrm{seq}}$**,** $\mathrm{LLPO}_{\mathrm{seq}}$ WLPO and LLPO for a whole double sequence of questions at once, Definition 13.8. Weak König's lemma is equivalent to $\mathrm{LLPO}_{\mathrm{seq}}$ with no choice principle, Proposition 13.9.

$\mathrm{MP}^{\vee}$**,** $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ LLPO for binary sequences known not to vanish, and its form for a whole sequence of them, Definition 14.18. Dominated division by nonzero points costs exactly $\mathrm{MP}^{\vee}$, with unique choice in the converse, Theorem 14.28. As a sentence of intuitionistic second-order arithmetic, Definition 7.39; its place among the prices, Theorem 7.45.

Unique choice for Π^0_1 **relations** Collecting, for each i, the one bit that passes a decidable sequence of tests, Definition 14.20; a case of the reading of maps.

$\mathrm{Dec}, \mathrm{Dec}_2, \mathrm{Z}, \mathrm{E}, \mathrm{A}, \mathrm{Eq}, \mathrm{Ev}, \mathrm{Od}, \mathrm{Cong}, \mathrm{Desc}, \mathrm{D}, \mathrm{Odd}, \mathrm{Even}$ The abbreviations of Definition 7.18 for sentences of intuitionistic second-order arithmetic about a set X or a relation R read through the pairing $J(x,y) = (x+y)\cdot(x+y)+x$. The six versions of the classification, $\mathsf{V}_{\mathrm{none}}, \dots, \mathsf{V}_{\mathrm{EM}}$, are (7.11); Theorem 7.33 derives each one's equivalence with its principle.

$[\![\varphi]\!]_{\rho,\sigma}$ The value of a formula in a system of truth values under an assignment, Definition 7.19.

$\Gamma \vdash_{\mathrm{PA}} \varphi$ Derivability in first-order Peano arithmetic, presented as a separate classical calculus in Chapter 8. Theorem 8.11 identifies it with the derivations whose ledger records neither LPO nor MP.

$S^n(0)$ The numeral with n successors. Its value in a structure M is $\nu(n) = S_M^n(0_M)$. Chapter 8.

φ^N The negative translation of φ, built from $=$, $\bot$, $\to$ and $\forall$. Section 20.3.

φ^G Friedman's translation of φ by a formula G: each equation $t = u$ becomes $(t = u) \vee G$, and $\bot$ becomes G. Definition 20.61.

N.3 Typed constructions and finite descriptions

$\Pi x : A\, B(x)$ The type of functions taking $x : A$ to an element of $B(x)$. If the output type does not depend on x, write $A \to B$.

$\mathrm{Id}_A(t,u)$ The identity type between two terms of type A in the language of Chapter 9.

$\mathbf{0}$ Its empty type. This is a type, not the orbit's starting element 0.

$\lambda x.t$ A function whose output at x is t. Application is written $f(x)$ or by juxtaposition.

$\mathsf{I}, \mathsf{R}, \mathsf{E}$ Iterator, indexed recursor and dependent eliminator, in sans-serif type. Their successive inputs are compared in Chapter 9; Appendix D repeats their computation equations.

$D(x)$ The optional type family distinguishing the start from a successor in the specified typed language.

$\mathscr{I}^-, \mathscr{R}^-, \mathscr{E}^-, \mathscr{E}^+$ The iterator, indexed and dependent fragments without the start discriminator, and the dependent fragment with it. Theorem 9.10.

U The universal program of Section 20.3. On input $\langle e, n\rangle$ it searches for a computation record of program e on input n accepted by a fixed record test, and returns the output written there. It is distinct from the unitary U of Part V.

$\mathrm{App}(e,n,y)$, $G(a,b,z)$ The arithmetic formulas saying that program e returns y on input n, and that $z = \langle a, b\rangle$. Section 20.3.3.

$e \simeq_A f$, $H_A(e)$ Hereditary extensional equality of program codes at a finite type, and admissibility $e \simeq_A e$. Section 20.3.5.

$C_\Theta(R)$, $R \approx_\Theta R'$ The condition that the environment R holds admissible values for the type context Θ, and the relation of two environments that hold related values. Section 20.3.5.

$c(t)$, $E_t(R, y)$, $\mathrm{Rd}_R(P)$ The program compiled from a term, the formula saying that it returns y at the environment R, and the reading of a formula as an arithmetic formula about programs. Sections 20.3.5 and 20.3.6.

$\mathrm{Ty}(A)$, $|t|$, Adm_T The type of $\mathscr{E}^+$ expressing a negative formula, the erasure of a typed term, and the reading of a type of $\mathscr{E}^+$ in the language of functions. Section 20.3.10.

$\mathrm{Con}(\mathrm{PA})$, $\mathscr{F}_U$ The consistency sentence of a fixed proof checker for Peano arithmetic, and $\mathscr{E}^+$ extended by one universe. Section 20.3.11.

$I\Delta_0$, $\mathrm{ZF}^*_{\mathrm{fin}}$, φ^a Bounded arithmetic, finite set theory with infinity negated and transitive containment, and the Ackermann reading of a set-theoretic sentence by binary membership. Section 20.3.11.

$\mathrm{Prov}(x)$, $\mathrm{Pr}(\ulcorner\varphi\urcorner)$, $\ulcorner\varphi\urcorner$ The provability formula of Appendix M.2; the sentence $\mathrm{Prov}(\ulcorner\varphi\urcorner)$, which says that Peano arithmetic derives φ and which Appendix M writes $\mathrm{Pr}(\varphi)$; and the code of φ, inside a formula its numeral. Section 20.3.12.

$\Gamma \Rightarrow \varphi$, Pr_Γ, $\mathrm{Con}(\mathrm{PA} + \Gamma)$ For a finite list $\Gamma = (\gamma_1, \dots, \gamma_n)$ of sentences: the implication $\gamma_1 \to (\gamma_2 \to \cdots \to (\gamma_n \to \varphi))$, the sentence $\mathrm{Pr}_\Gamma(\ulcorner\varphi\urcorner) = \mathrm{Pr}(\ulcorner\Gamma \Rightarrow \varphi\urcorner)$, and $\neg\,\mathrm{Pr}_\Gamma(\ulcorner\bot\urcorner)$. For Γ empty this is $\neg\,\mathrm{Pr}(\ulcorner\bot\urcorner)$, which is not the checker sentence $\mathrm{Con}(\mathrm{PA})$ above; (D5) derives it from $\mathrm{Con}(\mathrm{PA})$. Section 20.3.12.

T_k, A_k The tower $T_0 = \mathrm{PA}$, $T_{k+1} = T_k + \mathrm{Con}(T_k)$, and the sentence $A_k = \mathrm{Pr}_{T_k}(\ulcorner\bot\urcorner)$, so that $\mathrm{Con}(T_k)$ is $\neg A_k$. Theorems 20.53 and 20.54.

Σ_1 **sentence** A sentence $\exists x_1 \cdots \exists x_k\, B$ whose matrix B is bounded: built from equations and $\bot$ by $\wedge$, $\vee$, $\to$ and the bounded quantifiers $\forall y \le t$ and $\exists y \le t$. Section 20.3.12.

$\equiv$ Conversion under that language's computation rules. It is not an unstated equality-reflection principle. In Chapter 16, $A \equiv B$ means instead that the calculus derives both $A \vdash B$ and $B \vdash A$.

F_n The finite stage of tallies within the budget n. Definition 10.1 fixes its boundary convention.

$[\ell_n, u_n]$ The rational enclosure requested at precision index n. Its width is at most $1/(n+1)$, and later intervals lie inside earlier ones.

$v(p)$ The displayed classical real value of interval protocol p. Chapter 11 specifies the classical comparison.

$p \simeq q$ Observational equality: the two protocols' intervals overlap at every matching precision. It is weaker than equality of their finite descriptions.

$\mathbb{R}_\delta$ The line built from ratios: regular sequences of ratio records taken up to $\approx$. Definition 12.14.

$s \approx t$ The regular sequences s and t name the same point: for every k they eventually agree within $1/(k+1)$.

$\eta(q)$ The point of $\mathbb{R}_\delta$ named by the constant sequence of a ratio record q.

within $1/(k+1)$ For points x, y of $\mathbb{R}_\delta$: $x \le y + \eta(1/(k+1))$ and $y \le x + \eta(1/(k+1))$. Section 12.9.

certified real A sequence (a_n) of integers with $|a_m/2^m - a_n/2^n| \le 2^{-m} + 2^{-n}$ for all m, n; the integer a_n is its answer at precision n, Definition 15.1.

$g \Vdash_0 A$, $g \Vdash_1 A$ The object g of the context belongs to the reading of the formula A of the calculus of δ's objects at the earlier or at the later stage, Table 16.1; the computable calculus and its model of named objects are Definitions 16.18 and 16.20.

N.4 Physical records and finite matrices

$r : W \to \mathscr{R}$ In Chapter 17, the record produced from a world or preparation. A reader d recovers a target f exactly when $d \circ r = f$.

B^m The capacity of m answer positions from an alphabet of size B. Adaptive stopping is represented by padded transcripts whose padding is ignored by the reader.

ε An error tolerance. Each theorem specifies the distance or observable to which it applies.

$A^*, A^\dagger$ The adjoint, or conjugate transpose, of a complex matrix. The two notations denote the same operation.

I The identity matrix on the space indicated by the equation. In a block display, its dimension is fixed by that block.

$\mathrm{tr}, \|\cdot\|, \|\cdot\|_1$ Trace, operator norm and trace norm. The state-error convention uses the full trace norm, without a factor of one half.

ρ A state matrix: positive with trace one. A pure state can be specified by a unit column vector.

$[A, B]$ The commutator $AB - BA$.

$\otimes$ Tensor product. In Part V it retains the coordinates of both named systems; it does not authorize discarding a factor.

$C_y, K_{y,z}$ A Choi block for instrument outcome y and its Kraus matrices. The additional index z labels the finite factorization. Chapter 18.

V, U An isometry and a unitary, respectively, in the dilation construction. Their input and output spaces are stated there.

H_0, N, P_c Native core energy, occupation and the projection whose collective count is derived from the two conserved charges. Chapter 19 and Appendix B.

r, e **in the recorder theorem** The number of native recorder cells and the uniform event-probability error. Here r is an integer resource count, not the reading map used in earlier chapters.

N.5 Glossary

Additive congruence. An equivalence relation preserved by the addition operation. A recognition rule with this property remains consistent when records are extended. Chapter 7.

Ambient mathematics. The surrounding language and reasoning in which a construction, formal calculus or interpretation is studied. It must be distinguished from the object language defined inside it. Chapters 1, 5 and 9.

Bishop compact. A set of points with internal nets, for every k a finite list of members within $1/(k+1)$ of every member, that is complete: every Cauchy sequence of members converges to a member. The nets exist without being handed over. Every presented set is Bishop compact exactly when every Cauchy sequence of points converges, with no choice either way, Theorem 13.37. Definition 13.36.

Cancellative congruence. An additive congruence from which a common summand can be removed: $x + z \sim y + z$ implies $x \sim y$. Exactly these extend to the signed records. Theorem 7.12.

Canonical reading. The map determined by the start and step equations into a specified target system. Its uniqueness does not imply that it is injective or onto that target. Chapter 2.

Categoricity. Uniqueness of a structure up to the specified isomorphism among structures satisfying the stated conditions. The faithfulness and generation conditions are part of the result here. Chapter 2.

Certified real. A sequence of integers a_n, one for each precision n, with $|a_m/2^m - a_n/2^n| \leq 2^{-m} + 2^{-n}$ for all m and n; Chapter 15.

Characteristic zero. A property of a field: no sum $1 + \cdots + 1$ of one or more terms equals 0. Theorem 4.15.

Commutative semiring. A set with two associative and commutative operations, a neutral element for each, multiplication by the additive neutral element giving that element, and distributivity. The orbit maps uniquely into every one. Proposition 3.11.

Complete positivity. Positivity of an operation after adjoining any finite retained reference. It preserves positivity for correlated inputs as well as for the source considered alone. Chapter 18.

Computable calculus. The calculus of δ's objects with its carriers presented and its terms and maps computed by programs; it keeps every rule. Definition 16.18.

Computation record. A finite table of evaluation facts for a program on an input, each justified through an evaluation clause by facts later in the table, with a second table of the values each justification uses. Acceptance and Record soundness, the two facts about the fixed test for such records, are stated in Section 20.3 and derived in Theorem 20.5.

Conservativity. A richer system is conservative over a smaller one, for a class of statements, when every statement of that class it proves the smaller one also proves. Theorem 20.32 proves the form used here, for negative sentences.

Cotransitivity. The property of the strict order of the line built from ratios: if $x < y$, then $x < z$ or $z < y$ for every z. Proposition 12.19.

Countable choice. The principle that a relation with a witness for every natural number has a sequence of witnesses. It selects witnesses and decides nothing, Theorem 12.22. Definition 12.21.

Countermodel. An interpretation that satisfies the retained rules and fails a proposed conclusion. It separates resources only after both rule soundness and expressibility have been checked. Chapter 9.

Cover-compact. A set K of points such that every ratio cover of K has an N with every member of K in one of the first N intervals. That $[0, 1]$, or every presented set, is cover-compact is equivalent to FAN_Δ with no choice principle, Theorems 13.6 and 13.25. Definition 13.5.

Decidable predicate. A predicate for which each supplied input admits a positive or negative answer. Decidability of individual inputs does not supply an unbounded search result. Chapter 6.

Dependent elimination. A construction whose output type may vary with the orbit point. Its step transports data from the preceding type to the next one. Chapter 9.

Effective name. One finite program that returns a valid enclosure at every requested precision under the specified protocol. The existence of separate programs for finite prefixes is weaker. Chapter 10.

Equational over constant carriers. A formula of the calculus of δ's objects with no atom, whose equations compare terms of carriers that are the same at both stages and whose quantifiers range over such carriers. The full calculus derives every true one; Proposition 16.15.

Eventually constant. A formula in one searched variable whose truth value is the same for all values beyond some bound. Every quantifier-free arithmetic formula has this property, for each assignment of its other variables, with a bound computed from that assignment. Lemma 6.8.

Exact price. A logical resource sufficient for a stated result and recoverable from that result over the stated base. The reverse direction establishes necessity within that comparison. Two exact prices differ when a model satisfies one and refutes the other; the six prices of the recognition classification all differ, Theorem 7.38. Chapters 7 and 12.

Faithful generated system. A pointed step system whose step is injective, whose start is not a step, and whose entire carrier is generated from that start in the stated inductive sense. Definition 2.4. These are exactly the systems whose canonical reading is bijective, Proposition 2.8.

Fan theorem. The statement that every bar of the binary tree is uniform: one depth works for every path. Its decidable form is equivalent, with no choice principle, to Heine–Borel for $[0, 1]$; Chapter 13.

Finite instrument. A finite family of quantum operations with a retained outcome mark, whose sum preserves trace. It describes the returned state as well as the probabilities of marks. Chapter 18.

Full calculus. The calculus of δ's objects in which a term may carry any map of carriers. It has no recursive model (Theorem 16.14) and derives every true sentence that is equational over constant carriers (Proposition 16.15); its computable form is the computable calculus.

Generation. The condition that the carrier has no extra elements outside the structure reached from the designated start. It is equivalent to surjectivity of the canonical reading, Proposition 2.8. Restricted to properties expressible by first-order formulas, it no longer excludes extra elements. Chapter 2.

Group completion. The extension of an additive structure to a commutative group through which every additive map into a group factors uniquely. Theorem 4.13.

Hereditary extensional equality. Equality of program codes at a finite type: at O, equality of numbers; at $A \to B$, the two programs return outputs on every pair of hereditarily equal inputs, and the outputs are hereditarily equal. Section 20.3.5.

Heyting arithmetic. The intuitionistic first-order theory of 0, S, $+$ and $\cdot$, with induction for every formula. The base rules with induction for every formula present it: from every finite context, the posit-free derivations prove exactly its theorems, and they are the translations of its derivations, up to the way the injectivity of the successor is written. Corollary 8.12 and Remark 6.12.

Initiality. Existence of exactly one structure-preserving map from the designated object into every target of the specified kind. The target need not itself be faithful. Theorem 2.3.

Interpretation. An assignment of objects and operations to the language that respects its formation and inference rules. A formula's truth is evaluated within an interpretation. Chapters 5, 8 and 9.

Intuitionistic second-order arithmetic. Heyting arithmetic with a second kind of variable for sets of numbers, comprehension for every formula and induction stated for every set; usually called HAS. The base over which the six principles of Chapter 7 are separated. Definition 7.14.

Last exit. For a decidable set D of strings and a binary sequence p, a number n with $D(\bar{p}(m)) = 1$ for every $m \geq n$ such that for every $j < n$ some $m \geq j$ has $D(\bar{p}(m)) = 0$. When the c-set of D is a bar it exists and is unique, Lemma 13.17, so unique choice selects it, and FAN_c is then the uniform continuity of pointwise continuous maps from binary sequences to numbers, Corollary 13.19. Definition 13.16.

Ledger. The recorded additional resources used by a particular derivation. It describes that proof, not automatically the least resources needed for its conclusion. Chapter 6.

Located cut. A pair of inhabited, disjoint sets of ratios, the lower one closed downward without a greatest element and the upper one closed upward without a least element, such that whenever $q < r$, either q lies in the lower set or r in the upper set. Section 12.9.

Model of truth values. An interpretation in which a formula's value lies in a complete system of truth values with an implication, such as the open sets of a space or a chain. Every sentence derivable in intuitionistic second-order arithmetic has the top value, so a sentence with another value is not derivable there. Definition 7.19 and Lemma 7.21.

Negative formula. A formula built from equations, $\bot$, $\to$ and $\forall$. Section 20.3.

Numeral. The term $S^n(0)$, or its value $S_M^n(0_M)$ in a structure M. Chapter 8.

Observational equality. The equivalence of interval protocols defined by overlap at every precision. In the stated classical comparison, it agrees with equality of displayed values. Chapter 11.

Peano arithmetic. First-order classical arithmetic: six axioms for the start, the step, addition and multiplication, induction for every formula, and excluded middle. The rules of Chapter 5 with the EM rule, and without LPO and MP, prove exactly its theorems. Theorem 8.11.

Posit-free. Using none of EM, LPO or MP as added arithmetic rules in the ledger of this book. It does not exclude the base rules or induction on quantified formulas. The posit-free derivations prove exactly the theorems of Heyting arithmetic, and they are the translations of its derivations, up to the way the injectivity of the successor is written, Corollary 8.12. Chapter 6.

Positive cone. A subset of a ring closed under addition and multiplication that contains exactly one of x and $-x$ for each nonzero x, and does not contain 0. Proposition 4.16.

Presented compact set. A set of points given with its nets: for each n, a finite list of named members such that every member is within 2^{-n} of one of them, and completeness for named Cauchy families; Definition 13.22.

Presented uniform map. A function on the points of a set K given with a transformer, which sends each name of a member of K to a name of its value, and a modulus μ, such that members within $1/(\mu(k)+1)$ have values within $1/(k+1)$. Positive lower bounds and infima of such maps on presented sets are equivalent to FAN_Δ with no choice principle, Theorem 13.28 and Corollary 13.30. Definition 13.27.

Quantifier prefix. A string of quantifiers placed before a quantifier-free formula. The prefixes that go up into every structure obeying the six arithmetic rules are exactly the purely existential ones, and those that come down from every such structure are exactly the purely universal ones. Proposition 8.9.

Quotient. A carrier formed by treating equivalent representatives as one element. Operations descend to the quotient only when they respect the equivalence relation. Chapter 4.

Recorder. A rule that reads the finite record of the first examinations and either announces an answer or reads on; Definition 17.12.

Reduces to. A rule reduces to a set of rules when each instance of its schema has a derivation, from no hypotheses, whose additional rules all belong to that set; the base rules are always available. The rule can then be removed from any derivation. Section 6.7.

Region certificate. A finite check on one computed trajectory of a differential equation: at level n_0 every vector of the trajectory lies in the ball of radius $M - 2^{-(n_0+2)}$. It is a hypothesis about computed ratios, not about an exact solution. Definition 15.18.

Regular sequence. A sequence (s_n) of ratio records with $|s_m - s_n| \le 1/(m+1) + 1/(n+1)$ for all m, n. Its points form the line $\mathbb{R}_\delta$, Definition 12.14.

Retained reference. A system kept in the error comparison while the operation acts on another system. It exposes errors in correlations that a comparison of isolated source states can miss. Chapter 18.

Sharp event. An event represented by an orthogonal projection. The projection identities are used in the recorder noise bound. Chapter 19.

Soundness. Every derivable judgment is valid in the specified interpretation under the stated hypotheses. It does not say that every true statement is derivable. Chapter 5.

Two-stage carrier. An earlier set, a later set and a passage map from the first to the second: the objects of the calculus of Chapter 16, Definition 16.1.

Uniformity. One supplied object or procedure works for every input in the theorem's stated range. In the recorder theorem, one fixed program state and apparatus must satisfy the error bound for every source input. Chapters 10 and 19.

Well-definedness. Independence from choices that a construction has discarded, such as the representative of a quotient class. Chapter 4.

N.6 The cost, physics and learning (Parts VI to VIII)

Parts VI to VIII keep the notation above and add the entries below. Some letters change meaning there, and each chapter fixes its reading at first use. For example, φ is the golden ratio, but it also names the partial computable functions φ_e of Chapter 29, the sign laws of Section 26.8, formulas in Chapters 28 and 29, and the area factors φ_h of hinge costs in Chapter 27; J is the cost, not the pairing (7.10); G is the cost's logarithmic form, a token's generator G_t or Newton's constant; D is a debit count, a relative entropy, a drift or the dimension of the sphere S^D; T is Gödel's system, a period, a temperature or a statement, as in "T costs exactly P", never a theory of the tower T_k; and κ is the apparatus constant of Definition 22.10, the Einstein coupling of Proposition 22.40, the frequency of the Binet clock in Chapter 25, or the background of the lattice in Chapter 27. In Section 24.11, Q is an additive charge and R a writing rule, not the form and the matrix of Definition 24.13; φ is the function of a read; and T is the tick of a tick register or a number of ticks.

Maps, data forms and prices

map In Parts VI to VIII, a rule given as data: a proof that every input has exactly one output does not by itself give a map. Section 21.1. Under the default reading of Appendix C, which has unique choice, each data form below is its principle, Remark 21.2.

the LPO, WLPO and WEM functions One map d from binary sequences to $\{0,1\}$ with $d(\alpha) = 1$ exactly when α has an entry 1, respectively exactly when α is all 0; and one map d from propositions to $\{0,1\}$ with $d(P) = 1$ exactly when $\neg\neg P$. Definition 21.1.

zero test, apartness test Maps $z, a : \mathbb{R}_\delta \to \{0,1\}$ with $z(x) = 1$ exactly when $x = 0$, and $a(x) = 1$ exactly when $x \# 0$, that is, when $x < 0$ or $x > 0$. Definition 21.1. They exist exactly when the WLPO function, respectively the LPO function, does, Theorem 28.3.

free, costs exactly, classical Free: proved with no omniscience principle and no choice principle. "T costs exactly P": each implies the other with no choice principle. A result proved in classical mathematics is a theorem about the classical real numbers, or about complex matrices or measures, with excluded middle and the axiom of choice available, and carries no price on $\mathbb{R}_\delta$. Section 21.1.

exact price, between, below, open For a step S of the forcing chain: $S \Leftrightarrow P$ with no choice principle; $S \Rightarrow P$ and $Q \Rightarrow S$; only $Q \Rightarrow S$, the last two being *bracketed*; or none of these proved. Definition 22.1.

FZ, PZ, SZ, AGG Zero tests for functions, for functions of two points and for sequences of points, and aggregation of a sequence into one point. Definition 21.28.

$\mathrm{SE}_{\mathbb{Z}}$, DEP Strong extensionality for integer-valued maps: $P(x) \neq P(y)$ gives $x \# y$ for every map $P : \mathbb{R}_\delta \to \mathbb{Z}$; and the deformation-erasure principle, an integer pairing observable that tells the pair from the split. Definition 22.22; MP implies $\mathrm{SE}_{\mathbb{Z}}$, Theorem 22.24.

The cost and the ledger

$J(x)$, $G(t)$ The cost $J(x) = \frac{1}{2}(x + x^{-1}) - 1 = (x-1)^2/(2x)$ of a comparison at the ratio $x > 0$, and its logarithmic form $G(t) = \cosh t - 1$, with $J(e^t) = G(t)$. Definition 21.11; for a matrix ratio, Definition 27.1.

composition law, calibration A comparison cost C obeys the law when $C(s+t) + C(s-t) = 2C(s)C(t) + 2C(s) + 2C(t)$, and is calibrated when $C(0) = 0$ and for every k some j has $|C(t) - t^2/2| \leq t^2/(k+1)$ whenever $|t| \leq 1/(j+1)$. Definition 21.10; the two leave exactly $C = G$, Theorem 21.14.

log, cos, sin, π_δ, q The logarithm, the circle functions and π_δ on the line, Definitions 21.4 and 21.6. A quarter period is a point q with $0 \leq q \leq 2$ and $\cos q = 0$; there is exactly one, $q = \pi_\delta/2$, Theorem 21.8, and Chapter 25 writes periods as $4q$.

φ, $\mathbb{Z}[\varphi]$ The golden ratio $\varphi = (1 + \sqrt{5})/2$, with $\varphi^{-3} = \sqrt{5} - 2$, and the ring of formal sums $a + b\varphi$ with integers a, b and $\varphi^2 = \varphi + 1$, ordered by an integer sign test at no price. Definition 21.18, Lemma 21.19.

census, d, c, W_k, b_N A census counts a recognition's independent steps and closures; the join $(i, c) \mapsto (i + c, i)$ multiplies its golden value $c + i\varphi$ by φ. The words W_k in the debit and credit letters d, c define the recognition history, and b_N is the balance, debits minus credits, of its first N letters. Definition 21.24; Theorems 21.25 and 21.26.

rate A word w read over $t + 1$ ticks has the rate $\langle \mathrm{bal}(w); t + 1 \rangle$: balances are the integers $\mathbb{Z}_\delta$ and rates the ratios $\mathbb{Q}_\delta$, Theorem 21.27.

The forcing chain and three dimensions

T1–T4, C−2 to C41, P8, R2-01 to R2-35, S1–S15 The rows of Table 22.1: the first rung, the numbered sequence (Cn is the step Recognition Science numbers Tn there), its period bound, the second rung and the constants. Section 22.10.

S^D, native chain, $H_k(S^D \setminus K) = 0$ The native D-sphere of $\mathbb{R}_\delta^{D+1}$ with the sup distance; chains are finite lists of integer multiples of uniformly continuous simplices, equivalent under explicit moves; $H_k(S^D \setminus K) = 0$ says that every k-cycle bounds there, and no quotient group is formed. Definitions 23.1–23.4.

linked, data form An embedded circle is linked when some loop of its complement bounds no 2-chain there, Definition 23.5; this happens only in dimension 3, Theorem 23.8. The data form of $H_1(S^1) = \mathbb{Z}$ is an integer winding label on all 1-chains, Definition 23.39.

Ledgers, the light cone and the register

(D, C), t, x, τ The debit and credit counts of an account's record; its cone coordinates $t = D + C$ and $x = D - C$; and the proper time $\tau(w) = 2\sqrt{D(w)C(w)}$ of a posting word, a name for a function of two counts. Reading t, x and τ as time, position and clock time are identifications. Definition 24.1.

$R(t, x)$, Q, K, Λ_A The Hermitian matrix $R(t, x) = \frac{1}{2}(t\,1 + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3)$ with coordinates (t, x); the Minkowski form $Q = t^2 - |x|^2$; the future cone K of the points with $t \geq 0$ and $Q \geq 0$ ($\mathscr{K}$ in Chapter 27); and the matrix, on coordinates, of the coherent change $\rho \mapsto A\rho A^\dagger$. A coherent record is a Hermitian matrix ρ with $v^\dagger \rho v \geq 0$ for every v, and $R(t, x)$ is one exactly when $(t, x) \in K$. Definition 24.13; the record cone and the Lorentz group, Theorems 24.14 and 24.15.

step cost, $c_W(w)$ The charge of G on each change of count, and the comparison cost of a leg w against the whole word W. Definitions 24.18 and 24.23.

channel, triad, B In Chapter 21 a channel is binary. In Chapters 24 and 30 a channel has eight states; Chapter 30 labels them $0, \ldots, 7$, read as three bits, and the low-bit partner of a label flips its last bit. There the triad $(i, j, 7)$ carries the generator B, and $e^{\theta B}$ rotates the triad by the angle θ about $e_i + e_j + e_7$. Definitions 21.35, 24.32 and 30.19.

token, word, drive, S_k, F_l A token carries a skew-Hermitian generator G_t, a word is a list of tokens, and its drive is $\sum_t \|G_t\|$ in units of $\hbar$ (Chapter 21 calls this sum the generator cost and $\hbar$ times it the drive action). On eight-state channels S_k signs a configuration

by the state of channel k, F_l flips the low bit of channel l, and the entangling token has generator $\frac{\pi}{4}\mathrm{i}S_kF_l$; on the binary channels of Chapter 21 it has generator $\frac{\pi}{4}\mathrm{i}X$, with X an involution that flips channel l up to a sign. Definitions 21.35, 24.32 and 30.19.

holding, paid word, drift, certain change A self-adjoint holding H acts while each token runs for a time ε; the paid word W is the product of the paid tokens, its drift is $|w|\varepsilon\|H\|$ (divided by $\hbar$ in Chapter 30), and it makes a certain change when $|\langle e_v, We_u\rangle| = 1$ for some $v \neq u$. Definitions 21.35 and 30.19.

parity pattern, item time The parity pattern $p(L) \in \{0,1\}^d$ of a ledger of d accounts, $p_k(L) = (D_k - C_k) \bmod 2$, of which a post flips one bit, Proposition 24.44; a process has a uniform item time when, from every tick, the shortest pass that contains every pattern has one and the same length, Definition 24.46.

tick register, S, $\mathscr{C}$ A real vector space with a linear map T, the tick. A recognizer whose state is a function on its eight patterns has $\mathbb{R}^8$, the functions on $\mathbb{Z}/8$, with the shift $(Sf)(t) = f(t+1)$, and its core $\mathscr{C}$ holds the states with $f(t+4) = -f(t)$. Section 24.11, before Theorem 24.49.

(R1)–(R5) The read premises of Definition 24.51: a read $\mu(f, A)$ of a state f on a set A of positions is one function φ of the amplitude at each position (R1), adds over positions (R2), vanishes at amplitude 0 (R3), grows with the magnitude (R4) and does not see the phase made of ticks (R5). Theorem 24.53.

additive charge, writing rule A map $Q : \{0,1\}^3 \to G$ into a commutative monoid with $Q(000) = 0$ and $Q(x \vee y) = Q(x) + Q(y)$ when x and y have no common 1, Proposition 24.45; and a rule that gives the increment $R(s,t,n) \in \mathbb{Z}^d$ of the balance vector when account s is compared with t n times, defined before Theorem 24.56.

Orbits

polynomial field, motion A field with polynomial components and ratio coefficients, with the formal Lie derivative $\mathscr{L}_pF$; a motion is a map with $p(x(s))$ as a uniform derivative on every bounded interval. Definitions 25.1 and 25.2.

(x, v, u, m), E, L, A A Kepler state: position, velocity, a coordinate u that carries $1/|x|$, and the gravitational parameter m; the energy, the angular momentum and the Laplace–Runge–Lenz vector. Definition 25.8.

window, U_c, A_t, D_m, r The orbit field $u' = w$, $w' = a - u + bu^2$ for ratios a and b; its window $a > 0$, $b \geq 0$ and $100ab \leq 1$, with $a \leq 2u_0 \leq 3a$ for the initial value u_0; the centre U_c and half-amplitude A_t of the inverse radius, with $U_c + A_t = u_0$; and the clock constants D_m and r, with $D_m^2(1 - 2bU_c) = 1$. Definition 25.31, Lemmas 25.33 and 25.34.

H, Z, λ, (θ, T, Φ) For a window orbit of the proper-time field: the energy H, $Z = -2H$, $\lambda = am = m^2/L^2$, and a radial period, after which the radial state returns and the polar angle has grown by Φ. Definition 25.38.

p, e The semi-latus rectum and the eccentricity, read off the two apsidal inverse radii: $p(u_0 + u_1) = 2$ and $e(u_0 + u_1) = u_0 - u_1$. Theorem 25.42.

Heat, the tick and the spine

$D(q \,\|\, p)$ The relative entropy $\sum_{q_i>0} q_i \log(q_i/p_i)$, Definition 26.9; on the line $D_\eta(q \,\|\, \ell)$, with levels ℓ_i for $p_i = e^{\ell_i}$, Definition 26.1.

s, π, P, $D_n(a)$ A voxel state in $\mathbb{Z}^3$, the T9 law $\pi(s) = \varphi^{-|s|_1}/\varphi^9$, a voxel tick, and $D(P^n(a,\cdot) \,\|\, \pi)$. Definition 26.19.

the tick, (T1)–(T5) Recognition Science's tick, the voxel tick with positive holds and hold-return balance, and five numerical properties of it, stated as inputs that the book does not prove and used only where named. Definition 26.20 and the list after it.

e_n, $r(P)$ The even-return excess $P^{2n}(0,0) - \pi(0)$ and the exact rate $\sup_n e_{n+1}/e_n$, the pace. Definition 26.34.

(R) The return-probability hypothesis about the output of a finite interval computation at ticks 800 and 802; not proved here, and used only where named. Section 26.6.

(B), (M) Ball dominance and the monotone mean of one voxel's walk from the empty record: after every tick n each cost ball keeps at least its equilibrium mass, $G_n(B_j) \geq \pi(B_j)$, and the mean imbalance cost never falls, $m_n \leq m_{n+1}$. Named and not proved; Proposition 26.44 uses them only where it says so. Section 26.6.

spine, TC A finite set of voxels started at the vacuum, and the total correlation of a joint law: its relative entropy against the product of its marginals. Section 26.6, Definition 26.40.

sign law A law in one quantity built from comparisons with ratio thresholds by "and" and "or"; its price is read off seven bits at each threshold, the seventh, F, telling whether the law is visibly refuted at the threshold itself. Definition 26.51.

Gravity

J_2, FP(h, K) For a strain h, a real symmetric 4×4 matrix, at a wave K: J's long-wave form $J_2(h) = \frac{1}{2}\operatorname{tr}(h^2)$, and the Fierz–Pauli form of linearized gravity. Definition 27.1.

Π_K, **TT** The transverse projector $1 - KK^{\mathsf{T}}/|K|^2$; h is transverse-traceless at K when $hK = 0$ and $\operatorname{tr} h = 0$. Definition 27.1; on TT strains $\mathrm{FP}(h, K) = |K|^2 \cdot 2J_2(h)$, Theorem 27.3.

C_β The completion with DeWitt's β; $\beta = -1$ is Fierz–Pauli. Definition 27.9.

g_κ, R_1, E, $R(\kappa)$ The background $1 - \kappa \mathbf{1}\mathbf{1}^{\mathsf{T}}$ of the Kuhn lattice (Definition 27.19); Regge's tables at $\kappa = 1$ and at the Euclidean background $\kappa = 0$ (Table 27.1); and the table $R(\kappa)$ for every $\kappa \notin \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$, whose plane-wave form has its only pole at $\kappa = \frac{1}{4}$ (Theorem 27.28).

β_{PN} The post-Newtonian parameter (not DeWitt's β), read from the second-order static field. For J's cost on the Euclidean Kuhn lattice, the second-order equation for the static field built from two harmonic modes has no solution, so β_{PN} does not exist there; reading that field as a static mass's is an identification. Theorem 27.24, proved in classical mathematics.

definite record, locality clause Records definite in one of five senses, Definition 27.38; and the clause that each post is fixed by its site's record and one partner's, Definition 27.43.

The library and learning

ordered field, derivative convention, MON Classical structures by content, with stored decisions and $0^{-1} = 0$; the derivative where one exists and 0 elsewhere; and monotonicity from a nonnegative pointwise derivative. Definitions 28.13 and 28.22, Section 28.7.

Π^0_2 **weak excluded middle, the shift** $\neg\Pi_M \vee \neg\neg\Pi_M$ for every double sequence of bits M, where Π_M says that every row has an entry 1; the shift asks it only when every row is not not inhabited. Definition 28.25; as sentences of HAS, Definition 28.31.

Gödel's T**,** val(t) Programs built by application from K, S, 0, Succ and the recursors at every finite type, and their values. Definition 28.47.

record frame, IPC, LC A preordered set of stages with forcing, and intuitionistic and Gödel–Dummett logic. Definitions 29.1 and 29.5.

learner, certain learner A rule that names an index of a prior at every finite record, identifying a world in the limit when it settles on a correct index; and one that may stay silent and announces only correct indices. Definition 29.23.

EX, BC, BC***,** BC^k**,** EX* Criteria for learners that name codes of partial computable functions: settling on one correct code, eventually only correct codes, and these up to finitely many or k errors. Definition 29.34.

place learner, D^j**,** $C_k(\mathbf{b})$ A learner that names a place of a catalogue W at every finite record, its changes of mind counted; the derived catalogues, D^0 every place and $i \in D^{j+1}$ when $i \in D^j$ and, for every N, some place of D^j holding a world other than W_i agrees with W_i below N; and the catalogues $C_k(\mathbf{b})$ built from nested sums of tests. Definitions 29.57 and 29.63.

Σ_2**-LEM,** Σ_2**-LPO** For every binary double sequence, some row is all 0 or none is; every row has an entry 1 or some row is all 0. One principle, Proposition 29.50; Definitions 29.26 and 29.49.

LPO$_T$ LPO for the binary worlds that programs of Gödel's T compute. Definition 29.43.

Appendix O

Selected solutions

Each heading gives the chapter, the question number and the result in the text that the question concerns. Within a chapter the solutions follow the order of those results in the text, so a solution may come before one with a smaller question number. The solutions show the decisive steps and point back to their hypotheses. Try the question before reading its solution; in particular, identify the object of an induction and the direction of each map yourself.

O.1 Distinction and rigidity

Chapter 1, question 1: Proposition 1.2

Separation says that an expression with outer constructor S cannot equal the expression with outer constructor 0. This uses the difference between the two constructor forms. Injectivity compares two expressions with the same outer constructor: if $S(a) = S(b)$, their enclosed expressions must be equal, so $a = b$. The two facts concern different comparisons. One excludes an equality across constructors; the other recovers an equality from within a constructor.

Chapter 1, question 2: Theorem 1.4

For $r(e(n)) = n$, induct on the natural number n. At a successor,

$$r(e(n+1)) = r(S(e(n))) = r(e(n)) + 1 = n + 1.$$

The final equality uses the induction hypothesis. For $e(r(a)) = a$, induct on the orbit expression a. Its step is

$$e(r(S(a))) = e(r(a) + 1) = S(e(r(a))) = S(a).$$

The variables over which the two conclusions quantify determine the objects of induction. Both start cases follow by unfolding the two definitions at zero.

Chapter 2, question 1: Theorem 2.3

For a competing structure-preserving map f, the induction hypothesis is $f(a) = R_M(a)$. At the step,

$$f(S(a)) = s(f(a)) = s(R_M(a)) = R_M(S(a)).$$

The first equality uses preservation of the target step; the second uses the induction hypothesis; the third is the defining equation of R_M. None of the three faithful-generation conditions is needed. This proves uniqueness of a reading into any pointed step system, including one whose step merges distinct inputs.

Chapter 2, question 3: Lemma 2.6

Apply target induction to

$$P(y) : \quad \exists a \in \mathbb{N}_\delta \, R_M(a) = y.$$

At the target start b, use the witness $a = 0$. If a witnesses $P(y)$, then $S(a)$ witnesses $P(s(y))$, because $R_M(S(a)) = s(R_M(a)) = s(y)$. Target generation makes this property hold at every target element. Step injectivity and separation from the start are not used in this surjectivity argument.

Chapter 2, question 5: Proposition 2.8

The step is injective: distinct orbit elements have distinct successors, and $s(*) = 0_\delta$ is the successor of no orbit element. The start is a step, since $s(*) = 0_\delta$, so the second condition fails. Target induction fails for the property $y \neq *$: it holds at 0_δ and is preserved by s, which never returns $*$, yet it fails at $*$. The reading is the inclusion of the orbit, injective but not surjective. This is the first assertion of Proposition 2.8, and it shows that Lemma 2.5 needs its two conditions only at reached elements.

O.2 Arithmetic and its extensions

Chapter 3, question 1: Theorem 3.3

Addition is defined recursively in its second argument. In the commutativity step, the definition and induction hypothesis give

$$a + S(b) = S(a + b) = S(b + a).$$

The required answer is $S(b) + a$. The earlier lemma for a step in the first argument gives $S(b) + a = S(b + a)$, so it supplies the final equality in the reverse direction. Without that lemma, the recursive definition alone does not rewrite the desired first argument.

Chapter 3, question 2: Proposition 3.4

Prove right cancellation by induction on the common addend c. At $c = 0$, the hypothesis $a + 0 = b + 0$ is $a = b$. At $c = S(d)$, the hypothesis becomes $S(a + d) = S(b + d)$. Step injectivity gives $a + d = b + d$, and the induction hypothesis gives $a = b$. Left cancellation then follows by commutativity. This proof stays entirely within the constructed expressions; it does not pass through the reading r.

Chapter 3, question 6: Corollary 3.12

Let σ exchange 1 and 2 and fix every other natural number, and put $\varphi = \sigma \circ r$. Define $a \oplus b = \varphi^{-1}(\varphi(a) + \varphi(b))$ and $a \otimes b = \varphi^{-1}(\varphi(a)\varphi(b))$. Every semiring law transfers along the bijection φ. The neutral elements are $\varphi^{-1}(0) = 0$ and $\varphi^{-1}(1) = S(S(0))$. Since $\varphi(1) = 2$, we get $1 \oplus 1 = \varphi^{-1}(4)$, the tally with four steps. The hypothesis $a \oplus u = S(a)$ fails already at $a = 0$: there $0 \oplus u = u = S(S(0))$, while $S(0)$ is the tally with one step. The start is kept; the step is not.

Chapter 4, question 1: Proposition 4.2

Given $a + d = c + b$ and $c + f = e + d$, rearrange additions to obtain

$$(a + f) + d = (a + d) + f = (c + b) + f = (c + f) + b = (e + d) + b = (e + b) + d.$$

Cancellation gives $a + f = e + b$, the desired balance. All terms remain in the nonnegative carrier. Truncated subtraction discards the part below zero and cannot be treated as an additive inverse. The balance equation retains both sides until cancellation is justified.

Chapter 4, question 3: Proposition 4.8

For ratio representatives $(u, d), (v, e), (w, f)$, the two supplied cross equalities are $ue = vd$ and $vf = we$. Their consequence is

$$(uf)e = (wd)e.$$

The middle denominator e is nonzero. After the faithful integer reading, nonzero integer cancellation yields $uf = wd$; faithfulness returns this equality to the constructed integers. This is the exact point where the denominator condition is used. Allowing a zero middle denominator would remove the cancellation step.

Chapter 4, question 5: Theorem 4.13

With $\iota'(n) = 2n$, an additive g on the integers with $g(\iota'(n)) = r(n)$ must satisfy $g(2) = r(1) = 1$. Additivity gives $g(2) = g(1) + g(1)$, so $2g(1) = 1$, which has no integer solution. The carrier is the ordinary integers; the failure comes from the map $n \mapsto 2n$.

O.3 Rules and logical resources

Chapter 5, question 1: Display (5.1)

The hole formula is $1 + S(0) = S(x_0)$. The two terms are $1 + 0$ and 1. Substituting the first produces $1 + S(0) = S(1 + 0)$, supplied by the successor clause for addition. The zero clause supplies $1 + 0 = 1$. Equality substitution then fills the hole with the second term, giving $1 + S(0) = S(1)$. No induction is used.

Chapter 5, question 2: Theorem 5.5

Universal introduction lifts the old context before introducing the new variable. The truth of those lifted hypotheses under an extended assignment depends only on the old assignment, not on its new first entry. That entry may therefore be any natural number. The premise derivation establishes the body at every such entry, which is the universal conclusion.

In existential elimination, a witness supplies the new entry and makes the existential body true. The lifted old hypotheses remain true. The conclusion in that premise is also lifted, so its truth does not depend on the witness entry. It therefore descends to the original assignment as the required conclusion.

Chapter 5, question 7: Theorem 5.8

Injectivity fails: $S(0) = 1 = S(1)$, while $0 \neq 1$. Every other base rule, induction included, holds in the Boolean structure, by the third case of Theorem 5.8. A derivation from no hypotheses that avoids injectivity would therefore make $\neg(S(S(0)) = S(0))$ true there. But $S(S(0))$ and $S(0)$ both have value 1, so the formula is false. With injectivity the derivation is short: assume $S(S(0)) = S(0)$, apply injectivity to obtain $S(0) = 0$, refute that by separation, and discharge the assumption.

Chapter 5, question 5: Section 5.7

The rule at $y = a$ reads $x + S(a) = S(x + a)$, and $S(a) = a$, so $x + a = S(x + a)$: the sum must be a fixed point of S, that is, a or b. With only the new element a, every sum $x + a$ would be a, so $0 + a = a$ and the law $\forall x(0 + x = x)$ would hold. The second fixed point b allows $0 + a = b \neq a$. The remaining checks are those listed in the text.

Chapter 5, question 6: Corollary 5.9

The zero clause gives $0 + 0 = 0$. The successor clause gives $0 + S(0) = S(0 + 0)$, and substitution turns it into $0 + S(0) = S(0)$. The successor clause again gives $0 + S(S(0)) = S(0 + S(0))$, and substitution gives $0 + S(S(0)) = S(S(0))$. No induction occurs. An induction-free derivation of $\forall x(0 + x = x)$ would make the formula true in W by Theorem 5.8, but $0 + a = b \neq a$ there.

Chapter 6, question 1: Definition 6.3

The EM instance directly concludes $(0 = 0) \vee \neg(0 = 0)$ and has ledger (true, false, false; false). In the other derivation, reflexivity gives $0 = 0$ with an empty ledger, and left disjunction introduction preserves that empty ledger. Thus the displayed EM use is avoidable for this conclusion. Looking only at that proof would establish what it used, not what every proof must use.

Chapter 6, question 3: Definition 6.3

The induction formula is $\varphi(x) = \forall y(y = y)$, independent of x. It contains a quantifier, so the induction rule contributes the fourth entry. The base, step and all other rules use no EM, LPO or MP. The ledger is therefore (false, false, false; true). It is posit-free and nonempty. The formula tested for the induction entry is φ, not merely the final universally quantified conclusion.

Chapter 6, question 5: Lemma 6.8

The left side $x \cdot x$ has coefficients $(0, 0, 1)$ and the right side $y + y$ has $(2y)$, so $B = 1 + 1 + 2y = 2y + 2$. For $y = 8$, $B = 18$, and the search over $n = 0, \ldots, 18$ finds $n = 4$. For $n \geq 18$ the digit lists $(0, 0, 1)$ and $(16, 0, 0)$ differ. The formula $\exists z\,(x = z + z)$ holds at every even x and fails at every odd x, so its truth value never settles. The lemma's proof described each side by finitely many coefficients; a quantifier hides the value of z, so no such description exists.

Chapter 6, question 6: Theorem 6.10

The EM rule at $\exists x\varphi$ gives $(\exists x\varphi) \vee \neg\exists x\varphi$. In the first case, introduce the left side of the LPO conclusion. In the second, assume φ at a fresh variable; existential introduction, with that variable as the witness, gives $\exists x\varphi$, which contradicts $\neg\exists x\varphi$. Discharging the assumption and generalizing gives $\forall x\neg\varphi$, the right side. Implication introduction then adds the premise of the LPO instance, which is never used. Existential introduction occurs only in the passage from $\neg\exists x\varphi$ to $\forall x\neg\varphi$. The ledger is (true, false, false; false).

Chapter 6, question 7: Theorem 6.10

The only inductions are those in the proof of (6.9): one on the quantified formula $\forall y(x = y \vee \neg(x = y))$, and two quantifier-free ones on y that split into the zero and successor cases. For the equation $x = 0$, let β be $x = 0 \vee \neg(x = 0)$ under a quantifier whose variable does not occur in it. Assuming $\neg\beta$ gives $\neg(x = 0)$, since $x = 0$ would give β; that in turn gives β, a contradiction. So $\neg\neg\beta$ holds, and with it $\neg\neg\exists z\,\beta$. The MP rule at the quantifier-free β gives $\exists z\,\beta$, and existential elimination returns β. The ledger is (false, false, true; false).

Chapter 6, question 8: Theorem 6.11

The two results concern different sets of derivations: the proposition permits no induction, and the theorem permits induction on every formula. A reduction without induction remains a reduction with it, so the positive reductions of the proposition persist; its separations need not. Induction on quantified formulas decides equality, as in (6.9), and derives every MP instance, by Theorem 6.9. Once the restricted rule costs nothing, the reduction of EM to LPO and MP becomes a reduction of EM to LPO, and the four strengths collapse to two: the MP rule falls to the level of no rule, and EM to the level of LPO. The same observation shows that induction fails in the two-stage model, since the formula $\forall x(x = 0 \vee \neg(x = 0))$, which induction proves, fails in that model. Induction on quantifier-free formulas already proves that formula, but it does not supply the collapse. In the model of Proposition 6.14 quantifier-free induction holds, yet $\forall x\forall y(x = y \vee \neg(x = y))$ and an MP instance fail. With only that induction, MP keeps a level of its own, and the four strengths remain.

O.4 Recognition and representation

Chapter 7, question 2: Theorem 7.6

Take the supplied collision $a \sim b$ with $a < b$. Its positive difference $b - a$ is already a period at a, so it bounds the search for the least positive period. Once that period p is selected, $a \sim a + p$ supplies an upper bound a for the search for the least index. Pairwise decidability decides each candidate relation in these two bounded searches. The collision is the finite witness that prevents either search from being an unbounded existence task.

Chapter 7, question 6: Display (7.5)

Suppose (x, y) and (x', y') are both first collisions. If $y' < y$, the related pair (x', y') violates the leastness of y; if $y < y'$, the reverse holds. So $y = y'$. If $x' < x$, then $x' \sim y$ violates the second leastness condition of (x, y); if $x < x'$, the reverse holds. So $x = x'$.

For $\sim_{i,p}$ the pair $(i, i + p)$ is related. Any related pair $x < y$ has $i \leq x$ and $x + p \leq y$, so no related pair has its larger element below $i + p$, and no $x' < i$ is related to $i + p$. Hence $(i, i + p)$ is the first collision, and the event of γ sits at $2(i + p) + (i \bmod 2)$. The position determines its half y, the larger element, and y determines the smaller element as the least count related to it. Only the parity of that element remains to be recorded.

Chapter 7, question 5: Theorem 7.9

If θ has no witness, the second alternative of (7.4) never applies, so $x \sim_\theta y$ gives $x = y$. If $\sim_\theta$ is equality and $\theta(n)$ held, then $n \sim_\theta n + 1$, since $n \leq \min(n, n + 1)$, and this would give $n = n + 1$. Hence no n satisfies θ.

In the forward direction, WLPO is applied to $\beta_\sim$. In its all-zero branch a related pair $x \sim y$ must be shown equal. Decide whether $x = y$. If $x \neq y$, then $x < y$ or $y < x$, and by symmetry the smaller count is related to the larger, so the search at $\max(x, y)$ succeeds: $\beta_\sim(\max(x, y)) = 1$. So only the answer $x = y$ remains. Equality and order of two counts are decided by comparing them, a finite procedure; no principle about an unbounded search is involved.

Chapter 7, question 4: Theorem 7.10

Define $x \sim_P y$ by $x = y \vee P$. Its equivalence and addition properties follow by cases on supplied proofs of these disjunctions; defining the relation does not require deciding P.

If the catalogue says this relation is equality, then P would give $0 \sim_P 1$, hence $0 = 1$, a contradiction. This branch supplies $\neg P$. If the catalogue supplies an index and positive period, it gives a related pair i and $i + p$. These are unequal, so their relation $i = i + p \vee P$ supplies P. The two catalogue branches therefore decide the original proposition.

Chapter 7, question 7: Theorem 7.12

Adding 2 to 0 and to 3 gives 2 and 5. Both are at least 2 and both leave remainder 2 modulo 3, so $0 + 2 \sim_{2,3} 3 + 2$. But 0 and 3 are different and $0 < 2$, so they are not related. The relation is not cancellative.

For $\sim_{0,3}$, relation (7.8) gives $[a, b] \approx [c, d]$ exactly when $(a + d) \bmod 3 = (c + b) \bmod 3$. Since $I([c, d]) - I([a, b]) = (c + b) - (a + d)$, this says that 3 divides the difference of the integer readings: $\approx$ is congruence modulo 3 of the readings $I(u)$ and $I(v)$.

Chapter 7, question 11: Proposition 7.43

Each $\{n\}$ is open and closed, so $\mathrm{Dec}(X)$ and $\mathrm{A}(X)$ have the value $\top$. An open set containing ∞ contains all but finitely many natural numbers, so the value of $\mathrm{Z}(X)$, the interior of $\{\infty\}$, is empty, and $\neg\mathrm{Z}(X)$ is the whole space. The value of $\mathrm{E}(X)$ is $\mathbb{N}$. The value of $\mathrm{Ev}(X)$ is the

interior of the complement of the even numbers, which is the set of odd numbers, and $\mathrm{Od}(X)$ has the value of the even numbers. So the instances of MP, LLPO and $\mathrm{MP}^{\vee}$ at A all have the value $\mathbb{N}$, which misses ∞.

At B, $\mathrm{Dec}(Y)$ has the value $\top$, since $\{n\}$ and $\emptyset$ are open and closed. The value of $\mathrm{E}(Y)$ is the set of even numbers; its closure adds ∞, and the interior of that closure is again the set of even numbers, because every neighbourhood of ∞ contains odd numbers. So $\neg\neg\mathrm{E}(Y)$ has the value of the even numbers. The value of $\exists n\,(n \in X \wedge \neg\, n \in Y)$ is the union of the sets $\{n\} \cap \neg B_n$, which is the set of odd numbers, and its double negation is the same set. The instance of $\mathrm{W}(X)$ at B therefore has the value $\mathbb{N}$. Hence the value of $\mathrm{W}(X)$ lies inside $\mathbb{N}$, the value of $\mathrm{E}(X)$, and the instance of WMP at A is $\top$: at this sequence the hypothesis of WMP already fails at ∞.

Chapter 7, question 12: Proposition 7.46

The values of $\mathrm{Dec}(X)$ and $\mathrm{A}(X)$ are $\top$. The residue classes R_0, R_1, R_2 modulo 3 divide $\mathbb{N}$ into three disjoint sets, so each of the ultrafilters $U_{2,0}$ and $U_{2,1}$ contains exactly one of them. Some class R_i therefore belongs to neither, and its complement belongs to both, hence to F_2. The set $\{\infty\} \cup (\mathbb{N} \setminus R_i)$ is then open, contains ∞, and misses every $A_{3j+i} = \{3j + i\}$, so it lies in the value of $\mathrm{C}^3_i(X)$.

Which i it is depends on the two ultrafilters, and their construction does not say: $U_{2,0}$ contains the even numbers, which meet all three classes, and the ultrafilter chooses among them. That is no obstacle. The value of LLPO_3 is computed in classical mathematics, where it is enough that some i works at each point; the model is an outside witness that $\mathrm{LLPO}_3 \to \mathrm{LLPO}$ has no derivation, not a procedure for naming the class.

Chapter 7, question 13: Proposition 7.48

The open sets of the cone are $\emptyset$, $\{a\}$, $\{b\}$, $\{a, b\}$ and the whole space. So $\neg\{a\} = \{b\}$ and $\neg\{b\} = \{a\}$. For $n = 0$, $A_0 \cup \neg A_0 = \{a, b\}$; for $n \geq 1$, $A_n \cup \neg A_n$ is the whole space. The value of $\mathrm{Dec}(X)$ is $\{a, b\}$, that of $\mathrm{E}(X)$ is $\{a\}$, and that of $\mathrm{Z}(X)$ is $\neg A_0 = \{b\}$. The instance of LPO is $\{a, b\} \Rightarrow \{a, b\} = \top$. The instance of WEM is $\neg A_0 \cup \neg\neg A_0 = \{b\} \cup \{a\}$, which misses c.

At the focal point the set is not decidable, so LPO says nothing about it there; LPO only decides sets that are decidable. WEM speaks about every proposition, decidable or not, and it is exactly at such a point that it fails.

Chapter 8, question 1: Theorem 8.4

In the list induction step, the head of a nonempty list must be the one available mark. It is therefore the same head that expansion places in front of the recursively expanded tail. The induction hypothesis identifies the tails, completing the round trip.

If two lists have equal lengths, expanding that equality and using the round trip identifies the lists. This injectivity is used to reflect equality in the term and formula interpretation. With more than one possible mark, length would discard the head's identity, and that argument would fail.

Chapter 8, question 5: Theorem 8.8

For $x = n > 0$ and $y = a$: $S(a) = a$, so $n \cdot S(a) = n \cdot a = b$, since neither factor is 0 and one of them is a; and $n \cdot a + n = b + n = b$. For $x = y = a$: $a \cdot S(a) = a \cdot a = b$, and $a \cdot a + a = b + a = b$, because $x + a = b$ for every x.

Each numeral instance $\neg(S(S^n(0)) = S^n(0))$ is a quantifier-free sentence true in $\mathbb{N}$, so it holds in W by the first conclusion of Theorem 8.8. The universal sentence ranges over all of W, including a, where $S(a) = a$. The instances speak only about numerals, and the numerals miss a.

Chapter 8, question 6: Theorem 8.8

Every equation holds in the one-element carrier, so injectivity of the step and the four recursion equations hold. Separation fails, because the only element is zero and is also its own successor. Then $\nu(0) = \nu(1)$, so ν is not injective, and the first conclusion of the theorem fails: the quantifier-free sentence $S(0) = 0$ is true in the carrier and false in $\mathbb{N}$.

Chapter 8, question 7: Theorem 8.11

Assume $S(t) = S(u)$ as a hypothesis, apply the injectivity rule to obtain $t = u$, and discharge the hypothesis by implication introduction. No additional principle or induction occurs, so the ledger is empty. In the other direction, a use of the injectivity rule on a derivation of $S(t) = S(u)$ becomes implication elimination, applied to the axiom $S(t) = S(u) \to t = u$ and to the translated derivation of $S(t) = S(u)$.

Chapter 8, question 8: Theorem 8.11

Take the usual derivation: the base $0 + 0 = 0$ is the axiom $t + 0 = t$; the step assumes $0 + x = x$, uses the axiom $0 + S(x) = S(0 + x)$ and equality substitution, discharges the assumption and generalizes; induction on $0 + x = x$ concludes. It uses no excluded middle, and its only induction formula is quantifier-free. By the final clause of Theorem 8.11, the EM and induction entries of its translation are false, and the LPO and MP entries are false for every translation. The ledger is empty. The translation is the derivation (5.5) of Chapter 5, whose empty ledger Chapter 6 records: its conclusion contains a quantifier, but its induction formula does not.

Chapter 8, question 4: Theorem 8.13

The quotient relation is $x \sim y$ exactly when $r(x) = r(y)$. Consequently $[x] \mapsto r(x)$ is independent of the representative. If two outputs agree, the same kernel definition identifies their classes, proving injectivity. An element of the range is, by its definition, some $r(x)$; its preimage is $[x]$. A value in a larger ambient target need not have such an x, so that larger surjectivity does not follow.

O.5 Iteration and effective description

Chapter 9, question 2: Theorem 9.2

The recursor's step for predecessor is $g = \lambda x.\lambda y.x$. After applying the successor equation, the two function evaluations are

$$\mathrm{pred}(S(x)) = g(x, R(O, 0, g, x)) = (\lambda y.x)(R(O, 0, g, x)) = x.$$

The first argument of g is the preceding orbit point x. The second, which the function ignores, is the preceding recursive value. An ordinary iterator step receives only that second kind of input; the indexed recursor supplies the position explicitly.

Chapter 9, question 3: Theorem 9.6

Write a point of the two-chain model as (i, n), with start $(0, 0)$ and successor $(i, n + 1)$. The maps are $c(i, n) = (0, n)$ and $q(i, n) = (i, n)$. Both send the start to itself. For c,

$$c(S(i, n)) = (0, n + 1) = S(c(i, n));$$

for q, step preservation is immediate from its identity definition. At $(1, 0)$ their values are $(0, 0)$ and $(1, 0)$, so they differ. The recursive equations do not force equality on the chain that the designated start never reaches.

Chapter 10, question 1: Theorem 10.3

Let L be the length of the supplied tally x. Its own-budget representative belongs to F_L. Define the global output using that representative and the map for F_L.

If the same tally is supplied as an element of F_n, then $L \leq n$. Use the inclusion $F_L \hookrightarrow F_n$. Compatibility says that applying the map at F_n after this inclusion gives the same output as the map at F_L. This proves agreement with every permitted presentation. The own-budget construction gives the output; the compatibility equation proves its independence from a larger budget.

Chapter 10, question 4: Definition 10.10

The program can witness the finite-prefix assertion for requests below N, provided the supplied endpoint pairs form the required prefix. A uniform name has to halt and return the correct endpoints of one valid protocol at every request. Printing zeros afterward supplies termination, but does not by itself establish those endpoint conditions. A correct finite prefix leaves the infinitely many remaining requests as a separate obligation.

O.6 Enclosures and exact comparison

Chapter 11, question 1: Display (11.1)

If $a \leq b$, nesting gives $\ell_a \leq \ell_b \leq u_b$. If $b \leq a$, it gives $\ell_a \leq u_a \leq u_b$. Both prove $\ell_a \leq u_b$. Choosing $k = \max(a, b)$ combines the two arguments as $\ell_a \leq \ell_k \leq u_k \leq u_b$.

Chapter 11, question 2: Theorem 11.5

Overlap bounds the gap between one protocol's endpoint and the other protocol's corresponding interval. Adding the two width estimates then bounds the absolute difference of the displayed values by $2/(n+1)$. To apply the reciprocal-squeeze lemma in its stated form, request $n = 2m+1$. Then $2/(n+1) = 1/(m+1)$ for every m, so the displayed values agree. Overlap supplies the comparison; shrinking widths force its gap to vanish.

Chapter 12, question 2: Lemma 12.5

Each jump stage inspects only its own finite prefix. Either it finds a first event within that prefix or it does not; both results are decidable by finite search. This defines every stage without a decision about the whole sequence.

When a particular first event N has been supplied, 2^{N+1} is an explicitly constructed later stage. It exposes the nonzero jump at a precision fine enough for the contradiction in Lemma 12.5. The supplied N determines the later stage by a finite calculation.

Chapter 12, question 3: Theorem 12.6

For the zero test, WLPO decides whether every comparison test in the relevant binary sequence vanishes. That is enough to decide equality to zero as a proposition: zero or not zero. In the reverse direction, applying the proposed zero test to the jump construction decides whether the original binary sequence is all zero.

For trichotomy, the nonzero alternative must provide a strict sign. The proof uses LPO to obtain an actual event in the comparison sequence, hence a finite strict-separation witness. WLPO's negative answer to "all zero" does not by itself display that event. This is the additional step needed for the stronger comparison. Proposition 12.13 isolates that step. It is the implication $\neg Z(x) \to P(x) \vee N(x)$, and by Theorem 12.11 its exact price is $\mathrm{MP}_{\mathrm{bin}}$. Together with Theorem 12.6 this recovers $\mathrm{LPO} \iff \mathrm{WLPO} \wedge \mathrm{MP}_{\mathrm{bin}}$.

Chapter 12, question 5: Lemma 12.10

Assume $\neg P(x)$ and $\neg N(x)$, and fix n. The comparison $|q_n|(n+1) \le 2 \cdot 2^n$ is decidable, so it suffices to refute its failure. Suppose $2 \cdot 2^n < |q_n|(n+1)$. The sign of the integer q_n is decidable. If $q_n \ge 0$, then $|q_n| = q_n$ and n is a positive witness. If $q_n < 0$, then $|q_n| = -q_n$, so $q_n(n+1) < -2 \cdot 2^n$ and n is a negative witness. Either contradicts a hypothesis, so the bound holds at n. Decidability is used only at the single stage n: once for the zero bound and once for the sign of q_n. The proof never decides $P(x)$; it only refutes a witness at the stage under consideration, which is what the hypothesis $\neg P(x)$ provides, and no search is made.

Chapter 12, question 6: Corollary 12.12

Let x' have numerators $-q_n$. Condition (12.1) involves $|q_m 2^n - q_n 2^m|$, which negating both numerators leaves unchanged. A negative witness for x is a positive witness for x', and $x'' = x$. So stability of N for every presentation is stability of P for every presentation, and Corollary 12.12 for P gives it for N.

Chapter 12, question 7: Theorem 12.45

If $t_n = \eta(1)$, an event occurs at some $k \le n$, and the same event gives $t_{n+1} = \eta(1)$. If $t_n = \eta(0)$, then $t_n \le t_{n+1}$ because $\eta(0) \le \eta(0)$ and $\eta(0) \le \eta(1)$. Every term lies between $\eta(0)$ and $\eta(1)$, so the sequence is bounded. The limit L and the index N come from existential statements, and they are used inside a proof of one proposition: LPO's alternative for the given α. Using a witness of an existential statement to prove a proposition is ordinary existential elimination. No infinite family of selections is assembled into a sequence, and the test sequence itself is defined by bounded searches, not selected.

Chapter 12, question 8: Theorem 12.48

Cotransitivity applied to $\eta(0) < \eta(1)$ places L strictly above $\eta(0)$ or strictly below $\eta(1)$. Each case excludes one alternative of WLPO, and neither requires knowing which of $\eta(0)$ and $\eta(1)$ the point L equals. In the case $\eta(0) < L$ the argument gives only that the bits are not all zero. If every ratio below L is exceeded by a term, apply this at the ratio 0: some b_n exceeds $\eta(0)$. Then $b_n = \eta(1)$, since $b_n = \eta(0)$ would give $\eta(0) < \eta(0)$, and so $\alpha(n) = 1$. The case $L < \eta(1)$ still gives that every bit is zero, so the two cases give LPO. This is the first part of Theorem 12.49.

O.7 Compactness, algebra and computation

Chapter 13, question 1: Theorem 13.3

Let $D(s) = 1$ exactly when $|s| = 1$. Every binary sequence p has $\bar{p}(1)$ in D, so D is a decidable bar. Its complement contains the string 00 but not its initial segment 0, so the complement is not a tree. No string of length at least 1 is unseen, so the tree A of unseen strings is $\{\varnothing\}$, and it has no member of length 1. In general, suppose A has no member of length N, and let p be a binary sequence. Then $\bar{p}(N)$ is seen: some $n \le N$ has $D(\bar{p}(N)\restriction n) = 1$. Since $\bar{p}(N)\restriction n = \bar{p}(n)$, the sequence p meets D at a length at most N. The detour through A is needed because the tree form speaks only of trees. The seen strings are closed under extension, so their complement is closed under initial segments; the complement of D itself need not be.

Chapter 13, question 3: Theorem 13.6

A string of length m with numeral v is caught when some $k < m$ has $a_k < (v-1)/2^m$ and $(v+2)/2^m < b_k$. At $m = 0$ nothing is caught, since there is no $k < 0$. At $m = 1$ only $I_0 = (-1, \frac{1}{2})$ may be used, and the widened pieces $[-\frac{1}{2}, 1]$ and $[0, \frac{3}{2}]$ do not lie inside it. At $m = 2$ the intervals I_0 and $I_1 = (\frac{1}{4}, \frac{3}{4})$ may be used, and the widened pieces are $[(v-1)/4, (v+2)/4]$ for $v \le 3$. Their right ends $\frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}$ are never below $\frac{1}{2}$, which rules out I_0; for I_1 the left end must exceed $\frac{1}{4}$, which needs $v = 3$, and then the right end $\frac{5}{4}$ exceeds $\frac{3}{4}$. So nothing is caught. At $m = 3$ the intervals I_0, I_1 and $I_2 = (\frac{1}{2}, 2)$ may be used, and the widened pieces are $[(v-1)/8, (v+2)/8]$. The interval I_0 catches the pieces with $v + 2 < 4$, that is $v \le 1$. The interval I_1 would need $v - 1 > 2$ and $v + 2 < 6$, which no v satisfies. The interval I_2 catches the pieces with $v - 1 > 4$, that is $v \ge 6$. So

$$H = \{\varnothing,\ 0,\ 1,\ 00,\ 01,\ 10,\ 11,\ 010,\ 011,\ 100,\ 101\},$$

the members of length 3 being the pieces with $v = 2, 3, 4, 5$. At $m = 4$ every one of the sixteen pieces is caught, as the opening computes, so no string of length 4 lies in H.

Each piece of length 4 is caught by exactly one of I_0, I_1 and I_2: I_0 for $v \leq 5$, I_1 for $6 \leq v \leq 9$ and I_2 for $v \geq 10$. At the borders, $v = 5$ misses I_1 because its left end $\frac{4}{16}$ is not above $\frac{1}{4}$; $v = 6$ misses I_0 because its right end $\frac{8}{16}$ is not below $\frac{1}{2}$; $v = 9$ misses I_2 because its left end $\frac{8}{16}$ is not above $\frac{1}{2}$; and $v = 10$ misses I_1 because its right end $\frac{12}{16}$ is not below $\frac{3}{4}$. The proof yields the first $N = 4$ intervals; since every catching index is below 3, the first three intervals already cover $[0, 1]$.

Chapter 13, question 2: Theorem 13.7

Every binary sequence leaves T, so T cannot be infinite: weak König's lemma would give a path through it, a sequence that never leaves it. So it is false that T has a member of every length. Whether T has no member of length n is decided by inspecting 2^n strings, so what we know is $\neg\neg\exists n\, Q(n)$ for a decidable property Q. Searching $n = 0, 1, 2, \ldots$ for an empty level would stop only if $\exists n\, Q(n)$ held. Passing from the double negation to the existence statement is Markov's principle for Q: knowing that the search cannot run forever does not show that it stops.

The padded tree avoids the passage. It is infinite outright, with no case distinction on T, so weak König's lemma applies to it and gives a path p. The hypothesis on T, applied to the binary sequence p, gives an n with $\bar{p}(n) \notin T$. This existence statement is supplied by the hypothesis, not by a search. Membership of $\bar{p}(n)$ in the padded tree at $k = n$ then leaves only the second alternative: T has no member of length n.

Chapter 13, question 4: Theorem 13.18

Let D be a decidable bar and p a binary sequence. Each test $D(\bar{p}(n)) = 1$ is decided, and the bar asserts that some n passes. A search through $n = 0, 1, 2, \ldots$ therefore stops, at the least passing n, and the stopping point is unique. The search reads only $p(0), \ldots, p(h(p) - 1)$, which is the continuity of the first hit used in Theorem 13.15. Collecting these values into a map is the step at which that theorem names unique choice.

A last exit n of p requires $D(\bar{p}(m)) = 1$ for every $m \geq n$. No finite inspection of $D(\bar{p}(0)), \ldots, D(\bar{p}(M))$ establishes this, since a later value 0 can always follow. When the c-set of D is a bar there is an N with $\bar{p}(N)$ in the c-set, but membership in the c-set, $D(\bar{p}(N)t) = 1$ for every string t, is again a statement about all extensions, so a search cannot recognize such an N. That is why Lemma 13.17 proves only that exactly one last exit exists. Under FAN_c one number N serves for all sequences: every p has $\bar{p}(n)$ in the c-set for some $n \leq N$, so $D(\bar{p}(m)) = 1$ for every $m \geq N$. The last exit is then one more than the last $m < N$ with $D(\bar{p}(m)) = 0$, or 0 if there is none, found by inspecting N values. This is the map L in the proof of Theorem 13.18.

Chapter 13, question 5: Theorem 13.29

Fix the point a of the proof of Theorem 12.60 and one regular sequence naming its clamp c. The conclusion drawn from a is a proposition, so one name serves. The function $h(x) = \max\big(\min(c, x - \frac{2}{3}), \min(-c, \frac{1}{3} - x)\big)$ is built from x by translations, negations, maxima and minima. Each operation has an explicit formula on regular sequences, so a name of $-h(x)$ is computed from a name of x and the fixed name of c; this is the transformer. By Lemma 12.52, item 5, points within $1/(k+1)$ have values of h, and hence of $-h$, within $1/(k+1)$, so $\mu(k) = k$ is a modulus. With the presentation of $[0,1]$ by dyadic grids, $-h$ is a presented uniform map on a presented set. Theorem 13.29 therefore constructs its infimum with no choice principle and no omniscience principle.

Suppose the theorem were strengthened to supply, for every presented uniform map on an inhabited presented set, a member at which the infimum is attained. Applied to $-h$ it gives x in $[0,1]$ with $-h(x) \le -h(y)$, that is $h(y) \le h(x)$, for every y in $[0,1]$. Then x is a maximum point of h, and the first part of the proof of Theorem 12.60 gives $a \ge \eta(0)$ or $a \le \eta(0)$. Since a is arbitrary, this is LLPO by Corollary 12.18. The infimum costs nothing; a point that attains it costs at least LLPO.

Chapter 14, question 1: Definition 14.1

A basis selection of the single vector (x) is one bit b_0, and a bit is decided. If $b_0 = 0$, nothing is selected, and spanning says that (x) is the empty combination; so $x = 0$. If $b_0 = 1$, independence applies to the combination with coefficient 1: were $x = 0$, it would vanish and give $1 = 0$. So $x \neq 0$; this is the first statement of Lemma 14.11. Spanning gives the first alternative and independence the second, and together they give $x = 0 \vee x \neq 0$ for every point, which is WLPO by Corollary 12.18.

Chapter 14, question 2: Lemma 14.8

Taking absolute values in $|x|\,c = x$ gives $|x|\,|c| = \big||x|c\big| = |x| = |x| \cdot 1$, using $|uv| = |u|\,|v|$ and $\big||x|\big| = |x|$. Since $x \neq 0$, also $|x| \neq 0$, and Lemma 14.4 cancels $|x|$: $|c| = 1$.

The proof of Lemma 14.8 uses only that c is a point. Cotransitivity of $-1 < 1$ at c gives $-1 < c$ or $c < 1$. In the first case a negative x would be invertible and force $c = -1$, so $x \ge 0$; in the second a positive x would force $c = 1$, so $x \le 0$. No bound on c is given in advance and none is used, and the lemma does not need $x \neq 0$. The bound $|c| = 1$ is a consequence when $x \neq 0$; it is what makes c apart from 0 in Lemma 14.9.

Chapter 14, question 3: Theorem 14.16

Write $y = g(j/2^n)$. The statement $\neg(y \le 0)$ is $\neg\neg(y > 0)$. If $y < 0$, then $y > 0$ is impossible, since $0 < y < 0$ would give $0 < 0$; so $\neg(y > 0)$, which contradicts $\neg\neg(y > 0)$. Hence $\neg(y < 0)$, that is $y \ge 0$.

The relation of the proof says: $\beta = 1$ and $y \le 0$, or $\beta = 0$ and $\neg(y \le 0)$. The two conditions exclude each other, so at most one bit satisfies it, and WLPO, through the third row of Table 12.2, gives one. So exactly one bit satisfies it, and unique choice applies. The bits in the proof of

Theorem 12.54 only promise $y \le 0$ when $\beta = 1$ and $y \ge 0$ when $\beta = 0$. When $y = 0$ both promises hold, so both bits are correct and the relation need not have exactly one witness; that proof therefore collects its bits with countable choice.

Chapter 14, question 4: Definition 14.23

Suppose $a \le bl$ and $br \le a$. Then $br \le bl$. Here $b \ge 0$, because $0 \le a \le b$, and $b \ne 0$, so Lemma 14.5 gives $r \le l$, contrary to $l < r$. So $a \le bl$ and $br \le a$ cannot both hold.

With a single midpoint m in place of $l < r$ there is no contradiction to reach: $a \le bm$ and $bm \le a$ together say only that $bm = a$, and $a = bm$ makes both hold. To name a side at one point one has to compare a with bm, which is the comparison of the fourth row of Table 12.2, priced LLPO. Two different points leave room. Because positivity belongs to the point, a search through stages for a witness of $a > bl$ stays silent exactly when $a \le bl$, and one for $br > a$ exactly when $br \le a$; so the two searches cannot both stay silent. This is what $\mathrm{MP}^{\vee}_{\mathrm{seq}}$ turns into a bit in Theorem 14.28.

Chapter 15, question 1: Proposition 15.4

A sum requested at precision n asks both summands for precision $n + 2$. Induction on h: if the inner sums of depth $h - 1$ ask their leaves for $(n + 2) + 2(h - 1)$, the leaves of a sum of depth h are asked for $n + 2h$. The protocol sum of Chapter 11 turns a request n into $2n + 1$, and h nestings turn it into $2^h(n+1)-1$, again by induction, since $2\big(2^{h-1}(n+1)-1\big)+1 = 2^h(n+1)-1$. The first is linear in h, the second exponential. The difference comes from the modulus: with 2^{-n} two errors of size $2^{-(n+2)}$ and a rounding of $2^{-(n+1)}$ fit into 2^{-n}, whereas with $1/(n + 1)$ halving an error requires doubling the index.

Chapter 15, question 3: Theorem 15.11

Forty bits within 2^{-40} require $2^{40} \le 120K^4$, that is $K^4 \ge 2^{40}/120 \approx 9.16 \cdot 10^9$. Since $309^4 = 9116621361$ is smaller and $310^4 = 9235210000$ is larger, at least 310 steps are needed. The Taylor solver uses 32 steps for the oscillator at every precision, of order 61 at 40 bits. The two equations differ, but both are linear, and the mechanism is visible in the proofs: one step of a fixed-order method has an error of fixed order in the step, $h^5/120$ here, so the error over $[0, 1]$ decreases only like K^{-4}; the Taylor step has truncation error $M2^{-N}$ (Lemma 15.16(b)), which the solver drives down by raising the order N with the precision instead of shrinking the step.

Chapter 15, question 2: Definition 15.18

The field has $d = 1$, $c = 0$, no A-terms and the single B-term $(-1, 0, 0)$, so $B(u, v) = -u_0 v_0$. Definition 15.24 gives $\gamma = 0$, $\alpha = 0$ and $\beta = \lceil |-1| \rceil = 1$. With $M = 1$, $R = 0 + 0 + 1 \cdot 1 + 1 = 2$, so $\sigma = \mathrm{bl}(8) + 1 = 5$ and $S = 1 \cdot 2^5 = 32$. Then $G = \lfloor 40^{32}/2^{160} \rfloor + 1 = \lfloor (5/4)^{32} \rfloor + 1 = 1263$, and $1 + 2S(M + 2) = 193$, so $\kappa = \mathrm{bl}(1263 \cdot 193) + 2 = \mathrm{bl}(243759) + 2 = 18 + 2 = 20$, because $2^{17} \le 243759 < 2^{18}$. These are the same σ, S and G as for the oscillator, since both fields have $R = 2$; only $M + 2$ differs. The region certificate at level 0 asks that the 33 numbers $y_0^{(0)}, \dots, y_{32}^{(0)}$, computed with Taylor order and working precision 20, all have absolute value at

most $1 - 2^{-2} = 3/4$. That is a finite computation with dyadic numbers; whether it succeeds is decided by carrying it out. The classical solution $1/(2+t)$ stays between $1/3$ and $1/2$, so if the certificate holds, Theorem 15.29 identifies the certified point with $1/3$.

Chapter 15, question 4: Theorem 15.23

As a coefficient field the oscillator has $d = 2$, $c = 0$, the A-term $(1, 1)$ for coordinate 0 and $(-1, 0)$ for coordinate 1, and no B-terms. So $c_A = 2 \cdot 2 = 4$ and $c_B = 0$, and the bounds are $\gamma = 0$, $\alpha = 1$ and $\beta = 0$, as in the text. Then $2\,\mathrm{step}(N) = 2N(4 + 12) + 4 + 2N(N + 1)$, that is $\mathrm{step}(N) = 16N + 2 + N(N + 1)$. At $m = 40$ with $n_0 = 0$ and $\kappa = 21$, $N = 61$, so $\mathrm{step}(61) = 976 + 2 + 3782 = 4760$ and $\mathrm{cost}(40) = 1 + 2 + 32 \cdot 4760 = 152323$. The bound of Proposition 15.25 is $(32 \cdot (4 + 0 + 16) + 3) \cdot 62^2 = 643 \cdot 3844 = 2471692$, about sixteen times larger. The gap comes from replacing $\mathrm{step}(N)$ by $(c_A + c_B + 8d)(N + 1)^2$; both grow like m^2.

Chapter 15, question 5: Definition 15.27

Here $d = 1$ and p has the single term $(1, (3))$, so $D = 3$. The variables are $w_0 = 1$, $w_1 = x$, $w_2 = x^2$ and $w_3 = x^3$, which is $(D + 1)^1 = 4$ variables, as Theorem 15.28(a) allows. For w_a with $a > 0$ the only term is $(a \cdot 1,\ a - 1,\ 3)$, so

$$w_0' = 0, \qquad w_1' = w_0 w_3, \qquad w_2' = 2 w_1 w_3, \qquad w_3' = 3 w_2 w_3,$$

with initial values $(1, x_0, x_0^2, x_0^3)$. For x^3 the identity reads $(x^3)' = 3x^2\ p(x) = 3x^2 \cdot x^3$, which is $3 w_2 w_3$ at the monomial vector. The field has $c_B = 3 \cdot 3 = 9 \le 3 d_Q t_p = 12$.

Chapter 16, question 1: Lemma 16.5

In the context Ω take the formula $\neg \in (\mathrm{map}(\mathrm{id}))$, which says "not p" of the truth value p. With only the first half of the earlier clause, the earlier object $(\mathrm{false}, \mathrm{true})$ of Ω satisfies it: the atom fails there at the earlier stage, so the implication to $\bot$ holds vacuously. The passage sends this object to the proposition true, at which the formula fails at the later stage. So the reading is no longer a predicate. In the proof of the lemma, the case of an implication used exactly the second half of the earlier clause, which is the later clause at $h(g)$; without it that case has nothing to appeal to.

Chapter 16, question 2: Theorem 16.14

A closed term of type $\mathbb{N}$ is a numeral, an application $t\,u$, or another term, and each term has exactly one of these forms. Put $V(\bar{n}) = n$; for an application $t\,\bar{n}$ to a numeral put $V(t\,\bar{n}) = J(t)(n)$; and give every other closed term of type $\mathbb{N}$ the value 0. Let $J(t)$ be the function carried by t when that function is computable, and the zero function otherwise. Then $J(t)$ is always computable, and conditions (1) and (2) hold by construction. Condition (3) fails at the instance $h_r\,\bar{n} = \overline{r(n)}$ for every n with $r(n) = 1$: the left side has the value $J(h_r)(n) = 0$, because r is not computable, and the right side the value 1.

Chapter 16, question 4: Proposition 16.21

Suppose a tracked map $F : \Omega^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ sent every classifier to its 0–1 function. The map $p \mapsto (n \mapsto p)$ from Ω to $\Omega^{\mathbb{N}}$ is tracked: the program that returns the program 0 on every input names every later object of $\Omega^{\mathbb{N}}$. Evaluation at the number 0 is tracked. Composing the three gives a tracked map $\Omega \to \mathbb{N}$ that sends the true proposition to 1 and the false one to 0, which part (c) of the proposition forbids. The proof of Theorem 16.27 meets the same obstacle: comprehension gives, for every binary matrix, the classifier "row i vanishes", and $\mathsf{WLPO}_{\mathsf{seq}}$ asks for its 0–1 function as a named sequence.

Chapter 16, question 3: Theorem 16.22

Let A and B be sentences. From a derivation of $A \vdash B$, cut $\top \wedge A \vdash A$, obtained by conjunction elimination from construction (2), with it to get $\top \wedge A \vdash B$, and apply implication introduction: $\top \vdash A \to B$. Conversely, from $\top \vdash A \to B$, cut $A \vdash \top$ with it to get $A \vdash A \to B$, and apply implication elimination with $A \vdash A$: $A \vdash B$. Only constructions common to both calculi are used, and no term or map appears, so the argument is the same in the computable calculus.

Chapter 16, question 5: Theorem 16.32

The path p is computable, since g is a finite string. By Lemma 16.29(d) it leaves T_K. The proof of that part gives a program e with $\{e\}(x) = p(e)$ for every x, halting on input 0 within some k steps, and shows that $\bar{p}(k + e + 1) \notin T_K$. So the first initial segment of p outside T_K has length at most $k + e + 1$; it exceeds N, since every initial segment of length at most N lies in T_K. The bound depends on the program e that the recursion theorem gives for this p, and so on N through p; the proof of the theorem shows that no single bound serves all computable sequences.

O.8 Records and physical construction

Chapter 17, question 1: Lemma 17.2

The collision lemma assumes one fixed reader whose entire input is the record $r(w)$. If it also receives a control setting $c(w)$, equal values of r no longer imply equal inputs to that reader. The complete record must include the setting, for example as the pair $(r(w), c(w))$. The capacity theorem then applies to this enlarged record space. A setting that carries target information cannot be omitted from the count.

Chapter 17, question 4: Lemma 17.8

For worlds with a common record, let y be the reader's common answer. The two accuracy bounds give $d(f(w), y) \le \varepsilon$ and $d(f(w'), y) \le \varepsilon$. Symmetry reverses the second distance. The triangle inequality then gives

$$d(f(w), f(w')) \le d(f(w), y) + d(y, f(w')) \le 2\varepsilon.$$

A selected family with pairwise distance strictly greater than 2ε therefore cannot contain a record collision. The non-strict bound just proved would not contradict a distance equal to 2ε; the strict margin is necessary for this particular argument.

Chapter 17, question 6: Theorem 17.15

The question Q says that $\alpha(m) = \alpha(n) = 1$ for some $m < n$. If $Q(\alpha)$ holds with witnesses $m < n$, every run that agrees with α up to $n + 1$ satisfies Q, so Q is open and confirmable by Theorem 17.13. It is not closed: the run of zeros fails Q, and its record of length n is also the record of the run with entries 1 at n and $n + 1$, which satisfies Q. So Q is not refutable.

The map $\Phi(\alpha) = (1, \alpha(0), \alpha(1), \ldots)$ has finite information, since each entry of $\Phi(\alpha)$ is an entry of α, and $Q(\Phi(\alpha))$ holds exactly when α has an entry 1. So the event test reduces to Q, and by Proposition 17.18(3) deciding Q for every run implies LPO. Conversely LPO decides Q: apply it to the binary sequence whose entry at n is 1 exactly when $\alpha(0), \ldots, \alpha(n)$ contain two entries 1. Deciding Q costs exactly LPO.

Chapter 17, question 7: Theorem 17.17

A decision for α gives either that the first entry 1 of α is at an even place, which exhibits an entry 1, or the negation of that statement. The negation is compatible with a first entry 1 at an odd place and gives no place: a search for it could fail only if α were all zero, and turning that impossibility into a place is the binary Markov principle $\mathrm{MP}_{\mathrm{bin}}$, which is what we want to prove. The shifted run moves every entry one place to the right, so its first entry 1 is at an even place exactly when that of α is at an odd place. A positive decision for the shifted run therefore exhibits an entry 1 of α as well, and if both decisions are negative, α is all zero, which contradicts the promise.

Chapter 17, question 8: Theorem 17.21

Consistency asks for $|y - 1| < 1$, $|y - 3/2| < 1/2$ and $|y - 13/8| < 1/4$, that is, $0 < y < 2$, $1 < y < 2$ and $11/8 < y < 15/8$. The consistent values form the open interval $(11/8, 15/8)$. The golden ratio $\varphi = 1.6180\ldots$ lies in it. The largest admissible δ is the distance from φ to the nearer end, $\varphi - 11/8 = 0.2430\ldots$; the other end is at distance $15/8 - \varphi = 0.2569\ldots$.

The rivals lie above φ, so they are consistent exactly when $\varphi^{-K} < 15/8 - \varphi$. Since $\varphi^{-1} = \varphi - 1$ and $\varphi^{-2} = 2 - \varphi$, the first two rivals are $2\varphi - 1 = 2.236\ldots$ and 2, both inconsistent. Next, $\varphi^{-3} = \sqrt{5} - 2 = 0.2360\ldots < 0.2569\ldots$, so $\varphi + \varphi^{-3} = 1.8541\ldots$ is consistent, and so is every later rival, because the powers φ^{-K} decrease. The least K is 3.

Chapter 18, question 1: Theorem 18.3

The identity $V^{\dagger}V = I$ gives the upper-left block of U^2. It also gives $(I - VV^{\dagger})V = 0$ and its adjoint, which make the off-diagonal blocks vanish. Finally $VV^{\dagger}$ is a projection, so the lower-right block is

$$VV^{\dagger} + (I - VV^{\dagger})^2 = VV^{\dagger} + I - VV^{\dagger} = I.$$

The complement is the part of the larger output space outside the embedded input. Including its action is what makes the construction a unitary on the whole displayed space.

Chapter 18, question 3: Lemma 18.7

For an accepted word, the exact entry discrepancy is bounded by word-evaluation error, the accepted rational discrepancy, and target-naming error. These are $e + r + e$, with $e = r/8$, giving $5r/4 \le 2r$. This proves accuracy of the word actually accepted.

To prove that the search terminates, first use density to obtain a word whose exact entries are within e of the target. Its rational discrepancy in absolute complex value is then bounded by the two evaluation errors and that true gap: $e + e + e = 3e < r/2$. The rational certificate criterion makes this candidate pass. The second argument proves that some candidate will be accepted; it does not assume that the search already found one.

Chapter 18, question 5: Theorem 18.10

Set $h = \|H\|$. Since $t \le \delta\hbar/[2(\ell h + 1)]$,

$$\frac{\ell t h}{\hbar} \le \frac{\delta}{2}\frac{\ell h}{\ell h + 1} \le \frac{\delta}{2}.$$

Also $t \le T/(\ell + 1)$ implies $\ell t \le \ell T/(\ell + 1) < T$. For each pulse, its scalar baseline contribution is $t(\hbar/t)\|G\| = \hbar\|G\|$. Summing gives $\hbar c(w)$. The amplitude scale $\hbar/t$ grows when duration shrinks; the integrated baseline does not disappear.

O.9 Conservation and the closing argument

Chapter 19, question 2: Lemma 19.2

With $D = M - P$, expand

$$(I - P)D(I - P) - PDP = D - DP - PD.$$

Substituting $D = M - P$ and using $P^2 = P$ gives $M + P - MP - PM$, the noise block. Apply the event-error bound to the two vectors $(I - P)x$ and Px. They are orthogonal and sum to x, because P is an orthogonal projection. Their squared norms therefore sum to $\|x\|^2$, leaving one factor of e in the noise estimate.

Chapter 19, question 4: Theorem 19.5

The centered total-charge vector is

$$(A - I)\psi \otimes p + \psi \otimes (R - bI)p.$$

Its cross inner product factors as $\langle (A-I)\psi, \psi\rangle\langle p, (R-bI)p\rangle$ and is zero by the source witness's centered mean. The two squared norms are $3/4$ and $\|(R - bI)p\|^2$, since both ψ and p are unit vectors. Their sum is the stated total spread.

The factorization of the cross term uses the initial product $\psi \otimes p$ between source and whole recorder. It makes no factorization assumption about the cells inside p. Internal recorder correlations remain permitted.

Chapter 20, question 1: Theorem 2.7

The target step's injectivity and its separation from the target start give injectivity of the canonical reading. Target generation gives surjectivity by the image predicate used in the solution to Chapter 2, question 3. Initiality then makes the resulting structure-preserving bijection unique. Each conclusion uses its own condition; the existence of the initial map alone supplies neither of the two bijection properties.

Chapter 20, question 3: Theorem 10.8

Faithful storage proves that decoding the stored record recovers the same program expression. Execution after recovery therefore has exactly the same behavior as execution before storage, including failure to halt. A uniform interval name has a stronger specification: for every precision request it must return a valid pair of endpoints. A program that fails to halt on one such request fails that specification even if its storage and recovery are exact.

Chapter 20, question 7: Theorem 20.39

The translation of $\forall x \neg(S(x) = 0)$ is $\forall x(\neg\neg(S(x) = 0) \to \bot)$, expressed by

$$\prod_{x : O}\Big(\big((\mathrm{Id}_O(Sx, 0) \to \mathbf{0}) \to \mathbf{0}\big) \to \mathbf{0}\Big).$$

The valuation gives the identity type the value true and $\mathbf{0}$ the value false. Then $\mathrm{Id} \to \mathbf{0}$ is false, its implication into $\mathbf{0}$ is true, the next implication into $\mathbf{0}$ is false, and the product over O, with $\nu(O)$ true, is false. Theorem 9.8 excludes a closed term. Peano arithmetic proves the sentence, so $\mathscr{E}^{-}$ does not prove the translation of every theorem of Peano arithmetic; the discriminator is needed.

O.10 The comparison cost and its forcing chain

Chapter 21, question 3: Theorem 21.8

By part 1 of the theorem, $\cos\alpha = 5\sin\alpha$ and $\cos\beta = 239\sin\beta$. Put $a = \sin\alpha$ and $b = \sin\beta$. The addition formulas give $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 24a^2$ and $\sin 2\alpha = 2\sin\alpha\cos\alpha = 10a^2$, then $\cos 4\alpha = (24^2 - 10^2)a^4 = 476a^4$ and $\sin 4\alpha = 2 \cdot 10 \cdot 24\, a^4 = 480a^4$. With $\cos(-\beta) = \cos\beta$ and $\sin(-\beta) = -\sin\beta$,

$$\cos(4\alpha - \beta) = 476 \cdot 239\, a^4 b + 480\, a^4 b = 114244\, a^4 b,$$
$$\sin(4\alpha - \beta) = 480 \cdot 239\, a^4 b - 476\, a^4 b = 114244\, a^4 b.$$

(This is $(5 + \mathrm{i})^4(239 - \mathrm{i}) = 114244(1 + \mathrm{i})$.) So with $\theta = 4\alpha - \beta$, $\cos 2\theta = \cos^2\theta - \sin^2\theta = 0$. Part 1 gives $\frac{4}{25} \le \alpha \le \frac{1}{5}$ and $\frac{4}{1195} \le \beta \le \frac{1}{239}$, so $2\theta = 8\alpha - 2\beta$ lies between $\frac{32}{25} - \frac{2}{239} > 1$ and $\frac{8}{5} < 2$. By part 3, $2\theta = \pi_\delta/2$, that is $\pi_\delta = 4(4\alpha - \beta)$.

Chapter 21, question 1: Theorem 21.14

The cost $F(x) = \frac{1}{2}(x-1)^2$ has $F(1) = 0$ and $F(x) - (x-1)^2/2 = 0$, so it is calibrated at 1 with any modulus. At $x = y = 2$ and $z = \frac{1}{2}$ the multiplicative law would read $F(4) + F(1) = 2F(2)^2 + 4F(2)$, that is $\frac{9}{2} = \frac{1}{2} + 2 = \frac{5}{2}$, which is false. Put $C(t) = F(e^t) = \frac{1}{2}(e^t - 1)^2$ and $g = \cosh - 1$. Both are calibrated (the passage from calibration at 1 to calibration at 0 is the estimate in the proof of Theorem 21.14, part 3), so hypothesis (i) of Theorem 21.13 holds, by the triangle inequality at index $2k+1$, and hypothesis (ii) holds for g by Proposition 21.12. What fails is the duplication law for C: at $s = \log 2$, $C(2s) = F(4) = \frac{9}{2}$, while $2C(s)^2 + 4C(s) = 2F(2)^2 + 4F(2) = \frac{5}{2}$. Accordingly $C \neq g$: $C(\log 2) = \frac{1}{2}$, while $g(\log 2) = J(2) = \frac{1}{4}$. Curvature at the reference does not determine a cost; the law does.

Chapter 21, question 2: Proposition 21.16

Write $G = \cosh - 1$, which obeys the duplication law $G(2) = 2G(1)^2 + 4G(1)$, and $\lambda = \frac{2}{3}$. At $s = t = 1$ the left side of the composition law for λG is $\lambda G(2) + \lambda G(0) = \frac{2}{3}\big(2G(1)^2 + 4G(1)\big) = \frac{4}{3}G(1)^2 + \frac{8}{3}G(1)$. The right side is $2\lambda^2 G(1)^2 + 4\lambda G(1) = \frac{8}{9}G(1)^2 + \frac{8}{3}G(1)$. The difference is $\frac{4}{9}G(1)^2$, and $G(1) \geq \frac{1}{2}e^{-1} > 0$, so the law fails. With $\kappa = 3$ the product term becomes $\kappa\lambda^2 G(1)^2 = \frac{4}{3}G(1)^2$, and the two sides agree: λG obeys the law with product coefficient 3, as part 3 of the proposition says.

Chapter 21, question 4: Theorem 21.22

By part 1 there is a natural number c with $C(x) = J(x^c)$, so $C(2) = J(2^c) = \frac{9}{8}$. Put $z = 2^c \geq 1$. Then $J(z) = \frac{9}{8}$ means $z + z^{-1} = \frac{17}{4}$, that is $4z^2 - 17z + 4 = (4z-1)(z-4) = 0$, and $z \geq 1$ gives $z = 4$, so $c = 2$. By part 2 the weights are $\frac{1}{2}$ at 2 and at -2 and 0 everywhere else: the law is the fair step ± 2, and $C(x) = J(x^2)$.

Chapter 21, question 5: Theorem 21.26

$W_5 = W_4 W_3 =$ dcddcdcd dcddc, so the history begins dcddcdcddcddc. With $\varphi^{-3} = \sqrt{5} - 2 = 0.2360679\ldots$ the balances and discrepancies $D_N = N\varphi^{-3} - b_N$, to three places, are as follows; the letter in row N is posting N, made after b_N is read.

N	letter	b_N	D_N	N	letter	b_N	D_N
0	d	0	0	7	d	1	0.652
1	c	1	−0.764	8	d	2	−0.111
2	d	0	0.472	9	c	3	−0.875
3	d	1	−0.292	10	d	2	0.361
4	c	2	−1.056	11	d	3	−0.403
5	d	1	0.180	12	c	4	−1.167
6	c	2	−0.584	13		3	0.069

All values lie strictly between $-2/\varphi = -1.236\ldots$ and $2/\varphi^2 = 0.764\ldots$; the closest approaches are $D_7 = 0.652$ to the upper bound and $D_{12} = -1.167$ to the lower one. The rule of part (e) says that posting N is a debit exactly when $D_N > -2\varphi^{-3} = -0.472\ldots$. The postings 1, 4, 6, 9 and 12 have $D_N \le -0.584$ and are credits; all others have $D_N \ge -0.403$ and are debits, as the table shows. Every margin exceeds 0.06, far more than the rounding error; exactly, each comparison is the sign test of an element of $\mathbb{Z}[\varphi]$, for instance $D_{12} + 2/\varphi = (12\varphi^{-3} - 4) + (2\varphi - 2) = 26\varphi - 42$, which passes the sign test because $s = -58$, $t = 26$ and $58^2 = 3364 < 3380 = 5 \cdot 26^2$.

Chapter 21, question 6: Theorem 21.29

Write $G = \cosh - 1$. The coded cost of the constant c is $C_c(t) = G(t) + c\big(\max(0, t-1) + \max(0, -t-1)\big)$, so $C_c(0) = 0$, $C_c(2) = G(2) + c$, $C_c(3) = G(3) + 2c$, $C_c(4) = G(4) + 3c$ and $C_c(6) = G(6) + 5c$. Using the duplication law $G(2s) = 2G(s)^2 + 4G(s)$, the defect $C(s+t) + C(s-t) - 2C(s)C(t) - 2C(s) - 2C(t)$ is

$$-c\big(1 + 4G(2) + 2c\big) \text{ at } s = t = 2, \qquad -c\big(3 + 8G(3) + 8c\big) \text{ at } s = t = 3.$$

If both vanish, four times the first minus the second gives $c\big(8G(3) - 16G(2) - 1\big) = 0$. With $u = \cosh 1$, the addition formulas for cosh give $\cosh 2 = 2u^2 - 1$ and $\cosh 3 = 4u^3 - 3u$, so $8G(3) - 16G(2) - 1 = 32u^3 - 32u^2 - 24u + 23 = 23 + 8u(2u-3)(2u+1)$. Since $e \ge 1 + 1 + \frac{1}{2} + \frac{1}{6} = \frac{8}{3}$ and $e \le 3$, $u = \frac{1}{2}(e + e^{-1}) \ge \frac{1}{2}(\frac{8}{3} + \frac{1}{3}) = \frac{3}{2}$, so the factor is at least 23, hence invertible, and $c = 0$. Conversely $C_0 = \cosh - 1$ obeys the law. Part 2 of the theorem gives more: for every function H, the coded cost obeys the law exactly when $H \equiv 0$; its proof uses the forcing theorem, which reduces the law to the single equation $C_H = \cosh - 1$.

Chapter 22, question 1: Theorem 22.16

By the computation in the proof, $J(1+r^2) = r^4/(2(1+r^2))$ for every point r. Hence $J(1+r^2) = 0$ exactly when $r^4 = 0$, that is, when $r = 0$ (a point apart from 0 has a positive fourth power); so $J(1 + r^2) \ne 0$ exactly when $r \ne 0$. And $J(1 + r^2) > 0$ exactly when $r^4 > 0$, which holds exactly when $r \# 0$: a positive power is apart from 0, so its factor r is. Now assume (ii) and let $r \ne 0$. Then $J(1 + r^2) \ne 0$, so (ii) gives $J(1 + r^2) > 0$, hence $r \# 0$, that is, $r < 0$ or $r > 0$. So every point $r \ne 0$ satisfies $r < 0 \vee r > 0$. With $y = 0$, this is the row "$x \ne y \to x < y \vee y < x$" of Table 12.2, whose price is $\mathrm{MP}_{\mathrm{bin}}$. No jump point is used; the jump point is hidden inside the proof of that row.

Chapter 22, question 2: Theorem 22.20

Put $F_{-1} = 1$, so that $F_1 = F_0 + F_{-1}$. Then $\ell(0) = (0, 1) = (F_0, F_{-1})$, and if $\ell(n) = (F_n, F_{n-1})$ then $\ell(n+1) = A(F_n, F_{n-1}) = (F_n + F_{n-1}, F_n) = (F_{n+1}, F_n)$. Hence $\nu(\ell(n)) = F_n\varphi + F_{n-1}$, which is φ^n by Proposition 22.37(d) for $n \ge 1$ and equals $1 = \varphi^0$ for $n = 0$. The hypotheses of Theorem 22.20(b) for $s_* = 0$: $r(n) = \varphi^n > 0$ because $\varphi > 0$; A is a homomorphism of $\mathbb{Z} \times \mathbb{Z}$ and ν is additive; $\nu(\ell(n)) = r(n)$ was just shown, and $\ell(n+1) = A(\ell(n))$ holds by definition; the step preserves positive values because $\nu(A(a, b)) = (a + b)\varphi + a = \varphi(a\varphi + b)$ by $\varphi^2 = \varphi + 1$; and the levels of 0 are $1, \varphi, \varphi^2$ with $\varphi^2 = \varphi + 1$, so $(a, b) = (0, 1)$ posts additively, and the join ratio φ satisfies $\varphi > 1$, so $J(\varphi) > 0$ (Proposition 22.15). The theorem then returns $r(n+1) = \varphi\, r(n)$, which is visibly true in the model.

Chapter 22, question 3: Theorem 22.31

Let $x, y \geq 1$ with $J(x) = J(y)$, so that $w = (x - y)(xy - 1) = 0$. Suppose $x \# y$, say $x > y$ (the case $y > x$ is symmetric). Then $x > y \geq 1$, so $xy \geq x > 1$ and $xy - 1 > 0$. A product of two points apart from 0 is apart from 0, against $w = 0$. Hence $\neg(x \# y)$, which is $x = y$. No principle is used. On $[1, \infty)$ the branch $xy = 1$ adds nothing: $x, y \geq 1$ and $xy = 1$ force $x = y = 1$. The price of the general statement comes from pairs on opposite sides of 1, where $xy = 1$ and $x = y$ are genuinely different conclusions. In the proof of the theorem the pair $x = \exp((u - v)/2)$, $y = \exp(-(u + v)/2)$ has $x/y = \exp(u)$ and $xy = \exp(-v)$; choosing the branch $x = y$ or $xy = 1$ is choosing which of u and v vanishes, and that is the domain law, whose price is LLPO.

Chapter 22, question 5: Theorem 22.42

Fix x, a natural number $T \geq |x|$ and k, and take j for the uniform derivative 0 on $[-T, T]$: $|h(t) - h(u)| \leq |t - u|/(k + 1)$ whenever $|t|, |u| \leq T$ and $|t - u| \leq 1/(j + 1)$. With $N = (T + 1)(j + 1)$ and $t_i = ix/N$,

$$|h(x) - h(0)| \leq \sum_{i<N} |h(t_{i+1}) - h(t_i)| \leq \frac{|x|}{k + 1}.$$

This holds for every k, so $h(x) = h(0)$: h is constant. For the second part, Theorem 22.42(b) gives $g(x) = \sigma \exp(-x/\sigma)$ and $g(-x) = \sigma \exp(x/\sigma)$, so $g(x)g(-x) = \sigma^2 \exp(0) = \sigma^2$ by $\exp(a + b) = \exp(a) \exp(b)$ (Proposition 12.88).

Chapter 22, question 4: Theorem 22.44

Section 22.9 gives $3.1415925 < \pi < 3.1415927$ and $0.4812 < \log \varphi < 0.4813$, so $1.9248 < 4 \log \varphi < 1.9252$. Hence

$$\frac{\pi}{4 \log \varphi} > \frac{3.1415925}{1.9252} > 1.6318 > 1.631, \qquad \frac{\pi}{4 \log \varphi} < \frac{3.1415927}{1.9248} < 1.6322 < 1.633.$$

Each comparison is a comparison of integers after clearing denominators, for instance, after multiplying by 10^8, $314159250 > 314154136$. Since $K < 3/2$ by Theorem 22.44(b), the ratio exceeds K by more than $1.631 - 1.5 = 0.131 > 1/10$. The check's second test compares the ratio with 1, from which it differs by more than 0.63, far above the tolerance 10^{-6}.

Chapter 22, question 6: Proposition 22.49

Let w have period $p > 0$ and reach every point of $\{0, 1\}^D$. By induction on q, $w(r + qp) = w(r)$, so every value of w is one of $w(0), \ldots, w(p - 1)$. Two points of $\{0, 1\}^D$ are equal exactly when they agree in each of the D coordinates, a finite check; so for each point x the least $r < p$ with $w(r) = x$ is found by testing $r = 0, 1, \ldots, p - 1$ in turn. No choice is made: the search is bounded and each test is decided. Since $w(r_x) = x$, different points have different first visits, so $x \mapsto r_x$ is an injective map from the 2^D points to $\{0, \ldots, p - 1\}$, and $2^D \leq p$ by the count in the proof of Proposition 22.28.

For the second part, define lists G_D of points by $G_1 = (0, 1)$ and let G_{D+1} be G_D with 0 put in front of each point, followed by G_D in reverse order with 1 put in front of each point.

By induction, G_D lists each of the 2^D points once, consecutive points differ in exactly one coordinate (inside each half by induction; at the junction the two points differ only in the new first coordinate), and its last point $10\cdots0$ differs from its first point $00\cdots0$ in the first coordinate only. The sequence $w(n) = G_D(n \bmod 2^D)$ has period 2^D, reaches every point, and changes one coordinate per step. For $D = 3$ it is $000, 001, 011, 010, 110, 111, 101, 100$; the cycle in the proof of Proposition 22.49(a) is the same construction with the coordinates read in the opposite order.

Chapter 23, question 2: Lemma 23.12

With $|x| = |p| = 1$ and $\langle x, p\rangle = c$, $|2tx + kp|^2 = 4t^2 + 4tkc + k^2$. Put $B = A - m$, so $k = B - 2tc$. Then

$$4t^2 + 4tc(B - 2tc) + (B - 2tc)^2 = 4t^2 + 4tcB - 8t^2c^2 + B^2 - 4tcB + 4t^2c^2 = 4t^2(1 - c^2) + B^2.$$

Since $4Am = 4t^2(1 + c)(1 - c) = 4t^2(1 - c^2)$, this is $(A - m)^2 + 4Am = (A + m)^2 = Q^2$.

For $D = 1$, $p = (1, 0)$, $x = (0, 1)$ and $t = \frac{1}{2}$: $c = 0$, $m = 1$, $A = \frac{1}{4}$, $Q = \frac{5}{4}$ and $k = \frac{1}{4} - 0 - 1 = -\frac{3}{4}$. Hence

$$H(\tfrac{1}{2}, x) = \frac{(0, 1) - \frac{3}{4}(1, 0)}{5/4} = \Big(-\frac{3}{5}, \frac{4}{5}\Big),$$

and $\frac{9}{25} + \frac{16}{25} = 1$, so it lies on the circle. Its squared distance from p is $(\frac{8}{5})^2 + (\frac{4}{5})^2 = \frac{16}{5}$, which agrees with $4m/Q = \frac{16}{5}$ and exceeds $|x - p|^2 = 2 = 2m$. In the sup distance, $d(H(\frac{1}{2}, x), p) = \frac{8}{5} > 1 = d(x, p)$.

Chapter 23, question 4: Lemma 23.16

Near a pole $u = |x'|^2$ is small. If $u < \kappa$, then $\phi(x) = u/\kappa < 1$ and $c^2 = 1 - u > 1 - \kappa \geq 0$, so c is apart from 0, and the first coordinate $(1 - \phi)c/\sqrt{n}$ of $H(1, x)$ is apart from 0: such a point stops short of the equator. The proof never needs those points. The cycle z has its simplices r-apart from both poles, and with $\kappa = r^2$ item 4 of the proof gives $u \geq \kappa$ at every value of z; there $\phi = 1$ and $H(1, x) = (0, x')/\sqrt{u}$ lies on the equator. The cut-off is there for a different reason: it makes $n \geq \kappa/8$ on the whole sphere, so that H is one uniformly continuous homotopy of all of S^{D+1}. Then the prism of Lemma 23.11 is an operator on all maps $\Delta^k \to S^{D+1}$, defined term by term, and it respects the moves of Definition 23.4; this is what makes $P(\partial z) \sim \emptyset$ for a cycle z.

Chapter 23, question 1: Lemma 23.17

Let $f : S^1 \to S^0$ be uniformly continuous with an inverse modulus ω'. The arc $u \mapsto z(u)$, $u \in [-1, 0]$, of Lemma 23.33 runs from $z(-1) = (-1, 0)$ to $z(0) = (1, 0)$ and is 32-Lipschitz, so $s \mapsto f(z(s - 1))$ is a uniformly continuous map from $[0, 1]$ to S^0. By Lemma 23.17 it is constant, so $f(1, 0) = f(-1, 0)$. Then $d(f(1, 0), f(-1, 0)) = 0 < \omega'(1)$, and the inverse modulus gives $d((1, 0), (-1, 0)) < 1$; but that distance is 2. So no native embedding into S^0 exists.

No sign of a point is decided. For v with $v^2 = 1$ we have $(v-1)(v+1) = 0$ and $(v+1)-(v-1) = 2$, so $|v-1| + |v+1| \geq 2 > 1$, and cotransitivity gives $|v-1| > 1/2$ or $|v+1| > 1/2$. The factor that is bounded away from 0 has an inverse, so the other factor is 0: $v = 1$ or $v = -1$. This disjunction is proved for each given v; it is not a decision procedure on the line, and it is used only inside the proof of an equation. For two points v, w of S^0 with $|v - w| < 1$, the four combinations of signs leave only $v = w$, since mixed signs give $|v - w| = 2$.

Chapter 23, question 3: Lemma 23.26

Each chord lies in one chart, and its swept angle is the difference of the diamond angles of its end points in that chart. Using $\theta_j(p) = j + y'/(x' + |y'|)$ with $(x', y') = \rho^j(p)$:

- $(1, 0) \to (\frac{3}{5}, \frac{4}{5})$ in chart 0: from 0 to $\frac{4/5}{7/5} = \frac{4}{7}$; swept $\frac{4}{7}$.
- $(\frac{3}{5}, \frac{4}{5}) \to (0, 1)$ in chart 1, where $\rho(p) = (y, -x)$: from $1 - \frac{3}{7} = \frac{4}{7}$ to 1; swept $\frac{3}{7}$.
- $(0, 1) \to (-\frac{3}{5}, \frac{4}{5})$ in chart 1: from 1 to $1 + \frac{3}{7}$; swept $\frac{3}{7}$.
- $(-\frac{3}{5}, \frac{4}{5}) \to (-1, 0)$ in chart 2, where $\rho^2(p) = (-x, -y)$: from $2 - \frac{4}{7}$ to 2; swept $\frac{4}{7}$.
- $(-1, 0) \to (-\frac{3}{5}, -\frac{4}{5})$ in chart 2: from 2 to $2 + \frac{4}{7}$; swept $\frac{4}{7}$.
- $(-\frac{3}{5}, -\frac{4}{5}) \to (0, -1)$ in chart 3, where $\rho^3(p) = (-y, x)$: from $3 - \frac{3}{7}$ to 3; swept $\frac{3}{7}$.
- $(0, -1) \to (\frac{3}{5}, -\frac{4}{5})$ in chart 3: from 3 to $3 + \frac{3}{7}$; swept $\frac{3}{7}$.
- $(\frac{3}{5}, -\frac{4}{5}) \to (1, 0)$ in chart 0: from $-\frac{4}{7}$ to 0; swept $\frac{4}{7}$.

The total is $4 \cdot \frac{4}{7} + 4 \cdot \frac{3}{7} = 4$: one full turn. At the shared end points the readings of neighbouring charts agree or differ by 4, as Lemma 23.26 says; for instance $(\frac{3}{5}, \frac{4}{5})$ reads $\frac{4}{7}$ in chart 0 and in chart 1.

Chapter 23, question 5: Theorem 23.40

Since $i(ix) = (-x_0, -x_1) = -x$, the colouring satisfies $q(-x) \equiv q(ix) + 1 \equiv q(x) + 2 \pmod 4$. If $q(x)$ is 0 or 1 modulo 4, then $q(-x)$ is 2 or 3, and conversely; so $b(-x) \neq b(x)$, and b is an antipodal two-colouring. Residues of integers modulo 4 are computed, so b is a map.

For the identity, let a, b, c, d be $G(x), G(ix), G(-x), G(-ix)$. Then $q(x) = k_0(3a + 2b + c)$, and since $i(ix) = -x$ and $i(-x) = -ix$, $q(ix) = k_0(3b + 2c + d)$. Hence

$$q(ix) - q(x) - 1 = k_0(b + c + d - 3a) - 1 = k_0(a + b + c + d) - 4k_0 a - 1 = -4k_0 a,$$

using $k_0(a + b + c + d) = 1$, which is the statement that the cycle of four quarter chords at x has label 1.

Chapter 23, question 6: Theorem 23.40

Let d be a WLPO function. The sequence $\bar{0}$ of zeros is identically 0, so $d(\bar{0}) = 1$. Continuity at $\bar{0}$ gives n such that $d(\beta) = 1$ for every β agreeing with $\bar{0}$ below n. The sequence β whose only 1 is at place n agrees with $\bar{0}$ below n, so $d(\beta) = 1$; but β is not identically 0, so $d(\beta) = 0$. This is a contradiction, so continuity on Cantor space refutes the WLPO function.

Classically both principles are false. For BU_2, let $b(x) = 1$ when $x_1 > 0$, or $x_1 = 0$ and $x_0 > 0$, and $b(x) = 0$ otherwise. For a point of the circle, if $x_1 \neq 0$ exactly one of x_1 and $-x_1$ is positive, and if $x_1 = 0$ then $x_0 = \pm 1$ and exactly one of x_0 and $-x_0$ is positive; so $b(-x) \neq b(x)$ for every x. For continuity on Cantor space, excluded middle decides for each α whether it is identically 0, and a classical map d records the answer; by the first paragraph it is not continuous at $\bar{0}$.

Chapter 24, question 2: Proposition 24.3

If $D \leq D'$ and $C \leq C'$ then $2D \leq 2D'$ and $C \leq C'$, and conversely; so $(D, C) \mapsto (2D, C)$ preserves and reflects the order. It is not onto, since $(1, 0)$ is not an image, so it is not an order bijection, and the proposition does not apply to it. If $(kD, C/k) \in \mathbb{N} \times \mathbb{N}$ for all records, the record $(1, 0)$ gives $k \in \mathbb{N}$ and the record $(0, 1)$ gives $1/k \in \mathbb{N}$, so $k = 1$. Theorem 24.14(f) concerns real coherent records: on the diagonal it sends real (D, C) to $(kD, C/k)$, keeping the determinant DC and the cone. It is a symmetry of the real record cone, not of integer counts, and integer counts single out the frame in which they are counts.

Chapter 24, question 1: Theorem 24.5

The legs dc, ddc and dcd have counts $(1, 1)$, $(2, 1)$ and $(2, 1)$, so their proper times are 2, $2\sqrt{2}$ and $2\sqrt{2}$, with sum $2 + 4\sqrt{2} \approx 7.657 < 2\sqrt{15} \approx 7.746$. The squared deficit is $60 - (2 + 4\sqrt{2})^2 = 24 - 16\sqrt{2} \approx 1.373$. In Proposition 24.6 the ratios are $r_1 = 1$ and $r_2 = r_3 = 2$. The pair $(1, 2)$ gives $2 \cdot 2 \cdot 2\sqrt{2} \cdot J(1/\sqrt{2})$, and $J(1/\sqrt{2}) = J(\sqrt{2}) = \frac{1}{2}(\sqrt{2} + 1/\sqrt{2}) - 1 = \frac{3}{4}\sqrt{2} - 1$, so the term is $8\sqrt{2}(\frac{3}{4}\sqrt{2} - 1) = 12 - 8\sqrt{2}$; the pair $(1, 3)$ gives the same, and the pair $(2, 3)$ gives $J(1) = 0$. The sum is $24 - 16\sqrt{2}$, as required. In the integer form of the proof, $4(\sqrt{p_{12}} - \sqrt{p_{21}})^2 = 4(1 - \sqrt{2})^2 = 12 - 8\sqrt{2}$ with $p_{12} = 1$ and $p_{21} = 2$. The legs ddc and dcd post in the same proportion, $D_2C_3 = D_3C_2 = 2$; merging them loses no proper time, since $\tau(ddcdcd) = 2\sqrt{4 \cdot 2} = 4\sqrt{2} = 2\sqrt{2} + 2\sqrt{2}$, by Theorem 24.5(b).

Chapter 24, question 3: Theorem 24.12

For $(5, (2, -1))$: $|2| + |-1| = 3 \leq 5$ and $5 - (2 - 1) = 4$ is even, so it is reachable. Post dd to account 1 and c to account 2, then the pair dc to account 1: the records are $(3, 1)$ and $(0, 1)$, with 5 posts and potentials $(2, -1)$. For $(4, (2, -1))$: $4 - 1 = 3$ is odd, so it is not reachable; the number of posts at each account has the parity of its potential. For $(6, (3, 3))$: $6 \leq 6$ and $6 - 6 = 0$ is even; post ddd to each account. This point lies on the boundary of the taxicab cone: every post moves a potential away from 0.

Chapter 24, question 4: Theorem 24.17

$P_4(n) = n + 4n(n-1)/2 = 2n^2 - n$, so $P_4(2) = 6$, $P_4(3) = 15$, $P_4(4) = 28$ and $P_4(6) = 66$. Multiplication would require $P_4(4) = P_4(2)^2 = 36$ and $P_4(6) = P_4(2)P_4(3) = 90$; both fail. In the proof of the theorem, the condition at $2 \cdot 2$ is $k^2 = 2k$, which fails at $k = 4$ because $16 \neq 8$.

Chapter 24, question 5: Theorem 24.31

By the record law, agreement at τ_0 gives $x_j \in \mathbb{Z}$, and agreement at $\frac{8}{3}\tau_0$ gives $3 \mid 8x_j$, hence $3 \mid x_j$ because $\gcd(3, 8) = 1$; likewise $\frac{8}{5}$ and $\frac{8}{7}$ give $5 \mid x_j$ and $7 \mid x_j$, so $105 \mid x_j$. Conversely 105 times each of the four times is an integer. The other four times, 8, 4, 2 and $\frac{4}{3}$, add nothing: for $x_j \in 105\mathbb{Z}$ the products $8x_j$, $4x_j$ and $2x_j$ are integers, and $\frac{4}{3}x_j$ is an integer because $3 \mid x_j$. The identity $1 + \frac{8}{3} + \frac{8}{7} - 3 \cdot \frac{8}{5} = \frac{1}{105}$ uses only the four times, so they already generate $\frac{1}{105}\mathbb{Z}$.

Chapter 24, question 6: Theorem 24.43

The closing coder of $\alpha = (0, 1, 0, 0, \ldots)$ posts d, then dc for $\alpha(0) = 0$, cd for $\alpha(1) = 1$, dc for $\alpha(2) = 0$ and begins dc for $\alpha(3) = 0$: the first eight posts are $ddcddcd$, with balances $1, 2, 1, 0, 1, 2, 1, 2$ after one to eight posts. It closes after four posts, inside the pair coding the entry $\alpha(1) = 1$, as the proof of Theorem 24.39 predicts.

The recorder with states $\{0, 1, 2\}$ is in state 2 exactly when the last bit read is 1. On the run of zeros it stays in state 1 from the first entry on and never answers, so it does not always answer. The proof of Theorem 24.43(d), with $s = 2$, finds the states $0, 1, 1, 1$ after $0, 1, 2, 3$ entries, a first repetition at stages $i = 1 < j = 2$, and the run that copies the zeros up to place 2 and then repeats place 1: the run of zeros itself, which never answers. For a recorder that always answers this would be a contradiction; here it only shows that the hypothesis fails. On a run whose first 1 is at place p the recorder answers at stage $p + 1$, which can be arbitrarily late, so no deadline exists, consistently with the theorem.

Chapter 24, question 7: Theorem 24.53

Here $f = f_{1,2}$ in the notation of the proof of Theorem 24.53, so $Pf = f_{3/\sqrt{2},\,1/\sqrt{2}}$, with entries $3/\sqrt{2}$, $1/\sqrt{2}$, $-3/\sqrt{2}$, $-1/\sqrt{2}$ at positions $0, 2, 4, 6$. Then $\sum_t f(t)^2 = 1 + 4 + 1 + 4 = 10$ and $\sum_t (Pf)(t)^2 = \frac{9}{2} + \frac{1}{2} + \frac{9}{2} + \frac{1}{2} = 10$, while $\sum_t |f(t)| = 6$ and $\sum_t |(Pf)(t)| = 8/\sqrt{2} = 4\sqrt{2} \approx 5.66$: the magnitude reads see the phase. For g, $\|g\|^2 = 4$ and $\kappa(g, \{t\}) = g(t)^2 + g(t)^2\big(g(t+1)^2 - g(t-1)^2\big)/8$ gives $\frac{9}{8}$, $\frac{7}{8}$, 0, 0, $\frac{9}{8}$, $\frac{7}{8}$, 0, 0 at positions $0, \ldots, 7$, with total $4 = \|g\|^2$. Next $(Pg)(t) = \big(g(t) + g(t+2)\big)/\sqrt{2}$ gives $Pg = \frac{1}{\sqrt{2}}(1, 1, -1, -1, -1, -1, 1, 1)$, with $\|Pg\|^2 = 4$. Every position of Pg has squared amplitude $\frac{1}{2}$ and so do both its neighbours, so $\kappa(Pg, \{t\}) = \frac{1}{2}$ everywhere and the total is again 4, as Proposition 24.55 requires. Positions 0 and 1 of g both carry amplitude 1, yet κ reads $\frac{9}{8}$ at the first and $\frac{7}{8}$ at the second, so no single φ gives $\kappa(g, \{t\}) = \varphi(g(t))$.

O.11 Classical physics on the line

Chapter 25, question 1: Theorem 25.9

The state has $I = u^2|x|^2 = 1$ and $|v|^2 = \frac{1}{2}$, so $E = \frac{1}{4} - 1 = -\frac{3}{4}$. Further $L = x \times v = (0, -\frac{1}{2}, \frac{1}{2})$, $|L|^2 = \frac{1}{2}$, and $v \times L = (\frac{1}{2}, 0, 0)$, so $A = v \times L - m\,x/|x| = (-\frac{1}{2}, 0, 0)$ and $|A|^2 = \frac{1}{4}$. Indeed $2E|L|^2 + m^2 = -\frac{3}{4} + 1 = \frac{1}{4}$.

By Theorem 25.9, $A \cdot x + mr = |L|^2$ on the orbit. Here $A \cdot x = -|A|\,r$ at the present point, since A points away from the body, so $r(m - |A|) = |L|^2$ and $r = \frac{1}{2}/\frac{1}{2} = 1$, the largest value $|L|^2/(m - |A|)$ that the conic allows: the body is at its aphelion. The perihelion lies on the other side, at $r = |L|^2/(m + |A|) = \frac{1}{2}/\frac{3}{2} = \frac{1}{3}$. The eccentricity is $|A|/m = \frac{1}{2}$, the semi-major axis is $\frac{1}{2}(1 + \frac{1}{3}) = \frac{2}{3} = m/(-2E)$, and Kepler's third law $T^2(-2E)^3 = (2\pi_\delta)^2 m^2$ gives $T^2 \cdot \frac{27}{8} = 4\pi_\delta^2$, so $T = 2\pi_\delta(\frac{2}{3})^{3/2} = \frac{4}{9}\sqrt{6}\,\pi_\delta \approx 3.420$. Every number is computed from ratios, so nothing is decided.

Chapter 25, question 4: Theorem 25.42

With $x = m/p = 2.66 \times 10^{-8}$ the first-order term is $6\pi_\delta x \approx 5.014 \times 10^{-7}$ radian per revolution. Over 415.2 revolutions this is 2.082×10^{-4} radian, and with 1 radian $= 206265''$ it is about $42.9''$ per century.

The second-order term is $\frac{3}{2}\pi_\delta(18 + e^2)x^2 \approx 6.0 \times 10^{-14}$ radian per revolution, or about $5 \times 10^{-6}{}''$ per century, and the remainder is at most $160\pi_\delta x^3 \approx 9.5 \times 10^{-21}$ radian per revolution, below $10^{-12}{}''$ per century. The theorem therefore pins the relativistic advance of such an orbit to $42.9''$ per century with every neglected part bounded explicitly. Reading Mercury's orbit as a window orbit of the proper-time field, and m/p, e and the revolution count as measured values, are identifications and observations; the theorem supplies only the bounds. The observed advance also contains the much larger Newtonian contributions of the other planets, which the field of this chapter does not include.

Chapter 25, question 2: Theorem 25.54

For $F = x_0^2 + x_1^2$,

$$\mathscr{L}_p F = 2x_0\,x_1 + 2x_1\,(-x_0 + x_0^2) = 2x_0^2 x_1.$$

The largest exponent of one variable in a term is $E = 2$, so the grid is $\{0, 1, 2\}^2$. At the grid point $(1, 1)$ the value is $2 \neq 0$, which witnesses that F is not a first integral. For G, $\partial G/\partial x_0 = x_0 - x_0^2$ and $\partial G/\partial x_1 = x_1$, so

$$\mathscr{L}_p G = (x_0 - x_0^2)\,x_1 + x_1\,(-x_0 + x_0^2) = 0$$

as a polynomial. It vanishes on every grid point, and the theorem makes G a first integral. It is the energy $\frac{1}{2}x_1^2 + V(x_0)$ with the cubic potential $V(x_0) = \frac{1}{2}x_0^2 - \frac{1}{3}x_0^3$.

Chapter 25, question 3: Theorem 25.56

The field has one term, x^2, so $t_p = 1$, $D = 2$ and $c = 1$. Then $X = 3$, $M = 3^2 + 1 = 10$, $s = \mathrm{bl}(10) + 3 = 4 + 3 = 7$, and

$$k = \mathrm{bl}(1 \cdot 2 \cdot 1 \cdot 10 \cdot 2^7) = \mathrm{bl}(2560) = 12,$$

since $2^{11} \le 2560 < 2^{12}$. The theorem certifies the state at time 2^{-12}, where the solution is $3/(1 - 3 \cdot 2^{-12}) = 12288/4093 \approx 3.0022$. That is $1/1365$ of the way to the blow-up time $\frac{1}{3}$, a cautious but explicit start.

No passing run reaches time $\frac{1}{3}$. Such a run, of a scaled field $2^{-j}p$ to an end time a with $a2^{-j} \ge \frac{1}{3}$, gives by Theorem 25.50 a motion of p on $[0, a2^{-j}]$ from 3, and a motion on a bounded interval is bounded by a natural number B (the first step of the proof of Theorem 25.3). For $t < \frac{1}{3}$ the curve $3/(1-3s)$ is a motion on $[0, t]$, so by Theorem 25.3 the two agree there; but $3/(1-3t) > B$ once $t > \frac{1}{3} - 1/B$. So the certificate fails at every level for every such run.

Chapter 25, question 5: Theorem 25.62

Here $\alpha = J_1 - J_0 = 2$ and $\beta = J_2 - J_0 = 3$, both positive, so (b) applies with $\min(\alpha, \beta) = 2$ and $\max(\alpha, \beta) = 3$. At the start $L_1^2 + L_2^2 = \frac{2}{25}$, so along the motion $2(L_1^2 + L_2^2) \le 3 \cdot \frac{2}{25}$, that is $L_1^2 + L_2^2 \le \frac{3}{25} = 0.12$.

The energy–Casimir form keeps $2L_1^2 + 3L_2^2 = \frac{2}{25} + \frac{3}{25} = \frac{1}{5}$. On that ellipse $L_1^2 + L_2^2$ lies between $\frac{1}{15}$ (all of it in L_2) and $\frac{1}{10}$ (all of it in L_1), so the exact bound is 0.1; the theorem's 0.12 is the price of replacing the ellipse by two circles.

To keep $L_1^2 + L_2^2 \le 10^{-2}$ it suffices, by the second part of (b), that $3(L_1^2 + L_2^2) \le 2 \cdot 10^{-2}$ at the start, that is $L_1^2 + L_2^2 \le \frac{2}{3} \cdot 10^{-2}$.

For $J = (2, 1, 3)$, $\alpha = -1$ and $\beta = 1$: the minimum is negative, the inequality of (b) bounds nothing, and $Q = -L_1^2 + L_2^2$ is indefinite. Here $J_1 < J_0 < J_2$, so axis 0 is the middle axis. Euler's field is unchanged when the indices are shifted cyclically, and the shift that moves axis 0 to position 1 turns the moments into $(3, 2, 1)$; Theorem 25.64 then shows that no bound of this kind exists.

Chapter 25, question 6: Theorem 25.64

The moments $(J_1 + v^2, J_1, J_1 - u^2) = (3, 2, 1)$ give $J_1 = 2$ and $u^2 = v^2 = 1$; take $u = v = 1$ and $h = \sqrt{2}$, so that $h^2 = u^2 + v^2$. Then $C = h^2 r^2 = 2r^2 = 200$ gives $r = 10$, and $\lambda = uvhr = 10\sqrt{2} \approx 14.14$. As Proposition 25.65 says, $\lambda^2 = 200 = (J_0 - J_1)(J_1 - J_2)\,C$.

With $\tau = 1$ the start satisfies $\sqrt{L_0^2 + L_2^2} \le 2e^{-10\sqrt{2}}\sqrt{C} \approx 1.44 \times 10^{-6} \cdot \sqrt{200}$: relative to $|L|$, the body starts within about 1.4×10^{-6} of the middle axis, at $L(0) \approx (-1.44 \times 10^{-5},\ 14.142,\ 1.44 \times 10^{-5})$. At time 1 it is at $L = (-10, 0, 10)$, perpendicular to the middle axis, having turned through it in one time unit.

In Theorem 25.64(b) take $\mu = \sqrt{(J_0 - J_1)(J_1 - J_2)} = 1$. While $L_1 \ge \ell_1$, W grows at least like $e^{2\mu\ell_1 t}$. At the start $L_1 \approx \sqrt{C} = 10\sqrt{2}$, so the rate $2\mu\ell_1$ is 2λ there, and along the separatrix

$W = uvr^2 \operatorname{sech}^2\phi = 100\operatorname{sech}^2\phi$ indeed grows like $e^{2\lambda t}$ while $\phi < 0$. The separatrix does not stay in $L_1 \geq \ell_1$ for any $\ell_1 > 0$ up to time 1, since $L_1(1) = 0$; this is the escape the theorem forces.

Chapter 26, question 4: Theorem 26.6

Convexity. Lemma 26.4, item 1, at a point m in place of ℓ says $-\eta(a) \geq a + am - e^m$ for every $a \geq 0$. Let $a, b \geq 0$, let $0 \leq t \leq 1$, and put $c = ta + (1-t)b$. If $c > 0$, take $m = \log c$ and add the inequality at a with weight t to the one at b with weight $1 - t$:

$$t\big(-\eta(a)\big) + (1-t)\big(-\eta(b)\big) \geq c + cm - e^m = cm = -\eta(c).$$

Without deciding whether $c > 0$, apply this to $a + \varepsilon$ and $b + \varepsilon$ for every point $\varepsilon > 0$, where the combination is $c + \varepsilon > 0$. Both sides are continuous in ε, so Lemma 26.2, applied to their difference, gives the inequality at $\varepsilon = 0$. Hence $-\eta$ is convex on $[0, \infty)$. The other terms of $g(\ell, q) = e^\ell - q - \eta(q) - q\ell$ are affine in q, so $g(\ell, \cdot)$ is convex.

The alternative gives LPO. Suppose that for all data of Theorem 26.5, $F(q) = -y$ or $F(q) > -y$. Let a be a point and $A \geq |a|$ a natural number, and take the two-level system $n = 2$, $E_1 = E_2 = 0$, $\beta = 0$, $Z = 2$, $y = \log 2$ and $q = (\frac{1}{2} + c, \frac{1}{2} - c)$ with $c = a/(2A+1)$. By Theorem 26.5, item 2, $F(q) = -y$ exactly when $2q_1 = 2q_2 = 1$, that is $c = 0$, that is $a = 0$; by item 3, $F(q) > -y$ exactly when $2q_1 \# 1$ or $2q_2 \# 1$, that is $2c \# 0$, that is $a \# 0$. So the alternative gives $a = 0 \vee a \# 0$ for every point a. Since $a \# 0$ means $a < 0$ or $a > 0$, this is the sign trichotomy on the line, which is LPO by Corollary 12.18.

Chapter 26, question 1: Theorem 26.15

The predecessor set of a configuration j is $\{i : K_{ij} \neq 0\}$. Column 0 of K has the entries $K_{00} = \frac{1}{2}$ and $K_{20} = 1$, column 1 only $K_{01} = \frac{1}{2}$, and column 2 only $K_{12} = 1$. So the predecessor sets are $\{0, 2\}$, $\{0\}$ and $\{1\}$. Configurations 0 and 2 feed a common configuration, and nothing else is merged: the merging classes of K are $\{0, 2\}$ and $\{1\}$.

By Theorem 26.15(a), $D(qK \,\|\, p) = D(q \,\|\, p)$ exactly when q/p is constant on $\{0, 2\}$, that is $q_0/\frac{1}{2} = q_2/\frac{1}{4}$, or $q_0 = 2q_2$. The distributions with this property are

$$q = (2t,\ 1 - 3t,\ t), \qquad 0 \leq t \leq \tfrac{1}{3}.$$

The value $t = \frac{1}{3}$ gives $q_0 = (\frac{2}{3}, 0, \frac{1}{3})$, and $t = \frac{1}{4}$ gives p itself.

The rows of K^2 are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$, $(1, 0, 0)$ and $(\frac{1}{2}, \frac{1}{2}, 0)$. Column 0 of K^2 has three nonzero entries, so all three configurations feed configuration 0 and K^2 has a single merging class. K^2 is stochastic and $pK^2 = p$, so Theorem 26.15(b) applied to K^2 gives $D(qK^2 \,\|\, p) < D(q \,\|\, p)$ for every distribution $q \neq p$. The criterion asks only that the merging relation connect all configurations; the zero entries $(K^2)_{11}$, $(K^2)_{12}$ and $(K^2)_{22}$ do not matter. For q_0 the two-step loss is visible in the opening table: $q_0K^2 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and D falls from 0.2877 to 0.0566.

Chapter 26, question 2: Theorem 26.16

The rows of K^4 are $(\frac{9}{16}, \frac{5}{16}, \frac{1}{8})$, $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $(\frac{5}{8}, \frac{1}{8}, \frac{1}{4})$. Dividing column j by p_j gives the ratios $(K^4)_{ij}/p_j$:

$$\begin{pmatrix} \frac{9}{8} & \frac{5}{4} & \frac{1}{2} \\ \frac{1}{2} & 1 & 2 \\ \frac{5}{4} & \frac{1}{2} & 1 \end{pmatrix}.$$

The largest δ with $(K^4)_{ij} \geq \delta p_j$ for all i, j is the least ratio, $\delta = \frac{1}{2}$. With $N = 4$ and $m = 2$ the theorem gives

$$D(q_0 K^8 \,\|\, p) \leq \left(\tfrac{1}{2}\right)^2 D(q_0 \,\|\, p) = \tfrac{1}{4} \log \tfrac{4}{3} \approx 0.0719.$$

Stepping the opening table on gives $q_0 K^8 = (\frac{95}{192}, \frac{51}{192}, \frac{46}{192})$, whose ratios to p are $\frac{95}{96}$, $\frac{51}{48}$ and $\frac{46}{48}$, so

$$D(q_0 K^8 \,\|\, p) = \tfrac{95}{192} \log \tfrac{95}{96} + \tfrac{51}{192} \log \tfrac{51}{48} + \tfrac{46}{192} \log \tfrac{46}{48} \approx 0.000726.$$

The bound holds with a factor of about 100 to spare. Doeblin's condition uses only the worst entry of one window, here the entries $\frac{1}{2}$ of K^4, and treats every window as if the distribution were at its worst.

Chapter 26, question 3: Theorem 26.24

Write u for any of the six unit vectors. On $\mathbb{Z}^3$ the Metropolis tick posts $\rho/6$ along each move out of the vacuum and holds there with probability $1 - \rho = \rho^2$. From u the inward post to the vacuum has probability $\frac{1}{6}$, and the five outward posts, each $\rho/6$, lead to states of cost 2, outside S. The wall turns them into a hold, so

$$P_S(0, u) = \tfrac{\rho}{6}, \quad P_S(0, 0) = \rho^2, \quad P_S(u, 0) = \tfrac{1}{6}, \quad P_S(u, u) = \tfrac{5}{6}, \quad P_S(u, u') = 0 \ (u' \neq u).$$

Each row sums to 1. The weights are 1 at the vacuum and ρ at each u, with total $1 + 6\rho$, and $\pi_S(0) P_S(0, u) = \rho/(6(1 + 6\rho)) = \pi_S(u) P_S(u, 0)$; detailed balance holds, and summing it gives $\pi_S P_S = \pi_S$. The vacuum and every unit vector feed the vacuum, so P_S has a single merging class, and by Theorem 26.23(b) one walled tick lowers $D(q \,\|\, \pi_S)$ strictly from every $q \neq \pi_S$.

For Theorem 26.24 take $R = 1$. The holds of the Metropolis tick on $\mathbb{Z}^3$ are least at states with three nonzero coordinates, where three posts are $\frac{1}{6}$ and three are $\rho/6$, so every hold is at least $h = 1 - (1 + \rho)/2 = \rho^2/2 \approx 0.191$. Every post is at least $e = \rho/6 \approx 0.103$. So $\min(h, e) = \rho/6$ and the theorem gives the constant $(\rho/6)^2 = \rho^2/36 \approx 0.0106$ at window 2.

The true largest constant is the least ratio $P_S^2(x, y)/\pi_S(y)$. For $u \neq u'$, $P_S^2(u, u') = P_S(u, 0) P_S(0, u') = \rho/36$ and $\pi_S(u') = \rho/(1 + 6\rho)$, a ratio of $(1 + 6\rho)/36 = (6\varphi - 5)/36 \approx 0.1308$. The other ratios are larger: $P_S^2(0, 0)/\pi_S(0) = (\rho^4 + \rho/6)(1 + 6\rho) \approx 1.172$, $P_S^2(0, u)/\pi_S(u) = P_S^2(u, 0)/\pi_S(0) = (\rho^2 + \frac{5}{6})(1 + 6\rho)/6 \approx 0.954$, and $P_S^2(u, u) \geq \frac{25}{36}$. So the largest Doeblin constant at window 2 is $(6\varphi - 5)/36$, about twelve times the constant the theorem guarantees. The theorem bounds every step of a path by the worst hold or post anywhere on $\mathbb{Z}^3$, while the walled holds at the unit vectors are $\frac{5}{6}$.

Chapter 26, question 5: Proposition 26.43

At the empty record G_0 is the point mass at the vacuum, so $m_0 = 0$, $H(G_0) = 0$ and $D_0(0) = -\log \pi(0) = 9 \log \varphi$. Part (b) gives

$$D(\mathsf{P}_0 \,\|\, \Pi_c) = 9 \log \varphi + \log Z + 9 \log \varphi = 3 \log \varphi + 3 \log(1 + \varphi^{-2}),$$

since $\log Z = 3 \log(1 + \varphi^{-2}) - 15 \log \varphi$. Now $\varphi^2 = \varphi + 1$, so $\varphi + \varphi^{-1} = 2\varphi - 1 = \sqrt{5}$ and $1 + \varphi^{-2} = \sqrt{5}/\varphi$. Hence $D(\mathsf{P}_0 \,\|\, \Pi_c) = 3 \log \sqrt{5} = \frac{3}{2} \log 5 \approx 2.414$. Part (c) gives $\mathrm{TC}(\mathsf{P}_0) = H(G_0) = 0$: at the empty record the two voxels are independent, both at the vacuum. Theorem 26.41(b) gives $D(\mathsf{P}_0 \,\|\, \pi \otimes \pi) = 2D_0(0) + 0 = 18 \log \varphi \approx 8.662$, and part (d) checks: $D(\mathsf{P}_0 \,\|\, \Pi_c) - \log Z = \frac{3}{2} \log 5 + 15 \log \varphi - \left(\frac{3}{2} \log 5 - 3 \log \varphi\right) = 18 \log \varphi$.

Both Π_A and Π_c live on the anti-diagonal, with $\Pi_c(s, -s) = \pi(s)^2 / Z$, so

$$D(\Pi_A \,\|\, \Pi_c) = \sum_s \pi(s) \log \frac{\pi(s)\, Z}{\pi(s)^2} = \log Z + \sum_s \pi(s)(9 + |s|_1) \log \varphi = \log Z + 15 \log \varphi,$$

using the mean cost 6 of the T9 law from the proof of part (c). This is $3 \log(1 + \varphi^{-2}) = \frac{3}{2} \log 5 - 3 \log \varphi \approx 0.9705$, the limit in part (e). So, with (T1) and (T2), the pair does relax, to its own equilibrium Π_A; the conditioned product law is a different law, because conditioning the product law on a zero total squares the one-voxel weights.

The floor of part (d) is $-\log Z = 15 \log \varphi - 3 \log(1 + \varphi^{-2}) = 18 \log \varphi - \frac{3}{2} \log 5 \approx 6.248$. It exceeds $9 \log \varphi \approx 4.331$ by $9 \log \varphi - \frac{3}{2} \log 5 \approx 1.917$ nats. The floor of Theorem 26.41(d) came from the bound $Z \le \pi(0) = \varphi^{-9}$; part (b) computes Z exactly.

Chapter 26, question 6: Theorem 26.57

Take $D = 2$. At the origin put rate 1 on the edges to e_0 and e_1 and rate -1 on the edges from $-e_0$ and $-e_1$; the source is $\sum_i \big(j(0, i) - j(-e_i, i)\big) = 4$, and every edge leaving the box of radius 0 carries $f(0) = 1$ outward.

Fix a shell $\|x\|_\infty = s$ with $s \ge 1$. Its sites receive $1/(2s - 1)$ along each outward edge of the box of radius $s - 1$ and must send $1/(2s + 1)$ along each outward edge of the box of radius s. Consider the side $x_0 = s$, the sites (s, y) with $|y| \le s$. A site with $|y| \le s - 1$ receives $1/(2s - 1)$ from $(s - 1, y)$ and sends $1/(2s + 1)$ to $(s + 1, y)$, an excess of

$$\varepsilon_s = \frac{1}{2s - 1} - \frac{1}{2s + 1} = \frac{2}{(2s - 1)(2s + 1)}.$$

The corner (s, s) receives nothing from inside and sends $1/(2s + 1)$ through each of its two outward edges, to $(s + 1, s)$ and $(s, s + 1)$, a deficit of $2/(2s + 1)$. Route the excess along the side toward the corners: give the edge from (s, y) to $(s, y + 1)$, for $0 \le y \le s - 1$, the rate $(2y + 1)\varepsilon_s/2$, and give the edge from $(s, -y - 1)$ to $(s, -y)$ the rate $-(2y + 1)\varepsilon_s/2$. Then $(s, 0)$ sends $\varepsilon_s/2$ each way, each $(s, \pm y)$ with $1 \le y \le s - 1$ passes on $(2y - 1)\varepsilon_s/2$ and adds its own ε_s, and the corner receives $(2s - 1)\varepsilon_s/2 = 1/(2s + 1)$ from this side. Rotating the same routing to the other three sides, each corner receives $1/(2s + 1)$ from each of its two sides, which is what it sends out. Every site off the origin is balanced, every rate is rational, and every outward edge of the box of radius r carries $f(r) = 1/(2r + 1)$. On the first shell: $(1, 0)$ receives 1, sends

$\frac{1}{3}$ outward and $\frac{1}{3}$ to each of $(1, \pm 1)$, and each corner receives $\frac{1}{3}$ from two axis sites and sends $\frac{1}{3}$ out through each of its two outward edges.

A multiple cj with c an integer has rate $c/(2r+1)$ on the outward edges of the box of radius r. Once $2r+1 > |c|$ this has absolute value below 1, so it is an integer only when $c = 0$. This is the argument of Theorem 26.57(a) in two dimensions: by (26.4) every balanced shell-uniform current on $\mathbb{Z}^2$ has $4(2r+1)f(r) = \operatorname{div} j(0)$ for every r, and integer rates force the source to be 0.

Chapter 27, question 1: Theorem 27.3

For $|\varepsilon| < 1$ the matrix $1 + \varepsilon h = \operatorname{diag}(1, 1+\varepsilon, 1-\varepsilon, 1)$ has positive eigenvalues, and $J(1+y) = y^2/(2(1+y))$ gives

$$J(1+\varepsilon h) = J(1+\varepsilon) + J(1-\varepsilon) = \frac{\varepsilon^2}{2(1+\varepsilon)} + \frac{\varepsilon^2}{2(1-\varepsilon)} = \frac{\varepsilon^2}{1-\varepsilon^2}.$$

Divided by ε^2 it tends to $1 = J_2(h) = \frac{1}{2}\operatorname{tr}(h^2)$. Since $he_0 = 0$ and $\operatorname{tr} h = 0$, the strain is TT at e_0, and $\mathrm{FP}(h, e_0) = 1 \cdot 2 - 0 + 0 - 0 = 2 = |K|^2 \cdot 2J_2(h)$, as part (b) says.

Now $h' = h + E_{03} + E_{30}$ adds the gauge mode $K\xi^{\mathsf{T}} + \xi K^{\mathsf{T}}$ with $K = e_0$, $\xi = e_3$. The block of $1 + \varepsilon h'$ on the coordinates $0, 3$ is $\left(\begin{smallmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{smallmatrix}\right)$, with eigenvalues $1 \pm \varepsilon$, so the eigenvalues are $1+\varepsilon$ and $1-\varepsilon$, each twice, and $J(1+\varepsilon h') = 2\varepsilon^2/(1-\varepsilon^2)$, with long-wave limit $2 = J_2(h')$. The gauge mode added $J_2(e_0 e_3^{\mathsf{T}} + e_3 e_0^{\mathsf{T}}) = (K \cdot \xi)^2 + |K|^2|\xi|^2 = 1$, as part (c) says. For Fierz–Pauli, $\operatorname{tr}(h'^2) = 4$, $h'e_0 = e_3$ so $|h'K|^2 = 1$, $\operatorname{tr} h' = 0$ and $K^{\mathsf{T}} h' K = 0$, so $\mathrm{FP}(h', e_0) = 4 - 2 = 2$: unchanged, as part (a) requires, since $\Pi_{e_0} h' \Pi_{e_0} = \Pi_{e_0} h \Pi_{e_0}$. The ledger charges the change of coordinates; gravity does not.

Chapter 27, question 2: Theorem 27.5

The dimension is $(D+1)(D-2)/2 = 2$. The vectors orthogonal to both $\mathbf{1}$ and $m = (1, -1, 0, 0)$ are (a, a, b, c) with $2a + b + c = 0$, spanned by $u = (1, 1, -1, -1)$ and $v = (0, 0, 1, -1)$. Every combination of uu^{T}, vv^{T} and $uv^{\mathsf{T}} + vu^{\mathsf{T}}$ is symmetric and kills $\mathbf{1}$ and m; the traces are $|u|^2 = 4$, $|v|^2 = 2$ and $2\,u \cdot v = 0$. So the TT strains at m are spanned by

$$H_1 = uu^{\mathsf{T}} - 2vv^{\mathsf{T}} = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 3 \\ -1 & -1 & 3 & -1 \end{pmatrix}, \qquad H_2 = uv^{\mathsf{T}} + vu^{\mathsf{T}} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -2 & 0 \\ -1 & -1 & 0 & 2 \end{pmatrix}.$$

Each has zero row sums, zero trace, and equal zeroth and first columns, so $H_i\mathbf{1} = 0 = H_i m$; they are independent. Their costs are $J_2(H_1) = 16$ and $J_2(H_2) = 8$, both positive, as part (b) says. In the proof's construction one takes $i = 0$, $j = 1$ and $R = \{2, 3\}$; then $a_k = b_k = \frac{1}{2}$ and $r_k = e_k - \frac{1}{2}(e_0 + e_1)$, and $u = -(r_2 + r_3)$, $v = r_2 - r_3$.

Chapter 27, question 3: Theorem 27.24

Both momenta have zeroth component 0, and $k_1 \cdot k_1 = 9 + 16 + (5i)^2 = 0$, $k_2 \cdot k_2 = 25 + 144 + (13i)^2 = 0$. Since $k_3 = -(k_1 + k_2)$, $\Xi = (k_1 + k_2)\zeta^{\mathsf{T}} + \zeta(k_1 + k_2)^{\mathsf{T}}$ with $\zeta = -\xi$, a gauge mode at $k_1 + k_2$.

The central inversion sends the slot ij to $(4 - j)(4 - i)$: $01 \leftrightarrow 34$, $02 \leftrightarrow 24$, $03 \leftrightarrow 14$, $12 \leftrightarrow 23$, while 04 and 13 are fixed; on slot numbers this is $(0\,9)(1\,8)(2\,6)(4\,7)$. So $000 \mapsto 999$, $444 \mapsto 777$ and $044 \mapsto 977$, which is the monomial 779. In Table 27.2 the form $128\,r$ has the coefficients 4 and 4, -32 and -32, -48 and -48 on these three pairs, and $128\,d$ has 2 and 2, -10 and -10, -24 and -24.

With V_r in place of V_d the test direction $H' = \Xi$ gives $-\frac{1}{4}\mathrm{FP}_1(k_1 + k_2; H_2, \Xi) = 0$ by gauge invariance and $V_r(\eta, \eta, \Xi; k_1, k_2, k_3) = 0$ by part (b), so the equation reads $0 = 0$. The test is a necessary condition for a second-order field: it can exclude one, as it does for J's extra vertex, but passing it constructs nothing. Regge's vertex passes it, so this test leaves Regge's second-order field possible.

Chapter 27, question 4: Theorem 27.26

At $\kappa = 1$, $\det G = 1 - 4\kappa = -3$. The Heron values are $4(1-2) = -4$ on 012, 123, 234; $8(1-3) = -16$ on 013, 023, 124, 134; $12(1-4) = -36$ on 014, 034; and $16(1-4) = -48$ on 024. For 012 directly: its edges have 1, 1 and 2 steps, with squared lengths 0, 0 and $2 - 4 = -2$, and Heron's expression is $-(-2)^2 = -4$. The diagonal of G^{-1} is $(2, 2, 2, (3\kappa - 1)/(4\kappa - 1)) = (2, 2, 2, \frac{2}{3})$, with off-diagonal entries -1. The vertices off 012 are $l = 3$, $m = 4$, and $W_{ij} = (G^{-1})_{ij}$ for $i, j \geq 1$, so

$$W_{33}W_{44} - W_{34}^2 = 2 \cdot \tfrac{2}{3} - 1 = \tfrac{1}{3} = \frac{-4}{4 \cdot (-3)} = \frac{Q_{012}}{4(1 - 4\kappa)}.$$

Chapter 27, question 5: Theorem 27.40

By part (c) the squared Bloch length is $3/(4\kappa - 1) = 3/4$, so the Bloch vector of a posting record has length $\sqrt{3}/2$: the record is mixed, as $\kappa > 1$ requires. A sharp test $(1, \vec{m})$ passes $(1, \vec{p})$ with probability P, $2P - 1 = \vec{m} \cdot \vec{p} \leq \sqrt{3}/2$, so the largest probability is $(2 + \sqrt{3})/4 \approx 0.933$, attained by $\vec{m}$ along $\vec{p}$. Then $\varepsilon = (2 - \sqrt{3})/4$, $1 - 2\varepsilon = \sqrt{3}/2$, and the bound of part (c) reads $\kappa \leq \frac{1}{4} + 3/(4 \cdot \frac{3}{4}) = \frac{5}{4}$, with equality. Distinct posting records have pairing $(2\kappa - 1)/(4\kappa - 1) = \frac{3}{8}$. By part (d) the depolarizing strength satisfies $r^2(2\kappa - 1) = 1$, so $r^2 = \frac{2}{3}$ and $r = \sqrt{6}/3 \approx 0.816$; a posting off the posting plane then has a record of trace $2\kappa/(2\kappa - 1) = \frac{5}{3}$.

Chapter 27, question 6: Theorem 27.44

Write an assignment as (r_0, r_1, r_2) with one-bit records; the law posts a debit when $r_0 + r_1 + r_2$ is even. For the partner $y = 0$ take $(0, 0, 0)$ and $(0, 1, 0)$; for $y = 1$ take $(0, 0, 0)$ and $(0, 0, 1)$; for $y = 2$ take $(0, 0, 0)$ and $(0, 1, 0)$. In each pair the two assignments agree at the site 0 and at y, and their sums are 0 and 1, so one posts a debit and the other a credit. No partner determines the post at 0, so the law is not pair-reading there. Each post is still a single debit or credit on one account, and its cost on the ledger is that of any unit post, whatever decided it.

O.12 The library, learning and open problems

Chapter 28, question 1: Theorem 28.9

In the proof of item 4, $u(x) = 1$ exactly when $\neg(x > 0)$. Since $u(x)$ is a bit, $u(x) = 0$ exactly when $\neg(u(x) = 1)$, that is, when $\neg\neg(x > 0)$. The relation $x \leq 0$ is $\neg(0 < x)$, so $\neg\neg(x > 0)$ is $\neg(x \leq 0)$, and $S(x) = 1$ exactly when $u(x) = 0$. In the same way $S(x) = -1$ exactly when $u(x) = 1$ and $u(-x) = 0$, which gives $\neg\neg(x < 0)$.

Now assume MP. If $|S(y)| > 0$, then $S(y) \neq 0$, and since the three cases of the construction are decided, $S(y) = 1$ or $S(y) = -1$. In the first case $\neg\neg(y > 0)$, in the second $\neg\neg(y < 0)$. MP in the form $\neg\neg(x > 0) \to x > 0$ (Corollary 12.18), applied to y or to $-y$, gives $y > 0$ or $y < 0$, that is, $y \# 0$. For $y \# 0$ we have $|S(y) - S(0)| = 1$, so S jumps at 0, and the jump carries a test. Part 2 of Theorem 28.8 gives the LPO function. So the WLPO function together with MP gives the LPO function.

Without MP the step from $\neg\neg(y > 0)$ to $y > 0$ is missing. The value $S(y) = 1$ records only that y is not ≤ 0. Corollary 12.18 prices the passage from that to positivity at MP, so the jump of S need not carry a test, and only part 1 of the rule applies. That part gives the WLPO function, which is the price of S.

Chapter 28, question 2: Theorem 28.45

The strict slacks are $a - 0$ and $b - a$. Expand:

$$(b-a)^2 + 3a(b-a) + 3a^2 = b^2 - 2ab + a^2 + 3ab - 3a^2 + 3a^2 = b^2 + ab + a^2,$$

and $(b-a)(b^2 + ab + a^2) = b^3 - a^3$. So, with $k_1 = 1$ and

$$\begin{aligned} k_2 = (b-a)\cdot\big((b-a)(b-a) + (a-0)(b-a) + (a-0)(b-a) + (a-0)(b-a) \\ + (a-0)(a-0) + (a-0)(a-0) + (a-0)(a-0)\big), \end{aligned}$$

the identity $k_1 \cdot (b\cdot b\cdot b - a\cdot a\cdot a) = k_2$ holds at every point, by the ring laws (Proposition 12.19). The expression k_2 is built from the strict slacks by sums and products, so it is a cone expression, and $k_1 = 1$ is one too. Part 1 of Theorem 28.45 gives $a\cdot a\cdot a < b\cdot b\cdot b$ at every point satisfying $0 < a$ and $a < b$.

Density applies to closed conclusions only. Part 1 of Theorem 28.42, with the strict atoms $0 < a$ and $a < b$ as hypotheses, would move the closed atom $a\cdot a\cdot a \leq b\cdot b\cdot b$, which holds at all ratios satisfying them, but not the strict inequality. A strict conclusion need not survive the limit (Proposition 28.43, item 1).

Chapter 28, question 3: Theorem 28.50

Here $\|d_n\| + 1 = 10n + 16$. For $n = 6$ this is $76 \geq 64 = 2^6$, and for $n = 7$ it is $86 < 128 = 2^7$, so the least n is 7. There $\|d_7\| = 85$, while $\|t_{d_7}\| \geq 4^8 = 65{,}536$.

For given C and k the proof puts $a = C\cdot 16^k$, $j = k+a+4$ and $n = 2^j - 1$. Since $10n+16 \leq 16(n+1)$, $C(\|d_n\|+1)^k \leq a(n+1)^k = a\,2^{jk}$. For $j \geq 4$, $j^2 \leq 2^j$: this holds at $j = 4$, and $(j+1)^2 = j^2 + 2j + 1 \leq$

$2j^2$ for $j \geq 3$. Moreover $j^2 = jk + ja + 4j \geq jk + a + 16$, so $a + jk \leq j^2 - 16 < 2^j - 1 = n$. Since $a < 2^a$ for every a,

$$C(\|d_n\| + 1)^k \leq a\, 2^{jk} < 2^{a+jk} \leq 2^n \leq \|t_{d_n}\|.$$

Chapter 28, question 4: Theorem 28.52

Suppose $\|F(d)\|^q \leq C(\|d\| + 1)^p$ for every d with $\mathrm{sh}(d) \leq 3$, where $2p < 5q$. Every formula concluded in B_K has shape at most 3, so B_K qualifies. For $K \geq 1$, $\|B_K\| + 1 = 14K^2 + 28K + 16 \leq 16(K+1)^2 \leq 64K^2$, so the reported count $5K^5 \leq 6\,\|F(B_K)\|$ gives

$$\Big(\frac{5K^5}{6}\Big)^q \leq \|F(B_K)\|^q \leq C \cdot 64^p K^{2p},$$

and $K^{5q-2p} \leq C \cdot 64^p (6/5)^q$ for every $K \geq 1$. Since $5q - 2p \geq 1$, this fails for large K.

For part 3 the family is useless: B_K keeps K hypotheses open, so $\mathrm{cx}(B_K) = K$, and for $K > c$ it leaves the class of part 3. Its growth comes from exactly the long hypothesis lists that part 3 excludes; there the lower bound comes from the size of the matrix instead.

Chapter 28, question 5: Theorem 28.53

If $x = t(x) + t(x)$, then $2t(x) = x$; if $x = t(x) + t(x) + 1$, then $2t(x) = x - 1$. Either way $2t(x) \leq x \leq 2t(x) + 1$, so $x/2 - 1 < t(x) \leq x/2$, and $t(x)$ is the one integer in that interval, $\lfloor x/2 \rfloor$.

The proof of the theorem uses the derivation h, the specification of the compiler in Theorem 28.51 (the matrix holds at $\mathrm{val}(t)(x)$ for every x), the arithmetic just given, and item 4 of Theorem 28.49 for the orbit. The specification already determines the value, because the matrix has exactly one witness for each x. The inside of the program, its size and its evaluation do not affect the value; they affect only what the value costs to compute.

Chapter 28, question 6: Theorem 28.56

Let φ preserve 0, 1, sums and products, and reason classically. For every real s, $\varphi(s^2) = \varphi(s)^2 \geq 0$, and every nonnegative real is a square, so φ maps nonnegative numbers to nonnegative numbers. If $x \leq y$, then $\varphi(y) - \varphi(x) = \varphi(y - x) \geq 0$, so φ is monotone. From $\varphi(1) = 1$ and additivity, $\varphi(n) = n$ for every integer n, and $\varphi(u/d) \cdot d = \varphi(u) = u$ gives $\varphi(q) = q$ for every rational q. If $\varphi(x) < x$, take a rational q with $\varphi(x) < q < x$; monotonicity gives $q = \varphi(q) \leq \varphi(x)$, a contradiction. In the same way $\varphi(x) > x$ is impossible, so $\varphi(x) = x$.

To deduce item 2, let D and D' have the properties of item 1. If $D(x) = D(y)$, then neither $x < y$ nor $y < x$, so $x \leq y \leq x$ and $x = y$ (Proposition 12.19, item 2). So D is injective, and it is onto, so classically it has an inverse D^{-1}, which also preserves 0, 1, sums and products. Then $D' \circ D^{-1}$ preserves them, so it is the identity, and $D' = D$.

For Proposition 28.57 at $n = 2$, expansion along the first column gives $\det M = M_{00} \det M^{(0,0)} - M_{10} \det M^{(1,0)} = M_{00}M_{11} - M_{10}M_{01}$. In the permutation sum, the identity contributes $M_{00}M_{11}$. The exchange $\pi(0) = 1$, $\pi(1) = 0$ has sign -1 and contributes $-M_{\pi(0)0}M_{\pi(1)1} = -M_{10}M_{01}$. The two agree.

Chapter 29, question 1: Theorem 29.4

In $H(\alpha)$ the atom p_0 is forced at stage n exactly when $\alpha(k) = 1$ for some $k \le n$, so here exactly at the stages $n \ge 3$. The formula $\neg p_0$ is forced at n only if no stage $u \ge n$ forces p_0; since stage $\max(n, 3)$ forces p_0, $\neg p_0$ is forced at no stage. Consequently $\neg\neg p_0$, which asks that no later stage force $\neg p_0$, is forced at every stage. At stage 0, p_0 is not forced (the record has not yet seen the event) and $\neg p_0$ is not forced (stage 3 refutes it), so em $= p_0 \vee \neg p_0$ fails there; wem $= \neg p_0 \vee \neg\neg p_0$ holds through its second disjunct. What makes wem true at stage 0 is the fact that α is not all zero, which the record at stage 0 does not show. Asking for wem at stage 0 in every history therefore asks to decide, for every α, whether it is all zero, and that is WLPO, as part 2 of the theorem states.

Chapter 29, question 2: Theorem 29.8

The law gives $\frac{1}{3}$ to each of 100, 010, 001; it is exchangeable. Its first two records have the law $P_2(00) = \frac{1}{3}$ (from 001), $P_2(10) = \frac{1}{3}$, $P_2(01) = \frac{1}{3}$ and $P_2(11) = 0$. Every record has one entry 1, so the coin mixture is the single coin of bias $\frac{1}{3}$: $Q_2(00) = \frac{4}{9}$, $Q_2(10) = Q_2(01) = \frac{2}{9}$, $Q_2(11) = \frac{1}{9}$. The sum of the absolute differences is $\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{4}{9}$, and the total variation distance is $\frac{2}{9}$. The theorem's bound is $2(1 - (3)_2/3^2) = 2(1 - \frac{6}{9}) = \frac{2}{3}$, and $m(m-1)/n = \frac{2}{3}$ as well; indeed $\frac{4}{9} \le \frac{2}{3}$. Sampling two of three records without replacement can never show 11, while the coin shows it with probability $\frac{1}{9}$: this is the whole discrepancy, spread over the four records.

Chapter 29, question 3: Theorem 29.15

The recognition history begins dcddcdcddcddcdcddcdcd, the word W_6 of Definition 21.24, so the golden record begins

$$0\,1\,0\,0\,1\,0\,1\,0\,0\,1\,0\,0\,1\,0\,1\,0\,0\,1\,0\,1\,0.$$

Its windows of length 3 are 001, 010, 100 and 101: four, which is 3 + 1, as Theorem 29.15 requires; 000 and 011 never occur, and 110, 111 would violate balance. The windows of length 2 are 00, 01 and 10. Of these, 10 is right special, since both 100 and 101 occur, while 00 continues only to 001 and 01 only to 010. So $p(3) = p(2) + 1$, as in the proof of Theorem 29.14(2).

Chapter 29, question 4: Theorem 29.31

Write e_x for the law that is 1 exactly at x. Every point x splits H, into $H_{x,1} = \{e_x\}$, of dimension 0, and the other three laws; so $\mathrm{ldim}(H) = 1 + 0 = 1$. On the target e_3 with the points 0, 1, 2, 3: at 0 the part with label 1 is $\{e_0\}$, of rank 1, and the part with label 0 has three laws and dimension 1, rank 2; the learner predicts 0, correctly. At 1 the same comparison gives 0, correctly. At 2 the two parts are $\{e_2\}$ and $\{e_3\}$, both of rank 1, so the learner predicts 1 and errs; only e_3 remains, and the prediction at 3 is right. One mistake, equal to the dimension. The halving learner also predicts 0, 0, 1, 1 here and errs once, but its guarantee from Theorem 29.29(1) is only $\log_2 4 = 2$; the dimension gives the exact worst case, which by part 2 of Theorem 29.31 no learner can improve.

Chapter 29, question 5: Theorem 29.39

For the records 1 and 4, $S = 5$ and $T = 1 + \frac{1}{4} = \frac{5}{4}$, so $m_*^2 = 4$ and $m_* = 2$, with $\tau = m_* T = \frac{5}{2}$. The least cost is $C(2) = \tau - 2 = \frac{1}{2}$, and directly $J(\frac{1}{2}) + J(2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. The learner uses $f_0 = 1$ for the first record and the centre $f_1 = 1$ of $\{1\}$ for the second, paying $J(1) + J(4) = 0 + \frac{9}{8}$; its regret is $\frac{9}{8} - \frac{1}{2} = \frac{5}{8}$. The identity of part 4 reads

$$\frac{9}{8} = C_2(f_2) + f_1 T_1 J(f_0/f_1) + f_2 T_2 J(f_1/f_2) = \frac{1}{2} + 1 \cdot 1 \cdot J(1) + 2 \cdot \frac{5}{4} \cdot J(\tfrac{1}{2}) = \frac{1}{2} + 0 + \frac{5}{8},$$

with $f_2 = 2$: the regret is the J-cost of the one change of frame, from 1 to 2, weighted by $\tau = \frac{5}{2}$.

Chapter 29, question 6: Theorem 29.47

For one world ω with its entry 1 at m, the learner of part 3 stays silent below $m + 1$ and then commits by the entry at m; it is correct on both worlds because ω is decided. A single learner for all of them fails by part 1: it commits on the zero world at some stage N to a program z, the world whose only entry 1 is at N has the same record of length N, so the learner commits to z on it too, and z computes the zero world. The same argument shows that the catalogue of the zero world and these worlds has no locking function: if g were one and the zero world had index i_0, the world with its only entry 1 at $g(i_0)$ would agree with the zero world below $g(i_0)$ without being equal to it. By Theorem 29.48(3) these two facts are the same fact; by part 4, the catalogue is nevertheless learned in the limit with one change of mind, at no price.

Chapter 30, question 1: Lemma 30.3

Here $\sqrt{5} - 2 \approx 0.23607$. For $k = 1, 2$ the nearest integer is $m = 0$, and $6k|k(\sqrt{5} - 2)|$ equals $6\sqrt{5} - 12 \approx 1.416$ and $24\sqrt{5} - 48 \approx 5.67$; the first is at least 1 because $6\sqrt{5} \geq 13$, that is $180 \geq 169$. For $k = 3, 4, 5$ the nearest integer is $m = 1$, and the distances are $7 - 3\sqrt{5} \approx 0.2918$, $9 - 4\sqrt{5} \approx 0.0557$ and $5\sqrt{5} - 11 \approx 0.1803$, giving 5.25, 1.337 and 5.41. The case $k = 4$ comes closest: $24(9 - 4\sqrt{5}) = 216 - 96\sqrt{5} \geq 1$ because $215^2 = 46225 \geq 46080 = 96^2 \cdot 5$. The proof of the lemma explains it: with $M = m + 2k = 9$, $5k^2 - M^2 = 80 - 81 = -1$, so $|4\sqrt{5} - 9| = 1/(4\sqrt{5} + 9)$, as small as the lemma allows up to the factor $6|k|/(4\sqrt{5} + 9) \approx 1.34$. The denominators $1, 4, 17, 72, \ldots$ of the convergents of $[0; 4, 4, 4, \ldots]$ are where such near equalities occur.

Chapter 30, question 2: Theorem 30.9

Here $i = 2$, $p = 4$, $j = 5$, $q = 6$, so $\gcd(4, 6) = 2$ and the join is $\sim_{2,2}$: counts at least 2 are identified when they have the same parity. Directly, $4M \equiv 2 \pmod 6$ has the solution $M = 2$, and for $t \geq 2$

$$t \sim_{2,4} t + 28 \sim_{5,6} t + 22 \sim_{2,4} t + 2.$$

Since $2 < 5$ and $4 \nmid 6$, the criterion fails. Take $a = 4$ and $b = 2$, related by the join. A count n with $n \sim_{5,6} 2$ must be 2 itself, because 2 is below the index 5; and $2 \sim_{2,4} 4$ would need $4 \mid 2$. So no common count exists.

Chapter 30, question 3: Theorem 30.13

The number $\sqrt{3}-1 \approx 0.732$ is a root of t^2+2t-2, whose other root $-1-\sqrt{3} \approx -2.732$ lies outside $[-1,1]$; so it is a rate. The recipe of the proof, with $a=1$, $b=2$, $c=-2$, gives $CD=1$, $DC=3$, $q=-6$, $DD=13$, $CC=7$, the join $\mathsf{d} \mapsto \mathsf{d}^{13}\mathsf{c}$, $\mathsf{c} \mapsto \mathsf{d}^{3}\mathsf{c}^{7}$; indeed $(1+t)^2-6(1-t^2)-3(1-t)^2 = 4(t^2+2t-2)$. A smaller join with the same CD, DC and $CC-DD$ is $\mathsf{d} \mapsto \mathsf{d}^{6}\mathsf{c}$, $\mathsf{c} \mapsto \mathsf{ddd}$, primitive at the second power: its Perron root is $3+2\sqrt{3}$, its Perron vector $(3, 2\sqrt{3}-3)$, so $\beta = 3+2\sqrt{3}$ and the rate is $(2+2\sqrt{3})/(4+2\sqrt{3}) = \sqrt{3}-1$. The number $\sqrt{3}/3$ is a root of $3t^2-1$; every integer quadratic with this root is a rational multiple of it, and the other root $-\sqrt{3}/3$ lies in $[-1,1]$. So $\sqrt{3}/3$ is not a rate.

Chapter 30, question 4: Theorem 30.14

Both joins have incidence counts $2,1,1,2$, Perron vector $(1,1)$ and rate 0, so the discrepancy is the balance. For the first, $\sigma^2(\mathsf{d}) = \mathsf{dcd\,cdc\,dcd}$, whose prefix balances for $m = 0,\dots,8$ are $0,1,0,1,0,1,0,1,0$, while the letter at position m is d for even m and c for odd m; with $\pi(\mathsf{d})=0$, $\pi(\mathsf{c})=1$, the coboundary asks exactly $\pi(\text{letter } m)-\pi(\mathsf{d})$, which is 0 or 1 as it should be. The word $\sigma^2(\mathsf{c}) = \mathsf{cdc\,dcd\,cdc}$ has prefix balances $0,-1,0,-1,\dots$ and letters $\mathsf{c},\mathsf{d},\mathsf{c},\dots$, and $\pi(\text{letter } m)-\pi(\mathsf{c})$ is 0 or -1 accordingly. For the second join, $\sigma^2(\mathsf{d}) = \mathsf{ddc\,ddc\,ccd}$; at $m=1$ the prefix d has balance 1, and the letter at position 1 is d, so a coboundary would need $1 = \pi(\mathsf{d})-\pi(\mathsf{d}) = 0$. None exists, and by the theorem the books are not balanced.

Chapter 30, question 5: Theorem 30.24

On the triad $(0,1,7)$ the cycle $e^{2\pi B/3}$ sends $e_0 \mapsto e_1 \mapsto e_7 \mapsto e_0$. So one braid token of angle $2\pi/3$ moves 7 to 0, and one of angle $-2\pi/3$, the inverse cycle, moves 0 to 7; each drives $2\pi/3$, which is optimal by the theorem. A token of angle $\pi/2$ on this triad has diagonal entries $(1+2\cos\frac{\pi}{2})/3 = \frac{1}{3}$ at the labels $0,1,7$ and 1 elsewhere. A certain change out of a label needs the diagonal entry there to vanish, since the column is a unit vector with an entry of modulus 1 elsewhere; so the token moves no label with certainty, in line with Theorem 30.22(b).

Chapter 30, question 6: Theorem 30.35

In the coordinates $(i,j,7)$ and with $c = 1/\sqrt{3}$, $A = \begin{pmatrix} 0 & -c\theta & c\theta \\ c\theta & 0 & -c\theta \\ -c\theta & c\theta & -\mathrm{i}\varepsilon \end{pmatrix}$. Write $\omega = (\omega_1,\omega_2,\omega_3)$. The first row gives $c\theta(\omega_3-\omega_2) = c\theta(-3x^2-\mathrm{i}\sqrt{3}\theta x) = -\sqrt{3}\theta x^2-\mathrm{i}\theta^2 x$, and $-\mathrm{i}x\,\omega_1 = -\mathrm{i}\theta^2 x-\sqrt{3}\theta x^2$. The second row gives $c\theta(\omega_1-\omega_3) = c\theta(3x^2-\mathrm{i}\sqrt{3}\theta x) = \sqrt{3}\theta x^2-\mathrm{i}\theta^2 x$, and $-\mathrm{i}x\,\omega_2 = -\mathrm{i}\theta^2 x+\sqrt{3}\theta x^2$. The third row gives $c\theta(\omega_2-\omega_1)-\mathrm{i}\varepsilon\omega_3 = 2\mathrm{i}\theta^2 x-\mathrm{i}\varepsilon(\theta^2-3x^2)$, while $-\mathrm{i}x\,\omega_3 = -\mathrm{i}\theta^2 x+3\mathrm{i}x^3$; they agree exactly when $3x^3-3\varepsilon x^2-3\theta^2 x+\varepsilon\theta^2 = 0$, that is $p(x)=0$.

Appendix P

Finding an argument

This guide follows the questions answered by the main proofs. Each entry points to the conclusion and to the earlier result on which it depends. It is a route back into the argument, with the hypotheses kept in view.

P.1 What determines arithmetic?

Theorem 1.4 identifies the finite orbit expressions with their natural-number readings in both directions. The initial-map theorem, Theorem 2.3, then constructs a unique reading into any pointed step system. To make that reading bijective, use the injectivity and generation conditions of Definition 2.4; the conclusion is Theorem 2.7. Proposition 2.8 shows that these conditions are also necessary for a bijective reading.

The arithmetic operations are developed in Chapter 3. Their order of proof matters: the preparation lemmas support addition's laws, and those laws support multiplication. For the passage to signed differences and ratios, start with the two equivalence relations in Chapter 4. The well-definedness arguments explain why an operation gives one answer despite a change of representatives. Theorem 3.10 and Corollary 3.12 show that the operations are forced once adding one is the step, and Proposition 3.11 gives the unique map into every commutative semiring. Section 4.9 proves that each extension is the unique solution of its task: Theorem 4.13, Corollary 4.14 and Theorem 4.15, with the forced order in Proposition 4.16. Theorem 12.38 does the same for the line built from ratios, with countable choice.

P.2 What does a proof use?

The rules are collected in Appendix D; their interpretation and soundness are in Theorem 5.5. For the additional rules, Theorem 6.6 states the surrounding principles that validate the recorded uses. Chapter 6 also separates a proof's actual ledger from a claim of necessary resources. Section 5.7 shows what separation, injectivity and induction each supply, using Theorem 5.8. With induction on every formula, the exact price of each additional rule is given by Theorems 6.9, 6.10 and 6.11; Proposition 6.13 gives the prices without induction, and Proposition 6.14 with induction on quantifier-free formulas only. Corollary 6.15 replaces the assumptions of Theorem 6.6 by the least ones: an MP entry needs no principle, and the derivations with an EM or LPO entry need, taken together, exactly the metatheoretic LPO for the predicates that formulas of the language define, which for these predicates is the same as excluded middle. Theorem 8.11 identifies the rules, with excluded middle and without LPO and MP, with Peano arithmetic, and Corollary 8.12 identifies the posit-free derivations with Heyting arithmetic, up to how the injectivity of the successor is written.

The constructive recognition catalogue begins with a decidable relation and an exhibited collision: Theorem 7.6. Its priced formulations and their reverse implications appear in

Theorems 7.7, 7.9 and 7.10, and Section 7.9 collects them. Theorem 7.11 carries the prices to readers, and Corollary 7.13 carries all six prices to signed records, with the parity question asked of the period. Theorem 7.38 shows that the six prices are different: Section 7.12 builds the four models of intuitionistic second-order arithmetic that separate them, with Lemma 7.21 turning each model into a failure of derivability. Section 7.13 goes between them: Theorem 7.45 splits Markov's principle into two halves, Proposition 7.46, with Lemma 7.41, gives a strictly descending chain of prices below LLPO, Theorem 7.47 two further intermediate prices, and Corollary 7.49 shows that the six prices are not all the prices. Theorem 8.6 shows that the two arithmetic presentations satisfy the same formulas, Theorem 8.8 what a one-way translation keeps, and Proposition 8.9 that no other quantifier prefix is kept. For a general reader, the quotient is identified with its range in Theorem 8.13.

P.3 Which constructions give stronger eliminations?

Begin with the exact language in Section 9.2. The separation results culminate in Theorem 9.10; the model argument they use is completed in Section 9.8. In particular, the countermodels must preserve substitution and conversion, and the separating families must be expressible. Those checks make the displayed models evidence about the specified language. That the strongest fragment proves the translation of every theorem of Peano arithmetic is the first assertion of Theorem 20.39; the second gives the converse, through the compiled programs of Section 20.3.

P.4 Does a richer language prove more about numbers?

Theorem 8.11 identifies Peano arithmetic with the derivations whose ledger records neither LPO nor MP. The language of functions of Section 20.3.2 is read through compiled programs. Follow the argument in four stages.

Computation inside arithmetic. Theorem 20.3 describes computations by an arithmetic formula and derives its functionality and closure laws; Theorem 20.5 derives the two facts about the record test, and Lemma 20.10 turns them into the law every compiled application uses. The routine cases are in Appendix K.

Reading formulas. Hereditary equality and its symmetry and transitivity are Proposition 20.11; the reading is exact by Theorem 20.12. Typed terms return related values, Theorem 20.15, through Lemmas 20.13 and 20.14.

Substitution and the rules. Substitution is Theorem 20.18, built on Lemma 20.16 and Theorem 20.17. The nineteen rules are assembled in Theorem 20.26, with conversion in Theorem 20.22, extensionality in Theorem 20.24 and induction in Theorem 20.25.

Back to arithmetic. Theorem 20.31 recovers a negative sentence from its reading, Theorem 20.32 gives conservativity, and Theorem 20.39 compares $\mathscr{E}^+$ with Peano arithmetic, using Lemma 20.33, Theorem 20.36 and Theorem 20.38. Consistency, Gödel's second incompleteness theorem, a universe, bounded arithmetic and finite set theory follow in Section 20.3.11, and Löb's theorem and the tower of consistency statements in Section 20.3.12. Section 20.4 converts classical proofs shape by shape, for the shapes of the book's arithmetic language:

Theorem 20.57 translates derivations, Theorems 20.59 and 20.63 and Corollary 20.65 give the free shapes, and Theorem 20.66 with Table 20.4 collects the prices.

P.5 When does finite information give one uniform procedure?

Theorem 10.3 assembles compatible finite-stage maps. Faithful program storage is Theorem 10.8. Neither result by itself supplies termination at every precision. Theorems 10.12 and 11.11 state the stronger naming property and the boundary between it and separate finite-prefix descriptions.

For interval protocols, existence of a displayed classical value and its uniqueness are established in Chapter 11. Equality of those values agrees with observational equality in Theorem 11.5. The exact logical costs of zero testing, trichotomy, the weaker sign decision and the sign of a nonzero quantity are Theorems 12.6, 12.7, 12.8 and 12.11. Theorem 12.17 and Corollary 12.18 carry them to the line built from ratios. Table 12.3 prices that line's completeness; the results from Theorem 12.25 to Proposition 12.62 prove its rows, in both directions wherever a row names a price, and Theorems 12.72, 12.75 and 12.77 price weak König's lemma, the pigeonhole principle and Ramsey's theorem for pairs. Theorem 12.22 shows that the countable choice those prices use decides none of the omniscience principles, and Theorem 12.63 that excluded middle, without unique choice, cannot replace it. Table 13.1 prices compactness: Heine–Borel for $[0,1]$ and for presented sets, and positive bounds and infima of presented maps, are each equivalent to the decidable fan theorem with no choice principle (Theorems 13.6, 13.25 and 13.28, Corollary 13.30). Table 14.1 prices linear algebra on the line (Theorems 14.13, 14.16 and 14.28), and Section 12.10 the calculus on the line (Table 12.4). Keep the chosen representation and the stated comparison relation attached to each result.

Certified computation turns precision into a count. A certified real has the modulus of Definition 15.1, and Proposition 15.4 keeps precision additive. The solution value of a quadratic field, and through its quadratization that of a polynomial field, is a certified real by Theorem 15.22, from the region certificate through Theorems 15.20 and 15.21; its arithmetic is counted in Theorem 15.23 and its operands in Proposition 15.26. Theorem 15.29 identifies the point with the classical solution under classical hypotheses, and Theorem 15.33 proves a bound $K(m+1)^{\lambda}$ for one fixed binary machine, with one exponent λ for all problems; its stage programs are constructed in Appendix I.

P.6 What can a calculus of δ's objects derive?

Chapter 16 writes statements about points, functions and compact sets as sentences of one calculus. Theorem 16.8 shows that every derivation holds at both stages of every carrier, and Theorem 16.11 that every derivation of the second-order system of Section 7.10 translates into it. Theorem 16.14 explains why the full calculus has no recursive reading, and Proposition 16.15 shows that it derives every true sentence built from equations over constant carriers, LPO and $\mathrm{WLPO}_{\mathrm{seq}}$ among them. Definitions 16.18 and 16.20 give the computable calculus and its model, Theorem 16.22 shows the model sound, and Theorem 16.25 compiles the second-order system into the computable calculus. In that calculus LPO decides each question without collecting

the answers (Theorem 16.27), and the fan theorems, weak König's lemma, and Heine–Borel, positive bounds and infima and Heine–Cantor for presented sets, written with equations, are not derivable, even from LPO and with excluded middle as a base rule (Theorems 16.32 to 16.36).

P.7 What can a physical record preserve?

The collision lemma leads to exact capacity in Theorem 17.4. Adaptive tests are handled by constructing their full transcripts before applying the same argument; the bound is Theorem 17.7. Approximate targets require the strict separation margin in Theorem 17.9. Theorem 17.13 says which questions a finite record confirms or refutes; Theorems 17.15 and 17.17 and Proposition 17.18 say what records cannot settle and what that costs, Theorem 17.20 treats revisable verdicts, and Theorem 17.21 shows that an exact value can be refuted but never confirmed.

For finite instruments, Theorem 18.3 constructs the whole-space unitary. The naming equivalence is Theorem 18.9, and the supplied-control physical construction is Theorem 18.11. Appendix A gives the finite-control argument. The construction retains the mark, returned state and any finite reference.

The conserving-recorder restriction has different apparatus conditions. Theorem 19.5 combines the noise estimate, the explicit source witness and the recorder spread bound. Its derived count uses the finite identity proved in Appendix B. Chapter 20 brings these conclusions together while retaining the conditions of each.

P.8 What does a comparison cost?

Chapter 21 starts from an arbitrary function on the line. Theorem 21.14 shows that the composition law and the calibration of Definition 21.10 leave exactly one cost, $\cosh t - 1$ in logarithmic form and J of Definition 21.11 for ratios. Its proof combines Proposition 21.12, the duplication argument of Theorem 21.13 and the logarithm of Proposition 21.5, and uses no omniscience and no choice. The functions it needs come first: Theorem 21.8 proves that the sine's period is $2\pi_\delta$. Proposition 21.20 restricts the forcing theorem to the golden orbit, and Theorems 21.22 and 21.23 characterize J, in classical mathematics, as the cost of one fair binary distinction. Theorem 21.25 identifies the ledger with $\mathbb{Z}[\varphi]$, ordered by the sign test of Lemma 21.19; Theorem 21.26 reads its history as a sigma-delta encoder of φ^{-3}; and Theorem 21.27 shows that balances and rates are δ's integers and ratios.

Whether a given cost obeys the law is a decision with a price. Under the reading of maps of Section 21.1 it is computed in Theorems 21.29, 21.30, 21.31 and 21.32, and under unique choice in Corollary 21.33; Table 21.1 collects the answers. Theorem 21.37 and Corollary 21.38 show, in classical mathematics, that a comparison which changes a record with certainty costs $J \geq \frac{1}{4}$, less a drift from the holding of the register, when the reference action is one cycle.

P.9 What does the forcing chain of Recognition Science cost?

Chapter 22 restates each step of the chain on $\mathbb{R}_\delta$ (Section 22.1) and prices it in the sense of Definition 22.1. Records compare for free and points do not, Proposition 22.2. The cost enters through a meter: Theorem 22.5 shows that every meter reads a ratio, and the native reporting theorem, Theorem 22.11, turns the apparatus law into J. Theorem 22.12 replaces continuity by reciprocity and order, and Theorem 22.14 shows that, given the WLPO function, ratio reports alone do not force J. That a ratio other than 1 has positive cost is exactly MP (Theorem 22.16). Golden scaling is free (Theorem 22.20), and Theorem 22.24 prices three readings of the first costly join: free for apart parts, MP for distinct parts and $\mathrm{SE}_{\mathbb{Z}}$ for an integer pairing. The second rung carries the prices of Theorems 22.31, 22.32 and 22.33; the constants are Section 22.9, with the units test in Theorem 22.44. Theorem 22.45 proves the restated numbered chain free, step by step, and Section 22.11 with Table 22.1 sorts all 130 rows: 119 free, ten with an exact price, one bracketed and none open.

P.10 Why three dimensions?

Theorem 23.8 says that a natively embedded circle can be linked in the native D-sphere exactly when $D = 3$, with no omniscience and no choice. That D is the dimension of space, and that a kept pair of records is a linked curve, are Recognition Science's readings in its forcing chain (Chapter 22); the theorem proves the topology, not the readings. Its proof, in Section 23.9, has two halves. For $D = 3$, Theorem 23.34 shows that the equatorial circle is linked: the loop of Lemma 23.33 sweeps four quarter turns, and the linking value of Definition 23.31 is a cocycle (Lemma 23.32). For $D \neq 3$, Proposition 23.23, through Lemmas 23.20, 23.21 and 23.22, shows that a linked circle needs a 3-cycle of the sphere that bounds nothing, and the suspension ladder, Proposition 23.19, shows that every 3-cycle bounds when $D = 2$ or $D \geq 4$; Proposition 23.24 settles dimensions 0 and 1. The tools that work in every dimension are Lemmas 23.9 to 23.14. Theorem 23.38 computes the first homology of the circle as a relation, free, and Theorem 23.40 with Theorem 22.65 prices its data form at exactly the WLPO function.

P.11 What does one account's ledger contain?

Chapter 24 reads the counts of a double-entry account. Theorem 24.2 shows that one account's records fill a light cone, Theorem 24.4 reads J as velocity and the Lorentz factor, and Theorem 24.5 proves proper time superadditive: for words w_1, w_2 whose counts are all positive, $\tau(w_1 w_2)^2$ exceeds $(\tau(w_1) + \tau(w_2))^2$ by $2\tau(w_1)\tau(w_2)$ times a value of J; Proposition 24.6 extends this to many legs. Theorems 24.8 and 24.10 follow the recognition history. Several accounts reach only the taxicab cone (Theorem 24.12). The cone of one time and three space dimensions is the record cone of a coherent binary distinction (Theorem 24.14), whose coherent changes give the Lorentz group (Theorem 24.15); Theorems 24.17, 24.19, 24.20 and 24.22 say where the distinction is carried. Theorem 24.25 identifies the comparison cost of a history with its proper-time deficit. The clock's resolution, its self-recognition band and the action floor of a native step are Theorems 24.29, 24.31 and 24.34. Questions about an unending balance

history are priced in Theorems 24.39, 24.41 and 24.43, on the record questions of Section 17.7. Readings of counts as time and position are identifications throughout.

P.12 What does classical physics cost on the line?

Chapter 25 moves mechanics onto $\mathbb{R}_\delta$. A motion is determined by its initial state (Theorem 25.3). For Kepler's problem, Theorem 25.9 gives the integrals and the conic, Theorem 25.10 solves Kepler's equation with no choice, and Theorems 25.20 and 25.21 show that every bound non-radial motion returns, free. The perihelion direction lies between LLPO and LLPO with countable choice: a perihelion direction for every planar state implies LLPO, and LLPO with countable choice provides it (Theorem 25.13). N bodies keep ten integrals (Theorem 25.25). The relativistic orbit runs from the orbit equation (Theorems 25.35 and 25.37) to the motion in proper time: Theorems 25.42 and 25.43 give the perihelion advance to second order, and Theorems 25.46 and 25.47 the periods in proper time and for a far observer. Certified runs are motions (Theorem 25.50), and Theorem 25.59 gives one conversion certificate for a whole polynomial theory. For the rigid body, Theorem 25.62 proves rotation about an extreme axis stable, and Theorem 25.63 prices the decision of which axis that is at exactly LPO.

Chapter 26 begins with entropy on the line: Theorem 26.5 is the Gibbs principle, Theorem 26.6 prices its decisions, and Theorem 26.7 is the H-theorem. Relative entropy never rises under a stochastic step (Theorem 26.10), and Theorems 26.11 and 26.12 say where and when it falls. Its theorems on voxel ticks are proved for every voxel tick with the properties they state. For Recognition Science's tick the chapter takes five numerical properties of its kernel, (T1)–(T5), as stated inputs that it does not prove, and each result names the ones it uses. From (T1)–(T4), Theorem 26.23 with Proposition 26.25 gives strict decrease on inward-closed sets, Corollary 26.29 a rate from a point start, Corollary 26.37 the pace, and Theorem 26.42 the spine's sharp cutoff; the sharper bounds on the pace and the cutoff hold under hypothesis (R), about the output of a finite computation, which the chapter states and does not prove. Theorem 26.31 is a second law with a rate that is free on the line. The lattice fields are the results from Theorem 26.46 to Theorem 26.50, and Theorem 26.52 prices every sign law.

P.13 What does the cost say about gravity?

Chapter 27 compares J with linearized gravity. On transverse-traceless strains the Fierz–Pauli form is the dispersion $|K|^2$ times twice J's long-wave form (Theorem 27.3), the calibration of J fixes the factor (Theorem 27.4), and such strains exist exactly in three or more space dimensions (Theorem 27.5). The identity holds over every commutative ring (Theorem 27.6), hence on $\mathbb{R}_\delta$, where Theorem 27.8 prices the decisions about a wave. J allows every completion (Theorem 27.10), and locality selects Fierz–Pauli (Theorems 27.11 and 27.12). Curvature and lattice deficits see exactly the non-gauge part (Theorems 27.15 and 27.20), and no nonzero local invariant form is nonnegative (Theorem 27.18). At third order J's cubic vertex breaks gauge invariance, so on the Euclidean Kuhn lattice the second-order equation for a static field built from two harmonic modes has no solution, and β_{PN} does not exist for J's cost (Theorem 27.24). Theorems 27.26, 27.27, 27.28 and 27.29 give Regge's table at every background $\kappa \notin \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$, with a plane-wave form whose only pole is the degenerate background $\kappa = \frac{1}{4}$, and its Hessian,

with true second derivatives, at every Euclidean background $\kappa < \frac{1}{4}$; Theorem 27.30 runs the near-flat chain on $\mathbb{R}_\delta$. The background $\kappa = 1$ is equivalent to a definite history record and is not derived (Theorem 27.41, with Theorems 27.39 and 27.42), and the locality clause is a named premise (Theorems 27.44 and 27.47).

P.14 What do a library's conventions cost?

Chapter 28 prices the conventions of a classical library, stored as data. Theorem 28.3 prices decisions on the line stored as maps. The convention rule, Theorem 28.8, bounds the price of a value set at an excluded point from below, and Theorem 28.9 with Table 28.1 prices eighteen conventions, $0^{-1} = 0$ at the LPO function among them. Choosing representatives costs exactly the WLPO function (Theorem 28.11); the classical linear order costs the WLPO function and the ordered field the LPO function (Theorems 28.14 and 28.15), and Theorem 28.19 prices the complete ordered field. The derivative convention costs exactly MP with a differentiability decision (Theorem 28.24), Theorem 28.26 isolates the Π^0_2 gap above the WLPO function, and monotonicity from a pointwise derivative holds at the WLPO function (Theorem 28.29). The separating models are Theorem 28.32, over HAS, and Theorem 28.39, in a calculus of streams. Theorems 28.42 and 28.45 move whole areas of the library. Section 28.11 compiles classical proofs to programs of Gödel's T (Theorem 28.49), with a compiler of polynomial degree five (Theorem 28.51).

P.15 What does learning cost?

Chapter 29 starts with the logic of records: Theorem 29.2 gives the laws of three shapes of time, Theorem 29.4 prices them at a history's start, and Theorem 29.6 shows that the record tree has intuitionistic logic and one history Gödel–Dummett logic. Chance comes from exchangeable records (Theorems 29.8 and 29.9), and deciding a fixed chance costs WLPO (Theorem 29.11). Brouwer's theorem on a triangle is free approximately (Theorem 29.21) and costs LLPO exactly, given countable choice (Theorem 29.22). Learning a pair of worlds costs $\mathrm{MP}_{\mathrm{bin}}$ or LPO (Theorem 29.24), and Theorem 29.27 prices revisable verdicts. Mistakes, dimensions and samples are counted from Theorem 29.29 to Theorem 29.32, and Theorem 29.35 turns a learner of the zero world and its flags into LPO. On the programs of Gödel's T, learning costs exactly LPO_T (Theorem 29.45), and Theorem 29.48 says which catalogues are learned. Over every system of programs, Theorem 29.51 prices the selection method at Σ_2-LPO, Theorem 29.53 separates decision tables from propositions by unique choice, and Theorem 29.58 reads revision bounds off derived catalogues.

P.16 Which conjectures are settled, and which stay open?

Chapter 30 is proved in classical mathematics throughout. Two golden probes fix a calibration with exponent exactly 2 (Theorem 30.4), and Theorems 30.5 and 30.7 lead to the simultaneous approximation constant. Theorems 30.9 and 30.11 treat two recognizers and the memory that recognition needs. Self-similar ledgers have the rates of Theorem 30.13, keep balanced books exactly as Theorem 30.14 says, and fluctuate as Theorems 30.15 and 30.18 describe. On the native register at zero holding, three golden comparisons change a record with certainty

and two never do, when the reference action is one cycle (Theorem 30.21); on a register of one channel, a certain change costs exactly a third of a turn (Theorem 30.24); and on two or more channels, where entangling tokens are available, a certain change costs exactly a quarter turn per changed channel exactly when every changed channel moves to its low-bit partner (Theorem 30.26). Table 30.1 collects these prices with the floors under holding, and Section 30.5 states every open problem exactly.

P.17 Headline results and what their proofs cite

The table lists, for each headline result, the numbered results that its statement and proof cite, in the order of citation, and, where they hold parts of the argument, the chapters, sections and tables that its proof cites. Following an entry back to its own row, where it has one, retraces the argument.

Result	Its statement and proof cite
Theorem 2.7	Lemma 2.5, Lemma 2.6, Theorem 2.3
Theorem 4.15	Proposition 3.11, Corollary 4.14
Theorem 6.11	Section 20.3.3, Section 20.3.4, Theorem 6.9, Theorem 6.10, Theorem E.9, Lemma E.6, Lemma E.7, Lemma E.11, Lemma E.8
Theorem 7.33	Proposition 7.25, Proposition 7.27, Proposition 7.32, Proposition 7.30
Theorem 7.38	Lemma 7.21, Theorem 7.33, Corollary 7.34
Theorem 7.45	Proposition 7.35, Lemma 7.41, Proposition 7.43, Theorem 7.42, Proposition 7.44, Corollary 7.36
Theorem 8.11	Theorem 6.5
Theorem 12.22	Proposition 10.6, Section F.1
Theorem 12.38	Appendix C, Section F.2, Proposition 12.19, Corollary 4.14, Lemma F.3, Lemma F.5, Lemma F.6, Lemma F.1, Lemma 12.32, Lemma F.2, Theorem 12.33
Theorem 12.63	Theorem 12.22, Lemma F.7, Theorem 12.33, Corollary 12.35, Theorem 12.56, Theorem 12.58
Theorem 12.66	Lemma 7.21
Theorem 12.87	Theorem 12.60, Theorem 12.54, Corollary 12.18, Lemma 12.81, Proposition 12.82
Theorem 13.6	Theorem 13.3, Lemma 12.52, Proposition 12.62
Theorem 13.25	Theorem 13.6, Lemma 13.24, Lemma 12.52, Lemma 13.21
Theorem 13.28	Lemma 13.26, Theorem 13.25, Lemma 13.24, Theorem 13.20
Theorem 14.13	Lemma 14.12, Lemma 14.11
Theorem 14.16	Theorem 14.15, Theorem 14.13, Theorem 14.14, Proposition 12.19, Theorem 12.54
Theorem 14.28	Lemma 14.8, Theorem 14.19, Theorem 14.27, Lemma 14.21, Corollary 12.18
Theorem 15.22	Theorem 15.21, Lemma 15.3
Theorem 15.33	Theorem I.209, Appendix I, Corollary I.59, Lemma 15.36, Lemma I.19, Lemma 15.35, Theorem 15.20, Proposition I.32, Theorem 15.22, Lemma H.5, Lemma I.205, Lemma 15.34
Theorem 16.8	Lemma 16.5
Theorem 16.11	Section J.1, Lemma 16.12, Section J.2

Result	Its statement and proof cite
Proposition 16.15	Section J.3, Remark 16.7, Lemma F.11, Theorem 16.27, Proposition 13.4, Theorem 12.72, Theorem 13.25, Theorem 13.28, Corollary 13.30, Theorem 13.32, Theorem 16.8
Theorem 16.27	Theorem 16.22, Remark 16.24, Lemma J.2, Proposition 16.21, Theorem 12.22, Lemma 12.15, Lemma F.11
Theorem 16.34	Lemma 16.31, Lemma 16.29, Proposition 16.21
Theorem 17.11	Lemma 17.2, Lemma 17.10
Theorem 18.11	Theorem 18.10, Proposition 18.4
Theorem 19.5	Lemma 19.3, Lemma 19.4
Theorem 20.26	Corollary 20.19, Proposition 20.21, Proposition 20.23, Theorem 20.22, Theorem 20.24, Theorem 20.25, Lemma 20.20, Theorem 20.18
Theorem 20.39	Theorem 8.11, Theorem 20.38, Theorem 20.36, Theorem 20.32, Lemma 20.33
Theorem 20.42	Theorem 20.41, Appendix M, Section M.1, Section M.5
Theorem 20.44	Appendix L, Section L.1, Section L.8, Theorem 20.39, Theorem 20.41, Proposition 20.43, Theorem 20.42
Theorem 20.46	Proposition 20.45, Appendix L, Section L.9, Section L.12
Theorem 20.47	Appendix M, Section M.6, Theorem 20.39
Theorem 20.54	Theorem 20.51, Theorem 20.53, Theorem 20.50
Theorem 20.66	Theorem 20.63, Theorem 6.11, Theorem 20.59, Corollary 20.65, Theorem 8.11, Theorem 6.10
Theorem 21.8	Lemma 21.7
Theorem 21.14	Proposition 21.12, Theorem 21.13, Proposition 21.5
Theorem 21.26	Theorem 21.25
Corollary 21.33	Theorem 21.31, Theorem 21.30, Theorem 21.32
Theorem 21.37	Lemma 21.36
Theorem 22.11	Theorem 22.5, Lemma 12.52
Theorem 22.16	Proposition 22.15, Corollary 12.18
Theorem 22.20	Lemma 22.19, Lemma 22.18
Theorem 22.24	Proposition 22.15, Lemma 22.18, Table 12.2, Lemma 22.23
Theorem 22.45	Proposition 22.47, Theorem 21.14, Theorem 22.12, Proposition 22.37, Theorem 22.20, Proposition 22.48, Proposition 22.49, Chapter 23, Proposition 24.44, Proposition 22.50, Proposition 22.15, Proposition 22.51, Theorem 22.52, Theorem 22.53
Theorem 23.8	Theorem 23.34, Proposition 23.19, Lemma 23.22, Proposition 23.23, Lemma 23.20, Lemma 23.21, Proposition 23.24
Theorem 23.34	Lemma 23.33, Remark 23.6, Lemma 23.32
Theorem 23.38	Lemma 23.28, Lemma 23.30, Lemma 23.32, Lemma 23.33, Lemma 23.10, Lemma 23.11, Lemma 23.26
Theorem 23.40	Theorem 23.38, Lemma 23.20, Lemma 23.26
Theorem 24.5	Theorem 24.4
Theorem 24.15	Theorem 24.14
Theorem 24.25	Theorem 24.5
Theorem 24.39	Section 17.7, Lemma 24.38, Proposition 7.35, Theorem 17.13, Theorem 17.15
Theorem 24.43	Lemma 16.29
Theorem 25.20	Theorem 25.3, Theorem 25.9, Theorem 25.11

Result	Its statement and proof cite
Theorem 25.13	Theorem 25.9, Corollary 12.18, Theorem 12.54
Theorem 25.42	Proposition 25.41, Theorem 25.37, Lemma 25.33, Lemma 25.34
Theorem 25.47	Lemma 25.6, Proposition 25.41, Lemma 25.45, Theorem 25.46, Table 25.2, Lemma 25.36
Theorem 25.63	Corollary 12.18, Chapter 12, Theorem 12.7, Lemma 12.5
Theorem 26.6	Theorem 26.5, Corollary 12.18
Theorem 26.23	Theorem 26.10, Theorem 26.15
Corollary 26.29	Theorem 26.28, Proposition 26.26, Theorem 26.10
Theorem 26.42	Theorem 26.36, Lemma 26.35, Corollary 26.37
Theorem 26.52	Table 12.2
Theorem 27.8	Theorem 27.6, Corollary 12.18, Proposition 12.19
Theorem 27.12	Theorem 27.11
Theorem 27.24	Table 27.2, Theorem 27.15, Theorem 27.6
Theorem 27.29	Theorem 27.26, Theorem 27.27, Theorem 27.28
Theorem 27.41	Table 27.4, Theorem 27.40, Theorem 27.39
Theorem 28.9	Proposition 12.19, Lemma 28.2, Theorem 28.3, Corollary 12.18, Section 21.2, Theorem 28.8
Theorem 28.11	Theorem 28.10, Theorem 28.3, Theorem 12.61, Lemma 12.15
Theorem 28.15	Table 28.1, Theorem 28.9, Theorem 28.3, Theorem 28.14, Corollary 12.18, Proposition 12.19
Theorem 28.24	Theorem 28.8, Theorem 28.23, Lemma 28.18, Theorem 28.3, Proposition 28.4
Theorem 28.29	Theorem 28.11, Lemma 28.28, Lemma 28.27
Theorem 29.6	Theorem 29.2
Theorem 29.11	Proposition 12.19, Theorem 29.9, Lemma 12.5, Corollary 12.18, Theorem 12.11
Theorem 29.22	Theorem 12.54, Theorem 29.21, Theorem 12.72
Theorem 29.45	Theorem 29.35
Theorem 29.48	Theorem 29.24, Theorem 29.47
Theorem 30.4	Lemma 30.2, Lemma 30.3
Theorem 30.21	Chapter 21, Chapter 24, Lemma 30.20
Theorem 30.24	Theorem 30.22, Lemma 30.23
Theorem 30.26	Lemma 30.20, Theorem 30.25

Appendix Q

The price table of the mathematics physics uses

This appendix lists, one statement to a row, what the mathematics that physics uses costs on the line $\mathbb{R}_\delta$. Each row reads one statement of classical analysis, algebra, topology or probability on $\mathbb{R}_\delta$ exactly as its statement says; where the reading narrows the classical statement (to an interval, to a dimension, to a uniform derivative, to the value half of an eigenvalue theorem, to a finite probability space), the statement says so. Each row gives its number, its title and statement, the physics that relies on it ("Used for"), and its price, followed by "exact" or, for an open row, by the bounds that are proved; where this book proves the row, a reference follows. Section 28.12 of Chapter 28 explains what the table shows.

How to read a price

free
: No omniscience principle, and no choice beyond the *free allowance*: countable choice at pairs of ratios. If for every natural number n some pair of ratios (p, q) has a property $R(n, p, q)$, the allowance gives a sequence of pairs (p_n, q_n) with $R(n, p_n, q_n)$ for every n. It makes every located cut the cut of a point (Theorem 12.33) and so every Cauchy sequence of points converge (Corollary 12.35). A natural number, a yes-or-no answer and a finite table can each be written as a pair of ratios, so the allowance also chooses one natural number or one answer for every n, whether or not the answer is unique: for a unique bit b it chooses the pair (b, b). It chooses a regular sequence for every point of a sequence of points, by choosing the rational approximations, and so collects points that are each determined uniquely into a sequence. The free rows whose proofs form a limit point say so after their price.

LPO, WLPO, LLPO, MP
: The omniscience principles of Chapter 12; MP is Markov's principle, written $\mathrm{MP}_{\mathrm{bin}}$ in Corollary 12.18.

EM, WEM
: Excluded middle, $P \vee \neg P$, and weak excluded middle, $\neg P \vee \neg\neg P$, for every proposition P.

MP$^\vee$
: The disjunctive Markov principle: if two Σ^0_1 statements are not both false, one of them is not false. Markov's principle gives it.

LPO function, WLPO function
: LPO or WLPO given as one map that decides every binary sequence (Definition 21.1).

FAN
: The decidable fan theorem of Definition 13.2. The *full fan theorem*, named in some bounds, is the fan theorem for bars given by arbitrary properties.

BD-N
: Ishihara's boundedness principle [60]: every countable set of natural numbers that is pseudobounded is bounded. A set S of natural numbers is *pseudobounded* when every sequence $s_0, s_1, \ldots$ of members of S has $s_n < n$ for all large n.

AC_ω
: Countable choice at every type: if for every n some object x has $R(n, x)$, there is a sequence (x_n) with $R(n, x_n)$ for every n. A price that names AC_ω means that the proof uses countable choice (an upper bound), for instance to pick a member of a set for every n, and that the table has not shown the free allowance to suffice; the choice is not shown to be needed.

exact
: Both directions hold: the statement implies the principles named with no choice principle, and the principles named, with the free allowance and with AC_ω where the price names it, imply the statement. This is exactness in the table's sense; "costs exactly" in Section 21.1 allows no choice in either direction. A row with a reference is proved in this book at the result it names; most rows without one are not proved in this book. For a row priced AC_ω alone, "exact" asserts nothing on the lower side: no principle is shown to be needed.

open
: No exact price among the letters above; the bounds that are proved follow the price, with any exact equivalence to a principle that is not one of the letters, which the row then defines.

The table has 1467 rows, one for each statement: 1166 are free, 279 have an exact price, and 22 are open.

Q.1 Conventions at excluded points

1. Total inverse with $0^{-1} = 0$, apartness guard. A total inverse function on the line that sends 0 to 0 and inverts every point apart from 0, that is, at a known positive distance from 0. *Used for:* dividing by distances, masses or times; the value at 0 is never used physically. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.9(1).

2. Total inverse with $0^{-1} = 0$, nonzero guard. A total inverse function on the line that sends 0 to 0 and inverts every point that is not equal to 0. *Used for:* dividing by distances, masses or times; the value at 0 is never used physically. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.9(1).

3. Division by zero convention, $a/0 = 0$. A total division on the line with $a/0 = 0$ for every a, and $(a/b)\cdot b = a$ whenever b is apart from 0. *Used for:* ratios such as speed as distance over time and density as mass over volume. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.9(1).

4. Negative powers with zero sent to zero. For each whole number $n \geq 1$, a total function on the line that sends 0 to 0 and every point x apart from 0 to $1/x^n$. *Used for:* inverse-power laws such as $1/r^2$ for gravity and the Coulomb force. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.9(1).

5. Logarithm convention, $\log 0 = 0$. A total logarithm on the line with $\log 0 = 0$ and $\exp(\log x) = |x|$ at every point x apart from 0. *Used for:* entropy $S = k \log W$, decibel scales and logarithmic potentials. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.9(2).

6. Zero to a real power, with $0^0 = 1$. A total function on the line giving 0 to the power y: the value 1 at $y = 0$ and the value 0 at every y not equal to 0. *Used for:* power laws evaluated at zero, and the term $0^0 = 1$ in power series. *Price:* **WLPO function**; exact (equivalent to WLPO as a function). Proved in Theorem 28.9(3).

7. Whole-number powers of any base. A power function on the line defined at every base x and whole number n, with $x^0 = 1$ and $x^{n+1} = x^n \cdot x$; no point

is excluded. *Used for:* polynomial laws such as kinetic energy $\frac{1}{2}mv^2$ and spring energy $\frac{1}{2}kx^2$. *Price:* **free**; exact. Proved in Theorem 28.9(5).

8. Square root with $\sqrt{x} = 0$ for $x \le 0$. A total function on the line that is never negative, squares to x at every $x \ge 0$, and equals 0 at every $x \le 0$. *Used for:* lengths of vectors, speeds from kinetic energies, standard deviations. *Price:* **free**; exact. Proved in Theorem 28.9(5).

9. Sign function with sign 0 = 0. A total function on the line equal to 1 at every positive point, −1 at every negative point and 0 at 0. *Used for:* direction of friction and drag forces opposing the velocity. *Price:* **WLPO function**; exact (equivalent to WLPO as a function). Proved in Theorem 28.9(4).

10. Derivative convention: zero where none exists. An operator on functions of the line that returns the derivative of f at x where one exists and returns 0 where f has no derivative at x. *Used for:* writing force as minus the slope of a potential, even at kinks. *Price:* **open**; at least the LPO function; at most MP together with the WEM function; exactly MP together with a differentiability decision, one map that decides, for every function f on the line and every point x, whether it is impossible that f has no derivative at x (a principle that is not one of the letters). Proved in Theorem 28.24.

11. Truncated subtraction of whole numbers. A subtraction on whole numbers with $(a - b) + b = a$ when $b \le a$, and $a - b = 0$ when $a \le b$. *Used for:* lowering occupation numbers of quantum states, which stop at zero. *Price:* **free**; exact. Proved in Theorem 28.9(5).

12. Integer to whole number, negatives sent to zero. A map from the integers to the whole numbers that keeps every integer $z \ge 0$ and sends every $z \le 0$ to 0. *Used for:* clamping counts at zero, such as particle numbers that cannot go negative. *Price:* **free**; exact. Proved in Theorem 28.9(5).

13. Floor function on the line. A function that assigns to every point x of the line an integer m with $m \le x < m + 1$. *Used for:* finding which lattice cell or histogram bin holds a position or reading. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.3(3).

14. Decidable equality of points, as stored data. A stored procedure that decides, for any two points of the line, whether they are equal. *Used for:* branching on whether two energy levels are equal, as in degenerate perturbation theory. *Price:* **WLPO function**; exact (equivalent to WLPO as a function). Proved in Theorem 28.3(2).

15. Decidable $\le$ between points, as stored data. A stored procedure that decides, for any two points x and y of the line, whether $x \le y$. *Used for:* piecewise laws, such as a potential given by one formula for $x \le a$. *Price:* **WLPO function**; exact (equivalent to WLPO as a function). Proved in Theorem 28.3(2).

16. Decidable < between points, as stored data. A stored procedure that decides, for any two points x and y of the line, whether $x < y$. *Used for:* threshold tests, such as whether an energy lies below a barrier. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.3(3).

17. Maximum and absolute value on the line. The maximum of two points of the line lies above both and below every common upper bound; $|x|$ is the maximum of x and $-x$. *Used for:* absolute errors, magnitudes, and ramp functions such as $\max(0, x)$. *Price:* **free**; exact.

Q.2 Limits, continuity and derivatives

18. Taylor's remainder bound. Let f be continuous on [0, 1], with derivatives of orders 1 to n that are continuous on [0, 1], and with an (n+1)-th derivative at every point strictly inside (0, 1) that is at most C in absolute value. Then for every x in [0, 1] the error of the Taylor polynomial of degree n at 0 is at most $C\,x^{n+1}/n!$. This is weaker than the classical Lagrange bound, which has $(n + 1)!$ in place of $n!$. *Used for:* Error bars for expansions in a small parameter: small angles, perturbation series, linearized equations of motion and lattice expansions, with an error of order x^{n+1}. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

19. Lagrange remainder at a point of the closed interval. Let f be a real function on [0, 1] whose derivatives up to order n + 1 exist uniformly on [0, 1]. Then the difference between f(1) and the value at 1 of its degree-n Taylor polynomial at 0 equals the (n + 1)-th derivative of f at some point ξ of the closed interval [0, 1] (an end point allowed) divided by the factorial $(n + 1)!$. *Used for:* The mean-value form of an expansion error when only some point of the closed interval is needed for the next derivative, as in worst-case bounds that take the largest value of that derivative. *Price:* **LLPO + $\mathbf{AC}_{\omega}$**; exact (the upper bound uses countable choice, not shown to be needed).

20. Rolle's theorem with an interior point. Let f be uniformly continuous on [0, 1], with a derivative

f′(x) at every point x strictly inside (0, 1). If f(0) = f(1), there is a point c strictly inside (0, 1), at a known positive distance from both ends, where f′(c) = 0. *Used for:* Finding an instant at which a quantity that ends where it started has zero rate of change, such as the turning point of a thrown body or an extremum of energy along a path; it is also the step behind the mean value theorem. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

21. Mean value theorem with an interior point. Let f be uniformly continuous on [0, 1], with a derivative f′(x) at every point x strictly inside (0, 1). Then there is a point c strictly inside (0, 1), at a known positive distance from both ends, where f′(c) = f(1) − f(0). *Used for:* Saying that the average velocity over a time interval equals the instantaneous velocity at some instant inside it, as in estimates of how far a moving body travels. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

22. Cauchy's mean value theorem. Let f and g be uniformly continuous on [0, 1], with derivatives f′(x) and g′(x) at every point x strictly inside (0, 1). Then there is a point c strictly inside (0, 1), at a known positive distance from both ends, where (g(1) − g(0))·f′(c) = (f(1) − f(0))·g′(c). *Used for:* Comparing two quantities that change together, such as two coordinates of a moving body traced against time; it is the classical proof of the rule for limits of ratios of vanishing quantities and of the Lagrange error formula for Taylor expansions. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

23. Taylor's theorem with Lagrange remainder. Let $f_0 = f, f_1, \ldots, f_{n+1}$ be real functions such that $f_0, \ldots, f_n$ are uniformly continuous on [0, 1] and, for each $i \le n$, $f_{i+1}(x)$ is the derivative of f_i at every point x strictly inside (0, 1). Then f(1) minus the value at 1 of the degree-n Taylor polynomial of f at 0, built from $f_0(0), \ldots, f_n(0)$, equals $f_{n+1}(\xi)/(n+1)!$ at some point ξ strictly inside (0, 1), at a known positive distance from both ends. *Used for:* The exact form of the error of a truncated expansion, with the next derivative taken at an unknown interior point: small-oscillation approximations of a potential near its minimum, perturbation series and finite-difference formulas of numerical simulation. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

24. Inverse function theorem in one variable. Let f be a real function with a strict derivative d at the point a (for every accuracy there is a ball around a on which |f(x) − f(y) − d·(x − y)| is at most that accuracy times |x − y|), and let d not be 0. Then f is one-to-one on some ball around a, the image of every small enough ball around a contains a ball around f(a), and every function g with g(f(x)) = x on some ball around a has the strict derivative 1/d at f(a). *Used for:* Inverting a response curve: solving an equation of state for pressure given volume or the reverse, Legendre transforms, and changes of variable that swap which quantity is the independent one. *Price:* **MP**; exact (limit points use the completeness of the line: countable choice of rational pairs).

25. Inverse function theorem with a known reciprocal slope. Let f be a real function with a strict derivative d at the point a (for every accuracy there is a ball around a on which |f(x) − f(y) − d·(x − y)| is at most that accuracy times |x − y|), and let d have a reciprocal e, d·e = 1. Then f is one-to-one on some ball around a, the image of every small enough ball around a contains a ball around f(a), and every function g with g(f(x)) = x on some ball around a has the strict derivative e at f(a). *Used for:* inverting a response curve, Legendre transforms and changes of variable when the slope is known to be bounded away from zero. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

26. Local left inverses are strictly differentiable. Let f be a real function with a strict derivative d at the point a, with d not equal to 0. Then every function g with g(f(x)) = x for all x in some ball around a has the strict derivative 1/d at f(a): for every accuracy there is a ball around f(a) on which |g(u) − g(v) − (u − v)/d| is at most that accuracy times |u − v|. *Used for:* Differentiating an inverted calibration or equation of state, for example the rate of change of temperature with respect to entropy once entropy as a function of temperature is known. *Price:* **MP**; exact (limit points use the completeness of the line: countable choice of rational pairs).

27. Derivative of an inverse function. Let g be a real function continuous at a, let f have the derivative d at g(a) with d not equal to 0, and let f(g(y)) = y for every y near a. Then g has the derivative 1/d at a. *Used for:* The derivative of any inverse function: temperature as a function of energy from energy as a function of temperature, response functions and the derivatives of inverse trigonometric and logarithmic functions. *Price:* **MP**; exact.

28. Derivative of an inverse with a known reciprocal. Let g be a real function continuous at a, let f have the derivative d at g(a), let d have a reciprocal e (d·e = 1), and let f(g(y)) = y for every y near a. Then g has the derivative e at a. *Used for:* derivative of an inverse function when the slope is known to be bounded away from zero: inverted calibrations and equations of state. *Price:* **free**; exact.

29. Implicit function theorem for a nonzero partial derivative. Let φ(x, y) be a real function of two real variables with a strict derivative at (a, b), with partial derivative p in x and d in y, and let d not be 0. Then there are intervals around a and around b such that for every x in the first there is exactly one y in the second with φ(x, y) = φ(a, b), and every function ψ that picks that y for each x has the strict derivative −p/d at a. *Used for:* Solving a constraint for one variable: an equation of state F(p, V, T) = 0 solved for V, constraint surfaces and stationarity conditions in mechanics. *Price:* **MP**; exact (limit points use the completeness of the line: countable choice of rational pairs).

30. Implicit function theorem for an invertible partial derivative. Let φ(x, y) be a real function of two real variables with a strict derivative at (a, b), with partial derivative p in x and d in y, and let d have a reciprocal e, d·e = 1. Then there are intervals around a and around b such that for every x in the first there is exactly one y in the second with φ(x, y) = φ(a, b), and every function ψ that picks that y for each x has the strict derivative −p·e at a. *Used for:* solving a constraint for one variable when its partial derivative is known to be bounded away from zero: equations of state, constraint surfaces, stationarity conditions. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

31. L'Hôpital's rule with a positive denominator slope. Let f and g be real functions on [0, 1] with derivatives f′ and g′ uniform on the closed interval [0, 1], with f(0) = g(0) = 0 and g′(x) positive, at a known positive distance from 0, for every x in (0, 1]. If f′(x)/g′(x) tends to a number l as x decreases to 0, then f(x)/g(x) tends to l as x decreases to 0. *Used for:* Limits of ratios of two vanishing quantities when it is known that the denominator's rate keeps one sign, as in small-angle and small-coupling limits of response functions. *Price:* **free**; exact.

32. L'Hôpital's rule at zero from the right. Let f and g be real functions on [0, 1] with derivatives f′ and g′ uniform on the closed interval [0, 1], with f(0) = g(0) = 0 and g′(x) not equal to 0 for every x in (0, 1]. If f′(x)/g′(x) tends to a number l as x decreases to 0, then f(x)/g(x) tends to l as x decreases to 0. *Used for:* Evaluating limits of ratios in which numerator and denominator both vanish, such as sin(x)/x near 0 or (1 − cos(x))/x^2, which appear in small-angle and small-amplitude approximations, in low-frequency and low-temperature limits of response functions, and at removable singularities of physical formulas. *Price:* $\mathbf{MP}^{\vee}$ **+** $\mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

33. Convexity from a nondecreasing derivative. If f is continuous on [0, 1], differentiable at every point strictly inside (0, 1), and its derivative is nondecreasing there, then f(a x + b y) is at most a f(x) + b f(y) whenever x and y lie in [0, 1], a and b are nonnegative and a + b = 1. *Used for:* Showing that a free energy is convex or an entropy concave (thermodynamic stability) and setting up Legendre transforms. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

34. Convexity from a nonnegative second derivative. If f is continuous on [0, 1], differentiable at every point strictly inside (0, 1), and its derivative is itself differentiable at every point strictly inside with nonnegative derivative there, then f(a x + b y) is at most a f(x) + b f(y) whenever x and y lie in [0, 1], a and b are nonnegative and a + b = 1. *Used for:* Thermodynamic stability (positive heat capacity and positive compressibility mean a convex energy) and convex potentials in variational problems. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

35. Limit of a quotient of sequences. If the sequences x(n) and y(n) tend to a and b and b is not 0, then x(n)/y(n) tends to a/b. *Used for:* Every measured ratio taken to a limit: susceptibilities, rates per unit time and normalised probabilities. *Price:* **MP**; exact.

36. Limit of the inverses of a sequence. If the sequence y(n) tends to b and b is not 0, then 1/y(n) tends to 1/b. *Used for:* Reciprocals of converging quantities: effective masses, resistances and normalisation factors. *Price:* **MP**; exact.

37. Alternating series test. If a_0, a_1, a_2, … is a sequence of real numbers that never increases and tends to 0, the alternating series $a_0 - a_1 + a_2 - a_3 + \dots$ converges. *Used for:* Alternating perturbation and virial series; the error of a partial sum is at most the first omitted term. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

38. Alternating series have Cauchy partial sums. If a_0, a_1, a_2, … is a sequence of real numbers that never increases and tends to 0, the partial sums of the alternating series $a_0 - a_1 + a_2 - a_3 + \dots$ form a Cauchy sequence. *Used for:* Controlling an alternating series without knowing its limit in advance: numerical summation of alternating expansions. *Price:* **free**; exact.

39. Even alternating partial sums lie below the sum. If a_0, a_1, a_2, … is a sequence of real numbers that never increases and the alternating series $a_0 - a_1 + a_2 - a_3 + \dots$ converges to L, then every partial sum with an even number of terms is at most L. *Used for:* One-sided error bars for alternating series: a

computed partial sum bounds the true value from one side. *Price:* **free**; exact.

40. Odd alternating partial sums lie above the sum. If a_0, a_1, a_2, ... is a sequence of real numbers that never increases and the alternating series $a_0 - a_1 + a_2 - a_3 + \ldots$ converges to L, then every partial sum with an odd number of terms is at least L. *Used for:* One-sided error bars for alternating series: a computed partial sum bounds the true value from the other side. *Price:* **free**; exact.

41. Alternating series test for nondecreasing terms. If a_0, a_1, a_2, ... is a sequence of real numbers that never decreases and tends to 0, the alternating series $a_0 - a_1 + a_2 - a_3 + \ldots$ converges. *Used for:* Alternating series whose terms approach zero from below, such as expansions with a negative leading sign. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

42. Cauchy partial sums for nondecreasing alternating terms. If a_0, a_1, a_2, ... is a sequence of real numbers that never decreases and tends to 0, the partial sums of the alternating series $a_0 - a_1 + a_2 - a_3 + \ldots$ form a Cauchy sequence. *Used for:* alternating series whose terms rise to zero, used without knowing the limit in advance. *Price:* **free**; exact.

43. Even partial sums lie above the sum, nondecreasing terms. If a_0, a_1, a_2, ... is a sequence of real numbers that never decreases and the alternating series $a_0 - a_1 + a_2 - a_3 + \ldots$ converges to L, then every partial sum with an even number of terms is at least L. *Used for:* one-sided error bars of alternating series. *Price:* **free**; exact.

44. Odd partial sums lie below the sum, nondecreasing terms. If a_0, a_1, a_2, ... is a sequence of real numbers that never decreases and the alternating series $a_0 - a_1 + a_2 - a_3 + \ldots$ converges to L, then every partial sum with an odd number of terms is at most L. *Used for:* one-sided error bars of alternating series. *Price:* **free**; exact.

45. Cauchy condensation test. Let f(0), f(1), f(2), ... be nonnegative real numbers that never increase from f(1) on. Then the series of the f(n) converges exactly when the series of the numbers $2^k \cdot f(2^k)$ converges. *Used for:* Convergence of sums such as the sum of $1/n^p$ over interaction terms, density-of-states sums and lattice sums. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

46. Polynomial times geometric decay tends to zero. For every natural number k and every real r with $|r| < 1$, the numbers $n^k \cdot r^n$ tend to 0 as n grows. *Used for:* Boltzmann weights beat any polynomial density of states; moments of geometric distributions such as Bose and Planck weights are finite. *Price:* **free**; exact.

47. Polynomial times nonnegative geometric decay tends to zero. For every natural number k and every real r with $0 \le r < 1$, the numbers $n^k \cdot r^n$ tend to 0 as n grows. *Used for:* Boltzmann weights beat any polynomial density of states; moments of geometric (Bose, Planck) distributions are finite. *Price:* **free**; exact.

48. Linear times geometric decay tends to zero. For every real r with $|r| < 1$, the numbers $n \cdot r^n$ tend to 0 as n grows. *Used for:* the mean occupation of a geometric (Bose, Planck) distribution is finite; Boltzmann weights beat a linearly growing density of states. *Price:* **free**; exact.

49. Linear times nonnegative geometric decay tends to zero. For every real r with $0 \le r < 1$, the numbers $n \cdot r^n$ tend to 0 as n grows. *Used for:* the mean occupation of a geometric (Bose, Planck) distribution is finite; Boltzmann weights beat a linearly growing density of states. *Price:* **free**; exact.

50. Dirichlet's test for series. Let f(0), f(1), f(2), ... be real numbers that never increase and tend to 0, and let z(0), z(1), z(2), ... be real numbers whose partial sums $z(0) + \ldots + z(n-1)$ are all at most b in absolute value. Then the partial sums of the series $f(0) \cdot z(0) + f(1) \cdot z(1) + f(2) \cdot z(2) + \ldots$ form a Cauchy sequence. *Used for:* real oscillating terms with slowly decreasing amplitudes: conditionally convergent lattice sums, Fourier cosine and sine series at a point, Abel summation. *Price:* **free**; exact.

51. Dirichlet's test for nondecreasing factors. Let f(0), f(1), f(2), ... be real numbers that never decrease and tend to 0, and let z(0), z(1), z(2), ... be real numbers whose partial sums $z(0) + \ldots + z(n-1)$ are all at most b in absolute value. Then the partial sums of the series $f(0) \cdot z(0) + f(1) \cdot z(1) + f(2) \cdot z(2) + \ldots$ form a Cauchy sequence. *Used for:* real oscillating terms with slowly varying amplitudes: conditionally convergent lattice sums, Abel summation. *Price:* **free**; exact.

52. Cauchy product of absolutely convergent series with known sums. If the series a(0) + a(1) + ... and b(0) + b(1) + ... both converge absolutely, with sums s and t, then the series with terms $c(n) = a(0)\,b(n) + a(1)\,b(n-1) + \ldots + a(n)\,b(0)$ converges to s t. *Used for:* Multiplying generating functions and partition functions of independent subsystems; convolving independent probability weights. *Price:* **free**; exact.

53. Sum of the Cauchy product of absolutely convergent series. If the series of the numbers |a(n)| and of the numbers |b(n)| converge, then the series of the real numbers a(n) and b(n) converge, to s and t, and their Cauchy product, whose n-th term is

a(0)·b(n) + a(1)·b(n − 1) + … + a(n)·b(0), converges to s·t. *Used for:* multiplying generating functions and partition functions of independent subsystems; convolving independent weights. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

54. Cauchy product converges absolutely. If the series a(0) + a(1) + … and b(0) + b(1) + … both converge absolutely, then the series with terms c(n) = a(0) b(n) + a(1) b(n−1) + … + a(n) b(0) converges absolutely. *Used for:* The product series can be summed in any order: composing independent weights without worrying about the order of the terms. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

55. Rearranging an unconditionally convergent series keeps its sum. Let σ be a one-to-one correspondence of the indices 0, 1, 2, … with themselves. A series of real numbers f(0), f(1), f(2), … has the unconditional sum s (the sums over finite sets of indices tend to s as the sets grow) exactly when the rearranged series f(σ(0)), f(σ(1)), f(σ(2)), … has the unconditional sum s. *Used for:* a lattice sum, once its sites are listed, has one value whether it is summed shell by shell or cube by cube. *Price:* **free**; exact.

56. Rearranging preserves unconditional summability. Let σ be a one-to-one correspondence of the indices 0, 1, 2, … with themselves. A series of real numbers f(0), f(1), f(2), … has an unconditional sum (the sums over finite sets of indices tend to some number as the sets grow) exactly when the rearranged series f(σ(0)), f(σ(1)), f(σ(2)), … has one. *Used for:* a lattice sum, once its sites are listed, that converges unconditionally in one order converges in every order. *Price:* **free**; exact.

57. Schlömilch condensation test. Let f(0), f(1), f(2), … be nonnegative real numbers that never increase from f(1) on, and let u(0) < u(1) < u(2) < … be positive whole numbers whose gaps grow at most by a fixed whole factor C ≥ 1: u(n + 2) − u(n + 1) ≤ C·(u(n + 1) − u(n)) for every n. Then the series of the f(n) converges exactly when the series of the numbers (u(k + 1) − u(k))·f(u(k)) converges. *Used for:* Convergence tests for sums over sparse shells, such as squares or powers, in lattice sums and density-of-states sums. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

58. Interior extremum of a function returning to its start. Let f be uniformly continuous on [0, 1] (one accuracy rate for all points of the interval) with f(0) = f(1). Then at some point c strictly inside (0, 1), at a known positive distance from both ends, f takes its minimum or its maximum over [0, 1]. *Used for:* Variational principles: an extremum strictly inside for a quantity that returns to its starting value, such as the energy along a closed path. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

59. Interior local extremum of a function returning to its start. Let f be uniformly continuous on [0, 1] (one accuracy rate for all points of the interval) with f(0) = f(1). Then at some point c strictly inside (0, 1), at a known positive distance from both ends, f has a local minimum or a local maximum: on some interval around c, f is never below f(c), or never above f(c). *Used for:* Equilibrium points of an energy that returns to its starting value: a stable or unstable equilibrium lies strictly inside. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

60. Uniform limits of continuous functions are continuous. Let f_0, f_1, f_2, … be real functions on the line that converge uniformly to f (for every accuracy there is an index beyond which every f_n is within that accuracy of f at every point), and suppose that beyond every index some f_n is continuous. Then f is continuous. *Used for:* Continuity of limits of approximations: perturbative and lattice approximants that converge uniformly keep the continuity of the exact quantity. *Price:* **free**; exact.

61. Uniform limits of uniformly continuous functions are uniformly continuous. Let f_0, f_1, f_2, … be real functions on the line that converge uniformly to f (for every accuracy there is an index beyond which every f_n is within that accuracy of f at every point), and suppose that beyond every index some f_n is uniformly continuous. Then f is uniformly continuous. *Used for:* continuity of limits of approximations: perturbative and lattice approximants converging uniformly. *Price:* **free**; exact.

62. Uniform limits keep uniform continuity on the unit interval. Let f_0, f_1, f_2, … be real functions that converge uniformly to f on [0, 1] (for every accuracy there is an index beyond which every f_n is within that accuracy of f at every point of [0, 1]), and suppose that beyond every index some f_n is uniformly continuous on [0, 1]. Then f is uniformly continuous on [0, 1]. *Used for:* continuity of limits of approximations on a bounded parameter range: perturbative and lattice approximants converging uniformly. *Price:* **free**; exact.

63. Weierstrass M-test for series of functions. Let g_0, g_1, g_2, … be real functions on the line with $|g_n(x)| \le M_n$ for every n and x, where the series $M_0 + M_1 + M_2 + \dots$ converges. Then the series $g_0(x) + g_1(x) + g_2(x) + \dots$ converges at every x, and uniformly: for every accuracy there is a number of terms beyond which every partial sum is within that accuracy of the

sum, at every x. *Used for:* Power series, Fourier series and lattice series of fields that converge uniformly, so that their sums can be differentiated and integrated term by term. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

64. Comparison test with absolute values. If a(0), a(1), a(2), ... are real numbers with $|a(n)| \le b(n)$ for every n, and the series b(0) + b(1) + b(2) + ... converges, then the series a(0) + a(1) + a(2) + ... converges. *Used for:* Comparing a physical series with a known one: bounding the terms of an expansion by a convergent series proves that the expansion has a sum. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

65. Continuity of the sum of a dominated series. Let g_0, g_1, g_2, ... be continuous real functions on the line with $|g_n(x)| \le M_n$ for every n and x, where the series $M_0 + M_1 + M_2 + \dots$ converges. Then the series $g_0(x) + g_1(x) + g_2(x) + \dots$ converges at every x, and its sum depends continuously on x. *Used for:* Continuity of sums over modes: partition functions and mode sums that depend on temperature or field, with terms bounded by a convergent series. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

66. Derivative of a difference. If f has derivative a at x and g has derivative b at x, then f − g has derivative a − b at x. *Used for:* Linearity of every rate computation: the rate of change of a difference of two quantities is the difference of their rates. *Price:* **free**; exact.

67. Derivative of a constant multiple. If f has derivative a at x and c is a constant, then c f has derivative c a at x. *Used for:* Linearity of every rate computation: scaling a quantity by a fixed charge, mass or conversion factor scales its rate. *Price:* **free**; exact.

68. Derivative of a finite sum. If each of f(0), ..., f(n−1) has a derivative at x, then the sum f(0) + ... + f(n−1) has at x the derivative equal to the sum of the derivatives. *Used for:* Differentiating mode sums and finite expansions term by term, as in the force on a particle from a finite sum of potentials. *Price:* **free**; exact.

69. Limit of a difference. If x(n) tends to L and y(n) tends to M, then x(n) − y(n) tends to L − M. *Used for:* Taking differences of converging quantities to a limit: energy differences, transition frequencies and the subtraction of a reference value in regularized calculations. *Price:* **free**; exact.

70. Limit of a negative. If x(n) tends to L, then −x(n) tends to −L. *Used for:* Reversing signs inside limits: charges and currents of opposite sign and potentials written with a minus sign. *Price:* **free**; exact.

71. Limit of a power. If x(n) tends to L, then $x(n)^k$ tends to L^k for every natural number k. *Used for:* Powers of converging quantities: kinetic energy from a converging velocity, intensities from converging amplitudes and the polynomial terms of series expansions. *Price:* **free**; exact.

72. Limit of an absolute value. If x(n) tends to L, then |x(n)| tends to |L|. *Used for:* Magnitudes of converging quantities: sizes of errors, speeds from velocities and distances in convergence estimates. *Price:* **free**; exact.

73. Limit of a maximum. If x(n) tends to L and y(n) tends to M, then the larger of x(n) and y(n) tends to the larger of L and M. *Used for:* Ramp and threshold quantities taken to a limit, such as max(0, x) in rectifiers and occupation cutoffs, and worst-case bounds built from two estimates. *Price:* **free**; exact.

74. Limit of a minimum. If x(n) tends to L and y(n) tends to M, then the smaller of x(n) and y(n) tends to the smaller of L and M. *Used for:* Clipped quantities and best-case bounds taken to a limit: saturation cutoffs and the smaller of two error estimates. *Price:* **free**; exact.

75. Continuous function of a convergent sequence. If x(n) tends to L and the function g is continuous at L, then g(x(n)) tends to g(L). *Used for:* Applying physical laws to converging inputs: the energy, pressure or force computed from a converging state variable converges to its value at the limiting state. *Price:* **free**; exact.

76. Sequences equal from some point share limits. If two sequences agree from some index on and one of them tends to L, then the other tends to L too. *Used for:* Ignoring transients: the first steps of an iteration, the warm-up of a simulation or finitely many corrected terms do not change a limit. *Price:* **free**; exact.

77. Limit of a constant sequence. A sequence whose terms all equal c tends to c. *Used for:* The base case of every limit argument: a fixed physical constant, or a quantity that has settled, stays at its value. *Price:* **free**; exact.

78. Limit of a constant plus a sequence. If x(n) tends to L, then c + x(n) tends to c + L for every fixed number c. *Used for:* Shifting a converging quantity by a fixed offset: zero-point energies, reference potentials and background subtraction. *Price:* **free**; exact.

79. Limit of a sequence plus a constant. If x(n) tends to L, then x(n) + c tends to L + c for every fixed number c. *Used for:* Adding a fixed offset to a converging quantity: a rest energy plus a converging

kinetic energy, or a converging correction added to a known value. *Price:* **free**; exact.

80. Limit of a sequence times a constant. If x(n) tends to L, then x(n) times c tends to L times c for every fixed number c. *Used for:* Unit conversions and fixed couplings applied to converging quantities: multiplying by a charge, a mass or a conversion factor. *Price:* **free**; exact.

81. Limit of a sequence divided by a constant. If x(n) tends to L and the fixed number y has a reciprocal, then x(n) / y tends to L / y. *Used for:* Normalizing converging quantities: dividing by a volume, a particle number or a reference scale. *Price:* **free**; exact.

82. Limits keep an inequality. If x(n) tends to L, y(n) tends to M, and x(n) ≤ y(n) from some index on, then L ≤ M. *Used for:* Bounds survive limits: comparing a converging quantity with a known bound, energy inequalities and monotone approximations in variational methods. *Price:* **free**; exact.

83. A lower bound passes to the limit. If x(n) tends to a and b ≤ x(n) from some index on, then b ≤ a. *Used for:* Positivity of limits: a limit of nonnegative energies, probabilities or densities is nonnegative. *Price:* **free**; exact.

84. An upper bound passes to the limit. If x(n) tends to a and x(n) ≤ b from some index on, then a ≤ b. *Used for:* Upper bounds of limits: probabilities stay at most 1, speeds stay at most the speed of light and efficiencies stay below their bound in the limit. *Price:* **free**; exact.

85. Powers of a number above 1 grow without bound. If r > 1, then r^n exceeds any given number once n is large enough. *Used for:* Exponential growth: chain reactions, populations, unstable modes and amplification factors grow without bound. *Price:* **free**; exact.

86. The counting numbers grow without bound. The numbers 0, 1, 2, 3, ... exceed any given real number from some point on. *Used for:* Every limit of large size or many steps: thermodynamic limits, large particle numbers and long times. *Price:* **free**; exact.

87. Unbounded plus convergent is unbounded. If u(n) grows without bound and v(n) tends to a limit C, then u(n) + v(n) grows without bound. *Used for:* A diverging leading term dominates a converging correction: energies that grow without bound while their corrections settle. *Price:* **free**; exact.

88. Unbounded plus a constant is unbounded. If u(n) grows without bound, then u(n) + c grows without bound for every fixed number c. *Used for:* A fixed offset cannot stop a divergence: shifted energies and potentials measured from a reference constant. *Price:* **free**; exact.

89. Product of unbounded sequences is unbounded. If u(n) and v(n) both grow without bound, then u(n) times v(n) grows without bound. *Used for:* Divergence of products: a total energy that grows with both size and density, or an area that grows with two lengths. *Price:* **free**; exact.

90. Reciprocal of an unbounded sequence tends to zero. If u(n) grows without bound and w(n) is the reciprocal of u(n) from some index on, then w(n) tends to 0. *Used for:* Quantities that vanish as a scale diverges: 1/N corrections in large systems, 1/T at high temperature and inverse distances far from a source. *Price:* **free**; exact.

91. Reciprocal 1/r tends to zero. For every tolerance there is a point beyond which, for every number r larger than that point, the reciprocal 1/r is within the tolerance of 0. *Used for:* Fields and potentials that vanish at infinity: Coulomb and gravitational potentials proportional to 1/r, and boundary conditions at infinity in scattering. *Price:* **free**; exact.

92. Sum of two convergent series. If one series sums to a and another sums to b, regardless of the order of their terms, then the series of termwise sums sums to a + b. *Used for:* Linearity of mode sums and partition functions: the contributions of independent sources or modes add. *Price:* **free**; exact.

93. Difference of two convergent series. If one series sums to a and another sums to b, regardless of the order of their terms, then the series of termwise differences sums to a − b. *Used for:* Subtracting a reference series: counterterms, vacuum subtraction and differences of spectral sums. *Price:* **free**; exact.

94. Negative of a convergent series. If a series sums to a regardless of the order of its terms, then the series of the negated terms sums to −a. *Used for:* Reversing signs in sums: opposite charges in lattice sums and sign-reversed contributions in mode sums. *Price:* **free**; exact.

95. Constant times a convergent series. If the terms f(i) sum to a regardless of their order, then the terms c times f(i) sum to c times a for every fixed number c. *Used for:* Pulling a constant out of a sum: a coupling constant, a charge or a unit factor that multiplies every mode. *Price:* **free**; exact.

96. Convergent series times a constant. If the terms f(i) sum to a regardless of their order, then the terms f(i) times c sum to a times c for every fixed number c. *Used for:* Factoring a common constant out of a sum: normalization constants and conversion factors in mode sums. *Price:* **free**; exact.

97. The sum of a series is unique. A series cannot sum to two different values when the sum does not depend on the order of its terms. *Used for:* A computed partition function, energy sum or expectation value is a single definite number. *Price:* **free**; exact.

98. Termwise smaller series have smaller sums. If f(i) ≤ g(i) for every i, the terms f(i) sum to a and the terms g(i) sum to b, both regardless of order, then a ≤ b. *Used for:* Comparing sums: bounding a partition function by a simpler one and estimating spectral sums by comparison. *Price:* **free**; exact.

99. The zero series sums to zero. The series whose terms are all 0 sums to 0. *Used for:* The base case of every sum: vanishing contributions add nothing to an energy or a partition function. *Price:* **free**; exact.

100. Differentiating a series term by term, continuously differentiable terms (variant). Let g_0, g_1, g_2, ... be real functions on [0, 1], each g_n with a derivative g_n' that is uniform on the closed interval [0, 1] (one accuracy rate for all points, end points included), and let $|g_n'(x)| \le u(n)$ on [0, 1] for numbers u(n) whose series converges. If the series $g_0(y) + g_1(y) + g_2(y) + \ldots$ converges at one point y of [0, 1], then at every point x of [0, 1] the series of the values $g_n(x)$ and the series of the derivatives $g_n'(x)$ both converge, and the sum of the first series has the sum of the second as its derivative, uniformly on [0, 1]. *Used for:* Differentiating Fourier series, power series and eigenfunction expansions term by term, for example to get velocities, fields or currents from a series solution of the wave or heat equation. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

101. Taylor's theorem with the Peano remainder. Let $F_0 = f, F_1, \ldots, F_n$ be real functions on [0, 1] such that, for each $i < n$, F_{i+1} is the derivative of F_i uniformly on the closed interval [0, 1] (one accuracy rate for all points, end points included), and F_n is uniformly continuous on [0, 1]. Then near 0 the difference between f(x) and its Taylor polynomial of degree n at 0, built from $F_0(0), \ldots, F_n(0)$, is small compared with x^n: for every tolerance $c > 0$ there is a distance $d > 0$ such that this difference is at most $c \cdot x^n$ in absolute value for every x in [0, 1] with $x \le d$. *Used for:* Justifying small-oscillation, harmonic and linear-response approximations: replacing a potential or a response by its Taylor polynomial near an equilibrium point leaves an error that vanishes faster than the last power kept. *Price:* **free**; exact.

102. Compound growth exceeds every bound. If $r > 0$, then $(1 + r)^n$ exceeds any given number once n is large enough. *Used for:* Growth quoted as a fixed fractional increase per step: a gain of r per pass in a laser cavity or an excess of r neutrons per generation in a chain reaction grows without limit. *Price:* **free**; exact.

103. Geometric decay stays positive. If $0 < r < 1$, then r^n tends to 0 while every r^n stays strictly positive. *Used for:* Remaining fractions after repeated attenuation: the share of nuclei left after n half-lives or of light left after n absorbing layers shrinks toward zero without ever reaching it. *Price:* **free**; exact.

104. Powers of a number smaller than 1 in size vanish. If r is a real number with $|r| < 1$, then r^n tends to 0, also when r is negative. *Used for:* Damped responses that flip sign at every step: echoes inverted at each reflection, overshooting relaxation in feedback loops and iterations whose error alternates in sign. *Price:* **free**; exact.

105. Exponentials beat polynomials. If $r > 1$ and k is a whole number, then n^k / r^n tends to 0. *Used for:* Exponential suppression outweighs polynomial growth: Boltzmann factors beat polynomially growing numbers of states at high energy, and an exponentially converging method beats any polynomial error growth. *Price:* **free**; exact.

106. A constant over a power of n tends to zero. For every number r and every whole number k of at least 1, r / n^k tends to 0 as n grows. *Used for:* Power-law corrections that disappear in large systems: $1/N$ and $1/N^2$ finite-size corrections and discretization errors that fall like a power of the number of grid points. *Price:* **free**; exact.

107. Factorials beat exponentials. For every number c, $c^n / n!$ tends to 0 as n grows. *Used for:* High-order terms of factorial series vanish: the probability of very many events in a Poisson process and the late terms of the series for time evolution and Boltzmann factors. *Price:* **free**; exact.

108. n factorial is small next to n to the n. The ratio $n! / n^n$ tends to 0 as n grows. *Used for:* Counting in statistical mechanics: the number of orderings of n particles is negligible next to the number of ways to place n particles in n cells, the comparison behind Stirling-type entropy estimates. *Price:* **free**; exact.

109. Growth by a fixed factor is unbounded. If $v(0) > 0$, $c > 1$ and $v(n+1)$ is at least c times v(n) for every n, then v(n) exceeds any given number for large n. *Used for:* Instability criteria: a perturbation amplified at least by a fixed factor each period, as in parametric resonance or an unstable mode of a numerical scheme, grows without limit. *Price:* **free**; exact.

110. Discrete Gronwall bound from above. If c is at least 0 and $u(k+1)$ is at most $c\ u(k)$ for every k below n, then u(n) is at most $c^n\ u(0)$. *Used for:* Error growth in step-by-step methods: if each time step multiplies the error by at most c, the error after n

steps is at most c^n times the initial error. *Price:* **free**; exact.

111. Discrete Gronwall bound from below. If c is at least 0 and c u(k) is at most u(k+1) for every k below n, then c^n u(0) is at most u(n). *Used for:* Guaranteed amplification: a signal or population multiplied by at least c in every step is at least c^n times its initial size after n steps. *Price:* **free**; exact.

112. Fibonacci ratios approach the golden ratio. The golden ratio, the positive number φ with $\varphi^2 = \varphi + 1$, exists, and the ratio F(n+1)/F(n) of consecutive Fibonacci numbers tends to it. *Used for:* The golden ratio in physics: self-similar scaling, quasicrystal tilings, the most irrational winding number in the stability of orbits. *Price:* **free**; exact.

113. Inverse Fibonacci ratios approach one over φ. The ratio F(n)/F(n+1) of consecutive Fibonacci numbers tends to φ − 1, which equals 1/φ, where φ is the golden ratio. *Used for:* Rational approximations of the golden mean: the fractions F(n)/F(n+1) approach the golden-mean rotation number of the last invariant orbit to break in a perturbed system. *Price:* **free**; exact.

114. A damped linear recursion settles. If $0 \le a < 1$ and u(n+1) = a u(n) + b, then u(n) tends to b / (1 − a), whatever the starting value. *Used for:* Steady states of linear relaxation: repeated dilution and refilling, a capacitor charged in steps, and any first-order discrete process with constant input. *Price:* **free**; exact.

115. Error bound for geometrically converging iterations. If the consecutive terms of a sequence u of real numbers satisfy $|u(n) - u(n+1)| \le C\, r^n$ with $0 \le r < 1$ and the sequence tends to a, then $|u(n) - a| \le C\, r^n / (1 - r)$. *Used for:* Stopping rules for iterative solvers: fixed-point and self-consistent-field iterations estimate their remaining error from the size of the current step. *Price:* **free**; exact.

116. Truncation error of a geometrically bounded series. If real numbers f(n) satisfy $|f(n)| \le C\, r^n$ with $0 \le r < 1$ and the series of the f(n) sums to a, then the sum of the first n terms differs from a by at most $C\, r^n / (1 - r)$. *Used for:* Truncating perturbation and multipole expansions: deciding how many terms to keep for a required accuracy when the terms fall off geometrically. *Price:* **free**; exact.

117. Summable steps make a Cauchy sequence. If a sequence u of real numbers satisfies $|u(n) - u(n+1)| \le d(n)$ for every n and the numbers d(n) have a finite sum, then the terms of u get arbitrarily close to one another. *Used for:* Convergence of paths with finite total length: an iteration or sampled trajectory whose step lengths add up to a finite total cannot keep wandering. *Price:* **free**; exact.

118. Geometric series with ratio smaller than 1 in size. If $|r| < 1$, then $1 + r + r^2 + r^3 + \ldots$ sums to 1 / (1 − r), in any order of the terms, also for negative r. *Used for:* Sums of multiple reflections with sign changes: image charges between two plates, transmission through a cavity with a phase flip at each pass, and alternating feedback sums. *Price:* **free**; exact.

119. The sum of n times r to the n. If r is a real number with $|r| < 1$, then the sum over n of $n\, r^n$ equals $r / (1 - r)^2$, in any order of the terms. *Used for:* Mean occupation numbers: the average number of quanta in a thermal harmonic oscillator (the Planck distribution) and the mean of a geometric distribution. *Price:* **free**; exact.

120. The exponential series converges. For every number x, the terms $x^n / n!$ for n = 0, 1, 2, ... have a finite sum, regardless of the order of the terms. *Used for:* Convergence checks for series with factorial denominators, such as the Poisson distribution and the Taylor series of the exponential, made before the exponential function itself is constructed. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

121. A sum is the limit of its partial sums. If a series of real numbers sums to a regardless of the order of its terms, then the sums of its first n terms tend to a. *Used for:* Computing sums by truncation: partition functions, lattice sums and mode sums evaluated by adding terms in order until the result settles. *Price:* **free**; exact.

122. Ratio test for divergence. If $r > 1$, the real terms of a series do not all vanish from some index on, and from some index on each term is at least r times the previous one in size, then the series has no finite sum. *Used for:* Recognizing divergent expansions: perturbation series whose terms grow geometrically, and partition functions of systems whose number of states grows exponentially with energy above a limiting temperature. *Price:* **free**; exact.

123. Ratio test with limit above 1. If the ratio of the sizes of consecutive terms of a series of real numbers tends to a limit $l > 1$, then the series has no finite sum. *Used for:* Limits of power-series methods: a perturbative or virial expansion diverges beyond the point where the ratio of successive terms exceeds 1. *Price:* **free**; exact.

124. Bounded series of nonnegative terms converge. If every term f(n) is nonnegative and every partial sum $f(0) + \ldots + f(n-1)$ is at most a fixed number c, then the series has a finite sum, independent of the order of its terms. *Used for:* Normalizable probabilities and bounded totals: nonnegative probabilities that never add up to more than 1, or

nonnegative contributions bounded by a conserved total, have a definite sum. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

125. Bounded finite sums give a convergent series. If real numbers f(0), f(1), f(2), ... are all nonnegative and their sum over every finite set of indices is at most a fixed number c, then the series has a finite sum. *Used for:* Sums over states with no natural ordering: occupation sums and lattice sums of nonnegative terms converge as soon as every finite collection of states contributes at most a fixed amount. *Price:* **LPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

126. A dominated series has a smaller sum. If real numbers f(i) and g(i), indexed by the whole numbers i, satisfy $|f(i)| \le g(i)$ for every i, the f(i) sum to a and the g(i) sum to b, then $|a| \le b$. *Used for:* Bounding a sum by a majorant: estimating the size of a perturbation series or of a Green's function expansion by a simpler series of positive terms. *Price:* **free**; exact.

127. Derivatives take every intermediate value (Darboux). Let f be uniformly continuous on [0, 1] (one accuracy rate for all points of the interval), with a derivative f′(x) at every point x of [0, 1] (one-sided at the two ends). If $f'(0) < m < f'(1)$, then $f'(c) = m$ at some point c strictly between 0 and 1, even when the derivative f′ is not continuous. *Used for:* Rates of change cannot jump over values: a velocity, slope or force that is the derivative of a position, height or potential passes through every value between two of its values, so a force that pulls one way at one end of a path and the other way at the other end balances exactly at some point in between. *Price:* **LLPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

128. Darboux's theorem for a decreasing change of rate. Let f be uniformly continuous on [0, 1] (one accuracy rate for all points of the interval), with a derivative f′(x) at every point x of [0, 1] (one-sided at the two ends). If $f'(1) < m < f'(0)$, then $f'(c) = m$ at some point c strictly between 0 and 1, even when f′ is not continuous. *Used for:* A rate that starts above a value and ends below it must equal that value somewhere in between: a body whose velocity falls from above a given velocity to below it has exactly that velocity at some instant, whether or not its acceleration is well behaved. *Price:* **LLPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

129. Derivatives reach every value between their end values. Let f be uniformly continuous on [0, 1] (one accuracy rate for all points of the interval), with a derivative f′(x) at every point x of [0, 1] (one-sided at the two ends). If $f'(0) \le m \le f'(1)$, then $f'(c) = m$ at some point c of [0, 1]. *Used for:* Guaranteeing an exact match of a rate with a target value when the rates at the two ends bracket the target, for example a point where a force exactly balances a given load or where a growth rate equals a given rate, with the match allowed at an end of the interval. *Price:* **LLPO + AC**$_\omega$; exact (the upper bound uses countable choice, not shown to be needed).

130. Differentiating a series term by term. Take a series of real functions g(0), g(1), g(2), ... on [0, 1], each continuous on [0, 1] and differentiable at every point strictly inside, with |derivative of the n-th term| at most u(n) there and the sum of the u(n) finite. If the series converges at one point of [0, 1], it converges at every point of [0, 1], the series of derivatives converges at every point strictly inside, and the sum is differentiable at every point strictly inside, with derivative equal to the sum of the derivatives. *Used for:* Differentiating Fourier series, power series and eigenfunction expansions term by term, for example to get velocities, fields or currents from a series solution of the wave or heat equation. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of representatives).

131. First derivative test for a minimum. If a function f is continuous at b and differentiable on intervals (a, b) and (b, c) on either side of b, with $f' \le 0$ on (a, b) and $f' \ge 0$ on (b, c), then f has a local minimum at b. *Used for:* Locating a stable equilibrium from the direction of the force alone: a force that points back toward b from both sides marks a minimum of potential energy, including flat minima such as a potential proportional to x^4, where the curvature vanishes and a second-derivative check says nothing. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

132. First derivative test for a maximum. If a function f is continuous at b and differentiable on intervals (a, b) and (b, c) on either side of b, with $f' \ge 0$ on (a, b) and $f' \le 0$ on (b, c), then f has a local maximum at b. *Used for:* Finding peaks where a quantity stops rising and starts falling, such as the highest point of a projectile's path or the peak of a resonance curve, even when the peak is flat and the second derivative is zero. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

133. First derivative test for a minimum, local form. If a function f is continuous at b, differentiable at the points near b other than b, with $f' \le 0$ at all points in some interval just left of b and $f' \ge 0$ at all points in some interval just right of b, then f has a local minimum at b. *Used for:* Classifying an

equilibrium by checking the sign of the force only very close to it on each side: a restoring force from both sides, however weak, makes the point a minimum of the potential energy. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

134. First derivative test for a maximum, local form. If a function f is continuous at b, differentiable at the points near b other than b, with $f' \geq 0$ at all points in some interval just left of b and $f' \leq 0$ at all points in some interval just right of b, then f has a local maximum at b. *Used for:* Recognising a point of unstable balance from the local sign of the force: a force pushing away from the point on both sides makes it a maximum of the potential energy, so the system leaves it under any small disturbance. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

135. Second derivative test for a minimum. If a function f is continuous at a point x_0 and differentiable at every point near x_0, x_0 included, with derivative f', where $f'(x_0) = 0$ and f' has a positive derivative $f''(x_0) > 0$ at x_0, then f has a local minimum at x_0: $f(x_0) \leq f(y)$ for every y close enough to x_0. *Used for:* Deciding that an equilibrium is stable: where the force vanishes and the potential energy curves upward, the system rests at a minimum of energy, which is the starting point of the small-oscillation (harmonic) approximation for molecules, crystal lattices and orbits. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

136. Second derivative test for a maximum. If a function f is continuous at a point x_0 and differentiable at every point near x_0, x_0 included, with derivative f', where $f'(x_0) = 0$ and f' has a negative derivative $f''(x_0) < 0$ at x_0, then f has a local maximum at x_0: $f(y) \leq f(x_0)$ for every y close enough to x_0. *Used for:* Recognising unstable equilibria and barrier tops: where the force vanishes and the potential energy curves downward, as at the top of an activation barrier in a chemical reaction or an inverted pendulum, the system sits at a maximum of energy and small pushes grow. *Price:* **free**; exact (the upper bound chooses representatives of a sequence of points, which the free allowance contains).

137. A derivative that misses a value stays on one side. If a function f has a derivative $f'(x)$ at every point x of the interval $[0, 1]$ (one-sided at the two ends) and $f'(x)$ is not equal to m for every x in $[0, 1]$, then either $f'(x) < m$ for every x in $[0, 1]$, or $f'(x) > m$ for every x in $[0, 1]$, even when f' is not continuous. *Used for:* A rate that never takes a value cannot cross it: a velocity that is never zero never changes direction, and a growth rate, slope or force that never equals a given threshold stays on one side of it all along. *Price:* $\mathbf{MP} + \mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

138. L'Hôpital's rule at infinity. If f(x) and g(x) both tend to 0 as x grows without bound, f and g are differentiable for all large x, $g'(x)$ is never 0 there, and $f'(x)/g'(x)$ tends to a finite limit l as x grows, then $f(x)/g(x)$ also tends to l as x grows. Nothing else is assumed about f and g. *Used for:* Comparing two quantities that both vanish at large argument: the relative decay of fields and potentials at large distance (a screened potential against a Coulomb tail), ratios of correlation functions at large separation, the leading behaviour of response functions at high frequency, and ratios of quantities that vanish as the temperature or the energy grows without bound. *Price:* $\mathbf{MP}^{\vee} + \mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

139. L'Hôpital's rule approaching a point from below. If f(x) and g(x) both tend to 0 as x increases to a point b, f and g are differentiable at every point of an interval just below b, $g'(x)$ is never 0 there, and $f'(x)/g'(x)$ tends to a finite limit l as x increases to b, then $f(x)/g(x)$ also tends to l as x increases to b. Nothing else is assumed about f and g at b. *Used for:* Limits of ratios at the upper end of a physical range approached from below, where numerator and denominator vanish together: quantities followed as a temperature rises to a critical temperature, as a speed rises toward the speed of light, or as a coupling or an energy rises to a threshold value. *Price:* $\mathbf{MP}^{\vee} + \mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

140. L'Hôpital's rule at minus infinity. If f(x) and g(x) both tend to 0 as x decreases without bound, f and g are differentiable for all sufficiently negative x, $g'(x)$ is never 0 there, and $f'(x)/g'(x)$ tends to a finite limit l as x decreases without bound, then $f(x)/g(x)$ also tends to l as x decreases without bound. Nothing else is assumed about f and g. *Used for:* Ratios of quantities that vanish in the far past or at large negative argument: the early-time behaviour of signals that switch on smoothly from zero, and the comparison of two decaying solutions deep on the forbidden side of a potential barrier. *Price:* $\mathbf{MP}^{\vee} + \mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

141. L'Hôpital's rule from both sides. If f(x) and g(x) both tend to 0 as x approaches a point a from either side, f and g are differentiable at every point near a other than a itself, $g'(x)$ is never 0 at those points, and $f'(x)/g'(x)$ tends to a finite limit l as x approaches a, then $f(x)/g(x)$ also tends to l as x

approaches a from either side. Nothing is assumed about f and g at a itself. *Used for:* Removing apparent singularities of physical formulas at a point where numerator and denominator vanish together, such as sin(x)/x at x = 0, the response of a driven oscillator exactly at resonance, or a formula with a factor (E − E0) in numerator and denominator, approached from both sides of the special value. *Price:* $\mathbf{MP}^{\vee}$ + $\mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

142. Convexity means increasing chord slopes. A function f on the interval [0, 1] is convex (the chord between any two points of its graph lies on or above the graph) exactly when, for any three points x < y < z of [0, 1], the slope of the chord from x to y, (f(y) − f(x))/(y − x), is at most the slope of the chord from y to z, (f(z) − f(y))/(z − y). *Used for:* Testing convexity from three values at a time: checking that a free energy, an energy as a function of volume or entropy, or a potential is convex (the condition for thermodynamic and mechanical stability) from tabulated or measured points, by comparing successive chord slopes; the basis of convex interpolation and of the stability tests on a grid. *Price:* **free**; exact.

143. Secant slopes of a convex function increase. If f is convex on the interval [0, 1] and a is a point of [0, 1], then the slope of the chord from a to x, (f(x) − f(a))/(x − a), increases as x increases through the points of [0, 1] different from a. *Used for:* Bounding a convex function by chords through a fixed point: estimating the change of a convex free energy or potential from a reference state, the monotone growth of average rates such as the mean cost per unit change from a fixed starting point, and the slope comparisons behind Legendre transforms. *Price:* **free**; exact.

144. Concavity means decreasing chord slopes. A function f on the interval [0, 1] is concave (the chord between any two points of its graph lies on or below the graph) exactly when, for any three points x < y < z of [0, 1], the slope of the chord from y to z, (f(z) − f(y))/(z − y), is at most the slope of the chord from x to y, (f(y) − f(x))/(y − x). *Used for:* Testing concavity from three values at a time: checking that an entropy as a function of energy is concave (positive heat capacity, thermal stability), that a utility or a production curve shows diminishing returns, from successive chord slopes of measured or tabulated values. *Price:* **free**; exact.

145. Secant slopes of a concave function decrease. If f is concave on the interval [0, 1] and a is a point of [0, 1], then the slope of the chord from a to x, (f(x) − f(a))/(x − a), decreases as x increases through the points of [0, 1] different from a. *Used for:* Bounding a concave function by chords through a fixed point: average gains such as entropy gained per unit of energy added from a reference state shrink as more is added, giving upper and lower estimates of concave entropies and of diminishing returns. *Price:* **free**; exact.

146. A convex function lies above its tangent line. If f is convex on the interval [0, 1], x < y are points of [0, 1] and f has derivative f′(x) at x, then f′(x) is at most the chord slope (f(y) − f(x))/(y − x); equivalently f(y) is at least f(x) + f′(x)(y − x), so the graph lies above the tangent line at x. *Used for:* Tangent-line lower bounds: the supporting-line inequality behind Legendre transforms and convex duality (a convex free energy is the upper envelope of its tangent lines), and linearized estimates that are guaranteed to underestimate a convex energy, free energy or cost. *Price:* **free**; exact.

147. A convex function lies above its tangent, right end. If f is convex on the interval [0, 1], x < y are points of [0, 1] and f has derivative f′(y) at y, then the chord slope (f(y) − f(x))/(y − x) is at most f′(y); equivalently f(x) is at least f(y) + f′(y)(x − y). *Used for:* Tangent-line lower bounds taken at the far end of an interval: bounding a convex energy, free energy or cost on the left of a point by its tangent line at that point, as in backward error estimates and in comparing a convex potential with its linearization. *Price:* **free**; exact.

148. Tangent below a convex function, derivative taken inside the interval. If f is convex on the closed interval [0, 1], x < y are points of [0, 1] and f has derivative f′(x) at x computed only from points of [0, 1], then f′(x) is at most the chord slope (f(y) − f(x))/(y − x). *Used for:* Tangent-line bounds at the edge of a domain, where the function is only defined on one side: a convex free energy or potential at the boundary of its allowed range (zero density, a wall, a maximal filling) is still bounded below by its one-sided tangent. *Price:* **free**; exact.

149. Chord slope below the right-end derivative, inside the interval. If f is convex on the closed interval [0, 1], x < y are points of [0, 1] and f has derivative f′(y) at y computed only from points of [0, 1], then the chord slope (f(y) − f(x))/(y − x) is at most f′(y). *Used for:* Bounding a convex quantity near the upper edge of its allowed range by its one-sided tangent there, for example a convex energy at full filling or at the largest allowed volume. *Price:* **free**; exact.

150. A convex function's right derivative is below the chord slope. If f is convex on the interval [0, 1], x < y are points of [0, 1] and f has a right derivative d at x (computed from points to the right of x only), then d is at most the chord slope (f(y) − f(x))/(y − x). *Used for:* Convex functions with kinks: at a first-order phase transition a free energy has

different left and right derivatives (two values of density, magnetization or entropy); the right derivative still bounds the free energy from below on the right, which fixes the coexisting phases and the latent heat. *Price:* **free**; exact.

151. A convex function's left derivative is above the chord slope. If f is convex on the interval [0, 1], $x < y$ are points of [0, 1] and f has a left derivative d at y (computed from points to the left of y only), then the chord slope $(f(y) - f(x))/(y - x)$ is at most d. *Used for:* Convex functions with kinks: at a phase transition the left derivative of a free energy bounds it from below on the left, the companion of the right-derivative bound that brackets the jump of density or entropy across the transition. *Price:* **free**; exact.

152. The derivative of a convex function increases. If f is convex on the interval [0, 1] and has a derivative $f'(x)$ at every point x of [0, 1], then $f'(x) \le f'(y)$ whenever $x \le y$ are points of [0, 1]. *Used for:* Monotone response from convexity: the temperature, the chemical potential or minus the pressure, obtained as the derivative of a convex energy or free energy, increases with its conjugate variable (positive compressibility and heat capacity), which makes Legendre transforms one to one and equations of state invertible. *Price:* **free**; exact.

153. The derivative of a convex function increases, edges included. If f is convex on the closed interval [0, 1] and has a derivative $f'(x)$ at every point x of [0, 1], computed only from points of [0, 1] (so one-sided at the two end points), then $f'(x) \le f'(y)$ whenever $x \le y$ are points of [0, 1]. *Used for:* Monotone response up to the boundary of the allowed range: a convex energy or free energy defined only on a closed range (densities between zero and full packing, fractions between 0 and 1) still has an increasing derivative, so its conjugate variable is ordered on the whole range. *Price:* **free**; exact.

154. Convex functions of one variable are continuous inside (variant). If f is convex on the closed interval [0, 1], then f is continuous at every point strictly between 0 and 1. *Used for:* Continuity of thermodynamic potentials for free: a free energy, entropy or energy known only to be convex or concave in one variable is automatically continuous away from the edges of its range, so it can be evaluated through limits and approximated from measured values; phase transitions show up as kinks, never as jumps, inside the range. *Price:* **free**; exact.

155. L'Hôpital's rule with continuity at the endpoint. If f and g are continuous on the interval [0, 1) with $f(0) = g(0) = 0$, differentiable at every point strictly between 0 and 1 with $g'(x)$ never 0 there, and $f'(x)/g'(x)$ tends to a limit l as x decreases to 0, then $f(x)/g(x)$ also tends to l as x decreases to 0. *Used for:* Limits of ratios of quantities that start from zero together at the beginning of a process, such as the ratio of two displacements, charges or populations that both begin at zero at time a, read off from the ratio of their initial rates of change. *Price:* $\mathbf{MP}^{\vee}$ + $\mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

156. L'Hôpital's rule from below with continuity at the endpoint. If f and g are continuous on the interval (0, 1] with $f(1) = g(1) = 0$, differentiable at every point strictly between 0 and 1 with $g'(x)$ never 0 there, and $f'(x)/g'(x)$ tends to a limit l as x increases to 1, then $f(x)/g(x)$ also tends to l as x increases to 1. *Used for:* Limits of ratios of two quantities that reach zero together at the end of a process, such as two signals or reserves that run out together at the final time b, compared through the ratio of their final rates of change. *Price:* $\mathbf{MP}^{\vee}$ + $\mathbf{AC}_{\omega}$; exact (the upper bound uses countable choice, not shown to be needed).

Q.3 Exponential, logarithm, trigonometric and complex functions

157. Exponential is at least one plus x. For every real number x, $x + 1 \le e^x$. *Used for:* Gibbs' inequality, non-negativity of relative entropy, the Gibbs variational principle. *Price:* **free**; exact.

158. Logarithm is at most x minus one. For every $x > 0$, $\ln x \le x - 1$. *Used for:* Upper bounds on the information $-\ln p$ of an outcome and on free-energy differences. *Price:* **free**; exact.

159. Logarithm is at least one minus 1/x. For every $x > 0$, $1 - 1/x \le \ln x$. *Used for:* Lower bounds on the information of an outcome and entropy estimates. *Price:* **free**; exact.

160. Derivative of the logarithm is 1/x. At every $x > 0$ the natural logarithm has derivative 1/x, uniformly for pairs of nearby points (strict differentiability). *Used for:* Entropy from heat taken in at a given temperature, and the chemical potential of a dilute solute from its concentration. *Price:* **free**; exact.

161. Strict Gibbs inequality for the exponential. If $x \neq 0$ then $x + 1 < e^x$. *Used for:* Strict positivity of

relative entropy and of the free energy of a state out of equilibrium. *Price:* **MP**; exact.

162. Strict Gibbs inequality for the logarithm. If $x > 0$ and $x \neq 1$ then $\ln x < x - 1$. *Used for:* The relative entropy of two distinct probability laws is strictly positive. *Price:* **MP**; exact.

163. Exponential of the logarithm of a nonzero number. If $x \neq 0$ there is a number L with $e^L = |x|$ (so L is the logarithm of $|x|$). *Used for:* Logarithms of magnitudes: ln of a wave amplitude in the WKB method, ln of a Jacobian determinant, decibels. *Price:* **MP**; exact.

164. Exponential decay beats every power. For every natural number n, $x^n e^{-x}$ tends to 0 as x grows without bound. *Used for:* Every moment of a Boltzmann or Planck weight is finite. *Price:* **free**; exact.

165. Error bound for the exponential series. If $|x| \leq 1$ and $n \geq 1$, the error of the n-term partial sum of the series for e^x is at most $|x|^n \cdot (n + 1)/(n! \cdot n)$. *Used for:* Error bounds for every numerical value of the exponential. *Price:* **free**; exact.

166. Mercator series for the logarithm. If $|x| < 1$, the series $x + x^2/2 + x^3/3 + \dots$ converges unconditionally to $-\ln(1 - x)$. *Used for:* Free energies of Bose and two-level systems, $-\ln(1 - e^{-\beta\varepsilon})$, mixing entropies, virial expansions. *Price:* **free**; exact.

167. Derivative of the sine. The sine has derivative the cosine at every point. *Used for:* Harmonic oscillator and wave equation. *Price:* **free**; exact.

168. Derivative of the cosine. The cosine has derivative minus the sine at every point. *Used for:* Harmonic oscillator and wave equation. *Price:* **free**; exact.

169. Sine is continuous. The sine function is continuous. *Used for:* Continuous phase observables. *Price:* **free**; exact.

170. Cosine is continuous. The cosine function is continuous. *Used for:* Continuous phase observables. *Price:* **free**; exact.

171. Pythagorean identity for sine and cosine. For every x, $\cos^2 x + \sin^2 x = 1$. *Used for:* Energy conservation of an oscillator and normalisation of qubit probabilities. *Price:* **free**; exact.

172. Sine addition formula. $\sin(x + y) = \sin x \cos y + \cos x \sin y$. *Used for:* Composing rotations, adding phases, interference. *Price:* **free**; exact.

173. Cosine addition formula. $\cos(x + y) = \cos x \cos y - \sin x \sin y$. *Used for:* Composing rotations, adding phases, beats. *Price:* **free**; exact.

174. Cosine is at most one. For every x, $\cos x \leq 1$. *Used for:* Amplitude bounds and probabilities $\cos^2(\theta/2)$ staying in $[0, 1]$. *Price:* **free**; exact.

175. Cosine is at least minus one. For every x, $-1 \leq \cos x$. *Used for:* Amplitude bounds. *Price:* **free**; exact.

176. Sine is at most one. For every x, $\sin x \leq 1$. *Used for:* Amplitude bounds. *Price:* **free**; exact.

177. Sine is at least minus one. For every x, $-1 \leq \sin x$. *Used for:* Amplitude bounds. *Price:* **free**; exact.

178. Cosine of a quarter turn is zero. $\cos(\pi/2) = 0$. *Used for:* The quarter-period phase lag between displacement and velocity of an oscillator. *Price:* **free**; exact.

179. Sine of a half turn is zero. $\sin \pi = 0$. *Used for:* Nodes of a string fixed at both ends, which quantise its modes. *Price:* **free**; exact.

180. Cosine of a half turn is minus one. $\cos \pi = -1$. *Used for:* Phase reversal and destructive interference. *Price:* **free**; exact.

181. Sine of a full turn is zero. $\sin(2\pi) = 0$. *Used for:* Closed orbits and periodic phase. *Price:* **free**; exact.

182. Cosine of a full turn is one. $\cos(2\pi) = 1$. *Used for:* Closed orbits and periodic phase. *Price:* **free**; exact.

183. Cosine has period 2π. For every x, $\cos(x + 2\pi) = \cos x$. *Used for:* Closed orbits, phase wrap in Bragg scattering, single-valued $e^{im\varphi}$. *Price:* **free**; exact.

184. Sine has period 2π. For every x, $\sin(x + 2\pi) = \sin x$. *Used for:* Closed orbits, phase wrap in Bragg scattering, single-valued $e^{im\varphi}$. *Price:* **free**; exact.

185. Small-angle bound for the cosine. If $|x| \leq 1$ then $|\cos x - (1 - x^2/2)| \leq |x|^4 \cdot 5/96$. *Used for:* Long-wave expansion of lattice links: $1 - \cos k = k^2/2 + O(k^4)$. *Price:* **free**; exact.

186. Small-angle bound for the sine. If $|x| \leq 1$ then $|\sin x - (x - x^3/6)| \leq |x|^4 \cdot 5/96$. *Used for:* Anharmonic correction of the pendulum and small-angle expansions. *Price:* **free**; exact.

187. Sine is below its argument. If $x > 0$ then $\sin x < x$. *Used for:* A chord is shorter than its arc; small-angle bounds. *Price:* **free**; exact.

188. Sine is at most its argument. If $x \geq 0$ then $\sin x \leq x$. *Used for:* Small-angle bound on a restoring force. *Price:* **free**; exact.

189. Cosine lower bound $1 - x^2/2$. For every x, $1 - x^2/2 \leq \cos x$. *Used for:* The bound $1 - \cos k \leq k^2/2$ on the symbol of a lattice Laplacian. *Price:* **free**; exact.

190. Sine is positive between 0 and π. If $0 < x < \pi$ then $\sin x > 0$. *Used for:* The first standing-wave mode is positive; a rotating qubit's probability before the half turn. *Price:* **free**; exact.

191. Jordan's inequality for the sine. If $0 \leq x \leq \pi/2$ then $(2/\pi)\, x \leq \sin x$. *Used for:* Jordan's lemma for contour and Fourier integrals; transition amplitudes up to a quarter turn. *Price:* **free**; exact.

192. Power series of the cosine. For every x, the series $\Sigma\ (-1)^n\ x^{2n}/(2n)!$ converges unconditionally to cos x. *Used for:* Small-angle expansions and every phase factor. *Price:* **free**; exact.

193. Power series of the sine. For every x, the series $\Sigma\ (-1)^n\ x^{2n+1}/(2n+1)!$ converges unconditionally to sin x. *Used for:* Small-angle expansions and every phase factor. *Price:* **free**; exact.

194. π is greater than 3. $3 < \pi$. *Used for:* Numerical values of constants that contain π, and solid angles. *Price:* **free**; exact.

195. π is less than 4. $\pi < 4$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

196. π is at most 4. $\pi \le 4$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

197. π is greater than 3.14. $3.14 < \pi$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

198. π is less than 3.15. $\pi < 3.15$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

199. π is greater than 3.1415. $3.1415 < \pi$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

200. π is less than 3.1416. $\pi < 3.1416$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

201. π is greater than 3.141592. $3.141592 < \pi$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

202. π is less than 3.141593. $\pi < 3.141593$. *Used for:* Numerical values of constants that contain π. *Price:* **free**; exact.

203. Hyperbolic Pythagorean identity. For every x, $\cosh^2 x = \sinh^2 x + 1$. *Used for:* Lorentz boosts: $\gamma^2 - (\gamma\beta)^2 = 1$. *Price:* **free**; exact.

204. Addition formula for the hyperbolic cosine. $\cosh(x + y) = \cosh x \cosh y + \sinh x \sinh y$. *Used for:* Rapidities add when relativistic velocities compose. *Price:* **free**; exact.

205. Addition formula for the hyperbolic sine. $\sinh(x + y) = \sinh x \cosh y + \cosh x \sinh y$. *Used for:* Rapidities add when relativistic velocities compose. *Price:* **free**; exact.

206. Hyperbolic cosine is positive. For every x, $\cosh x > 0$. *Used for:* $\gamma \ge 1$ and positive weights in the recognition cost. *Price:* **free**; exact.

207. Hyperbolic cosine is at least one. For every x, $\cosh x \ge 1$. *Used for:* $\gamma \ge 1$, and the recognition cost $\cosh t - 1$ is never negative. *Price:* **free**; exact.

208. Derivative of the hyperbolic sine. The hyperbolic sine has derivative the hyperbolic cosine at every point. *Used for:* The boost generator and the slope of the recognition cost. *Price:* **free**; exact.

209. Derivative of the hyperbolic cosine. The hyperbolic cosine has derivative the hyperbolic sine at every point. *Used for:* The boost generator and the slope of the recognition cost. *Price:* **free**; exact.

210. Hyperbolic tangent is below one. For every x, $\tanh x < 1$. *Used for:* $\beta = \tanh(\text{rapidity}) < 1$, so speeds stay below c; the magnetisation of a paramagnet stays below saturation. *Price:* **free**; exact.

211. Hyperbolic cosine is bounded by a Gaussian. For every x, $\cosh x \le e^{x^2/2}$. *Used for:* Hoeffding's lemma: concentration bounds for sums of ± 1 spins. *Price:* **free**; exact.

212. Strict Gibbs inequality, order-defined strictness. If $x \ne 0$ then $x + 1 < e^x$, where $a < b$ means $a \le b$ and not $b \le a$. *Used for:* Strict positivity of relative entropy, read with the strictness a partial order defines. *Price:* **free**; exact.

213. Strict Gibbs inequality for the logarithm, order-defined strictness. If $x > 0$ and $x \ne 1$ then $\ln x < x - 1$, where the conclusion $a < b$ means $a \le b$ and not $b \le a$. *Used for:* The relative entropy of two distinct probability laws is strictly positive, read with the strictness a partial order defines. *Price:* **free**; exact.

214. Polar form with a principal angle (away from zero). For every complex number z with $|z| > 0$ there are a number $r \ge 0$ with $r^2 = (\text{Re } z)^2 + (\text{Im } z)^2$ and an angle θ with $-\pi < \theta \le \pi$ such that $r \cos\theta = \text{Re } z$ and $r \sin\theta = \text{Im } z$, that is, $z = r(\cos\theta + i \sin\theta)$. *Used for:* The phase of a wave function or of any complex amplitude, the phase angle of an impedance, the direction of a vector in a plane, each read in the window from $-\pi$ to π. *Price:* **LPO**; exact.

215. Polar form with a principal angle (every point). For every complex number z, zero included, there are a number $r \ge 0$ with $r^2 = (\text{Re } z)^2 + (\text{Im } z)^2$ and an angle θ with $-\pi < \theta \le \pi$ such that $r \cos\theta = \text{Re } z$ and $r \sin\theta = \text{Im } z$, that is, $z = r(\cos\theta + i \sin\theta)$. *Used for:* The phase of a wave function or of any complex amplitude, the phase angle of an impedance, the direction of a vector in a plane, each read in the window from $-\pi$ to π at every point, zero included. *Price:* **LPO**; exact.

216. Every nonzero complex number is an exponential. For every complex number $z \ne 0$ there is a complex number w with $\exp(w) = z$: every nonzero complex number is a value of the exponential. *Used for:* Writing a complex amplitude as the exponential of a log-amplitude plus i times a phase (WKB, eikonal and scattering phases); logarithms of impedances and propagators. *Price:* **MP**; exact.

217. Every complex number away from zero is an exponential. For every complex number z with $|z| > 0$ there is a complex number w with $\exp(w) = z$. *Used for:* Writing a complex amplitude known to stay away from zero as the exponential of a log-amplitude plus i times a phase; logarithms of impedances and propagators. *Price:* **free**; exact.

218. Principal complex logarithm. For every complex number $z \neq 0$ there is a complex number w with $\exp(w) = z$ and $-\pi < \operatorname{Im} w \leq \pi$. *Used for:* The principal branch of the logarithm of an amplitude, $\log \psi = \log |\psi| + i \arg \psi$: scattering phase shifts read in one window, logarithms of impedances on the principal sheet. *Price:* **LPO**; exact.

219. Principal logarithm away from zero. For every complex number z with $|z| > 0$ there is a complex number w with $\exp(w) = z$ and $-\pi < \operatorname{Im} w \leq \pi$. *Used for:* The principal branch of the logarithm of an amplitude known to stay away from zero, $\log \psi = \log |\psi| + i \arg \psi$: scattering phase shifts read in one window, logarithms of impedances on the principal sheet. *Price:* **LPO**; exact.

220. Every number between −1 and 1 is a cosine. For every real number c with $-1 \leq c \leq 1$ there is an angle θ with $0 \leq \theta \leq \pi$ and $\cos \theta = c$. *Used for:* The angle between two vectors from their dot product; scattering, bond and dihedral angles. *Price:* **free**; exact.

221. Complex exponential turns sums into products. For all complex numbers z and w, $\exp(z + w) = \exp(z) \cdot \exp(w)$. *Used for:* Composing time evolutions, so that the phases of successive steps add; the shift rule of the Fourier transform. *Price:* **free**; exact.

222. Complex exponential of zero is one. The complex exponential of 0 is 1: $\exp(0) = 1$. *Used for:* Evolution over zero time is the identity. *Price:* **free**; exact.

223. Real part of a phase factor is the cosine. For every real number x, the real part of $\exp(i\,x)$ is $\cos x$. *Used for:* Phasors in AC circuits and optics; interference fringes. *Price:* **free**; exact.

224. Imaginary part of a phase factor is the sine. For every real number x, the imaginary part of $\exp(i\,x)$ is $\sin x$. *Used for:* Phasors in AC circuits and optics; the quadrature component of a wave. *Price:* **free**; exact.

225. Real part of the complex exponential. For every complex number z, the real part of $\exp(z)$ is $\exp(\operatorname{Re} z) \cdot \cos(\operatorname{Im} z)$. *Used for:* Damped oscillations $e^{-\gamma t} \cos(\omega t)$. *Price:* **free**; exact.

226. Imaginary part of the complex exponential. For every complex number z, the imaginary part of $\exp(z)$ is $\exp(\operatorname{Re} z) \cdot \sin(\operatorname{Im} z)$. *Used for:* Damped oscillations $e^{-\gamma t} \sin(\omega t)$. *Price:* **free**; exact.

227. Exponential of the negative is the inverse. For every complex number z, $\exp(z) \cdot \exp(-z) = 1$, and $\exp(-z)$ is the only complex number w with $\exp(z) \cdot w = 1$. *Used for:* Time reversal; the inverse of a phase factor or of a growth factor. *Price:* **free**; exact.

228. Exponential of n times z is the n-th power. For every complex number z and every natural number n, $\exp(n\,z) = \exp(z)^n$. *Used for:* Floquet and Bloch factors over n periods or n lattice cells. *Price:* **free**; exact.

229. Euler's formula for complex angles. For every complex number x, $\exp(i\,x) = \cos x + i \sin x$, where the complex cosine and sine are $\cos x = (\exp(i\,x) + \exp(-i\,x))/2$ and $\sin x = (\exp(-i\,x) - \exp(i\,x)) \cdot i/2$. *Used for:* Plane waves, phasors and rotating frames. *Price:* **free**; exact.

230. De Moivre's formula. For every complex number z and every natural number n, $(\cos z + i \sin z)^n = \cos(n\,z) + i \sin(n\,z)$, where the complex cosine and sine are $\cos z = (\exp(i\,z) + \exp(-i\,z))/2$ and $\sin z = (\exp(-i\,z) - \exp(i\,z)) \cdot i/2$. *Used for:* Multiple-angle formulas; n successive rotations, each by one fixed angle θ, make a rotation by $n\,\theta$. *Price:* **free**; exact.

231. Complex cosine extends the real cosine. For every real number x, the complex cosine $(\exp(i\,x) + \exp(-i\,x))/2$ equals the real cosine $\cos x$. *Used for:* Continuing real oscillation formulas to complex frequencies and damped modes. *Price:* **free**; exact.

232. Complex sine extends the real sine. For every real number x, the complex sine $(\exp(-i\,x) - \exp(i\,x)) \cdot i/2$ equals the real sine $\sin x$. *Used for:* Continuing real oscillation formulas to complex frequencies and damped modes. *Price:* **free**; exact.

233. Size of the complex exponential. For every complex number z, $|\exp(z)| = \exp(\operatorname{Re} z)$. *Used for:* Growth and decay rates of damped or unstable modes. *Price:* **free**; exact.

234. A phase factor has size one. For every real number x, $|\exp(i\,x)| = 1$. *Used for:* Phase factors conserve probability; unitarity of phase rotations. *Price:* **free**; exact.

235. Complex exponential never vanishes. For every complex number z, $\exp(z) \neq 0$. *Used for:* Dividing by exponential factors; the phase factor $e^{i\,S/\hbar}$ of a path integral never vanishes. *Price:* **free**; exact.

236. Complex exponential is apart from zero (stronger form). For every complex number z, $|\exp(z)|^2 > 0$, so $\exp(z)$ lies a positive distance away from 0. *Used for:* Dividing by exponential factors such as Boltzmann weights and the phase factors $e^{i\,S/\hbar}$ of a path integral. *Price:* **free**; exact.

237. Exponential series sums to the complex exponential. For every complex number z, the series $\sum z^n/n!$ over n = 0, 1, 2, … converges unconditionally to $\exp(z) = \exp(\mathrm{Re}\ z) \cdot (\cos(\mathrm{Im}\ z) + i \sin(\mathrm{Im}\ z))$, in its real part and in its imaginary part. *Used for:* The exponential series summed term by term: rotating phasors, damped oscillations and perturbation expansions of exponential factors. *Price:* **free**; exact.

238. Real part of the exponential, from its series. For all complex numbers z and w, if the real parts of the terms $z^n/n!$ (n = 0, 1, 2, …) sum unconditionally to Re w, then $\mathrm{Re}\ w = \exp(\mathrm{Re}\ z) \cdot \cos(\mathrm{Im}\ z)$. *Used for:* Damped oscillations $e^{-\gamma t}\cos(\omega t)$ computed from the exponential series. *Price:* **free**; exact.

239. Imaginary part of the exponential, from its series. For all complex numbers z and w, if the imaginary parts of the terms $z^n/n!$ (n = 0, 1, 2, …) sum unconditionally to Im w, then $\mathrm{Im}\ w = \exp(\mathrm{Re}\ z) \cdot \sin(\mathrm{Im}\ z)$. *Used for:* Damped oscillations $e^{-\gamma t}\sin(\omega t)$ computed from the exponential series. *Price:* **free**; exact.

240. When the complex exponential equals one. For every complex number z, $\exp(z) = 1$ exactly when $z = 2\pi i\, n$ for some integer n. *Used for:* Single-valued wave functions $e^{im\varphi}$ and quantized angular momentum; Bloch periodicity; the period lattice of every phase factor. *Price:* **free**; exact.

241. Zeros of the sine are the integer multiples of π. For every real number x, $\sin x = 0$ exactly when $x = n\pi$ for some integer n. *Used for:* Quantization of standing waves: a string of length L fixed at both ends has modes with $k\,L = n\,\pi$; the nodes of a particle in a box. *Price:* **free**; exact.

242. Zeros of the cosine are the odd multiples of $\pi/2$. For every real number θ, $\cos\theta = 0$ exactly when $\theta = (2k + 1)\pi/2$ for some integer k. *Used for:* Modes of a pipe or a string with one open end, $k\,L = (2n + 1)\,\pi/2$; quarter-wave resonances. *Price:* **free**; exact.

243. Cosine equals one exactly at whole turns. For every real number x, $\cos x = 1$ exactly when $x = 2\pi n$ for some integer n. *Used for:* Single-valued phases $e^{im\varphi}$ force an integer m (quantized angular momentum); constructive interference at whole-wavelength path differences. *Price:* **free**; exact.

244. Sine vanishes only at 0 between $-\pi$ and π. For every real number x with $-\pi < x < \pi$, $\sin x = 0$ exactly when $x = 0$. *Used for:* The pendulum's only equilibrium within a half turn of the bottom; the first node of a standing wave. *Price:* **free**; exact.

245. Cosine equals one only at 0 within a full turn. For every real number x with $-2\pi < x < 2\pi$, $\cos x = 1$ exactly when $x = 0$. *Used for:* A rotation by less than a full turn that returns every vector is no rotation; the central interference maximum. *Price:* **free**; exact.

246. Sine of a quarter turn is one. $\sin(\pi/2) = 1$. *Used for:* The peak displacement of an oscillator a quarter period after it passes equilibrium. *Price:* **free**; exact.

247. Sine vanishes at every integer multiple of π. For every integer n, $\sin(n\pi) = 0$. *Used for:* The nodes of a string fixed at both ends. *Price:* **free**; exact.

248. Cosine is one at every whole number of turns. For every integer n, $\cos(2\pi n) = 1$. *Used for:* A rotation by a whole number of turns changes nothing; periodic phases. *Price:* **free**; exact.

249. Cosine is positive within a quarter turn of 0. For every real number x with $-\pi/2 < x < \pi/2$, $\cos x > 0$. *Used for:* The positive projection of a vector tilted by less than a right angle; Malus's law and forward scattering. *Price:* **free**; exact.

250. Cosine is nonnegative within a quarter turn of 0. For every real number x with $-\pi/2 \le x \le \pi/2$, $\cos x \ge 0$. *Used for:* Nonnegative projections and obliquity factors. *Price:* **free**; exact.

251. Sine is nonnegative on the closed half turn. For every real number x with $0 \le x \le \pi$, $\sin x \ge 0$. *Used for:* The first standing-wave mode has one sign; the area element $r^2 \sin\theta$ of spherical coordinates is nonnegative. *Price:* **free**; exact.

252. Sine is negative between $-\pi$ and 0. For every real number x with $-\pi < x < 0$, $\sin x < 0$. *Used for:* The restoring force of a displaced pendulum points back toward equilibrium. *Price:* **free**; exact.

253. Sine is strictly increasing from $-\pi/2$ to $\pi/2$. For all real numbers x and y with $-\pi/2 \le x < y \le \pi/2$, $\sin x < \sin y$. *Used for:* An angle within a quarter turn of 0 is fixed by its sine (Snell's law, the arcsine). *Price:* **free**; exact.

254. Cosine is strictly decreasing from 0 to π. For all real numbers x and y with $0 \le x < y \le \pi$, $\cos y < \cos x$. *Used for:* A scattering angle is fixed by its cosine (the Compton and Rutherford formulas). *Price:* **free**; exact.

255. Equal cosines: angles agree up to sign and whole turns. For all real numbers x and y, $\cos x = \cos y$ exactly when $y = 2\pi k + x$ or $y = 2\pi k - x$ for some integer k. *Used for:* Diffraction maxima and equal phases of two waves; reading an angle from its cosine. *Price:* **LLPO**; exact.

256. Equal sines: angles agree up to reflection and whole turns. For all real numbers x and y, $\sin x = \sin y$ exactly when $y = 2\pi k + x$ or $y = (2k + 1)\pi - x$ for some integer k. *Used for:* The Bragg condition and equal phases of two waves; reading an angle from its sine. *Price:* **LLPO**; exact.

257. Uniqueness of power series coefficients. Let $r > 0$. If, at every real x with $|x| \le r$, the power series $a_0 + a_1x + a_2x^2 + \dots$ and $b_0 + b_1x + b_2x^2 + \dots$ with real coefficients both converge unconditionally and their sums are equal, then $a_n = b_n$ for every n. *Used for:* Matching coefficients order by order in perturbation theory, the Frobenius method for the radial equations of atoms, reading off counts from the generating functions of statistical mechanics. *Price:* **free**; exact.

258. Identity theorem for a power series at a point. Let $r > 0$ and let f be a function on the real line with $f(w + x) = a_0 + a_1x + a_2x^2 + \dots$ (converging unconditionally) for every x with $|x| \le r$. Then these are equivalent: (1) f vanishes at points other than w arbitrarily close to w, in the sense that no $\varepsilon > 0$ makes $f(z) \neq 0$ for every $z \neq w$ with $|z - w| < \varepsilon$; (2) there is an $\varepsilon > 0$ with $f(z) = 0$ for every z with $|z - w| < \varepsilon$. *Used for:* The identity theorem behind analytic continuation (here for a function of a real variable): a response function or a wavefunction given by a convergent expansion that vanishes at points accumulating at a regular point vanishes on a whole neighbourhood of that point. *Price:* **free**; exact.

259. Power series of a locally zero function vanishes. Let $r > 0$ and let f be a function on the real line with $f(w + x) = a_0 + a_1x + a_2x^2 + \dots$ (converging unconditionally) for every x with $|x| \le r$. If there is an $\varepsilon > 0$ with $f(z) = 0$ for every z with $|z - w| < \varepsilon$, then every coefficient a_n is 0. *Used for:* A quantity that vanishes for all small values of a coupling has no perturbative contribution at any order. *Price:* **free**; exact.

260. First power series coefficient is the derivative. Let $r > 0$ and let f be a function on the real line with $f(x + y) = a_0 + a_1y + a_2y^2 + \dots$ (converging unconditionally) for every y with $|y| \le r$. Then f has derivative a_1 at x. *Used for:* Linear response: the first-order coefficient of a convergent expansion in a coupling or a field is the derivative of the response at the expansion point (susceptibilities, first-order perturbation energies). *Price:* **free**; exact.

261. Sine of the arcsine. For every x with $-1 \le x \le 1$ there is an angle θ with $-\pi/2 \le \theta \le \pi/2$ and $\sin\theta = x$ (θ is the arcsine of x, so $\sin(\arcsin x) = x$). *Used for:* Angles read from a sine: the refraction angle in Snell's law, the Bragg angle from $\sin\theta = n\lambda/2d$, the amplitude angle of a pendulum. *Price:* **free**; exact.

262. Arcsine of the sine. If $-\pi/2 \le x \le \pi/2$, then every angle θ with $-\pi/2 \le \theta \le \pi/2$ and $\sin\theta = \sin x$ equals x (so $\arcsin(\sin x) = x$). *Used for:* The refraction angle in Snell's law and the Bragg angle are single-valued: the arcsine undoes the sine within a quarter turn of 0. *Price:* **free**; exact.

263. Arccosine of the cosine. If $0 \le x \le \pi$, then every angle θ with $0 \le \theta \le \pi$ and $\cos\theta = \cos x$ equals x (so $\arccos(\cos x) = x$). *Used for:* A scattering angle is fixed by its cosine (Compton and Rutherford formulas); a bond angle is fixed by the dot product of its bonds. *Price:* **free**; exact.

264. Tangent of the arctangent. For every real x there is an angle θ with $-\pi/2 < \theta < \pi/2$ and $\tan\theta = x$, where $\tan\theta = \sin\theta / \cos\theta$ and $\cos\theta > 0$ on that window (θ is the arctangent of x, so $\tan(\arctan x) = x$). *Used for:* Turning a ratio into an angle: the phase lag of a driven oscillator, the Hall angle, the Brewster angle $\arctan(n_2/n_1)$, the direction of a vector in the plane. *Price:* **free**; exact.

265. Arctangent of the tangent. If $-\pi/2 < x < \pi/2$, then every angle θ with $-\pi/2 < \theta < \pi/2$ and $\tan\theta = \tan x$ equals x (so $\arctan(\tan x) = x$). *Used for:* The arctangent undoes the tangent within a quarter turn of 0: a phase lag or a slope angle is single-valued. *Price:* **free**; exact.

266. Arctangent of zero. The angle 0 lies strictly between $-\pi/2$ and $\pi/2$ and has tangent 0, and every angle strictly between $-\pi/2$ and $\pi/2$ with tangent 0 is 0 (so $\arctan 0 = 0$). *Used for:* An undamped oscillator driven far below resonance moves in phase with the drive; a horizontal line has slope angle 0. *Price:* **free**; exact.

267. Arctangent of one is $\pi/4$. The angle $\pi/4$ lies strictly between $-\pi/2$ and $\pi/2$ and has $\tan(\pi/4) = 1$, and every angle strictly between $-\pi/2$ and $\pi/2$ with tangent 1 is $\pi/4$ (so $\arctan 1 = \pi/4$). *Used for:* The 45° phase lag of the current in a driven series RLC circuit at a half-power frequency; a beam split equally between two perpendicular directions. *Price:* **free**; exact.

268. Hyperbolic sine of the inverse hyperbolic sine. For every real x, $\sinh(\operatorname{arsinh} x) = x$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Rapidity from momentum: $\eta = \operatorname{arsinh}(p/mc)$ with $p = mc\sinh\eta$; the shape of a hanging chain. *Price:* **free**; exact.

269. Inverse hyperbolic sine of the hyperbolic sine. For every real x, $\operatorname{arsinh}(\sinh x) = x$, where $\operatorname{arsinh} y = \ln(y + \sqrt{(1 + y^2)})$. *Used for:* The rapidity is recovered from the momentum it produces; rapidities of boosts along one line add. *Price:* **free**; exact.

270. Hyperbolic cosine of the inverse hyperbolic cosine. For every $x \ge 1$, $\cosh(\operatorname{arcosh} x) = x$, where $\operatorname{arcosh} x = \ln(x + \sqrt{(x^2 - 1)})$. *Used for:* Rapidity from the Lorentz factor: $\eta = \operatorname{arcosh}\gamma$ with $\gamma = \cosh\eta$; the eccentric anomaly of a hyperbolic orbit. *Price:* **free**; exact.

271. Inverse hyperbolic cosine of the hyperbolic cosine. For every $x \ge 0$, $\operatorname{arcosh}(\cosh x) = x$, where $\operatorname{arcosh} y = \ln(y + \sqrt{(y^2 - 1)})$ for $y \ge 1$. *Used for:* A nonnegative rapidity is recovered from its Lorentz

factor $\gamma = \cosh \eta$. *Price:* **free**; exact.

272. Tangent with a value at every real number. For every real x there is a number w such that $\cos x \cdot w = 1$ whenever $\cos x \neq 0$, and $w = 0$ whenever $\cos x = 0$; the tangent $\sin x \cdot w$ is then defined at every real x. *Used for:* Physics reads the tangent where the cosine is away from 0, as in a phase angle $\tan \varphi = \omega L/R$ or the slope of a line, and there it costs nothing. A tangent with a value at every real number, including the odd multiples of $\pi/2$ where the cosine vanishes, is a convention of formal mathematics that physics does not use. *Price:* **LPO**; exact.

273. Zeros of the sine are where the cosine is ±1. For every real x, $\sin x = 0$ if and only if $\cos x = 1$ or $\cos x = -1$. *Used for:* At a node of a standing wave the conjugate quadrature is at its full amplitude, $+1$ or -1; an oscillator's velocity vanishes exactly at its turning points. *Price:* **free**; exact.

274. Zeros of the cosine are where the sine is ±1. For every real x, $\cos x = 0$ if and only if $\sin x = 1$ or $\sin x = -1$. *Used for:* An oscillator passes its equilibrium point at full speed: where the displacement quadrature vanishes the velocity quadrature is $+1$ or -1. *Price:* **free**; exact.

275. Cosine of a whole multiple of π. For every integer k, $|\cos(k\pi)| = 1$. *Used for:* Full amplitude at whole half-periods: a cosine wave sampled at multiples of half its period is $+1$ or -1, the alternating signs $(-1)^k$ of standing-wave modes at a boundary. *Price:* **free**; exact.

276. Cosine is an even function. For every real x, $\cos |x| = \cos x$. *Used for:* Quantities built from the cosine do not see the sign of a displacement or a phase: a pendulum's potential energy $1 - \cos \theta$, the tight-binding band $-2t \cos k$. *Price:* **free**; exact.

277. Hyperbolic cosine is an even function. For every real x, $\cosh |x| = \cosh x$. *Used for:* The Lorentz factor $\gamma = \cosh \eta$ depends only on the size of the rapidity, not on the direction of motion; a hanging chain is symmetric about its lowest point. *Price:* **free**; exact.

278. Absolute value of the hyperbolic sine. For every real x, $|\sinh x| = \sinh |x|$. *Used for:* The size of the momentum $p = mc \sinh \eta$ is fixed by the size of the rapidity η, whatever the direction of the boost. *Price:* **free**; exact.

279. Equal cosines of two angles. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For all real θ and ψ, $\cos \theta = \cos \psi$ if and only if θ and ψ are equal as angles or θ and $-\psi$ are equal as angles. *Used for:* Phase matching of waves: equal cosines fix a phase only up to its sign and whole turns (interference fringes, standing waves, the two solutions of a cosine condition). *Price:* **LLPO**; exact.

280. Equal sines of two angles. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For all real θ and ψ, $\sin \theta = \sin \psi$ if and only if θ and ψ are equal as angles or $\theta + \psi$ and π are equal as angles. *Used for:* The two solutions θ and $\pi - \theta$ of a sine condition (Snell's law, Bragg's law), up to whole turns; phase matching of waves. *Price:* **LLPO**; exact.

281. Zeros of the sine of an angle. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For every real θ, $\sin \theta = 0$ if and only if θ is equal as an angle to 0 or to π. *Used for:* The nodes of a wave on a ring or of a standing wave sit at phase 0 or π, up to whole turns. *Price:* **free**; exact.

282. Zeros of the cosine of an angle. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For every real θ, $\cos \theta = 0$ if and only if θ is equal as an angle to $\pi/2$ or to $-\pi/2$. *Used for:* Quarter-wave points: where the cosine quadrature of a wave vanishes the phase is a quarter turn from 0, one way or the other, up to whole turns (quadrature detection, a driven oscillator at resonance). *Price:* **free**; exact.

283. Angles with equal doubles. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For all real ψ and θ, 2ψ and 2θ are equal as angles if and only if ψ and θ are equal as angles or ψ and $\theta + \pi$ are equal as angles. *Used for:* Directions of lines counted modulo a half turn: a polarisation axis or a liquid-crystal director is fixed by twice its angle. *Price:* **free**; exact.

284. Angles equal to their own negative. Read an angle as a real number up to adding whole turns of 2π: θ and ψ are equal as angles when $\theta - \psi = 2\pi k$ for some integer k. For every real θ, θ and $-\theta$ are equal as angles if and only if θ is equal as an angle to 0 or to π. *Used for:* Phases left unchanged by a reflection or by time reversal are 0 or π, up to whole turns: real amplitudes carry a sign and nothing more. *Price:* **free**; exact.

285. When a real power is below one. For $x > 0$ and every real y, with $x^y = e^{y \ln x}$: $x^y < 1$ if and only if ($1 < x$ and $y < 0$) or ($x < 1$ and $0 < y$). *Used for:* Telling decay from growth for a Boltzmann factor or a transmission coefficient x^y when the sign of the exponent is known strictly. *Price:* **free**; exact.

286. When a real power is above one. For $x > 0$ and every real y, with $x^y = e^{y \ln x}$: $1 < x^y$ if and only if ($1 < x$ and $0 < y$) or ($x < 1$ and $y < 0$). *Used for:* Gain versus loss for an amplification factor or a growth law

x^y when the sign of the exponent is known strictly. *Price:* **free**; exact.

287. When a real power is at most one. For $x > 0$ and every real y, with $x^y = e^{y \ln x}$: $x^y \le 1$ if and only if ($1 \le x$ and $y \le 0$) or ($x \le 1$ and $0 \le y$). *Used for:* Bounding a Boltzmann factor or a transmission coefficient x^y by 1 when the base may be exactly 1 or the exponent exactly 0. *Price:* **LLPO**; exact.

288. Size of the sine from the cosine. For every real x, if $r \ge 0$ and $r^2 = 1 - \cos^2 x$, then $r = |\sin x|$: a nonnegative square root of $1 - \cos^2 x$ can only be $|\sin x|$. *Used for:* The transverse factor $\sin\theta = \sqrt{(1 - \mu^2)}$ of a direction with $\mu = \cos\theta$: spherical coordinates, the associated Legendre functions, the transverse momentum of a scattered particle. *Price:* **free**; exact.

289. Size of the cosine from the sine. For every real x, if $r \ge 0$ and $r^2 = 1 - \sin^2 x$, then $r = |\cos x|$: a nonnegative square root of $1 - \sin^2 x$ can only be $|\cos x|$. *Used for:* The cosine of an angle from its sine: the cosine of a refraction angle from Snell's law in the Fresnel formulas, the longitudinal factor of a direction. *Price:* **free**; exact.

290. The primitive n-th root of unity. For every natural number $n \ge 1$, $\zeta = e^{2\pi i/n}$ is a primitive n-th root of unity: $\zeta^n = 1$, and $\zeta^l = 1$ for a natural number l only when n divides l. *Used for:* The generator of the n-fold cyclic symmetry of a ring of n sites: Bloch phases of a periodic lattice, the n-fold rotation axis of a molecule or a crystal, the clock operator of an n-level system. *Price:* **free**; exact.

291. Roots of unity have modulus one. If z is a complex number with $z^n = 1$ for some natural number $n \ge 1$, then $|z| = 1$. *Used for:* A phase factor that returns to 1 after n steps conserves probability: the eigenvalues of a unitary symmetry of finite order have modulus one. *Price:* **free**; exact.

292. Discrete Fourier orthogonality on the integers mod N. (The standard characters of the integers mod N.) For natural numbers $N \ge 1$ and m, the sum of $e^{2\pi i \cdot jm/N}$ over $j = 0, 1, \ldots, N - 1$ equals N if N divides m, and equals 0 otherwise. *Used for:* Every lattice Fourier transform inverts because of this identity: Bloch's theorem on a ring of N sites, phonon and tight-binding bands, the discrete Fourier transform of a periodic signal, the quantum Fourier transform. *Price:* **free**; exact.

293. The n-th roots of unity are n phases. For every natural number $n \ge 1$ and every complex number x with an inverse: $x^n = 1$ if and only if $x = e^{2\pi i \cdot k/n}$ for some natural number $k < n$. *Used for:* The eigenvalues of an n-fold cyclic symmetry (a ring of n sites, an n-fold rotation axis, the clock operator of an n-level system) are exactly the n phases $e^{2\pi i k/n}$, so every state carries a quasi-momentum label k. *Price:* **free**; exact.

294. Powers of a primitive root of unity sum to zero. (The case of complex numbers.) If $k > 1$ and ζ is a complex primitive k-th root of unity ($\zeta^k = 1$, and $\zeta^l = 1$ for a natural number l only when k divides l), then $1 + \zeta + \zeta^2 + \ldots + \zeta^{k-1} = 0$. *Used for:* The phases of a ring of k sites cancel: a uniform superposition has zero overlap with every nonzero quasi-momentum; the k phasors of balanced polyphase power, such as three-phase current, sum to zero. *Price:* **free**; exact.

295. Termwise derivative of a power series inside its interval. For real numbers: if $f(x + z) = \Sigma\, a_n\, z^n$ (an unconditionally convergent sum) for every real z with $|z| < r$, where $r > 0$, then at every real y with $|y| < r$ the termwise derivative series $\Sigma\, (n + 1)\, a_{n+1}\, y^n$ converges to a number d, and f has derivative d at $x + y$. *Used for:* Differentiating a convergent expansion in one real variable anywhere inside its interval of convergence, and getting a convergent expansion of the derivative: forces from a potential given by its Taylor series, group velocities from a dispersion relation, the response of a quantity expanded in a coupling or a temperature to that parameter. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

296. Strict logarithm bound, partial-order strictness throughout. For every real x with $0 < x$ and $x \ne 1$, where $a < b$ means $a \le b$ and not $b \le a$ in both places, there is a number L with $e^L = x$ (so L is the logarithm of x) and $L < x - 1$. *Used for:* The relative entropy of two distinct probability laws is strictly positive, read with the strictness that a partial order defines in both the hypothesis and the conclusion. *Price:* **MP**; exact.

297. Liouville's theorem for entire power series. Liouville's theorem for functions given by an everywhere convergent power series: if $f(z) = a_0 + a_1 z + a_2 z^2 + \ldots$ converges unconditionally at every complex number z and $|f(z)|^2 \le M^2$ for one number M and every z, then $a_n = 0$ for every $n \ge 1$, so f is the constant a_0. *Used for:* An analytic function that stays bounded on the whole complex plane is constant, so a nonconstant one such as exp, sin or cos must grow without bound somewhere off the real axis; the standard proof that every polynomial has a complex root applies it to the reciprocal of a polynomial that had none. *Price:* **free**; exact.

298. Cauchy's estimate for power series coefficients. Cauchy's estimate for functions given by a power series: if $R > 0$ and $f(z) = a_0 + a_1 z + a_2 z^2 + \ldots$ converges unconditionally (in its real and imaginary parts) at every complex number z with $|z| = R$, with $|f(z)| \le C$ at each of those points, then $|a_n|\, R^n \le C$ for every n. *Used for:* Bounds the Taylor

coefficients of an analytic function by its size on a circle: the standard tool for the error of truncating a Taylor expansion and, in the form where the function is analytic inside the circle, for the radius of convergence of a perturbation series. *Price:* **free**; exact.

299. Machin's formula for π. If θ_1 and θ_2 are angles strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta_1 = 1/5$ and $\tan\theta_2 = 1/239$ (θ_1 is the arctangent of 1/5 and θ_2 the arctangent of 1/239), then $4\theta_1 - \theta_2 = \pi/4$. *Used for:* Computing π to any number of digits, quickly and with a certified error, for every physical formula that contains π: 4π in Coulomb's and Gauss's laws, 2π in angular frequency, the reduced Planck constant $h/2\pi$. *Price:* **free**; exact.

300. Addition formula for the arctangent. For real x and y with $x \cdot y < 1$, if θ, φ and ψ are angles strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta = x$, $\tan\varphi = y$ and $\tan\psi = (x + y)/(1 - x \cdot y)$, then $\theta + \varphi = \psi$ (so $\arctan x + \arctan y = \arctan((x + y)/(1 - x \cdot y))$). *Used for:* Adding the phase lags of two stages in series, such as two first-order filters or two driven oscillators in a chain; composing slope angles; identities such as $\arctan(1/2) + \arctan(1/3) = \pi/4$ that turn arctangent series into fast formulas for π. *Price:* **free**; exact.

301. Arctangent of a reciprocal. For every real $x > 0$, if θ and ψ are angles strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta = x$ and $\tan\psi = 1/x$, then $\psi = \pi/2 - \theta$ (so $\arctan(1/x) = \pi/2 - \arctan x$). *Used for:* Complementary angles of a right triangle; the phases across the resistor and across the capacitor of an RC divider add to $\pi/2$; converting an angle measured from one axis into the angle measured from the other. *Price:* **free**; exact.

302. Gregory's series for the arctangent. For every real x with $|x| < 1$ and every angle θ strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta = x$, the series $x - x^3/3 + x^5/5 - \ldots = \sum_n (-1)^n x^{2n+1}/(2n + 1)$ has sum θ, unconditionally: the sums over finite sets of indices tend to θ, so the order of the terms does not matter (so the series sums to $\arctan x$). *Used for:* Computing arctangents, and with Machin's formula π itself, to any precision with a certified remainder (after N terms the error is at most $|x|^{2N+1}$); small-angle expansions of phase lags, Brewster and Hall angles, and scattering angles. *Price:* **free**; exact.

303. Derivative of the arctangent. For every real x, the arctangent has derivative $1/(1 + x^2)$ at x: for every k there is j such that whenever $|y - x| \le 1/(j + 1)$, the angles θ and φ strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta = x$ and $\tan\varphi = y$ satisfy $|\varphi - \theta - (y - x)/(1 + x^2)| \le |y - x|/(k + 1)$. *Used for:* The Lorentzian (Cauchy) line shape $1/(1 + x^2)$ of a resonance is the derivative of the arctangent; it gives the phase response of a damped oscillator near resonance and the integrals behind Rutherford scattering and Landau–Zener transitions. *Price:* **free**; exact.

304. Derivative of the tangent (cosine away from zero). If $\cos x$ is apart from 0 (known to be positive or known to be negative), then the tangent has derivative $1/\cos^2 x$ at x: for every k there is j such that $\cos y$ is apart from 0 whenever $|y - x| \le 1/(j + 1)$, and for those y, $|\tan y - \tan x - (y - x)/\cos^2 x| \le |y - x|/(k + 1)$, where tan = sin/cos. *Used for:* The sensitivity of a slope to its angle, $d\tan\theta/d\theta = 1/\cos^2\theta = 1 + \tan^2\theta$: an inclined plane, a deflection read on a flat screen, small-angle error propagation in optics, and the substitution $u = \tan\theta$ in Lorentzian integrals. *Price:* **free**; exact.

305. Derivative of the tangent (cosine not zero). If $\cos x \neq 0$ (only the negation of $\cos x = 0$ is known), then the number $1/\cos^2 x$ exists, and for every k there is j such that $\cos y$ is apart from 0 (known to be positive or known to be negative) whenever $|y - x| \le 1/(j + 1)$, and for those y, $|\tan y - \tan x - (y - x)/\cos^2 x| \le |y - x|/(k + 1)$, where tan = sin/cos. *Used for:* The slope sensitivity $d\tan\theta/d\theta = 1/\cos^2\theta$ for an angle known only not to be a right angle; turning that negative knowledge into a usable bound on the slope is exactly Markov's principle. *Price:* **MP**; exact.

306. Derivative of the arcsine inside (−1, 1). If $-1 < x < 1$, then the arcsine has derivative $1/\sqrt{(1 - x^2)}$ at x: for every k there is j such that every y with $|y - x| \le 1/(j + 1)$ also lies strictly between −1 and 1, and for those y the angles θ and φ in $[-\pi/2, \pi/2]$ with $\sin\theta = x$ and $\sin\varphi = y$ satisfy $|\varphi - \theta - (y - x)/\sqrt{(1 - x^2)}| \le |y - x|/(k + 1)$. *Used for:* How fast an angle read from its sine moves: the angle of a pendulum or a rotating vector from its projection, the arcsine position distribution of a harmonic oscillator, and the substitution $x = \sin\theta$ behind the integrals of $dx/\sqrt{(1 - x^2)}$ in oscillator and orbital periods. *Price:* **free**; exact.

307. Derivative of the arcsine (point not equal to ±1). If $-1 \le x \le 1$, $x \neq -1$ and $x \neq 1$ (only the negations of $x = -1$ and $x = 1$ are known), then the number $1/\sqrt{(1 - x^2)}$ exists, and for every k there is j such that every y with $|y - x| \le 1/(j + 1)$ lies strictly between −1 and 1, and for those y the angles θ and φ in $[-\pi/2, \pi/2]$ with $\sin\theta = x$ and $\sin\varphi = y$ satisfy $|\varphi - \theta - (y - x)/\sqrt{(1 - x^2)}| \le |y - x|/(k + 1)$. *Used for:* The rate at which an angle read from its sine moves, for a sine known only not to be ±1; turning that negative knowledge into a bound on the rate is exactly Markov's principle. *Price:* **MP**; exact.

308. Derivative of the arccosine inside (−1, 1). If $-1 < x < 1$, then the arccosine has derivative $-1/\sqrt{(1 - x^2)}$ at x: for every k there is j such that every y with $|y - x| \le 1/(j + 1)$ also lies strictly between −1 and 1, and for those y the angles θ and φ in $[0, \pi]$ with $\cos\theta = x$ and $\cos\varphi = y$ satisfy $|\varphi - \theta + (y - x)/\sqrt{(1 - x^2)}| \le |y -$

$x|/(k + 1)$. *Used for:* How fast an angle read from its cosine moves: a scattering angle from its cosine in the Compton and Rutherford formulas, a bond angle from a dot product, the polar angle of a spin or a rotating vector from its projection on an axis. *Price:* **free**; exact.

309. Adding exponents multiplies powers. For $x > 0$ and all real y and z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^{y+z} = x^y \cdot x^z$. *Used for:* Composing scaling factors: growth by x^y followed by growth by x^z is growth by x^{y+z}, as for decay or amplification over successive intervals. *Price:* **free**; exact.

310. Subtracting exponents divides powers. For $x > 0$ and all real y and z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^{y-z} = x^y / x^z$. *Used for:* Ratios of power-law quantities, such as dividing a scaling law by its value at a reference scale. *Price:* **free**; exact.

311. Power of a power, positive base. For $x > 0$ and all real y and z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^{y \cdot z} = (x^y)^z$. *Used for:* Rewriting a power law in another variable, for example Kepler's law $T^2 \propto a^3$ as $T \propto a^{3/2}$. *Price:* **free**; exact.

312. A negative exponent gives the reciprocal. For $x > 0$ and every real y, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^{-y} = 1/x^y$. *Used for:* Inverse-power laws such as the r^{-2} of gravity and Coulomb's law, written as reciprocals of positive powers. *Price:* **free**; exact.

313. Real powers of positive numbers are positive. For $x > 0$ and every real y, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y > 0$. *Used for:* Positivity of Boltzmann factors, partition-function terms and power-law densities. *Price:* **free**; exact.

314. Logarithm of a real power. For $x > 0$ and every real y, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $\ln(x^y) = y \cdot \ln x$. *Used for:* Reading an exponent as the slope of a log-log plot; decibel and entropy scales of power-law quantities. *Price:* **free**; exact.

315. Exponential of a product as a power. For all real x and y, with real powers of a positive base b defined by $b^y = e^{y \ln b}$: $e^{x \cdot y} = (e^x)^y$. *Used for:* Rewriting a Boltzmann factor or a growth law $e^{r\,t}$ as a power $(e^r)^t$ of a fixed base. *Price:* **free**; exact.

316. Real powers of e are exponentials. For every real x, with real powers of a positive base b defined by $b^y = e^{y \ln b}$: $(e^1)^x = e^x$. *Used for:* Identifying the natural exponential with powers of Euler's number in decay and growth laws. *Price:* **free**; exact.

317. Zeroth power of a positive base is one. For $x > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^0 = 1$. *Used for:* The exponent-zero member of a family of power laws is the constant 1, as in normalising a scaling family. *Price:* **free**; exact.

318. First power of a positive base. For $x > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^1 = x$. *Used for:* Linear laws as the exponent-one member of a family of power laws. *Price:* **free**; exact.

319. Every real power of one is one. For every real y, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $1^y = 1$. *Used for:* A ratio equal to 1 stays 1 under any scaling exponent; normalising a power law at its reference point. *Price:* **free**; exact.

320. Power of a product of positive numbers. For $x > 0$, $y > 0$ and every real z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $(x \cdot y)^z = x^z \cdot y^z$. *Used for:* Splitting a power law of a product of quantities, as in dimensional analysis and the scaling of composite quantities. *Price:* **free**; exact.

321. Power of a reciprocal, positive base. For $x > 0$ and every real y, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $(1/x)^y = 1/x^y$. *Used for:* Moving a power between numerator and denominator, as with inverse-distance and inverse-temperature laws. *Price:* **free**; exact.

322. Power of a quotient of positive numbers. For $x > 0$, $y > 0$ and every real z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $(x/y)^z = x^z / y^z$. *Used for:* Powers of dimensionless ratios such as $(r/r_0)^n$ and $(T/T_0)^\beta$ in scaling laws. *Price:* **free**; exact.

323. Comparing powers of a base above one. For $x > 1$ and all real y and z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y \le x^z$ if and only if $y \le z$. *Used for:* Ordering growth factors by their exponents; comparing levels of a geometric hierarchy of scales. *Price:* **free**; exact.

324. Strictly comparing powers of a base above one. For $x > 1$ and all real y and z, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y < x^z$ if and only if $y < z$. *Used for:* Recovering the order of two exponents from measured growth factors, as in fitting power laws. *Price:* **free**; exact.

325. Larger exponent gives a larger power above one. For $x > 1$ and real numbers $y < z$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y < x^z$. *Used for:* Growth with a larger exponent is larger: populations, compound interest, chained amplifier stages. *Price:* **free**; exact.

326. Larger exponent gives no smaller power from one up. For $x \ge 1$ and real numbers $y \le z$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y \le x^z$. *Used for:* Bounding a growth factor by the one with the larger exponent. *Price:* **free**; exact.

327. Larger exponent gives a smaller power below one. For $0 < x < 1$ and real numbers $z < y$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y < x^z$. *Used for:* Attenuation: a thicker absorber or a longer time gives a smaller transmitted or surviving fraction. *Price:* **free**; exact.

328. Larger exponent gives no larger power up to one. For $0 < x \le 1$ and real numbers $z \le y$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^y \le x^z$. *Used for:* Bounding a survival or transmission fraction raised to a larger exponent by the one with the smaller exponent. *Price:* **free**; exact.

329. A nonnegative power keeps the order of positive bases. For real numbers $0 < x \le y$ and every real $z \ge 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^z \le y^z$. *Used for:* Monotone scaling laws such as radiated power $\propto T^4$: a larger input never gives a smaller output. *Price:* **free**; exact.

330. A positive power keeps strict order of positive bases. For real numbers $0 < x < y$ and every real $z > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^z < y^z$. *Used for:* Strictly monotone scaling laws: a hotter black body radiates strictly more, a heavier star is strictly brighter on its mass-luminosity law. *Price:* **free**; exact.

331. A positive power preserves and reflects order. For $x > 0$, $y > 0$ and every real $z > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^z \le y^z$ if and only if $x \le y$. *Used for:* Inverting a monotone power law to compare inputs from outputs, as when comparing temperatures from radiated powers. *Price:* **free**; exact.

332. A positive power preserves and reflects strict order. For $x > 0$, $y > 0$ and every real $z > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^z < y^z$ if and only if $x < y$. *Used for:* Deciding which of two inputs is larger from the outputs of a power law, such as stellar masses from luminosities. *Price:* **free**; exact.

333. A negative power reverses strict order. For real numbers $0 < x < y$ and every real $z < 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $y^z < x^z$. *Used for:* Inverse-power laws fall with distance: gravity and Coulomb forces $\propto r^{-2}$, density profiles $\propto r^{-\alpha}$. *Price:* **free**; exact.

334. A nonpositive power reverses order. For real numbers $0 < x \le y$ and every real $z \le 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $y^z \le x^z$. *Used for:* Bounding an inverse-power law at a larger distance by its value at a smaller one. *Price:* **free**; exact.

335. Whole-number real powers of a positive base. For $x > 0$ and every whole number n, with real powers of a positive base defined by $x^y = e^{y \ln x}$: x raised to the real exponent n equals the product of n factors $x \cdot x \cdot \ldots \cdot x$ (and 1 when $n = 0$). *Used for:* Polynomial laws such as kinetic energy $\frac{1}{2}mv^2$ and spring energy $\frac{1}{2}kx^2$ as members of real-exponent scaling families. *Price:* **free**; exact.

336. Real square of a positive base. For $x > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: x raised to the real exponent 2 equals $x \cdot x$. *Used for:* Kinetic energy $\frac{1}{2}mv^2$ and inverse-square laws written with real exponents. *Price:* **free**; exact.

337. Square root as the one-half power. For $x > 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $x^{1/2} \ge 0$, $x^{1/2} \cdot x^{1/2} = x$, and every $r \ge 0$ with $r \cdot r = x$ equals $x^{1/2}$; so the square root $\sqrt{x}$ is $x^{1/2}$. *Used for:* Speeds from kinetic energies $v = (2E/m)^{1/2}$, root-mean-square values, diffusion lengths growing as $t^{1/2}$. *Price:* **free**; exact.

338. The n-th root raised to the n-th power. For $x > 0$ and every whole number $n \neq 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $(x^{1/n})^n = x$. *Used for:* n-th roots undo n-th powers: the side of a cube from its volume, Kepler's semi-major axis from the period. *Price:* **free**; exact.

339. The n-th root of the n-th power. For $x > 0$ and every whole number $n \neq 0$, with real powers of a positive base defined by $x^y = e^{y \ln x}$: $(x^n)^{1/n} = x$. *Used for:* Recovering a length from an area or a volume, and an amplitude from a power law of whole order. *Price:* **free**; exact.

340. Derivative of a power law at a positive point. For $x > 0$ and every real p, with real powers of a positive base defined by $x^y = e^{y \ln x}$: the function $y \mapsto y^p$ has derivative $p \cdot x^{p-1}$ at x. Precisely, for every k there is j such that every y with $|y - x| \le 1/(j + 1)$ is positive and, for all such y, $|y^p - x^p - (y - x) \cdot p \cdot x^{p-1}| \le (1/(k + 1)) \cdot |y - x|$. *Used for:* Forces from power-law potentials, the force being minus the derivative of the potential with respect to the distance; the exponent of a scaling law as its logarithmic slope; heat capacities from energy laws $E \propto T^p$. *Price:* **free**; exact.

341. Adding complex exponents multiplies powers. For the principal logarithm of x, that is, any complex number w with $\exp(w) = x$ and $-\pi < \operatorname{Im} w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: $x^{y+z} = x^y \cdot x^z$ for all complex numbers y and z. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. *Used for:* Combining complex power-law factors, such as the frequency factors $(i\omega)^\alpha \cdot (i\omega)^\beta = (i\omega)^{\alpha+\beta}$ of fractional (constant-phase) elements in impedance and viscoelastic models, and complex scaling exponents in critical phenomena. *Price:* **free**; exact.

342. Subtracting complex exponents divides powers. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for all complex numbers y and z and every complex v with $x^z \cdot v = 1$, $x^{y-z} = x^y \cdot v$ (that is, $x^{y-z} = x^y / x^z$). The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. *Used for:* Ratios of complex power-law factors, such as a transfer function $(i\omega)^\alpha / (i\omega)^\beta$ of two fractional elements, or a scaling law divided by its value at a reference scale. *Price:* **free**; exact.

343. Negative complex exponent gives the reciprocal. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for every complex number y, $x^y \cdot x^{-y} = 1$, and every complex v with $x^y \cdot v = 1$ equals x^{-y} (that is, $x^{-y} = 1/x^y$). The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. The case x = 0, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Inverting a complex power law, such as turning the admittance $(i\omega)^\alpha$ of a constant-phase element into its impedance $(i\omega)^{-\alpha}$, or a growth factor into the matching decay factor. *Price:* **free**; exact.

344. Complex power minus one is the reciprocal. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: $x \cdot x^{-1} = 1$, and every complex v with $x \cdot v = 1$ equals x^{-1} (that is, $x^{-1} = 1/x$). The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. The case x = 0, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Consistency of power-law formulas with the reciprocal, such as admittance as the reciprocal of impedance $Z^{-1} = 1/Z$ in an alternating-current circuit. *Price:* **free**; exact.

345. Zeroth complex power of a nonzero number. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: $x^0 = 1$. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. *Used for:* Normalization of power-law factors: a complex power-law factor with zero exponent is 1, as for a fractional element of order zero, which reduces to a plain resistor. *Price:* **free**; exact.

346. First complex power of a nonzero number. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: $x^1 = x$. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. *Used for:* Consistency of complex power laws with the ordinary first power, as when a fractional element of order one reduces to an ordinary capacitor or inductor. *Price:* **free**; exact.

347. Whole-number multiple of a complex exponent. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: $x^{n \cdot y} = (x^y)^n$ for every natural number n and every complex number y, where $(x^y)^n$ is the product of n copies of x^y. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. The case x = 0, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Repeating a complex scaling step n times, as for n identical fractional stages in a cascade, whose total factor is the n-th power of one stage's factor. *Price:* **free**; exact.

348. Complex power with a whole-number exponent. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for every natural number n, x raised to the complex exponent n equals x^n, the product of n copies of x. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. The case x = 0, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Agreement of complex power-law formulas with ordinary powers at whole-number exponents, such as $(i\omega)^2 = -\omega^2$ in the transfer function of a second-order (mass and spring) system. *Price:* **free**; exact.

349. Principal n-th root of a nonzero complex number. For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for every natural number $n \ge 1$, $(x^{1/n})^n = x$, so $x^{1/n} = \exp(w/n)$ is an n-th root of x (the principal n-th root). The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. *Used for:* Taking roots of complex quantities, such as the complex wave number $k = (\omega/c) \cdot \varepsilon^{1/2}$ in an absorbing medium, the momentum $(2mE)^{1/2}$ of a bound or decaying state, and the roots in Cardano's formula. *Price:* **free**; exact.

350. Principal n-th roots of complex numbers (n ≥ 2). For the principal logarithm of x, that is, any complex number w with exp(w) = x and $-\pi < \mathrm{Im}\ w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for every whole number $n \ge 2$, $(x^{1/n})^n = x$; for n = 2 this says $x^{1/2} = \exp(w/2)$ is a square root of x (the principal square root), and for n = 3 that $x^{1/3}$ is a cube root of x. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x

costs exactly LPO. The case $x = 0$, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Square roots of complex numbers throughout physics: the complex refractive index $(\varepsilon)^{1/2}$ of an absorbing medium, the two roots of a damped oscillator's characteristic equation, and the eigenvalues of a 2×2 non-Hermitian Hamiltonian. *Price:* **free**; exact.

351. The exponential is one-to-one on the principal strip. If two complex numbers a and b both have imaginary part in $-\pi < \mathrm{Im} \le \pi$ and $\exp(a) = \exp(b)$, then $a = b$. *Used for:* A phase in the window from $-\pi$ to π is determined by the phase factor it produces: reading scattering phase shifts, WKB phases and the logarithm of an amplitude on one sheet without ambiguity. *Price:* **free**; exact.

352. The logarithm undoes the exponential on the principal strip. For every complex number a with $-\pi < \mathrm{Im}\, a \le \pi$, every principal logarithm of $\exp(a)$, that is, every complex w with $\exp(w) = \exp(a)$ and $-\pi < \mathrm{Im}\, w \le \pi$, equals a (so $\log(\exp a) = a$). *Used for:* Recovering a log-amplitude and phase from a complex amplitude on the principal sheet, as when a scattering phase shift or a WKB phase in the window from $-\pi$ to π is read back from its phase factor. *Price:* **free**; exact.

353. Power of a complex power inside the principal strip. For the principal logarithm of x, that is, any complex number w with $\exp(w) = x$ and $-\pi < \mathrm{Im}\, w \le \pi$, and with complex powers defined by $x^y = \exp(w \cdot y)$: for complex numbers y and z with $-\pi < \mathrm{Im}(w \cdot y) \le \pi$, $x^{y \cdot z} = (x^y)^z$, where $(x^y)^z = \exp(v \cdot z)$ for every principal logarithm v of x^y. The logarithm w is given: the identity holds for every such w, and producing w from a nonzero x costs exactly LPO. The case $x = 0$, with the conventions $0^0 = 1$ and $0^y = 0$, is not priced here. *Used for:* Rewriting a complex power law in another variable, such as $(i\omega)^{\alpha\beta} = ((i\omega)^{\alpha})^{\beta}$ for nested fractional elements, valid as long as the intermediate phase stays in the window from $-\pi$ to π. *Price:* **free**; exact.

354. Double-angle formula for the arctangent. For every real x with $-1 < x < 1$, if θ and ψ are angles strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta = x$ and $\tan\psi = 2x/(1 - x^2)$, then $2\theta = \psi$ (so $2 \arctan x = \arctan(2x/(1 - x^2))$). *Used for:* Doubling a phase lag, as for two identical first-order stages (filters or relaxation processes) in series; the doubling steps in fast arctangent formulas for π. *Price:* **free**; exact.

355. Euler's arctangent formula for π. If θ_1 and θ_2 are angles strictly between $-\pi/2$ and $\pi/2$ with $\tan\theta_1 = 1/2$ and $\tan\theta_2 = 1/3$ (θ_1 is the arctangent of 1/2 and θ_2 the arctangent of 1/3), then $\theta_1 + \theta_2 = \pi/4$. *Used for:* Computing π with fast arctangent series, for every physical formula that contains π (4π in Coulomb's and Gauss's laws, 2π in angular frequency, the reduced Planck constant $h/2\pi$); the 45° total phase of two stages whose phase lags have tangents 1/2 and 1/3. *Price:* **free**; exact.

356. Pythagorean identity for complex angles. For every complex number z, $\cos^2 z + \sin^2 z = 1$, where for every complex number w the complex cosine and sine are $\cos w = (\exp(i\,w) + \exp(-i\,w))/2$ and $\sin w = (\exp(-i\,w) - \exp(i\,w)) \cdot i/2$. *Used for:* Puts the cosine and sine of a complex frequency or a complex propagation constant on an equal footing with real angles, as used for damped oscillators, lossy waveguides and leaky resonator modes. *Price:* **free**; exact.

357. Pythagorean identity, sine first, complex angles. For every complex number z, $\sin^2 z + \cos^2 z = 1$, where for every complex number w the complex cosine and sine are $\cos w = (\exp(i\,w) + \exp(-i\,w))/2$ and $\sin w = (\exp(-i\,w) - \exp(i\,w)) \cdot i/2$. *Used for:* Puts the sine and cosine of a complex frequency or a complex propagation constant on an equal footing with real angles, as used for damped oscillators, lossy waveguides and leaky resonator modes. *Price:* **free**; exact.

358. Sine addition formula for complex angles. For all complex numbers x and y, $\sin(x + y) = \sin x \cdot \cos y + \cos x \cdot \sin y$, where for every complex number w the complex cosine and sine are $\cos w = (\exp(i\,w) + \exp(-i\,w))/2$ and $\sin w = (\exp(-i\,w) - \exp(i\,w)) \cdot i/2$. *Used for:* Combining phases and shifting the argument of a mode with complex frequency or complex wavenumber, for example in damped oscillations and in waves that decay as they travel. *Price:* **free**; exact.

359. Cosine addition formula for complex angles. For all complex numbers x and y, $\cos(x + y) = \cos x \cdot \cos y - \sin x \cdot \sin y$, where for every complex number w the complex cosine and sine are $\cos w = (\exp(i\,w) + \exp(-i\,w))/2$ and $\sin w = (\exp(-i\,w) - \exp(i\,w)) \cdot i/2$. *Used for:* Combining phases and shifting the argument of a mode with complex frequency or complex wavenumber, for example in damped oscillations and in waves that decay as they travel. *Price:* **free**; exact.

360. Complex cosine is even. For every complex number z, $\cos(-z) = \cos z$, where for every complex number w the complex cosine and sine are $\cos w = (\exp(i\,w) + \exp(-i\,w))/2$ and $\sin w = (\exp(-i\,w) - \exp(i\,w)) \cdot i/2$. *Used for:* Symmetric (even) standing-wave modes, and the fact that a cosine mode does not depend on the sign of a complex wavenumber. *Price:* **free**; exact.

361. Complex sine is odd. For every complex number z, $\sin(-z) = -\sin z$, where for every complex

number w the complex cosine and sine are cos w = (exp(i w) + exp(−i w))/2 and sin w = (exp(−i w) − exp(i w)) · i/2. *Used for:* Antisymmetric (odd) standing-wave modes, and the sign change of a sine mode under reversal of a complex wavenumber. *Price:* **free**; exact.

362. Double-angle formula for the complex cosine. For every complex number z, cos(2z) = 2 cos²z − 1, where for every complex number w the complex cosine and sine are cos w = (exp(i w) + exp(−i w))/2 and sin w = (exp(−i w) − exp(i w)) · i/2. *Used for:* Second harmonics and intensities of waves with complex frequency: cos² of a damped oscillation written through the cosine at twice the argument. *Price:* **free**; exact.

363. Double-angle formula for the complex sine. For every complex number z, sin(2z) = 2 sin z · cos z, where for every complex number w the complex cosine and sine are cos w = (exp(i w) + exp(−i w))/2 and sin w = (exp(−i w) − exp(i w)) · i/2. *Used for:* Second harmonics of waves with complex frequency, and products of sine and cosine modes rewritten at twice the argument. *Price:* **free**; exact.

364. Hyperbolic cosine addition formula, complex arguments. For all complex numbers x and y, cosh(x + y) = cosh x · cosh y + sinh x · sinh y, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Composing Lorentz boosts (rapidities add) and transfer matrices of evanescent or tunnelling layers, with complex arguments. *Price:* **free**; exact.

365. Hyperbolic sine addition formula, complex arguments. For all complex numbers x and y, sinh(x + y) = sinh x · cosh y + cosh x · sinh y, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Composing Lorentz boosts (rapidities add) and transfer matrices of evanescent or tunnelling layers, with complex arguments. *Price:* **free**; exact.

366. Square of the complex hyperbolic cosine. For every complex number z, cosh²z = sinh²z + 1, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* The relation $\gamma^2 = (\gamma\beta)^2 + 1$ between the Lorentz factor and the momentum factor of a boost, and the energy bookkeeping of evanescent waves, here for complex arguments. *Price:* **free**; exact.

367. Hyperbolic Pythagorean identity, complex arguments. For every complex number z, cosh²z − sinh²z = 1, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* The invariant hyperbola of special relativity ($\gamma^2 - (\gamma\beta)^2 = 1$) and the parametrization of hyperbolic orbits, here for complex arguments. *Price:* **free**; exact.

368. Complex hyperbolic cosine is even. For every complex number z, cosh(−z) = cosh z, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Symmetric evanescent profiles, such as the even bound state of a double well or the field inside a symmetric barrier. *Price:* **free**; exact.

369. Complex hyperbolic sine is odd. For every complex number z, sinh(−z) = −sinh z, where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Antisymmetric evanescent profiles, such as the odd bound state of a double well. *Price:* **free**; exact.

370. Hyperbolic cosine plus sine is the exponential. For every complex number z, cosh z + sinh z = exp(z), where for every complex number w the complex hyperbolic cosine and sine are cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Splitting a growing or decaying exponential into its even and odd parts, as in barrier-penetration and transfer-matrix calculations. *Price:* **free**; exact.

371. Cosine of an imaginary argument is hyperbolic. For every complex number z, cos(z i) = cosh z, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Continuing a standing wave cos(kx) to an imaginary wavenumber k = iκ gives the evanescent profile cosh(κx) of tunnelling and of the field beyond total internal reflection. *Price:* **free**; exact.

372. Sine of an imaginary argument is hyperbolic. For every complex number z, sin(z i) = sinh(z) · i, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Continuing a standing wave sin(kx) to an imaginary wavenumber k = iκ gives the evanescent profile sinh(κx) of tunnelling through a barrier. *Price:* **free**; exact.

373. Hyperbolic cosine of an imaginary argument is the cosine. For every complex number z, cosh(z i) = cos z, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Turning a

hyperbolic (evanescent or boost) formula into an oscillating one by an imaginary argument, as in the Wick rotation between real and imaginary time. *Price:* **free**; exact.

374. Hyperbolic sine of an imaginary argument is the sine. For every complex number z, sinh(z i) = sin(z) · i, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Turning a hyperbolic (evanescent or boost) formula into an oscillating one by an imaginary argument, as in the Wick rotation between real and imaginary time. *Price:* **free**; exact.

375. Euler's formula, cosine plus i sine form. For every complex number z, cos z + i sin z = exp(i z), where for every complex number w the complex cosine and sine are cos w = (exp(i w) + exp(−i w))/2 and sin w = (exp(−i w) − exp(i w)) · i/2. *Used for:* Writing a combination of cosine and sine as one complex exponential: plane waves, phasors and rotating frames. *Price:* **free**; exact.

376. Complex hyperbolic cosine extends the real one. For every real number x, the complex hyperbolic cosine (exp(x) + exp(−x))/2, computed in the complex numbers, equals the real hyperbolic cosine cosh x. *Used for:* Continuing real evanescent and boost formulas to complex arguments without changing their values on the real axis. *Price:* **free**; exact.

377. Complex hyperbolic sine extends the real one. For every real number x, the complex hyperbolic sine (exp(x) − exp(−x))/2, computed in the complex numbers, equals the real hyperbolic sine sinh x. *Used for:* Continuing real evanescent and boost formulas to complex arguments without changing their values on the real axis. *Price:* **free**; exact.

378. Sine of x plus i y. For all complex numbers x and y, sin(x + y i) = sin x · cosh y + cos x · sinh y · i, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Splits a sine mode with a complex wavenumber or frequency into an oscillating factor and a growing or decaying factor, as for lossy waveguides, leaky modes and damped resonances. *Price:* **free**; exact.

379. Cosine of x plus i y. For all complex numbers x and y, cos(x + y i) = cos x · cosh y − sin x · sinh y · i, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Splits a cosine mode with a complex wavenumber or frequency into an oscillating factor and a growing or decaying factor, as for lossy waveguides, leaky modes and damped resonances. *Price:* **free**; exact.

380. Real and imaginary parts of the complex sine. For every complex number z with real part x and imaginary part y, sin z = sin x · cosh y + cos x · sinh y · i, with x and y taken as complex numbers on the real axis, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Gives the oscillating and the evanescent part of a sine mode at a complex frequency, the starting point for showing where the complex sine can vanish. *Price:* **free**; exact.

381. Real and imaginary parts of the complex cosine. For every complex number z with real part x and imaginary part y, cos z = cos x · cosh y − sin x · sinh y · i, with x and y taken as complex numbers on the real axis, where for every complex number w, cos w = (exp(i w) + exp(−i w))/2, sin w = (exp(−i w) − exp(i w)) · i/2, cosh w = (exp(w) + exp(−w))/2 and sinh w = (exp(w) − exp(−w))/2. *Used for:* Gives the oscillating and the evanescent part of a cosine mode at a complex frequency, the starting point for showing where the complex cosine can vanish. *Price:* **free**; exact.

382. Zeros of the complex sine are real. For every complex number θ, sin θ = 0 exactly when θ = kπ for some integer k, where sin θ = (exp(−iθ) − exp(iθ)) · i/2; so the complex sine vanishes nowhere off the real axis. *Used for:* A string or cavity of length L fixed at both ends: when the standing-wave condition is that the sine of the wavenumber times L vanishes, with a complex wavenumber, every solution is real, the wavenumber times L being nπ, so there are no extra evanescent solutions. *Price:* **free**; exact.

383. Zeros of the complex cosine are real. For every complex number θ, cos θ = 0 exactly when θ = (2k + 1)π/2 for some integer k, where cos θ = (exp(iθ) + exp(−iθ))/2; so the complex cosine vanishes nowhere off the real axis. *Used for:* A pipe or string of length L with one fixed and one free end: when the standing-wave condition is that the cosine of the wavenumber times L vanishes, with a complex wavenumber, every solution is real, the wavenumber times L being (2n + 1)π/2, so there are no extra evanescent solutions. *Price:* **free**; exact.

384. Sine shifted by a half turn. For every real x, sin(x + π) = −sin x. *Used for:* The sign flip of a wave delayed by half a period: reflection of a wave at a fixed end, antisymmetric normal modes, phase inversion of an alternating signal. *Price:* **free**; exact.

385. Sine shifted back by a half turn. For every real x, sin(x − π) = −sin x. *Used for:* The sign flip of a wave advanced by half a period: phase inversion of an alternating signal, the opposite displacement of a pendulum half a cycle earlier. *Price:* **free**; exact.

386. Sine of the supplementary angle. For every real x, $\sin(\pi - x) = \sin x$. *Used for:* Equal sines of supplementary angles: the two launch angles with equal range in projectile motion, the ambiguity of an angle read from its sine in diffraction and refraction. *Price:* **free**; exact.

387. Sine repeats after a full turn. For every real x, $\sin(x + 2\pi) = \sin x$. *Used for:* The periodicity of oscillations and rotating phasors: one full cycle returns a wave, a pendulum or a rotating vector to its starting value. *Price:* **free**; exact.

388. Sine of a full turn minus an angle. For every real x, $\sin(2\pi - x) = -\sin x$. *Used for:* Reflection of a phase angle about a full turn: the sine of a clockwise angle measured from a full cycle, used in the symmetry of Fourier coefficients and rotating frames. *Price:* **free**; exact.

389. Cosine shifted by a half turn. For every real x, $\cos(x + \pi) = -\cos x$. *Used for:* The sign flip of a wave delayed by half a period: destructive interference of waves half a wavelength apart, phase inversion of an alternating signal. *Price:* **free**; exact.

390. Cosine shifted back by a half turn. For every real x, $\cos(x - \pi) = -\cos x$. *Used for:* The sign flip of a wave advanced by half a period: destructive interference of waves half a wavelength apart, the opposite position of an oscillator half a cycle earlier. *Price:* **free**; exact.

391. Cosine of the supplementary angle. For every real x, $\cos(\pi - x) = -\cos x$. *Used for:* Opposite projections of supplementary angles: the reversed horizontal component of a reflected velocity, the law of cosines for an exterior angle. *Price:* **free**; exact.

392. Cosine repeats after a full turn. For every real x, $\cos(x + 2\pi) = \cos x$. *Used for:* The periodicity of oscillations and rotating phasors: one full cycle returns a wave, a pendulum or a rotating vector to its starting projection. *Price:* **free**; exact.

393. Cosine of a full turn minus an angle. For every real x, $\cos(2\pi - x) = \cos x$. *Used for:* A clockwise rotation by x and a counterclockwise rotation by $2\pi -$ x end at one point and have equal cosines, used in Fourier series and rotating frames. *Price:* **free**; exact.

394. Sine shifted by a quarter turn is the cosine. For every real x, $\sin(x + \pi/2) = \cos x$. *Used for:* The quarter-period lead of the cosine over the sine: the 90 degree phase between position and velocity in simple harmonic motion, and between current and voltage in a capacitor or inductor. *Price:* **free**; exact.

395. Sine shifted back by a quarter turn. For every real x, $\sin(x - \pi/2) = -\cos x$. *Used for:* The quarter-period lag of a driven oscillator at resonance behind its drive, and the 90 degree phase of the current in an inductor behind the voltage. *Price:* **free**; exact.

396. Sine of the complementary angle. For every real x, $\sin(\pi/2 - x) = \cos x$. *Used for:* Exchanging sine and cosine for complementary angles: the components of a vector measured from the other axis, the angle of incidence measured from the surface instead of the normal. *Price:* **free**; exact.

397. Cosine shifted by a quarter turn. For every real x, $\cos(x + \pi/2) = -\sin x$. *Used for:* The quarter-period lead of the velocity of a harmonic oscillator, and the 90 degree phase of the current in a capacitor ahead of the voltage. *Price:* **free**; exact.

398. Cosine shifted back by a quarter turn. For every real x, $\cos(x - \pi/2) = \sin x$. *Used for:* Turning a cosine wave into a sine wave by a quarter-period delay: quadrature signals, the in-phase and quadrature components of a modulated wave. *Price:* **free**; exact.

399. Cosine of the complementary angle. For every real x, $\cos(\pi/2 - x) = \sin x$. *Used for:* Exchanging cosine and sine for complementary angles: the components of a vector measured from the other axis, the grazing angle and the angle of incidence in optics. *Price:* **free**; exact.

400. Sine vanishes at natural-number multiples of π. For every natural number n, $\sin(n\pi) = 0$. *Used for:* The nodes of a string fixed at both ends and of a particle in a box: the wavelengths 2L/n satisfy the standing-wave condition. *Price:* **free**; exact.

401. Cosine is one after whole turns. For every natural number n, $\cos(n \cdot 2\pi) = 1$. *Used for:* The return of a rotating phasor to its start after whole cycles: a whole number of cycles satisfies the condition for constructive interference and the periodic boundary condition of a ring. *Price:* **free**; exact.

402. Cosine of 45 degrees. $\cos(\pi/4) = \sqrt{2}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The projection of a unit spin or polarisation vector at 45 degrees: Malus's law at 45 degrees passes half the intensity, and light polarised at 45 degrees splits into equal amplitudes on a polarising beam splitter. *Price:* **free**; exact.

403. Sine of 45 degrees. $\sin(\pi/4) = \sqrt{2}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The equal sine and cosine components of a vector at 45 degrees: equal superpositions of two quantum states, and the equal horizontal and vertical velocity components of a projectile launched at 45 degrees. *Price:* **free**; exact.

404. Cosine of 60 degrees. $\cos(\pi/3) = 1/2$. *Used for:* The 60 degree geometry of hexagonal and triangular lattices (graphene, close-packed crystals), and the projection of a spin or a polarisation vector at 60 degrees. *Price:* **free**; exact.

405. Sine of 30 degrees. $\sin(\pi/6) = 1/2$. *Used for:* The 30 degree geometry of hexagonal lattices and inclined planes: the component of gravity along a 30 degree slope is half the weight. *Price:* **free**; exact.

406. Cosine of 30 degrees. $\cos(\pi/6) = \sqrt{3}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The 30 degree geometry of hexagonal and triangular lattices (the nearest-neighbour vectors of graphene), and the component of a force at 30 degrees. *Price:* **free**; exact.

407. Sine of 60 degrees. $\sin(\pi/3) = \sqrt{3}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The height of an equilateral triangle: lattice spacings of hexagonal and triangular lattices and the geometry of benzene rings. *Price:* **free**; exact.

408. Tangent of 45 degrees. The cosine of $\pi/4$ is apart from 0 (known to be positive or known to be negative), so the tangent $\tan(\pi/4) = \sin(\pi/4)/\cos(\pi/4)$ is defined, and $\tan(\pi/4) = 1$. *Used for:* The slope of a 45 degree incline or ray, and the 45 degree phase shift of a first-order filter at its cutoff frequency. *Price:* **free**; exact.

409. Tangent of 30 degrees. The cosine of $\pi/6$ is apart from 0 (known to be positive or known to be negative), so the tangent $\tan(\pi/6) = \sin(\pi/6)/\cos(\pi/6)$ is defined, and $\tan(\pi/6) \cdot \sqrt{3} = 1$, that is $\tan(\pi/6) = 1/\sqrt{3}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The slope of a 30 degree incline or ray: the friction coefficient at which a block starts to slide on a 30 degree slope, and the angles of hexagonal lattices. *Price:* **free**; exact.

410. Tangent of 60 degrees. The cosine of $\pi/3$ is apart from 0 (known to be positive or known to be negative), so the tangent $\tan(\pi/3) = \sin(\pi/3)/\cos(\pi/3)$ is defined, and $\tan(\pi/3) = \sqrt{3}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The slope of a 60 degree incline or ray, and the Brewster angle of 60 degrees for a medium of refractive index $\sqrt{3}$ (the tangent of the Brewster angle is the index). *Price:* **free**; exact.

411. Cosine of 22.5 degrees. $\cos(\pi/8) = \sqrt{(2 + \sqrt{2})}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The 22.5 degree measurement angles of the CHSH Bell test, where the quantum winning probability is $\cos^2(\pi/8) = (2 + \sqrt{2})/4$, and spin-½ and polarisation projections at 22.5 degrees. *Price:* **free**; exact.

412. Sine of 22.5 degrees. $\sin(\pi/8) = \sqrt{(2 - \sqrt{2})}/2$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The 22.5 degree measurement angles of the CHSH Bell test and of quantum state discrimination, and polarisation projections at 22.5 degrees. *Price:* **free**; exact.

413. Half-angle formula for the cosine. For every real x with $-\pi \le x \le \pi$, $\cos(x/2) = \sqrt{((1 + \cos x)/2)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The amplitude of a spin-½ state rotated by an angle x, whose square $(1 + \cos x)/2$ is the probability of keeping the spin direction; Rabi oscillations and the half-angle of Bloch-sphere rotations. *Price:* **free**; exact.

414. Half-angle formula for the size of the sine. For every real x, $|\sin(x/2)| = \sqrt{((1 - \cos x)/2)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The amplitude of a spin-½ flip after a rotation by x, whose square $(1 - \cos x)/2$ is the flip probability; the chord length $2|\sin(x/2)|$ between two points of a unit circle at angle x. *Price:* **free**; exact.

415. Half-angle formula for the sine on a turn. For every real x with $0 \le x \le 2\pi$, $\sin(x/2) = \sqrt{((1 - \cos x)/2)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The spin-flip amplitude after a rotation by an angle between 0 and a full turn, and the chord of a circle subtending that angle. *Price:* **free**; exact.

416. Half-angle formula for the sine on a reversed turn. For every real x with $-2\pi \le x \le 0$, $\sin(x/2) = -\sqrt{((1 - \cos x)/2)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* The signed spin-flip amplitude after a rotation in the negative sense by up to a full turn. *Price:* **free**; exact.

417. Cosine from the sine on a half turn. For every real x with $-\pi/2 \le x \le \pi/2$, $\cos x = \sqrt{(1 - \sin^2 x)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* Recovering a cosine from a measured sine: the refraction angle in Snell's law, the transmitted component at an interface, the direction cosines of a vector. *Price:* **free**; exact.

418. Sine from the cosine on a half turn. For every real x with $0 \le x \le \pi$, $\sin x = \sqrt{(1 - \cos^2 x)}$, where $\sqrt{}$ is the nonnegative square root. *Used for:* Recovering a sine from a measured cosine: the transverse component of a vector from its projection, the scattering-angle factor $\sqrt{(1 - \cos^2\theta)}$ in cross sections and dipole radiation. *Price:* **free**; exact.

419. Hyperbolic sine of zero. $\sinh 0 = 0$. *Used for:* Zero rapidity carries zero momentum; the slope of a hanging chain vanishes at its lowest point. *Price:* **free**; exact.

420. Hyperbolic cosine of zero. $\cosh 0 = 1$. *Used for:* A particle at rest has Lorentz factor $\gamma = \cosh 0 = 1$. *Price:* **free**; exact.

421. Hyperbolic sine is odd. For every real x, $\sinh(-x) = -\sinh x$. *Used for:* Reversing a boost reverses the momentum it produces. *Price:* **free**; exact.

422. Hyperbolic cosine is even. For every real x, $\cosh(-x) = \cosh x$. *Used for:* Opposite rapidities give equal energies γmc^2; a hanging chain is symmetric about its lowest point. *Price:* **free**; exact.

423. Subtraction formula for the hyperbolic sine. For all real x and y, $\sinh(x - y) = \sinh x \cosh y - \cosh$

x sinh y. *Used for:* Relative rapidity of two frames along one line, and the momentum seen from a moving frame. *Price:* **free**; exact.

424. Subtraction formula for the hyperbolic cosine. For all real x and y, $\cosh(x - y) = \cosh x \cosh y - \sinh x \sinh y$. *Used for:* The relative Lorentz factor of two frames, $\gamma = \gamma_1\gamma_2(1 - \beta_1\beta_2)$. *Price:* **free**; exact.

425. Hyperbolic cosine plus sine is the exponential, real line. For every real x, $\cosh x + \sinh x = e^x$. *Used for:* The Doppler factor of a boost with rapidity η is $\cosh \eta + \sinh \eta = e^\eta$. *Price:* **free**; exact.

426. Hyperbolic sine plus cosine is the exponential. For every real x, $\sinh x + \cosh x = e^x$. *Used for:* The Doppler factor of a boost with rapidity η is $\sinh \eta + \cosh \eta = e^\eta$. *Price:* **free**; exact.

427. Exponential minus hyperbolic cosine. For every real x, $e^x - \cosh x = \sinh x$. *Used for:* Splitting a growing exponential into its odd part; momentum from the Doppler factor and the energy. *Price:* **free**; exact.

428. Exponential minus hyperbolic sine. For every real x, $e^x - \sinh x = \cosh x$. *Used for:* Splitting a growing exponential into its even part; energy from the Doppler factor and the momentum. *Price:* **free**; exact.

429. Hyperbolic cosine minus sine. For every real x, $\cosh x - \sinh x = e^{-x}$. *Used for:* The inverse Doppler factor $e^{-\eta}$ of a boost; the decaying part of a hyperbolic solution. *Price:* **free**; exact.

430. Hyperbolic sine minus cosine. For every real x, $\sinh x - \cosh x = -e^{-x}$. *Used for:* The inverse Doppler factor of a boost with its sign reversed; the decaying part of a hyperbolic solution. *Price:* **free**; exact.

431. Hyperbolic Pythagorean identity, difference form. For every real x, $\cosh^2 x - \sinh^2 x = 1$. *Used for:* The mass shell $E^2 - p^2c^2 = m^2c^4$ in rapidity form, and the invariance of the interval under boosts. *Price:* **free**; exact.

432. Square of the hyperbolic sine. For every real x, $\sinh^2 x = \cosh^2 x - 1$. *Used for:* Momentum from the Lorentz factor: $(\gamma\beta)^2 = \gamma^2 - 1$. *Price:* **free**; exact.

433. Double rapidity formula for the hyperbolic cosine. For every real x, $\cosh 2x = \cosh^2 x + \sinh^2 x$. *Used for:* Composing two equal boosts; the photon-number factors cosh 2r of squeezed light. *Price:* **free**; exact.

434. Double rapidity formula for the hyperbolic sine. For every real x, $\sinh 2x = 2 \sinh x \cosh x$. *Used for:* Composing two equal boosts; the correlations sinh 2r of a two-mode squeezed state. *Price:* **free**; exact.

435. Triple rapidity formula for the hyperbolic cosine. For every real x, $\cosh 3x = 4 \cosh^3 x - 3 \cosh x$. *Used for:* Solving cubic equations with one real root by a hyperbolic substitution, as for cubic equations of state. *Price:* **free**; exact.

436. Triple rapidity formula for the hyperbolic sine. For every real x, $\sinh 3x = 4 \sinh^3 x + 3 \sinh x$. *Used for:* Solving depressed cubic equations with one real root by the substitution $x = 2\sqrt{(p/3)} \sinh t$. *Price:* **free**; exact.

437. Hyperbolic tangent of zero. $\tanh 0 = 0$, where $\tanh x = \sinh x / \cosh x$. *Used for:* Zero rapidity is zero velocity; zero field gives zero magnetisation of a paramagnet. *Price:* **free**; exact.

438. Hyperbolic tangent is odd. For every real x, $\tanh(-x) = -\tanh x$, where $\tanh x = \sinh x / \cosh x$. *Used for:* Reversing a boost reverses the velocity; reversing the field reverses the magnetisation. *Price:* **free**; exact.

439. Hyperbolic tangent from exponentials. For every real x, $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$, where $\tanh x = \sinh x / \cosh x$. *Used for:* The thermal polarisation $\tanh(E/(k \cdot T))$ of a two-level system or a spin ½ from its two Boltzmann weights, with k Boltzmann's constant and T the temperature. *Price:* **free**; exact.

440. Hyperbolic sine is below hyperbolic cosine. For every real x, $\sinh x < \cosh x$. *Used for:* Momentum stays below energy for a massive particle: $\gamma\beta < \gamma$, that is $pc < E$. *Price:* **free**; exact.

441. Hyperbolic tangent is above minus one. For every real x, $-1 < \tanh x$, where $\tanh x = \sinh x / \cosh x$. *Used for:* Velocities stay above $-c$ for every rapidity; magnetisation stays above negative saturation. *Price:* **free**; exact.

442. Square of the hyperbolic tangent is below one. For every real x, $\tanh^2 x < 1$, where $\tanh x = \sinh x / \cosh x$. *Used for:* $\beta^2 < 1$ for every rapidity, so the Lorentz factor $1/\sqrt{(1 - \beta^2)}$ is always defined. *Price:* **free**; exact.

443. Size of the hyperbolic tangent is below one. For every real x, $|\tanh x| < 1$, where $\tanh x = \sinh x / \cosh x$. *Used for:* Speeds stay below the speed of light for every finite rapidity; mean-field magnetisation stays below saturation. *Price:* **free**; exact.

444. Hyperbolic sine of a logarithm. For every $x > 0$, $\sinh(\ln x) = (x - 1/x)/2$. *Used for:* Momentum from the Doppler factor k of a boost: $\gamma\beta = (k - 1/k)/2$. *Price:* **free**; exact.

445. Hyperbolic cosine of a logarithm. For every $x > 0$, $\cosh(\ln x) = (x + 1/x)/2$. *Used for:* The Lorentz factor from the Doppler factor k, $\gamma = (k + 1/k)/2$; the recognition cost $J(x) = (x + 1/x)/2 - 1$ equals $\cosh(\ln x) - 1$. *Price:* **free**; exact.

446. Exponential of half is a square root. For every real x, $e^{x/2} = \sqrt{(e^x)}$. *Used for:* Amplitudes as square roots of Boltzmann weights or probabilities: $e^{-E/(2k\cdot T)} = \sqrt{(e^{-E/(k\cdot T)})}$, with k Boltzmann's constant and T the temperature. *Price:* **free**; exact.

447. Hyperbolic sine is strictly increasing. For all real $x < y$, $\sinh x < \sinh y$. *Used for:* Momentum increases strictly with rapidity, so a measured momentum fixes the rapidity. *Price:* **free**; exact.

448. Hyperbolic sine is one to one. For all real x and y, $\sinh x = \sinh y$ implies $x = y$. *Used for:* A momentum determines its rapidity uniquely. *Price:* **free**; exact.

449. Equal hyperbolic sines have equal arguments. For all real x and y, $\sinh x = \sinh y$ if and only if $x = y$. *Used for:* Two particles of equal mass have equal momenta exactly when they have equal rapidities. *Price:* **free**; exact.

450. Hyperbolic sine preserves and reflects order. For all real x and y, $\sinh x \le \sinh y$ if and only if $x \le y$. *Used for:* Comparing the momenta of equal-mass particles is comparing their rapidities. *Price:* **free**; exact.

451. Hyperbolic sine preserves and reflects strict order. For all real x and y, $\sinh x < \sinh y$ if and only if $x < y$. *Used for:* A larger momentum means a larger rapidity and conversely, for particles of equal mass. *Price:* **free**; exact.

452. Hyperbolic sine vanishes only at zero. For every real x, $\sinh x = 0$ if and only if $x = 0$. *Used for:* A particle has zero momentum exactly when its rapidity is zero. *Price:* **free**; exact.

453. Hyperbolic sine is nonzero away from zero. For every real x, $\sinh x \neq 0$ if and only if $x \neq 0$. *Used for:* A particle has nonzero momentum exactly when its rapidity is nonzero. *Price:* **free**; exact.

454. Hyperbolic sine is positive exactly for positive arguments. For every real x, $0 < \sinh x$ if and only if $0 < x$. *Used for:* Momentum and rapidity point in one direction: the momentum $m\cdot c\cdot\sinh \eta$ of a particle is positive exactly when its rapidity η is positive. *Price:* **free**; exact.

455. Hyperbolic sine is nonnegative exactly for nonnegative arguments. For every real x, $0 \le \sinh x$ if and only if $0 \le x$. *Used for:* Momentum and rapidity point in one direction, including zero: the momentum $m\cdot c\cdot\sinh \eta$ of a particle is nonnegative exactly when its rapidity η is nonnegative. *Price:* **free**; exact.

456. Hyperbolic sine is negative exactly for negative arguments. For every real x, $\sinh x < 0$ if and only if $x < 0$. *Used for:* Momentum and rapidity point in one direction: the momentum $m\cdot c\cdot\sinh \eta$ of a particle is negative exactly when its rapidity η is negative. *Price:* **free**; exact.

457. Hyperbolic sine is nonpositive exactly for nonpositive arguments. For every real x, $\sinh x \le 0$ if and only if $x \le 0$. *Used for:* Momentum and rapidity point in one direction, including zero: the momentum $m\cdot c\cdot\sinh \eta$ of a particle is nonpositive exactly when its rapidity η is nonpositive. *Price:* **free**; exact.

458. Exponential of the inverse hyperbolic sine. For every real x, $e^{\operatorname{arsinh} x} = x + \sqrt{(1 + x^2)}$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* The Doppler factor from the momentum: $e^{\eta} = p/mc + \sqrt{(1 + (p/mc)^2)}$. *Price:* **free**; exact.

459. Inverse hyperbolic sine of zero. $\operatorname{arsinh} 0 = 0$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Zero momentum gives zero rapidity. *Price:* **free**; exact.

460. Inverse hyperbolic sine is odd. For every real x, $\operatorname{arsinh}(-x) = -\operatorname{arsinh} x$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Reversing the momentum reverses the rapidity. *Price:* **free**; exact.

461. Hyperbolic cosine of the inverse hyperbolic sine. For every real x, $\cosh(\operatorname{arsinh} x) = \sqrt{(1 + x^2)}$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* The Lorentz factor from the momentum: $\gamma = \sqrt{(1 + (p/mc)^2)}$. *Price:* **free**; exact.

462. Hyperbolic tangent of the inverse hyperbolic sine. For every real x, $\tanh(\operatorname{arsinh} x) = x/\sqrt{(1 + x^2)}$, where $\tanh y = \sinh y / \cosh y$ and where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Velocity from momentum: $v/c = (p/mc)/\sqrt{(1 + (p/mc)^2)}$. *Price:* **free**; exact.

463. Inverse hyperbolic sine is one to one. For all real x and y, $\operatorname{arsinh} x = \operatorname{arsinh} y$ implies $x = y$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* A rapidity determines its momentum uniquely. *Price:* **free**; exact.

464. Equal inverse hyperbolic sines have equal arguments. For all real x and y, $\operatorname{arsinh} x = \operatorname{arsinh} y$ if and only if $x = y$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Two equal-mass particles have equal rapidities exactly when they have equal momenta. *Price:* **free**; exact.

465. Inverse hyperbolic sine is strictly increasing. For all real $x < y$, $\operatorname{arsinh} x < \operatorname{arsinh} y$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Rapidity increases strictly with momentum. *Price:* **free**; exact.

466. Inverse hyperbolic sine preserves and reflects order. For all real x and y, $\operatorname{arsinh} x \le \operatorname{arsinh} y$ if and only if $x \le y$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* Comparing rapidities of equal-mass particles is comparing their momenta. *Price:* **free**; exact.

467. Inverse hyperbolic sine preserves and reflects strict order. For all real x and y, $\operatorname{arsinh} x < \operatorname{arsinh} y$ if and only if $x < y$, where $\operatorname{arsinh} x = \ln(x + \sqrt{(1 + x^2)})$. *Used for:* A larger momentum means a

larger rapidity and conversely, for particles of equal mass. *Price:* **free**; exact.

468. Inverse hyperbolic sine vanishes only at zero. For every real x, arsinh x = 0 if and only if x = 0, where arsinh x = ln(x + √(1 + x²)). *Used for:* Zero rapidity exactly at zero momentum. *Price:* **free**; exact.

469. Inverse hyperbolic sine is nonnegative exactly for nonnegative arguments. For every real x, 0 ≤ arsinh x if and only if 0 ≤ x, where arsinh x = ln(x + √(1 + x²)). *Used for:* Rapidity and momentum point in one direction, including zero: the rapidity arsinh(p/(m·c)) is nonnegative exactly when the momentum p is nonnegative. *Price:* **free**; exact.

470. Inverse hyperbolic sine is nonpositive exactly for nonpositive arguments. For every real x, arsinh x ≤ 0 if and only if x ≤ 0, where arsinh x = ln(x + √(1 + x²)). *Used for:* Rapidity and momentum point in one direction, including zero: the rapidity arsinh(p/(m·c)) is nonpositive exactly when the momentum p is nonpositive. *Price:* **free**; exact.

471. Inverse hyperbolic sine is positive exactly for positive arguments. For every real x, 0 < arsinh x if and only if 0 < x, where arsinh x = ln(x + √(1 + x²)). *Used for:* Rapidity and momentum point in one direction: the rapidity arsinh(p/(m·c)) is positive exactly when the momentum p is positive. *Price:* **free**; exact.

472. Inverse hyperbolic sine is negative exactly for negative arguments. For every real x, arsinh x < 0 if and only if x < 0, where arsinh x = ln(x + √(1 + x²)). *Used for:* Rapidity and momentum point in one direction: the rapidity arsinh(p/(m·c)) is negative exactly when the momentum p is negative. *Price:* **free**; exact.

473. Lower decimal bound for one over e. $0.36787944116 < e^{-1}$. *Used for:* A certified lower bound on the fraction 1/e left after one lifetime, one attenuation length or one RC time constant. *Price:* **free**; exact.

474. Upper decimal bound for one over e. $e^{-1} < 0.3678794412$. *Used for:* A certified upper bound on the fraction 1/e left after one lifetime, one attenuation length or one RC time constant. *Price:* **free**; exact.

475. Rational approximation of e to ten digits. $|e - 2244083/825552| \le 1/10^{10}$. *Used for:* An exact rational stand-in for e in numerical checks of decay and growth laws, good to ten digits. *Price:* **free**; exact.

476. Rational approximation of ln 2 to ten digits. $|\ln 2 - 287209/414355| \le 1/10^{10}$. *Used for:* Landauer's bound k·T·ln 2 per erased bit (k Boltzmann's constant, T the temperature) and half-lives t = τ ln 2, with an exact rational stand-in good to ten digits. *Price:* **free**; exact.

477. Cauchy's coefficient formula from equally spaced samples. Let $a_0, a_1, a_2, \ldots$ be complex numbers such that at every point z of the circle |z| = R the series $a_0 + a_1 z + a_2 z^2 + \ldots$ converges (real and imaginary parts unconditionally) to a sum of modulus at most M; let 0 < r < R, and let f be a function equal to the sum of this series at every point of the circle |z| = r. (Here the function is assumed to be given by such a series; that every complex-differentiable function is, which is the classical theorem of Cauchy, is not part of this row.) Then for every n the averages $(1/N) \cdot \Sigma$ over k < N of $\omega^{-kn} \cdot f(r\,\omega^k)$, over the N equally spaced points $r\,\omega^k$ of that circle ($\omega = e^{2\pi i/N}$), converge to $a_n\, r^n$ as N grows. These averages are the Riemann sums of Cauchy's circle integral for the n-th coefficient, so the coefficient formula holds with no integral needed. *Used for:* Computing Taylor coefficients of response functions and generating functions from equally spaced samples on a circle by the discrete Fourier transform. *Price:* **free**; exact.

478. Mean value property on a circle. Let $a_0, a_1, a_2, \ldots$ be complex numbers such that at every point z of the circle |z| = R the series $a_0 + a_1 z + a_2 z^2 + \ldots$ converges (real and imaginary parts unconditionally) to a sum of modulus at most M; let 0 < r < R, and let f be a function equal to the sum of this series at every point of the circle |z| = r. (Here the function is assumed to be given by such a series; that every complex-differentiable function is, which is the classical theorem of Cauchy, is not part of this row.) Let f(0) be the sum of the series at the center. Then the averages $(1/N) \cdot \Sigma$ over k < N of $f(r\,\omega^k)$, over the N equally spaced points $r\,\omega^k$ of the circle |z| = r ($\omega = e^{2\pi i/N}$), converge to f(0) as N grows. *Used for:* A two-dimensional potential that is the real part of a power series bounded on a larger circle (electrostatics, ideal fluid flow, steady heat flow): its value at the centre is the limit of its averages over equally spaced points of the circle. *Price:* **free**; exact.

479. Cauchy's coefficient formula from dyadic circle samples. Let $a_0, a_1, a_2, \ldots$ be complex numbers such that at every point z of the circle |z| = R the series $a_0 + a_1 z + a_2 z^2 + \ldots$ converges (real and imaginary parts unconditionally) to a sum of modulus at most M; let 0 < r < R, and let f be a function equal to the sum of this series at every point of the circle |z| = r. (Here the function is assumed to be given by such a series; that every complex-differentiable function is, which is the classical theorem of Cauchy, is not part of this row.) Then for every n and every tolerance 1/(j + 1) there is m_0 such that for every $m \ge m_0$ the average over the 2^m equally spaced points of the circle |z| = r, $(1/2^m) \cdot \Sigma$ over k < 2^m of $\omega^{-kn} \cdot f(r\,\omega^k)$ with $\omega = e^{2\pi i/2\hat{}m}$, is within 1/(j + 1) of $a_n\, r^n$ (in squared modulus). *Used for:* The radix-2 fast Fourier

transform of samples on a circle, used to compute Fourier and Taylor coefficients of analytic signals and response functions. *Price:* **free**; exact.

480. Shifting a monic polynomial to a new center. For a monic polynomial $p(z) = z^n + b_{n-1}z^{n-1} + \ldots + b_1 z + b_0$ with complex coefficients, $n \geq 1$, and any complex point z, there are complex numbers $c_0, c_1, \ldots, c_n$ with $c_0 = p(z)$ and $c_n = 1$ such that $p(z + h) = c_0 + c_1 h + \ldots + c_n h^n$ for every complex h. *Used for:* Expanding a polynomial potential or a characteristic polynomial about an operating point, the shift step of every polynomial root finder, and the expansion step in the proof that every polynomial has a root. *Price:* **free**; exact.

481. Approximate roots of a monic polynomial. For every monic polynomial $p(z) = z^n + b_{n-1}z^{n-1} + \ldots + b_0$ with complex coefficients and $n \geq 1$, and every k, there is a complex number z with $|p(z)|^2 \leq 1/(k + 1)$. *Used for:* Eigenvalue and zero-finding by iteration: every characteristic polynomial of a finite system has points where it is as small as one wishes, with no step that decides a sign or a zero. *Price:* **free**; exact.

482. Fundamental theorem of algebra for monic polynomials. Every monic polynomial $p(z) = z^n + b_{n-1}z^{n-1} + \ldots + b_1 z + b_0$ with complex coefficients and $n \geq 1$ has a complex root: there is z with $p(z) = 0$. *Used for:* Normal modes of coupled oscillators, energy levels of a finite quantum system, poles of a response function: every characteristic polynomial of a finite system has a root. *Price:* **free**; exact.

483. Polynomial roots when the leading coefficient is apart from zero. Every polynomial $a_0 + a_1 z + \ldots + a_n z^n$ with complex coefficients, degree $n \geq 1$ and leading coefficient of positive size, $|a_n| > 0$, has a complex root. *Used for:* Characteristic polynomials of normal modes, eigenvalues of a finite Hamiltonian and poles of a transfer function or propagator, when the leading coefficient is a known quantity of positive size (a mass, a coupling, a determinant computed to be positive). *Price:* **free**; exact.

484. Fundamental theorem of algebra with nonzero leading coefficient. Every polynomial $a_0 + a_1 z + \ldots + a_n z^n$ with complex coefficients, degree $n \geq 1$ and leading coefficient $a_n \neq 0$ has a complex root. *Used for:* Generalized eigenvalue problems $\det(\lambda B - A) = 0$ and pole-finding when all that is known of the leading coefficient (such as $\det B$) is that it is not zero. *Price:* **MP**; exact.

485. Principal square root of a complex number. For every complex number x there is a complex number s with $\mathrm{Re}\, s = \sqrt{((|x| + \mathrm{Re}\, x)/2)}$, with $\mathrm{Im}\, s = \sqrt{((|x| - \mathrm{Re}\, x)/2)}$ if $0 \leq \mathrm{Im}\, x$, and with $\mathrm{Im}\, s = -\sqrt{((|x| - \mathrm{Re}\, x)/2)}$ if $\mathrm{Im}\, x < 0$. (This s is x to the power 1/2 on the principal branch; a root is given for each x separately, and one function giving all of them is a stronger statement.). *Used for:* Complex momenta $k = \sqrt{(2m(E - V))}$ below a barrier (the principal branch gives the decaying solution), the complex refractive index $\sqrt{\varepsilon}$ of a passive lossy medium, and the plasma dispersion relation $kc = \sqrt{(\omega^2 - \omega_p{}^2)}$: the branch cut along the negative real axis fixes the sign. *Price:* **WLPO**; exact.

486. Derivative of the complex exponential. For every complex number z, the exponential has derivative exp z at z: for every k there is j such that every complex h with $|h| \leq 1/(j + 1)$ satisfies $|\exp(z + h) - \exp z - \exp z \cdot h| \leq |h|/(k + 1)$. *Used for:* Phasors and alternating-current circuits, damped oscillations $e^{(-\gamma + i\omega)\cdot t}$, plane waves $e^{i\cdot(k\cdot x - \omega\cdot t)}$, and the Schrödinger evolution $e^{-i\cdot E\cdot t/\hbar}$. *Price:* **free**; exact.

487. Derivative of the complex cosine. For every complex number z, the cosine has derivative $-\sin z$ at z: for every k there is j such that every complex h with $|h| \leq 1/(j + 1)$ satisfies $|\cos(z + h) - \cos z + \sin z \cdot h| \leq |h|/(k + 1)$. *Used for:* Waves with a complex wavenumber (damped and evanescent waves), AC circuit analysis, and analytic continuation of oscillator responses. *Price:* **free**; exact.

488. Derivative of the complex sine. For every complex number z, the sine has derivative cos z at z: for every k there is j such that every complex h with $|h| \leq 1/(j + 1)$ satisfies $|\sin(z + h) - \sin z - \cos z \cdot h| \leq |h|/(k + 1)$. *Used for:* Waves with a complex wavenumber, AC circuit analysis, and analytic continuation of oscillator responses. *Price:* **free**; exact.

489. Derivative of the complex hyperbolic cosine. For every complex number z, the hyperbolic cosine has derivative sinh z at z: for every k there is j such that every complex h with $|h| \leq 1/(j + 1)$ satisfies $|\cosh(z + h) - \cosh z - \sinh z \cdot h| \leq |h|/(k + 1)$. *Used for:* Hyperbolic functions of complex arguments in transmission lines, heat and diffusion kernels, and partition functions continued to complex temperature. *Price:* **free**; exact.

490. Derivative of the complex hyperbolic sine. For every complex number z, the hyperbolic sine has derivative cosh z at z: for every k there is j such that every complex h with $|h| \leq 1/(j + 1)$ satisfies $|\sinh(z + h) - \sinh z - \cosh z \cdot h| \leq |h|/(k + 1)$. *Used for:* Hyperbolic functions of complex arguments in transmission lines, heat and diffusion kernels, and partition functions continued to complex temperature. *Price:* **free**; exact.

491. Termwise derivative of a complex power series, derivative's sum given. For complex numbers: if $r > 0$ and $f(x + z) = \Sigma\, a_n\, z^n$ (real and imaginary parts converging unconditionally) for every complex z with $|z| < r$, and y is a complex

number with $|y| < r$ at which the series $\Sigma\ (n + 1)\ a_{n+1}\ y^n$ converges to d, then f has complex derivative d at $x + y$: for every k there is j such that every complex h with $|h| \le 1/(j + 1)$ has $|f(x + y + h) - f(x + y) - d \cdot h| \le |h|/(k + 1)$. *Used for:* Differentiating analytic response functions and Green's functions inside their disc of convergence, the complex potential of two-dimensional flow and electrostatics, and perturbation series in a complex coupling. *Price:* **free**; exact.

492. Termwise derivative of a complex power series. For complex numbers: if $r > 0$ and $f(x + z) = \Sigma\ a_n\ z^n$ for every complex z with $|z| < r$, then at every complex y with $|y| < r$ the termwise derivative series $\Sigma\ (n + 1)\ a_{n+1}\ y^n$ converges to a complex number d, and f has complex derivative d at $x + y$. *Used for:* Differentiating a convergent complex expansion anywhere inside its disc of convergence and getting a convergent expansion of the derivative: analytic response functions, complex potentials, perturbation series in a complex coupling. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

493. Cauchy product of absolutely convergent complex series. If the moduli $|a_n|$ and $|b_n|$ of two complex series have convergent sums, then $\Sigma\ a_n$ and $\Sigma\ b_n$ converge, to s and t, and the series of Cauchy coefficients $c_n = a_0b_n + a_1b_{n-1} + \ldots + a_nb_0$ converges (unconditionally) to $s \cdot t$. *Used for:* Multiplying two convergent expansions order by order: products of Taylor series, the exponential law for series, generating functions in statistical mechanics, and perturbation series. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

494. Lower decimal bracket of e. $2.7182818283 < e$. *Used for:* A ten-digit lower bound for e, used to bound the 1/e survival fraction after one lifetime and Boltzmann factors at one thermal energy. *Price:* **free**; exact.

495. Upper decimal bracket of e. $e < 2.7182818286$. *Used for:* A ten-digit upper bound for e, used to bound the 1/e survival fraction after one lifetime and Boltzmann factors at one thermal energy. *Price:* **free**; exact.

496. Lower decimal bracket of the natural logarithm of 2. $0.6931471803 < \ln 2$. *Used for:* A ten-digit lower bound for ln 2, used to bound Landauer's entropy k ln 2 per erased bit and the entropy of a two-state system. *Price:* **free**; exact.

497. Upper decimal bracket of the natural logarithm of 2. $\ln 2 < 0.6931471808$. *Used for:* A ten-digit upper bound for ln 2, used to bound Landauer's entropy k ln 2 per erased bit and the entropy of a two-state system. *Price:* **free**; exact.

498. Hyperbolic cosine exceeds one at every nonzero number. For every real x, $1 < \cosh x$ if and only if $x \neq 0$, where $x \neq 0$ says only that $x = 0$ is false. *Used for:* A moving particle has Lorentz factor $\gamma = \cosh \eta$ strictly above one; the recognition cost $\cosh(\ln x) - 1$ is strictly positive at every ratio $x \neq 1$. *Price:* **MP**; exact.

499. Hyperbolic cosine exceeds one for numbers apart from zero. For every real x, $1 < \cosh x$ if and only if x is apart from 0 ($x < 0$ or $x > 0$ is known). *Used for:* The Lorentz factor of a particle whose rapidity is known to be positive or known to be negative is strictly above one; the recognition cost $\cosh(\ln x) - 1$ is strictly positive at every ratio x known to be above one or known to be below one. *Price:* **free**; exact.

500. Hyperbolic cosine increases on the nonnegative half-line. For all real x and y with $0 \le x$, $0 \le y$ and $x < y$, $\cosh x < \cosh y$. *Used for:* The Lorentz factor $\gamma = \cosh \eta$ grows strictly with the rapidity $\eta \ge 0$, so the energy of a particle grows strictly with its speed. *Price:* **free**; exact.

501. Hyperbolic cosine compares the sizes of its arguments. For all real x and y, $\cosh x \le \cosh y$ if and only if $|x| \le |y|$. *Used for:* Comparing the Lorentz factors of two particles is comparing the sizes of their rapidities, whatever their directions of motion. *Price:* **free**; exact.

502. Strict comparison of hyperbolic cosines by size. For all real x and y, $\cosh x < \cosh y$ if and only if $|x| < |y|$. *Used for:* A larger Lorentz factor means a larger rapidity in size and conversely, for particles moving in either direction. *Price:* **free**; exact.

503. Hyperbolic sine versus its argument: at least when nonnegative. For every real x, $x \le \sinh x$ if and only if $0 \le x$. *Used for:* For a forward boost the momentum in units of mc, $\gamma\beta = \sinh \eta$, is at least the rapidity η. *Price:* **free**; exact.

504. Hyperbolic sine versus its argument: at most when nonpositive. For every real x, $\sinh x \le x$ if and only if $x \le 0$. *Used for:* For a backward boost the momentum in units of mc, $\gamma\beta = \sinh \eta$, is at most the rapidity η. *Price:* **free**; exact.

505. Hyperbolic sine exceeds its argument exactly for positive arguments. For every real x, $x < \sinh x$ if and only if $0 < x$. *Used for:* Relativistic momentum in units of mc, $\gamma\beta = \sinh \eta$ exceeds the rapidity η for every forward boost, so the small-rapidity approximation $\sinh \eta \approx \eta$ always undershoots. *Price:* **free**; exact.

506. Hyperbolic sine is below its argument exactly for negative arguments. For every real x, $\sinh x < x$ if and only if $x < 0$. *Used for:* Relativistic momentum in units of mc, $\gamma\beta = \sinh \eta$ lies below the

rapidity η for every backward boost. *Price:* **free**; exact.

507. Cosine is nonnegative from $-\pi/2$ to $\pi/2$. For every real x with $-\pi/2 \le x \le \pi/2$, $0 \le \cos x$. *Used for:* The in-phase component of a wave is nonnegative within a quarter period of a crest; the projection of a spin or a polarisation within 90° of an axis is nonnegative. *Price:* **free**; exact.

508. Cosine is negative between quarter and three-quarter turns. For every real x with $\pi/2 < x < \pi + \pi/2$, $\cos x < 0$. *Used for:* A cosine wave is below its rest position strictly between a quarter and three quarters of a period; the projection onto an axis more than 90° away is negative. *Price:* **free**; exact.

509. Cosine is nonpositive between quarter and three-quarter turns. For every real x with $\pi/2 \le x \le \pi + \pi/2$, $\cos x \le 0$. *Used for:* A wave's displacement is on the trough side from a quarter to three quarters of a period; projections at 90° or more from an axis are not positive. *Price:* **free**; exact.

510. Sine is nonpositive on the negative half turn. For every real x with $-\pi \le x \le 0$, $\sin x \le 0$. *Used for:* An oscillator that crosses zero going upward at time zero was on the negative side for the half period before; the sign of a wave's displacement over half a cycle. *Price:* **free**; exact.

511. Sine is nonnegative on the interval from 0 to π. For every real x with $0 \le x \le \pi$, $0 \le \sin x$. *Used for:* The fundamental mode $\sin(\pi x/L)$ of a string fixed at both ends is nonnegative along the string; the positive half of an alternating signal. *Price:* **free**; exact.

512. Sine is positive inside the first half turn. For every real x with $0 < x < \pi$, $0 < \sin x$. *Used for:* The ground state $\sin(\pi x/L)$ of a particle in a box and the fundamental mode of a string have no node inside the interval. *Price:* **free**; exact.

513. Sine preserves order within a quarter turn of zero. For all real x and y with $-\pi/2 \le x \le y \le \pi/2$, $\sin x \le \sin y$. *Used for:* Comparing angles through their sines within a quarter turn either side of zero, as in Snell's law $n_1 \sin\theta_1 = n_2 \sin\theta_2$ and Bragg's law. *Price:* **free**; exact.

514. Cosine reverses order on the first half turn. For all real x and y with $0 \le x \le y \le \pi$, $\cos y \le \cos x$. *Used for:* A wider scattering angle gives a smaller cosine: the Compton factor $1 - \cos\theta$ grows with the angle up to a half turn, and a polariser's transmitted amplitude $\cos\theta$ falls as it turns from 0 to a half turn. *Price:* **free**; exact.

515. Cosine strictly decreases on the first quarter turn. For all real x and y with $0 \le x < y \le \pi/2$, $\cos y < \cos x$. *Used for:* Tilting a polariser or a spin axis further from alignment strictly lowers the projected amplitude within a quarter turn. *Price:* **free**; exact.

516. Sine is strictly increasing within a quarter turn of zero. For all real x and y with $-\pi/2 \le x \le \pi/2$, $-\pi/2 \le y \le \pi/2$ and $x < y$, $\sin x < \sin y$. *Used for:* An angle within a quarter turn either side of zero is fixed by its sine, as when Snell's law is solved for the refraction angle. *Price:* **free**; exact.

517. Cosine is strictly decreasing on the first half turn. For all real x and y with $0 \le x \le \pi$, $0 \le y \le \pi$ and $x < y$, $\cos y < \cos x$. *Used for:* A scattering angle between 0 and π is fixed by its cosine, as in Compton scattering and the Rutherford cross-section. *Price:* **free**; exact.

518. Sine is injective within a quarter turn of zero. For all real x and y with $-\pi/2 \le x \le \pi/2$ and $-\pi/2 \le y \le \pi/2$, $\sin x = \sin y$ implies $x = y$. *Used for:* A measured sine fixes an angle within a quarter turn either side of zero, as for refraction and diffraction angles. *Price:* **free**; exact.

519. Cosine is one to one on the first half turn. For all real x and y with $0 \le x \le \pi$ and $0 \le y \le \pi$, $\cos x = \cos y$ implies $x = y$. *Used for:* A measured cosine fixes a scattering or bond angle between 0 and π. *Price:* **free**; exact.

520. Tangent is positive on the open first quarter turn. For every real x with $0 < x < \pi/2$, $0 < \tan x$, where $\tan x = \sin x / \cos x$ (the cosine is positive there). *Used for:* A positive inclination has a positive slope: inclined planes, the friction angle with $\tan\theta = \mu$, a positive Hall angle. *Price:* **free**; exact.

521. Tangent is nonnegative on the first quarter turn where defined. For every real x with $0 \le x \le \pi/2$, $0 \le \tan x$ wherever the tangent $\tan x = \sin x / \cos x$ is defined, that is wherever $\cos x$ is known to be apart from 0. *Used for:* Slopes and friction coefficients read from angles between zero and a right angle are never negative. *Price:* **free**; exact.

522. Tangent is negative on the open negative quarter turn. For every real x with $-\pi/2 < x < 0$, $\tan x < 0$, where $\tan x = \sin x / \cos x$ (the cosine is positive there). *Used for:* A downward inclination has a negative slope; a negative Hall angle or phase angle has a negative tangent. *Price:* **free**; exact.

523. Tangent is nonpositive on the negative quarter turn where defined. For every real x with $-\pi/2 \le x \le 0$, $\tan x \le 0$ wherever the tangent $\tan x = \sin x / \cos x$ is defined, that is wherever $\cos x$ is known to be apart from 0. *Used for:* Slopes read from downward angles down to a right angle are never positive. *Price:* **free**; exact.

524. Tangent is strictly increasing within a quarter turn of zero. For all real x and y with $-\pi/2 < x < y < \pi/2$, $\tan x < \tan y$, where $\tan x = \sin x / \cos x$

(the cosine is positive there). *Used for:* A steeper angle gives a steeper slope; phase lags of a driven oscillator below resonance and Hall angles are ordered by their tangents. *Price:* **free**; exact.

525. Tangent is strictly increasing on the first quarter turn. For all real x and y with $0 \le x < y < \pi/2$, tan x < tan y, where tan x = sin x / cos x (the cosine is positive there). *Used for:* A steeper inclination between level and vertical gives a steeper slope, as for inclined planes and friction angles. *Price:* **free**; exact.

526. Tangent is strictly increasing on the open central half turn. For all real x and y strictly between $-\pi/2$ and $\pi/2$ with x < y, tan x < tan y, where tan x = sin x / cos x (the cosine is positive there). *Used for:* An angle strictly within a quarter turn either side of zero is fixed by its slope: Brewster's angle, the phase of a driven oscillator below resonance, the Hall angle. *Price:* **free**; exact.

527. Tangent is injective on the open central half turn. For all real x and y strictly between $-\pi/2$ and $\pi/2$, tan x = tan y implies x = y, where tan x = sin x / cos x (the cosine is positive there). *Used for:* A measured slope fixes the angle strictly within a quarter turn either side of zero: Brewster's angle from $\tan\theta = n_2/n_1$, the phase lag of a driven oscillator below resonance. *Price:* **free**; exact.

528. Equal tangents give equal angles in the central half turn. For all real x and y with $-\pi/2 < x < \pi/2$ and $-\pi/2 < y < \pi/2$, tan x = tan y implies x = y, where tan x = sin x / cos x (the cosine is positive there). *Used for:* Two angles strictly within a quarter turn either side of zero with equal slopes are equal, as in reading a Hall angle or a phase lag from a measured ratio. *Price:* **free**; exact.

529. Every complex number has n-th roots. For every complex number x and every whole number $n \ge 1$ there is a complex number z with $z^n = x$. *Used for:* n-th roots of complex impedances, propagation constants and transfer functions, and the roots behind fractional powers of complex quantities when no particular branch is required. *Price:* **free**; exact.

530. Every complex number is a square. For every complex number x there is a complex number z with $x = z \cdot z$. *Used for:* Square roots of discriminants when solving quadratic dispersion relations and characteristic equations, where either root may be used because the other is its negative. *Price:* **free**; exact.

531. Eigenvalues of a 2 × 2 complex matrix. For every 2 × 2 matrix with complex entries a, b (first row) and c, d (second row) there are complex numbers λ_1 and λ_2 with $(w - a)(w - d) - b\cdot c = (w - \lambda_1)(w - \lambda_2)$ for every complex number w, so λ_1 and λ_2 are the roots of the characteristic polynomial counted with multiplicity; their sum is the trace, $\lambda_1 + \lambda_2 = a + d$, and their product is the determinant, $\lambda_1 \cdot \lambda_2 = a\cdot d - b\cdot c$. *Used for:* Two-level quantum systems (the two energy levels of a 2 × 2 Hamiltonian, with sum equal to the trace and product equal to the determinant), two coupled oscillators, and the trace and determinant of a two-dimensional linear system as the sum and product of its eigenvalues. *Price:* **free**; exact.

532. Monic polynomials split into linear factors. For every monic polynomial $p(w) = w^n + b_{n-1}w^{n-1} + \dots + b_1 w + b_0$ with complex coefficients ($n \ge 0$) there are complex numbers $z_1, \dots, z_n$ such that $p(w) = (w - z_1)(w - z_2)\cdots(w - z_n)$ for every complex number w (for n = 0 the product is 1). *Used for:* Normal modes and eigenvalues of a finite system: the characteristic polynomial is a product of n linear factors, so it has n roots counted with multiplicity; the pole-zero factorization of rational transfer functions and propagators. *Price:* **free**; exact.

533. Polynomial factorization when the leading coefficient is apart from zero. For every polynomial $p(w) = a_0 + a_1 w + \dots + a_n w^n$ with complex coefficients whose leading coefficient has positive size, $|a_n| > 0$, there are complex numbers $z_1, \dots, z_n$ such that $p(w) = a_n(w - z_1)(w - z_2)\cdots(w - z_n)$ for every complex number w. *Used for:* Factoring a characteristic or dispersion polynomial into its roots when the leading coefficient is a known quantity of positive size (a mass, a coupling, a determinant computed to be positive): eigenvalues with multiplicity, poles and zeros of a response function. *Price:* **free**; exact.

534. Polynomial factorization with nonzero leading coefficient. For every polynomial $p(w) = a_0 + a_1 w + \dots + a_n w^n$ with complex coefficients and leading coefficient $a_n \ne 0$ there are complex numbers $z_1, \dots, z_n$ such that $p(w) = a_n(w - z_1)(w - z_2)\cdots(w - z_n)$ for every complex number w. *Used for:* Factoring a generalized characteristic polynomial $\det(\lambda B - A)$ or the denominator of a transfer function into its roots when all that is known of the leading coefficient is that it is not zero. *Price:* **MP**; exact.

535. Cauchy product of absolutely convergent complex series, sums given. If $\Sigma\, a_n$ and $\Sigma\, b_n$ are complex series with sums s and t whose moduli $|a_n|$ and $|b_n|$ have convergent sums, then the series of Cauchy coefficients $c_n = a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0$ converges (unconditionally) to $s \cdot t$. *Used for:* Multiplying two convergent expansions order by order: products of Taylor series, the exponential law for series, generating functions in statistical mechanics, and perturbation series. *Price:* **free**; exact.

Q.4 Integrals and measure

536. Constant multiple of an integral. On [0, 1], if a function has the integral I, as a limit of Riemann sums, then c times the function has the integral c·I. *Used for:* Pulling normalization constants and couplings out of integrals. *Price:* **free**; exact.

537. Integral of a difference. On [0, 1], if f and g have the integrals I and J, as limits of Riemann sums, then f − g has the integral I − J. *Used for:* Work and energy differences. *Price:* **free**; exact.

538. Integral of a finite sum. On [0, 1], if each of finitely many functions has an integral, as a limit of Riemann sums, then their sum has as its integral the sum of their integrals. *Used for:* Superposing modes; finite expansions. *Price:* **free**; exact.

539. Integral of the absolute value bounds the integral. On [0, 1], if f has the integral I and |f| has the integral J, as limits of Riemann sums, then $|I| \le J$. *Used for:* Error and perturbation bounds. *Price:* **free**; exact.

540. Uniform limits pass under the integral. If functions with integrals over [0, 1] converge uniformly on [0, 1] to a function with an integral, their integrals converge to the integral of the limit. *Used for:* Limits of approximate solutions. *Price:* **free**; exact.

541. Term-by-term integration of a series. If the partial sums of a series of functions converge uniformly on [0, 1], and every term and the limit have integrals over [0, 1], then the partial sums of the term integrals converge to the integral of the limit. *Used for:* Fourier and perturbation series. *Price:* **free**; exact.

542. Monotone convergence theorem, value of the limit. On [0, 1], the integrals of an increasing sequence of uniformly continuous functions whose integrals stay bounded converge to a number. *Used for:* Partition functions and normalizations as limits of truncations. *Price:* **LPO + AC**$_\omega$; exact.

543. Improper integral of a nonnegative function. A nonnegative continuous function on the half-line whose integrals up to each bound stay bounded has an integral over the whole half-line. *Used for:* Normalizing over the half-line without a decay rate. *Price:* **LPO + AC**$_\omega$; exact.

544. Comparison test for improper integrals. On the half-line [0, ∞), a continuous function whose size is bounded by a continuous function with a convergent integral over the half-line also has a convergent integral there. *Used for:* Dominated tails such as exponential or inverse-square decay. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

545. Differentiation under the integral sign. On [0, 1], if f(s, x) has a derivative in the parameter s that is uniform in s and x together, then the integral of f(s, x) over x has as its derivative in s the integral of that derivative. *Used for:* Hellmann–Feynman, Green's functions, parameter tricks. *Price:* **free**; exact.

546. Continuity of a parameter integral. If f(s, x) is uniformly continuous on the unit square [0, 1] × [0, 1], its integral over x from 0 to 1 is a uniformly continuous function of s. *Used for:* Potentials depending continuously on their sources. *Price:* **free**; exact.

547. Mean value theorem for integrals. A uniformly continuous function on [0, 1] equals its integral over [0, 1], its mean value, at some point of the closed interval [0, 1]. *Used for:* Mean-value estimates. *Price:* **LLPO + AC**$_\omega$; exact.

548. Mean value theorem for integrals, interior point. A uniformly continuous function on [0, 1] equals its integral over [0, 1], its mean value, at some point strictly between 0 and 1. *Used for:* Mean-value estimates. *Price:* **LPO + AC**$_\omega$; exact.

549. Cauchy–Schwarz inequality for integrals. On [0, 1], if f·g, f^2 and g^2 have the integrals I, A and B, as limits of Riemann sums, then $I^2 \le A{\cdot}B$. *Used for:* Uncertainty relations, overlap bounds. *Price:* **free**; exact.

550. Fubini's theorem on the unit square. For a uniformly continuous function on the unit square [0, 1] × [0, 1], the integral taken first in one variable and then in the other equals the integral taken in the opposite order. *Used for:* Every iterated two- and three-dimensional integral. *Price:* **free**; exact.

551. Zero integral of a nonnegative function. A uniformly continuous function that is nonnegative on [0, 1] and has integral zero over [0, 1] is zero at every point of [0, 1]. *Used for:* A state of zero norm is the zero state. *Price:* **free**; exact.

552. Dini's theorem for integrals. If uniformly continuous functions on [0, 1] increase at every point of [0, 1] to a uniformly continuous limit, their integrals over [0, 1] converge to the integral of the limit. *Used for:* Removing a cutoff or regulator under an integral. *Price:* **open**; equivalent, over countable choice, to the measure form of the Heine–Borel theorem (the lengths of a countable cover of the interval by open intervals add up to at least one); the direction from the theorem to the principle uses no choice; not one of the price letters; also at most LLPO + AC$_\omega$ and at most the decidable fan theorem with countable choice.

553. Bounded convergence theorem for integrals. If uniformly continuous functions on [0, 1], all bounded in size by one constant, converge at every point of [0, 1] to a uniformly continuous limit, their integrals over [0, 1] converge to the integral of the limit. *Used for:* Limits under the integral without monotonicity or a rate. *Price:* **open**; equivalent over countable choice to continuity of measure from above in sequence form (for decreasing open sets with no common point, any prescribed finite parts of the later sets eventually have small measure); the direction from the theorem to the principle uses no choice; not one of the price letters; at most LPO + AC_ω; at least the measure form of the Heine–Borel theorem.

554. Integrals over adjacent pieces of [0, 1] add. On the line, let f be uniformly continuous on [0, 1] and $0 \le a \le c \le b \le 1$. If the integrals of f from a to c, from c to b and from a to b exist, the first two add up to the third. *Used for:* Splitting a path, a time interval or a region into pieces. *Price:* **free**; exact.

555. Fundamental theorem of calculus, first half, on [0, 1]. On the line, if f is uniformly continuous on [0, 1] and F(x) is the integral of f from 0 to x for every x in [0, 1], then F has derivative f on [0, 1], uniformly in the point. *Used for:* Recovering a rate from an accumulated quantity (charge from current, work from force). *Price:* **free**; exact.

556. Reversing the interval negates the integral. The integral from b to a is minus the integral from a to b. *Used for:* Orientation of line integrals and time reversal. *Price:* **free**; exact.

557. Linear change of variable in an integral. Substituting $x = c\,t + d$ turns the integral of f over the image interval into the integral of c times $f(c\,t + d)$ over the original interval. *Used for:* Rescaling units and shifting origins. *Price:* **free**; exact.

558. Integrability on every subinterval of [0, 1]. On the line, a function f uniformly continuous on [0, 1] has an integral from a to b for all points a and b of [0, 1]. *Used for:* Every definite integral of a continuous quantity. *Price:* **free**; exact (a modulus of uniform continuity is chosen by countable choice of natural numbers; the Riemann sums then form a Cauchy sequence with a known modulus, and its limit point uses the completeness of the line: countable choice of rational pairs, Corollary 12.35).

559. Change of variables by a differentiable map. On the line, let f be uniformly continuous on [0, 1], let g have the uniform derivative g' on [0, 1] and map [0, 1] into itself, and let a and b lie in [0, 1]. If f has the integral I from g(a) to g(b), then $f(g(x))\cdot g'(x)$ has the integral I from a to b. *Used for:* Changing coordinates along a path, in time or in energy. *Price:* **free**; exact.

560. Integration by parts on an interval. On the line, let u and v have uniform derivatives u' and v' on [0, 1], and let a and b lie in [0, 1]. If $u'\cdot v$ has the integral J from a to b, then $u\cdot v'$ has the integral $u(b)\cdot v(b) - u(a)\cdot v(a) - J$ from a to b. *Used for:* Moving derivatives between factors: weak forms, virial and expectation-value identities. *Price:* **free**; exact.

561. Mean value theorem for integrals on an interval. On the line, let f be uniformly continuous on [0, 1] and $0 \le a \le b \le 1$. If f has the integral I from a to b, then $(b - a)\cdot f(c) = I$ at some point c with $a \le c \le b$. *Used for:* Mean-value estimates over any interval. *Price:* **LLPO + AC_ω**; exact.

562. Mean value theorem for integrals, interior point, on an interval. On the line, let f be uniformly continuous on [0, 1] and let $a < b$ be points of [0, 1]. If f has the integral I from a to b, then $(b - a)\cdot f(c) = I$ at some point c strictly between a and b, at a known positive distance from both. *Used for:* Mean-value estimates over any interval. *Price:* **LPO + AC_ω**; exact.

563. Integrals preserve order on an interval. For points $0 \le a \le b \le 1$ of the line, if $f \le g$ at every point of [a, b] and f and g have the integrals I and J from a to b, then $I \le J$. *Used for:* Comparison bounds on energies, probabilities and errors. *Price:* **free**; exact.

564. Bound on an integral by a bound on the integrand. For points $0 \le a \le b \le 1$ of the line, if $|f(x)| \le C$ at every point x of [a, b] and f has the integral I from a to b, then $|I| \le C\cdot(b - a)$. *Used for:* Error estimates and a-priori bounds. *Price:* **free**; exact.

565. Cauchy–Schwarz inequality for integrals on an interval. For any points a and b of the line, if $f\cdot g$, f^2 and g^2 have the integrals I, A and B from a to b, then $I^2 \le A\cdot B$. *Used for:* Uncertainty relations and overlap bounds on a finite range. *Price:* **free**; exact.

566. Fubini's theorem on a rectangle. For points $0 \le a \le b \le 1$ and $0 \le c \le d \le 1$ of the line and a function f(x, y) uniformly continuous on the unit square, suppose the integrals of f over x from a to b (for each y in [c, d]) and over y from c to d (for each x in [a, b]) exist, and so do the two iterated integrals over the rectangle [a, b] × [c, d]. Then the iterated integral taken first in x and then in y equals the iterated integral taken first in y and then in x. *Used for:* Every iterated integral over a rectangular region. *Price:* **free**; exact.

567. Differentiation under the integral sign on an interval. For points $0 \le a \le b \le 1$ of the line, let f(s, x) have the derivative $f'(s, x)$ in the parameter s, uniformly for s in [0, 1] and x in [a, b] together (one accuracy rate for all of them). If, for every s in [0, 1], the function $x \mapsto f(s, x)$ has the integral G(s) and $x \mapsto f'(s, x)$ has the integral D(s) from a to b, then G has the derivative D, uniformly on [0, 1]. *Used for:*

Hellmann–Feynman, Green's functions, parameter tricks on a finite range. *Price:* **free**; exact.

568. Continuity of a parameter integral on an interval. For points $0 \le a \le b \le 1$ of the line and a function f(s, x) uniformly continuous on the unit square $[0, 1] \times [0, 1]$, if G(s) is the integral of $x \mapsto f(s, x)$ from a to b for every s in [0, 1], then G is uniformly continuous on [0, 1]. *Used for:* Potentials and responses that vary continuously with their sources. *Price:* **free**; exact.

569. Gamma function at whole numbers. The integral from 0 to infinity of t to the power n times e to the minus t equals n factorial, for every whole number n. *Used for:* Partition functions, Bose and Fermi integrals, Laplace transforms, radial normalizations. *Price:* **free**; exact.

570. Beta integral at whole numbers. For whole numbers a and b: if t to the power a times (1 − t) to the power b has an integral I from 0 to 1 (its Riemann sums come within every tolerance of I), then (a + b + 1)! times I equals a! times b!, so I = a! b!/(a + b + 1)!. *Used for:* Order statistics and Bayesian priors, phase-space integrals, binomial weights. *Price:* **free**; exact.

571. Laplace transform of a power. For every whole number n and every $s > 0$, the integral from 0 to infinity of t to the power n times e to the minus s t equals n! divided by s to the power n + 1. *Used for:* Response functions, relaxation and decay laws, partition functions at temperature 1/s. *Price:* **free**; exact.

572. Gamma function at real arguments. For every real $s > 0$, the integral of t to the power s − 1 times e to the minus t exists over every closed interval inside (0, infinity), and these integrals converge as the interval grows to (0, infinity). *Used for:* Partition functions and densities of states in any dimension, Maxwell–Boltzmann moments, radial integrals with non-integer exponents. *Price:* **free**; exact (the integral over each closed interval is that of a uniformly continuous function, as in row 1467; the parts below ε and beyond R are at most ε^s/s and a constant times $e^{-R/2}$, so the integrals converge with a known modulus; the limit point uses the completeness of the line: countable choice of rational pairs).

573. Gamma recurrence at real arguments. For every real $s > 0$, Gamma of s + 1 equals s times Gamma of s. *Used for:* Factorial-type normalizations at non-integer order, thermal averages in d dimensions. *Price:* **free**; exact.

574. Gamma function is positive. For every real $s > 0$, Gamma of s is strictly positive. *Used for:* Positivity of partition functions and normalization constants. *Price:* **free**; exact.

575. Beta integral at real arguments. For real a, $b > 0$, the integral of t to the power a − 1 times (1 − t) to the power b − 1 exists over every closed interval inside (0, 1), and these integrals converge as the interval grows to (0, 1). *Used for:* Beta distributions and Bayesian priors, Dirichlet and phase-space integrals, scattering amplitudes. *Price:* **free**; exact (the integral over each closed interval is that of a uniformly continuous function, as in row 1467; the parts within ε of 0 and of 1 are at most a constant times ε^a/a and ε^b/b, so the integrals converge with a known modulus; the limit point uses the completeness of the line: countable choice of rational pairs).

576. Beta function is symmetric. For real a, $b > 0$, the Beta integral with exponents a, b equals the one with exponents b, a. *Used for:* Exchange symmetry of two-body phase space and of Beta priors. *Price:* **free**; exact.

577. Gamma and Beta functions are linked. For real a, $b > 0$, Gamma of a times Gamma of b equals Gamma of a + b times the Beta integral with exponents a and b. *Used for:* Dirichlet and phase-space integrals, Beta-distribution normalizations, Veneziano-type amplitudes. *Price:* **free**; exact.

578. Recurrence of the Beta integral in its first argument. For real a, $b > 0$, (a + b) times the Beta integral at (a + 1, b) equals a times the Beta integral at (a, b). *Used for:* Moment recursions of Beta distributions; reduction of phase-space integrals. *Price:* **free**; exact.

579. Gamma function is log-convex. For real y, $d > 0$ and θ between 0 and 1, Gamma at y + θd is at most Gamma at y to the power 1 − θ times Gamma at y + d to the power θ. *Used for:* Bounds on partition functions and thermodynamic convexity; ratio bounds for Gamma functions. *Price:* **free**; exact.

580. Gamma at one. The integral of e to the minus t from 0 to infinity equals 1. *Used for:* Normalization of exponential decay and waiting-time laws. *Price:* **free**; exact.

581. Euler–Gauss limit for the Gamma function. For real $s > 0$, n to the power s times n factorial divided by $s(s + 1)\cdots(s + n)$ tends to Gamma of s as n grows. *Used for:* Numerical evaluation of Gamma; infinite-product forms of partition functions. *Price:* **free**; exact.

582. Bohr–Mollerup characterization of the Gamma function. A positive function on the positive half-line with f(1) = 1, f(y + 1) = y f(y) and log-convex values equals the Gamma function at every positive point. *Used for:* Identifying factorial-type normalizations by their recurrence and

convexity. *Price:* **free**; exact.

583. Gamma at one half. Gamma of one half is positive and its square equals π. *Used for:* Gaussian and error-function normalizations, half-integer partition functions, the volume of balls. *Price:* **free**; exact.

584. Legendre duplication formula. For real $s > 0$, Gamma of s times Gamma of $s + 1/2$ equals Gamma of 2s times 2 to the power $1 - 2s$ times the square root of π. *Used for:* Half-integer Gamma values in d-dimensional integrals and Bessel normalizations. *Price:* **free**; exact.

585. Euler's reflection formula. For real s strictly between 0 and 1, Gamma of s times Gamma of $1 - s$ times the sine of π s equals π. *Used for:* Fermi–Dirac and Bose–Einstein integrals, the Casimir and Planck constants' Gamma–zeta factors, analytic continuation in field theory. *Price:* **free**; exact.

586. Euler's product for the sine. For every real x, π x times the product over $j < n$ of $(1 - x^2/(j + 1)^2)$ tends to the sine of π x as n grows. *Used for:* Spectra of strings and drums (zeros of sine), partition functions of oscillator chains. *Price:* **free**; exact.

587. Bose–Einstein integral as Gamma times zeta. For real $s > 1$, the integral from 0 to infinity of t to the power $s - 1$ divided by (e to the t minus 1) equals Gamma of s times zeta of s. *Used for:* Black-body radiation and the Stefan–Boltzmann law, Bose gases and Bose–Einstein condensation. *Price:* **free**; exact.

588. Fermi–Dirac integral through Gamma and zeta. For real $s > 1$, the integral from 0 to infinity of t to the power $s - 1$ divided by (e to the t plus 1) equals $(1 - 2$ to the power $1 - s)$ times Gamma of s times zeta of s. *Used for:* Degenerate electron gases, white dwarfs and metals, relic neutrino densities. *Price:* **free**; exact.

589. Convergence of the p-series for real exponents above one. For real $s > 1$, the sums of 1 over k to the power s, for k from 1 to n, converge as n grows (this limit is zeta of s). *Used for:* Zeta values in thermal physics, lattice sums, Casimir energies. *Price:* **free**; exact (the sum beyond n is at most $n^{1-s}/(s - 1)$, a known Cauchy modulus; the limit point uses the completeness of the line: countable choice of rational pairs).

590. Basel problem: the sum of inverse squares. The sum of 1 over k squared, for k from 1 to infinity, equals π squared over 6. *Used for:* Thermal and Casimir sums, string and lattice zero-point energies, Debye models. *Price:* **free**; exact. Proved in Theorem 12.97.

591. Sum of inverse fourth powers. The sum of 1 over k to the fourth power, for k from 1 to infinity, equals π to the fourth over 90. *Used for:* Black-body radiation, Stefan–Boltzmann constant, photon gas energy density. *Price:* **free**; exact. Proved in Theorem 12.97.

592. Black-body integral of the Planck spectrum. The integral from 0 to infinity of t cubed divided by (e to the t minus 1) equals π to the fourth over 15. *Used for:* The Stefan–Boltzmann law and the energy density of thermal radiation. *Price:* **free**; exact.

Q.5 Differential equations and dynamics

593. Grönwall's bound as one function. There is one function $g(\delta, K, \varepsilon, x)$ equal to $\delta + \varepsilon x$ when $K = 0$ and satisfying $K\cdot g = K\delta e^{Kx} + \varepsilon(e^{Kx} - 1)$ for every K; it is $\delta e^{Kx} + \varepsilon x\cdot\psi(Kx)$ with $\psi(y) = \Sigma\, y^m/(m+1)!$. *Used for:* The envelope that bounds the growth of every perturbation and every numerical error in a differential equation. *Price:* **free**; exact.

594. Grönwall's bound without feedback. When the growth rate K is 0, Grönwall's bound $\delta e^{Kx} + \varepsilon x\cdot\varphi(Kx)$, where $\varphi(u) = \Sigma\, u^m/(m+1)!$ over $m \geq 0$, so that $\varepsilon x\cdot\varphi(Kx) = (\varepsilon/K)(e^{Kx} - 1)$ when $K \neq 0$, is $\delta + \varepsilon x$. *Used for:* Error growth from a steady error source in a system with no amplification. *Price:* **free**; exact.

595. Closed form of Grönwall's bound. K times Grönwall's bound $g(x)$ equals $K\delta e^{Kx} + \varepsilon(e^{Kx} - 1)$, so $g(x) = \delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$ whenever K is apart from 0. *Used for:* The explicit exponential envelope of perturbation growth in a system with amplification rate K. *Price:* **free**; exact.

596. Derivative of Grönwall's bound. Grönwall's bound $g(x) = \delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$ has derivative $Kg(x) + \varepsilon$ at every x. *Used for:* Showing the envelope solves the comparison equation $g' = Kg + \varepsilon$ that bounds a growing error. *Price:* **free**; exact.

597. Derivative of the shifted Grönwall bound. The function $x \mapsto g(x - a)$, with g Grönwall's bound $\delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$, has derivative $Kg(x - a) + \varepsilon$ at every x. *Used for:* The comparison envelope for an error that starts growing at time a. *Price:* **free**; exact.

598. Grönwall's bound at time zero. Grönwall's bound $\delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$ equals δ at $x = 0$. *Used for:* The envelope starts at the initial size of the perturbation. *Price:* **free**; exact.

599. Grönwall's bound with no error source. With $\varepsilon = 0$, Grönwall's bound is δe^{Kx}. *Used for:*

Exponential separation of two exact motions of a law with Lipschitz constant K. *Price:* **free**; exact.

600. Grönwall's bound from nothing is nothing. With $\delta = 0$ and $\varepsilon = 0$, Grönwall's bound is 0 for every x and K. *Used for:* Uniqueness: two motions that start together never separate. *Price:* **free**; exact.

601. Grönwall's bound is continuous in the error rate. For fixed δ, K and x, Grönwall's bound $\delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$ depends continuously on ε. *Used for:* Passing from error rates slightly above ε to ε itself in perturbation bounds. *Price:* **free**; exact.

602. Grönwall's bound only grows. For δ, ε, $K \geq 0$, Grönwall's bound $\delta e^{Kx} + (\varepsilon/K)(e^{Kx} - 1)$ is nondecreasing in x. *Used for:* Error envelopes never shrink, so a bound at a later time covers every earlier time. *Price:* **free**; exact.

603. Grönwall comparison for a real function. Let f be a real function on [a, b] with $f(a) \leq \delta$, and suppose that for every tolerance there is one closeness such that every forward difference quotient $(f(t) - f(s))/(t - s)$ with $a \leq s < t \leq b$ and t that close to s is at most $f'(s)$ plus the tolerance. If $f'(x) \leq K \cdot f(x) + \varepsilon$ for $a \leq x < b$, then $f(x) \leq \delta \cdot \exp(K(x - a)) + (\varepsilon/K)(\exp(K(x - a)) - 1)$ on [a, b] ($\delta + \varepsilon(x - a)$ when K = 0). *Used for:* Bounding energies, Lyapunov functions and other scalar quantities that grow at most in proportion to themselves. *Price:* **free**; exact.

604. Grönwall's inequality for the size of a motion. Let f(t) be a motion of states with d real coordinates (size of a state: its largest absolute coordinate) on [a, b], with right derivative f' uniform on [a, b] (one accuracy rate for all times). If $\|f(a)\| \leq \delta$ and $\|f'(t)\| \leq K\|f(t)\| + \varepsilon$ for $a \leq t < b$, then $\|f(t)\| \leq \delta \cdot \exp(K(t - a)) + (\varepsilon/K)(\exp(K(t - a)) - 1)$ on [a, b] ($\delta + \varepsilon(t - a)$ when K = 0). *Used for:* Growth of perturbations, stability estimates and error bounds for every differential equation. *Price:* **free**; exact.

605. A state that starts at zero stays at zero. Let f(t) be a motion of states with d real coordinates (size of a state: its largest absolute coordinate) on [a, b], with right derivative f' uniform on [a, b] (one accuracy rate for all times). If $f(a) = 0$ and $\|f'(t)\| \leq K\|f(t)\|$ for $a \leq t < b$, then $f(t) = 0$ for every t in [a, b]. *Used for:* A system at rest under a law that vanishes at rest stays at rest; the core of uniqueness. *Price:* **free**; exact.

606. Separation of approximate motions in a region. Let v(t, x) be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on sets S(t) for $a \leq t < b$. Let f and g stay in S(t), have right derivatives uniform on [a, b], and satisfy $\|f'(t) - v(t, f(t))\| \leq \varepsilon_1$, $\|g'(t) - v(t, g(t))\| \leq \varepsilon_2$ and $\|f(a) - g(a)\| \leq \delta$. Then $\|f(t) - g(t)\| \leq \delta \cdot \exp(K(t - a)) + ((\varepsilon_1 + \varepsilon_2)/K)(\exp(K(t - a)) - 1)$ on [a, b] ($\delta + (\varepsilon_1 + \varepsilon_2)(t - a)$ when K = 0). *Used for:* Error bounds for numerical integrators and for approximate solutions confined to a region. *Price:* **free**; exact.

607. Separation of approximate motions. Let v(t, x) be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on all states at every time. Let f and g have right derivatives uniform on [a, b] and satisfy $\|f'(t) - v(t, f(t))\| \leq \varepsilon_1$, $\|g'(t) - v(t, g(t))\| \leq \varepsilon_2$ and $\|f(a) - g(a)\| \leq \delta$. Then $\|f(t) - g(t)\| \leq \delta \cdot \exp(K(t - a)) + ((\varepsilon_1 + \varepsilon_2)/K)(\exp(K(t - a)) - 1)$ on [a, b] ($\delta + (\varepsilon_1 + \varepsilon_2)(t - a)$ when K = 0). *Used for:* Global error bounds for numerical integration of a law with a uniform Lipschitz constant. *Price:* **free**; exact.

608. Separation of exact motions in a region. Let v(t, x) be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on sets S(t) for $a \leq t < b$. Two motions f and g that stay in S(t), solve $f' = v(t, f)$ and $g' = v(t, g)$ with right derivatives uniform on [a, b], and start with $\|f(a) - g(a)\| \leq \delta$ satisfy $\|f(t) - g(t)\| \leq \delta \cdot \exp(K(t - a))$ on [a, b]. *Used for:* Sensitivity to initial data: nearby states separate at most exponentially, with rate at most K. *Price:* **free**; exact.

609. Separation of exact motions. Let v(t, x) be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on all states at every time. Two motions f and g that solve $f' = v(t, f)$ and $g' = v(t, g)$ with right derivatives uniform on [a, b] and start with $\|f(a) - g(a)\| \leq \delta$ satisfy $\|f(t) - g(t)\| \leq \delta \cdot \exp(K(t - a))$ on [a, b]. *Used for:* Bounding how fast initial uncertainty spreads in a deterministic system. *Price:* **free**; exact.

610. Uniqueness forward in time. Two motions of a law Lipschitz on a region containing both, with equal states at time a, agree at every time of [a, b]. *Used for:* Determinism: the present state and the law fix the future. *Price:* **free**; exact.

611. Uniqueness backward in time. Two motions of a law Lipschitz on a region containing both, with equal states at the final time b, agree at every time of [a, b]. *Used for:* Determinism: the present state and the law fix the past. *Price:* **free**; exact.

612. Uniqueness from an inner time. Let v(t, x) be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on sets S(t) for $a < t < b$. Two motions that stay in S(t) for $a < t < b$, solve $x' = v(t, x)$ with a two-sided derivative uniform on the closed interval [a, b] (ends included), and agree at one time t_0 with $a < t_0 < b$ agree at every time in [a, b]. *Used for:* Determinism in both directions of time from a state

given in the middle of an interval. *Price:* **free**; exact.

613. Uniqueness on an open time interval. Let $v(t, x)$ be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on sets $S(t)$ for $a < t < b$. Two motions that stay in $S(t)$, solve $x' = v(t, x)$ with derivatives uniform on every closed interval inside (a, b), and agree at one time t_0 in (a, b) agree at every time in (a, b). *Used for:* Determinism for motions defined only on an open time interval, such as up to a collision. *Price:* **free**; exact.

614. Uniqueness for a globally Lipschitz law. Two motions on $[a, b]$ of a law that is Lipschitz everywhere, with equal states at time a, agree on all of $[a, b]$. *Used for:* Determinism for laws with a uniform Lipschitz constant, such as linear laws. *Price:* **free**; exact.

615. Uniqueness of motions for all time. Let $v(t, x)$ be a law of motion for states with d real coordinates (size of a state: its largest absolute coordinate) that is K-Lipschitz in x, $K \geq 0$, on sets $S(t)$ at every time. Two motions defined at all times that stay in $S(t)$, solve $x' = v(t, x)$ with derivatives uniform on every bounded time interval, and agree at one time agree at every time. *Used for:* Determinism of a physical system over its whole history. *Price:* **free**; exact.

616. Picard–Lindelöf existence theorem. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law $f(t, y)$ be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, continuous in t uniformly there (one accuracy rate for all those times and states), and let $t_1 \leq t_0 \leq t_2$ with $L \cdot \max(t_2 - t_0, t_0 - t_1) \leq a - r$. Then from every state x within r of x_0 there is a motion through x at time t_0 on all of $[t_1, t_2]$, whose velocity $f(t, \text{state})$ is its derivative uniformly on $[t_1, t_2]$. *Used for:* Existence of the motion of a mechanical, electrical or chemical system from a given initial state. *Price:* **free**; exact: no omniscience principle; the proof chooses a modulus of continuity in time, a choice of natural numbers, and then proceeds as in row 617, whose only choice is of representatives; both choices are inside the free allowance.

617. Picard–Lindelöf with a known continuity modulus. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law $f(t, y)$ be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), and let $t_1 \leq t_0 \leq t_2$ with $L \cdot \max(t_2 - t_0, t_0 - t_1) \leq a - r$. Then from every state x within r of x_0 there is a motion through x at time t_0 on all of $[t_1, t_2]$, whose velocity $f(t, \text{state})$ is its derivative uniformly on $[t_1, t_2]$. *Used for:* Existence of the motion of a system whose law varies in time at a known rate, as integrators compute it. *Price:* **free**; exact: no omniscience principle; the only choice is one countable choice of representatives, which the free allowance contains.

618. Picard–Lindelöf from the centre of the region. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law $f(t, y)$ be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, continuous in t uniformly there (one accuracy rate for all those times and states), and let $t_1 \leq t_0 \leq t_2$ with $L \cdot \max(t_2 - t_0, t_0 - t_1) \leq a$. Then there is a motion through x_0 at time t_0 on all of $[t_1, t_2]$, whose velocity $f(t, \text{state})$ is its derivative uniformly on $[t_1, t_2]$. *Used for:* Existence of the motion of a system started at the point about which its law is known. *Price:* **free**; exact: no omniscience principle; this is row 616 with $r = 0$, and its choices, of a modulus of continuity and of representatives, are inside the free allowance.

619. Motion of a linear system. For every constant real $d \times d$ matrix A and every initial state x_0, there is a motion defined at every time that starts at x_0 at time 0 and whose velocity is A times its state, with the derivative uniform on every bounded time interval. *Used for:* Small oscillations, linear circuits, and the evolution of a finite quantum system. *Price:* **free**; exact.

620. Picard–Lindelöf for a law given by a formula. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law $f(t, y)$ be computed from approximations of its arguments, as every law written by a formula (sums, products, maxima, minima, rational constants) is. Let it be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), where a, r, L and K are nonnegative, $t_1 \leq t_0 \leq t_2$ and $L \cdot \max(t_2 - t_0, t_0 - t_1) \leq a - r$. Then from every state x within r of x_0 there is a motion through x at time t_0 on all of $[t_1, t_2]$, with its derivative uniform there. *Used for:* Existence of the motion of any mechanical, electrical or chemical system whose law is written as a formula. *Price:* **free**; exact.

621. Evolution of a linear system composes in time. For every constant real $d \times d$ matrix A there is one function $\Phi(x, t)$ of the initial state x and the time t such that, from every initial state x, the motion $t \mapsto \Phi(x, t)$ solves $x' = A\,x$ at all times (with the derivative uniform on every bounded time interval) and starts at $\Phi(x, 0) = x$; it satisfies $\Phi(x, s + t) = \Phi(\Phi(x, s), t)$, so evolving for time s and then for time t is evolving for $s + t$; and it is linear in the initial state. *Used for:*

Propagating small oscillations, linear circuits and a finite quantum system in real coordinates step by step in time. *Price:* **free**; exact.

622. Flow of a formula law, Lipschitz in the initial state. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law f(t, y) be computed from approximations of its arguments, as every law written by a formula (sums, products, maxima, minima, rational constants) is. Let it be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), where a, r, L and K are nonnegative, $t_1 \le t_0 \le t_2$ and $L\cdot\max(t_2 - t_0, t_0 - t_1) \le a - r$. Then there is one function $\alpha(x, t)$ of the initial state and the time that, for every x within r of x_0, is a motion through x at time t_0 on all of $[t_1, t_2]$ with its derivative uniform there, and one constant $L' \ge 0$ with $\|\alpha(x, t) - \alpha(y, t)\| \le L'\cdot\|x - y\|$ for all x and y within r of x_0 and every t in $[t_1, t_2]$. *Used for:* The flow of a mechanical or chemical system as a map of initial states, continuous in them: the basis of sensitivity analysis and of phase portraits. *Price:* **free**; exact.

623. Chain rule along a path. Let H be a real function of states with d real coordinates (size of a state: its largest absolute coordinate) whose gradient G exists uniformly on a region (one closeness for each accuracy, everywhere in the region) and is bounded there. If a path α stays in the region and has velocity α' uniformly on [a, b], with α' bounded, then $t \mapsto H(\alpha(t))$ has derivative $G(\alpha(t))\cdot\alpha'(t)$ uniformly on [a, b]. *Used for:* Rates of change of energy, entropy or any observable along a moving state. *Price:* **free**; exact.

624. Conserved quantities stay constant along motions. Let H be a real function of states with d real coordinates (size of a state: its largest absolute coordinate) whose gradient G exists uniformly on a region (one closeness for each accuracy, everywhere in the region) and is bounded there. Let α be a motion of a law f on [a, b], with velocity $f(t, \alpha(t))$ and the derivative uniform on [a, b], that stays in the region, with the speed $f(t, \alpha(t))$ bounded. If $G(\alpha(t))\cdot f(t, \alpha(t)) = 0$ for every t in [a, b], then $H(\alpha(s)) = H(\alpha(t))$ for all s and t in [a, b]. *Used for:* Energy conservation for Hamiltonian systems, conservation of momentum and of every first integral along the motion. *Price:* **free**; exact.

625. Lyapunov functions do not increase along motions. Let H be a real function of states with d real coordinates (size of a state: its largest absolute coordinate) whose gradient G exists uniformly on a region (one closeness for each accuracy, everywhere in the region) and is bounded there. Let α be a motion of a law f on [a, b], with velocity $f(t, \alpha(t))$ and the derivative uniform on [a, b], that stays in the region, with the speed $f(t, \alpha(t))$ bounded. If $G(\alpha(t))\cdot f(t, \alpha(t)) \le 0$ for every t in [a, b], then $H(\alpha(t)) \le H(\alpha(s))$ whenever $a \le s \le t \le b$. *Used for:* Stability arguments: energy that can only dissipate, entropy-like quantities that can only move one way, Lyapunov's direct method. *Price:* **free**; exact.

626. Local uniqueness of motions near a time. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law v(t, x) be K-Lipschitz in x, $K \ge 0$, on sets S(t) for every time t in some open interval around t_0. Let two motions f and g stay in S(t) and solve $x' = v(t, x)$ on open intervals around t_0, with derivatives uniform on every closed interval inside those intervals. If $f(t_0) = g(t_0)$, then f and g agree at every time of some open interval around t_0. *Used for:* Determinism over short times, where a law is only known to be well behaved near the present. *Price:* **free**; exact.

627. Time-independent laws meet the existence hypotheses. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law f(y) that does not depend on time be bounded by L and K-Lipschitz on the closed ball of radius a about x_0, where a, r, L and K are nonnegative, $t_1 \le t_0 \le t_2$ and $L\cdot\max(t_2 - t_0, t_0 - t_1) \le a - r$. Then f satisfies every hypothesis of the Picard–Lindelöf theorem on $[t_1, t_2]$, and since it does not change in time, any rule serves as its modulus of continuity in time. *Used for:* Autonomous systems: a pendulum, a planet, a chemical network whose rules do not change in time. *Price:* **free**; exact.

628. Existence of the flow of a formula law. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law f(t, y) be computed from approximations of its arguments, as every law written by a formula (sums, products, maxima, minima, rational constants) is. Let it be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), where a, r, L and K are nonnegative, $t_1 \le t_0 \le t_2$ and $L\cdot\max(t_2 - t_0, t_0 - t_1) \le a - r$. Then there is one function $\alpha(x, t)$ of the initial state and the time that, for every x within r of x_0, is a motion through x at time t_0 on all of $[t_1, t_2]$, with its derivative uniform there. *Used for:* The flow map of a system: every initial state evolved by one rule. *Price:* **free**; exact.

629. Continuity of the flow of a formula law. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law f(t, y) be computed from approximations of its arguments, as every law written by a formula (sums, products, maxima, minima, rational constants) is. Let it be

K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), where a, r, L and K are nonnegative, $t_1 \le t_0 \le t_2$ and $L \cdot \max(t_2 - t_0, t_0 - t_1) \le a - r$. Then there is one function $\alpha(x, t)$ that, for every x within r of x_0, is a motion through x at time t_0 on all of $[t_1, t_2]$ with its derivative uniform there, and α is uniformly continuous jointly in the initial state x (within r of x_0) and the time t (in $[t_1, t_2]$). *Used for:* Small changes in the initial state and the time give small changes in the evolved state: the basis of numerical prediction. *Price:* **free**; exact.

630. Motions with ordinary derivatives on the open interval. For states with d real coordinates (size of a state: its largest absolute coordinate), let a law f(t, y) be computed from approximations of its arguments, as every law written by a formula (sums, products, maxima, minima, rational constants) is. Let it be K-Lipschitz in y and bounded by L on the closed ball of radius a about x_0 for every time t in $[t_1, t_2]$, with a known modulus of continuity in t (a rule giving, for each accuracy, a time step that achieves it for all states of the ball), where a, r, L and K are nonnegative, $t_1 \le t_0 \le t_2$ and $L \cdot \max(t_2 - t_0, t_0 - t_1) \le a - r$. Then from every state x within r of x_0 there is a motion α through x at time t_0 that has the velocity $f(t, \alpha(t))$, as an ordinary two-sided derivative, at every time t strictly between t_1 and t_2. *Used for:* Existence of the motion with an ordinary derivative at every inner time, as textbooks state it. *Price:* **free**; exact.

631. The evolution of the zero law is the identity. For states with d real coordinates, every evolution map Φ of the law $x' = 0$ (any map with $\Phi(x, 0) = x$ whose motions $t \mapsto \Phi(x, t)$ solve the law, with derivatives uniform on every bounded time interval) leaves every state where it is: $\Phi(x, t) = x$ for every state x and time t. *Used for:* The trivial evolution: nothing moves when no force acts. *Price:* **free**; exact.

632. Linear evolution can be run backward. Let A be a constant real d × d matrix and $\Phi(x, t)$ any evolution map of the law $x' = A\,x$ (any map with $\Phi(x, 0) = x$ whose motions $t \mapsto \Phi(x, t)$ solve the law, with derivatives uniform on every bounded time interval). Then evolving for time t and then for time −t returns every state to itself, and so does evolving for time −t and then for time t: $\Phi(\Phi(x, t), -t) = x$ and $\Phi(\Phi(x, -t), t) = x$. *Used for:* Reversibility of linear dynamics: recovering an earlier state from a later one. *Price:* **free**; exact.

633. Decoupled modes evolve separately. Let A be a constant real d × d matrix whose entries off the diagonal are 0, and $\Phi(x, t)$ any evolution map of the law $x' = A\,x$ (any map with $\Phi(x, 0) = x$ whose motions $t \mapsto \Phi(x, t)$ solve the law, with derivatives uniform on every bounded time interval). Then coordinate c of $\Phi(x, t)$ is $\exp(t \cdot a)$ times coordinate c of x, where a is the diagonal entry of A in row c. *Used for:* Normal modes: independent oscillators or decaying channels each evolve by their own rate. *Price:* **free**; exact.

634. Linear evolution in another basis. Let A, U and V be constant real d × d matrices with $U \cdot V = V \cdot U = I$. Let $\Phi(x, t)$ be an evolution map of the law $x' = A\,x$ and $\Psi(x, t)$ one of the law $x' = (U \cdot A \cdot V)\,x$: maps with $\Phi(x, 0) = x$ and $\Psi(x, 0) = x$ whose motions in t solve their laws, with derivatives uniform on every bounded time interval. Then $\Psi(x, t) = U \cdot \Phi(V \cdot x, t)$ for every state x and time t. *Used for:* Solving a linear system by changing to the basis of its modes and back. *Price:* **free**; exact.

635. Exponential stability from a quadratic Lyapunov function. Let A and P be constant real d × d matrices, and let $\Phi(x, t)$ be an evolution map of the law $x' = A\,x$: a map with $\Phi(x, 0) = x$ whose motions $t \mapsto \Phi(x, t)$ solve the law at all times, with derivatives uniform on every bounded time interval. Suppose that $\lambda \cdot \|x\|^2 \le x^{\mathsf{T}} P\,x \le \Lambda \cdot \|x\|^2$ and $x^{\mathsf{T}} (A^{\mathsf{T}} P + P A)\,x \le -\varepsilon \cdot \|x\|^2$ for every state x, where $\|x\|$ is the Euclidean length, and let $c \ge 0$ with $c \cdot \Lambda \le \varepsilon$. Then $\lambda \cdot \|\Phi(x, t)\|^2 \le \Lambda \cdot \exp(-c \cdot t) \cdot \|x\|^2$ for every state x and every time $t \ge 0$. *Used for:* Proving that an equilibrium of a linear system (or the linearization of a nonlinear one) is stable with a decay rate: control design, damped oscillators, circuits. *Price:* **free**; exact.

636. Deciding whether a linear law is stable. Deciding, for every real number a, whether no motion of the law $x' = a\,x$ ever grows (that is, $\exp(a \cdot t) \le 1$ for every time $t \ge 0$, which holds exactly when $a \le 0$) or not, is exactly the weak limited principle of omniscience. *Used for:* Classifying an equilibrium as stable or unstable from a measured rate, without a margin. *Price:* **WLPO**; exact.

637. Deciding whether a linear law is asymptotically stable. Deciding, for every real number a, whether every motion of the law $x' = a\,x$ decays to 0 (that is, for every tolerance there is a time after which $\exp(a \cdot t)$ stays below it, which holds exactly when $a < 0$) or not, is exactly the limited principle of omniscience. *Used for:* Classifying an equilibrium as attracting or not from a measured rate, without a margin. *Price:* **LPO**; exact.

638. Duhamel's formula (variation of constants). Let A be a constant real d × d matrix and $\Phi(x, t)$ an evolution map of the law $x' = A\,x$: a map with $\Phi(x, 0) = x$ whose motions $t \mapsto \Phi(x, t)$ solve the law at all times, with derivatives uniform on every bounded time interval. Let b(t) be a forcing term and x(t) a motion of the driven law $x' = A\,x + b(t)$ at all times, with its derivative uniform on every bounded time

interval. Then for every time t, x(t) = Φ(x(0), t) + (the integral from 0 to t of Φ(b(s), t − s) ds), coordinate by coordinate, where the integral is the limit of Riemann sums on [0, 1] after the substitution s = t·u. *Used for:* Driven linear systems: forced oscillators, circuits with sources, linear response of a quantum system to a field. *Price:* **free**; exact.

Q.6 Linear algebra and spectra

639. Determinant of a product. For every size n and every two n × n matrices A and B on the line: det(AB) = det(A)·det(B). *Used for:* Jacobians and volume scaling under composed linear maps; the determinant conditions defining SO(n) and SL(n) are preserved by products. *Price:* **free**; exact.

640. Determinant of the identity matrix. For every size n, the n × n identity matrix on the line has determinant 1. *Used for:* The identity transformation scales every volume by exactly 1: normalisation of Jacobians and of expansions of det(1 + εA). *Price:* **free**; exact.

641. Laplace expansion along a column. For every n ≥ 0, every (n + 1) × (n + 1) matrix A on the line and every column i: $\det(A) = \sum_k (-1)^{k+i} \cdot A_{ki} \cdot$ det(A with row k and column i deleted), the sum running over all rows k. *Used for:* Expanding secular determinants and Slater determinants by cofactors. *Price:* **free**; exact.

642. Two equal columns give determinant zero. For every size n and every n × n matrix A on the line: if two different columns of A are equal entry by entry, then det(A) = 0. *Used for:* Pauli exclusion: a Slater determinant with one orbital placed in two columns vanishes. *Price:* **free**; exact.

643. Determinant is additive in one column. For every size n, every n × n matrix A on the line, every column j and every two column vectors u and v: the determinant of A with column j replaced by u + v equals the determinant of A with column j replaced by u plus the determinant of A with column j replaced by v. *Used for:* Linearity of a Slater determinant in each orbital: a superposed orbital splits the determinant into a sum. *Price:* **free**; exact.

644. Determinant scales with one column. For every size n, every n × n matrix A on the line, every column j, every number c and every column vector u: the determinant of A with column j replaced by c·u equals c times the determinant of A with column j replaced by u. *Used for:* Stretching one axis by a factor c scales a volume, a Jacobian or a Slater determinant by c. *Price:* **free**; exact.

645. Adjugate times matrix is the determinant. For every size n and every n × n matrix A on the line: adj(A)·A = det(A)·1. Here adj(A) is the adjugate: its (i, j) entry is $(-1)^{i+j}$ times the determinant of A with row j and column i deleted. *Used for:* Closed-form inverses A^{-1} = adj(A)/det(A), and Jacobi's formula for the derivative of a determinant. *Price:* **free**; exact.

646. Matrix times adjugate is the determinant. For every size n and every n × n matrix A on the line: A·adj(A) = det(A)·1, where adj(A) is the adjugate, whose (i, j) entry is $(-1)^{i+j}$ times the determinant of A with row j and column i deleted. *Used for:* Cofactor inverses of small matrices, such as a 3 × 3 inertia tensor or metric. *Price:* **free**; exact.

647. A one-sided inverse of a square matrix is two-sided. For every size n and every two n × n matrices A and B on the line: AB = 1 exactly when BA = 1. *Used for:* Checking a rotation needs only $O^TO = 1$, since $OO^T = 1$ then follows; a left inverse found by solving equations is also a right inverse. *Price:* **free**; exact.

648. Invertibility criterion through the determinant. For every size n, an n × n matrix A of real numbers (on the line) has a two-sided inverse, a matrix B with AB = 1 and BA = 1, exactly when det(A) has an inverse, a number e with det(A)·e = 1. *Used for:* Deciding when a change of frame, a transfer matrix or a linear response map can be undone. *Price:* **free**; exact.

649. An invertible determinant gives an inverse matrix. For every size n and every n × n matrix A on the line: if det(A) has an inverse, a number e with det(A)·e = 1, then A has a two-sided inverse, a matrix B with AB = 1 and BA = 1. *Used for:* Inverting a metric, a Jacobian or a coefficient matrix once its determinant is known to be invertible. *Price:* **free**; exact.

650. A nonzero determinant gives an inverse matrix. For every size n and every n × n matrix A on the line: if det(A) ≠ 0, then A has a two-sided inverse, a matrix B with AB = 1 and BA = 1. The hypothesis says only that det(A) is not zero; no inverse of det(A) is given. *Used for:* Inverting a matrix known only to be nonsingular: normal-mode problems, inverse metrics, linear response. *Price:* **MP**; exact.

651. Cayley–Hamilton theorem for square matrices. For every size n and every n × n matrix M on the line there are numbers $c_0, c_1, \ldots, c_n$ with $c_n = 1$ such that $\det(t\cdot 1 - M) = c_0 + c_1 t + \ldots + c_n t^n$ for every number t, and $c_0\cdot 1 + c_1 M + c_2 M^2 + \ldots + c_n M^n = 0$. So

every square matrix satisfies its own characteristic polynomial. *Used for:* Writing exp(t A) or any function of a transfer matrix or spin generator as a polynomial of degree below n. *Price:* **free**; exact.

652. Determinant of a product in either order. For every size n and every two n × n matrices A and B on the line: det(AB) = det(BA). *Used for:* The volume factor of two composed linear maps, such as successive frame changes or transfer matrices, does not depend on their order. *Price:* **free**; exact.

653. Determinant is unchanged by a change of basis. For every size n, every n × n matrix M on the line with a given two-sided inverse M^{-1} ($MM^{-1} = 1$ and $M^{-1}M = 1$) and every n × n matrix N: det(M N M^{-1}) = det(N). *Used for:* The determinant of an operator does not depend on the basis: similarity transforms and changes of frame. *Price:* **free**; exact.

654. Determinant is unchanged by the inverse change of basis. For every size n, every n × n matrix M on the line with a given two-sided inverse M^{-1} ($MM^{-1} = 1$ and $M^{-1}M = 1$) and every n × n matrix N: det(M^{-1} N M) = det(N). *Used for:* Basis independence of the volume factor for an operator written in new coordinates as $M^{-1}NM$. *Price:* **free**; exact.

655. Determinant of the negative of a matrix. For every size n and every n × n matrix A on the line: $\det(-A) = (-1)^n \det(A)$. *Used for:* Spatial inversion $x \mapsto -x$ reverses orientation in odd dimensions (determinant −1 in three dimensions); sign conventions of characteristic polynomials. *Price:* **free**; exact.

656. Determinant of a diagonal matrix. For every size n and all numbers $d_1, \ldots, d_n$ on the line: the diagonal matrix with diagonal entries $d_1, \ldots, d_n$ has determinant $d_1 \cdot d_2 \cdot \ldots \cdot d_n$. *Used for:* The determinant of a diagonalised operator is the product of its eigenvalues; Gaussian integrals over normal modes. *Price:* **free**; exact.

657. Determinant of an upper triangular matrix. For every size n and every n × n matrix A on the line that is upper triangular (every entry below the diagonal is 0): det(A) is the product of the diagonal entries $A_{11} \cdot A_{22} \cdot \ldots \cdot A_{nn}$. *Used for:* Reading a determinant off a triangular (Schur or LU) form; the volume of a sheared box. *Price:* **free**; exact.

658. Determinant of a lower triangular matrix. For every size n and every n × n matrix A on the line that is lower triangular (every entry above the diagonal is 0): det(A) is the product of the diagonal entries $A_{11} \cdot A_{22} \cdot \ldots \cdot A_{nn}$. *Used for:* Jacobians of triangular changes of variables, such as the Cholesky factor of a covariance matrix. *Price:* **free**; exact.

659. Cramer's rule equals the adjugate applied to the right side. For every size n, every n × n matrix A on the line and every vector b: the Cramer vector of A and b, whose i-th entry is the determinant of A with column i replaced by b, equals adj(A)·b, where adj(A) is the adjugate (the transposed matrix of signed cofactors). *Used for:* Solving small linear systems in closed form: Kirchhoff circuit equations, coupled oscillators with a drive. *Price:* **free**; exact.

660. Cramer's rule without division. For every size n, every n × n matrix A on the line and every vector b: A applied to the Cramer vector of A and b equals det(A)·b. The Cramer vector has as its i-th entry the determinant of A with column i replaced by b. *Used for:* Circuit currents and driven normal-mode amplitudes, up to the common factor det(A). *Price:* **free**; exact.

661. Unique solution of a linear system with invertible determinant. For every size n, every n × n matrix A on the line, every vector b and every number e with det(A)·e = 1: the vector x = e·adj(A)·b solves Ax = b, and every solution of Ax = b equals it. Here adj(A) is the adjugate, the transposed matrix of signed cofactors. *Used for:* The unique response of a linear system to a source: circuit currents, forced oscillation amplitudes, linear transport equations. *Price:* **free**; exact.

662. A nonzero determinant makes a linear system solvable. For every size n, every n × n matrix A on the line and every vector b: if $\det(A) \neq 0$, then the system Ax = b has a solution x. The hypothesis says only that det(A) is not zero; no inverse of det(A) is given. *Used for:* Solving linear equations known only to be nonsingular: normal-mode amplitudes, network equations, linearised equilibrium conditions. *Price:* **MP**; exact.

663. An orthogonal matrix has determinant squaring to one. For every size n and every n × n orthogonal matrix A on the line ($A^TA = 1$ and $AA^T = 1$): det(A)·det(A) = 1. *Used for:* Rigid motions preserve volume up to orientation: |det| = 1 for every rotation or reflection. *Price:* **free**; exact.

664. Square roots of one are plus or minus one. For every real number d (on the line): d·d = 1 exactly when d = 1 or d = −1. *Used for:* Eigenvalues of parity, of a reflection or of any operator squaring to the identity are +1 or −1. *Price:* **free**; exact.

665. An orthogonal matrix has determinant plus or minus one. For every size n and every n × n orthogonal matrix A on the line ($A^TA = 1$ and $AA^T = 1$): det(A) = 1 or det(A) = −1. *Used for:* Splitting the orthogonal group into proper rotations (determinant 1) and orientation-reversing motions such as parity (determinant −1). *Price:* **free**; exact.

666. Determinant of the conjugate transpose. For every size n and every n × n matrix A with

complex entries (real and imaginary parts on the line): det(A†) is the complex conjugate of det(A), where A† is the conjugate transpose of A. *Used for:* Determinants of adjoint operators; the determinant of a Hermitian matrix is real. *Price:* **free**; exact.

667. A unitary matrix has determinant of absolute value one. For every size n and every n × n unitary matrix U with complex entries (real and imaginary parts on the line), meaning U†U = 1 and UU† = 1: conj(det(U))·det(U) = 1 and det(U)·conj(det(U)) = 1, where conj is complex conjugation. So det(U) is a phase, a complex number of absolute value 1. *Used for:* A unitary evolution multiplies volumes by a pure phase; SU(n) is the set of unitaries whose phase det(U) is 1. *Price:* **free**; exact.

668. Determinant of a block upper triangular matrix. For all block sizes m and n and matrices on the line A (m × m), B (m × n) and D (n × n): the (m + n) × (m + n) block matrix with top block row A, B and bottom block row 0, D has determinant det(A)·det(D). *Used for:* One-way coupled subsystems: the characteristic polynomial of a block triangular generator, such as a rate matrix with an absorbing sector, is the product of the blocks' characteristic polynomials. *Price:* **free**; exact.

669. Determinant of a block lower triangular matrix. For all block sizes m and n and matrices on the line A (m × m), C (n × m) and D (n × n): the (m + n) × (m + n) block matrix with top block row A, 0 and bottom block row C, D has determinant det(A)·det(D). *Used for:* Cascaded sectors that feed forward without feedback, such as a decay chain's rate matrix split into parent and daughter blocks. *Price:* **free**; exact.

670. Determinant of a block diagonal matrix. For all n and m and all m × m matrices M_1, ..., M_n on the line: the block diagonal matrix with M_1, ..., M_n along its diagonal (size nm) has determinant $\det(M_1)\cdot\det(M_2)\cdot\ldots\cdot\det(M_n)$. *Used for:* A Hamiltonian or transfer matrix that splits into independent symmetry sectors has determinant and characteristic polynomial equal to products over the sectors. *Price:* **free**; exact.

671. Determinant of a Kronecker product. For all sizes n and m, every n × n matrix A and every m × m matrix B on the line: the Kronecker product A ⊗ B, of size nm, has determinant $\det(A)^m \cdot \det(B)^n$. *Used for:* Determinants of operators on a composite system: product evolutions and product frame changes on two subsystems. *Price:* **free**; exact.

672. Strict row dominance gives a nonzero determinant. For every size n and every n × n matrix A on the line that is strictly row dominant, meaning the sum of $|A_{kj}|$ over all columns j ≠ k is less than $|A_{kk}|$ for every row k: det(A) ≠ 0. *Used for:* Unique solvability of discretised operators: implicit diffusion steps, a finite-difference Laplacian plus a positive mass term. *Price:* **free**; exact.

673. Strict column dominance gives a nonzero determinant. For every size n and every n × n matrix A on the line that is strictly column dominant, meaning the sum of $|A_{ik}|$ over all rows i ≠ k is less than $|A_{kk}|$ for every column k: det(A) ≠ 0. *Used for:* Implicit time steps of master equations, whose matrices are strictly column dominant, have a unique solution. *Price:* **free**; exact.

674. Gershgorin's circle theorem for real eigenvalues. For every size n, every n × n matrix A on the line and every real number μ that is an eigenvalue in the sense that the equation Av = μv does not force v = 0: there is a row k with $|\mu - A_{kk}|$ at most the sum of $|A_{kj}|$ over all columns j ≠ k. *Used for:* Locating energy levels: every level lies within the total hopping of some on-site energy in a tight-binding Hamiltonian. *Price:* **LLPO**; exact.

675. Spectral theorem for complex Hermitian matrices. For every size n and every n × n Hermitian matrix H with complex entries (real and imaginary parts on the line, H† = H): there are an n × n matrix V with V†V = 1 and real numbers λ_1, ..., λ_n with HV = V·diag(λ_1, ..., λ_n). So H = V·diag(λ_1, ..., λ_n)·V† with V unitary: the columns of V form an orthonormal basis of eigenvectors with real eigenvalues. *Used for:* Diagonalising a finite Hamiltonian or observable: the energy eigenbasis, real measurement outcomes, normal modes. *Price:* **LLPO + AC**$_\omega$; exact.

676. Highest-energy eigenstate of a Hermitian matrix. For every size n ≥ 1 and every n × n Hermitian matrix H with complex entries (real and imaginary parts on the line, H† = H): there are a unit vector v (⟨v, v⟩ = 1) and a real number l with Hv = l·v such that the real part of ⟨w, Hw⟩ is at most l for every unit vector w. So v is an eigenvector for the largest energy. *Used for:* The state that maximises the expectation value of a finite observable, attained at its largest eigenvalue. *Price:* **LLPO + AC**$_\omega$; exact.

677. Ground state of a Hermitian matrix. For every size n ≥ 1 and every n × n Hermitian matrix H with complex entries (real and imaginary parts on the line, H† = H): there are a unit vector v (⟨v, v⟩ = 1) and a real number l with Hv = l·v such that l is at most the real part of ⟨w, Hw⟩ for every unit vector w. So v is a ground state, an eigenvector for the lowest energy. *Used for:* The ground state of a finite quantum system and its energy, the minimiser of the variational principle. *Price:* **LLPO + AC**$_\omega$; exact.

678. Trace of a Hermitian matrix from its eigenbasis. For every size n, all n × n matrices H and V with complex entries (real and imaginary parts on

the line) and all real numbers $\lambda_1, \ldots, \lambda_n$: if $V\dagger V = 1$ and $HV = V\cdot\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, so that the k-th column of V is an eigenvector of H with eigenvalue λ_k, then the trace of H is $\lambda_1 + \ldots + \lambda_n$. *Used for:* The sum of the energy levels of a finite Hamiltonian is its trace: sum rules and the high-temperature expansion of the partition function. *Price:* **free**; exact.

679. Determinant of a Hermitian matrix from its eigenbasis. For every size n, all n × n matrices H and V with complex entries (real and imaginary parts on the line) and all real numbers $\lambda_1, \ldots, \lambda_n$: if $V\dagger V = 1$ and $HV = V\cdot\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, so that the k-th column of V is an eigenvector of H with eigenvalue λ_k, then $\det(H) = \lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_n$. *Used for:* The product of the energy levels is the determinant: Gaussian integrals and determinants of finite Hermitian operators. *Price:* **free**; exact.

680. A singular matrix has a unit null vector. For every size $n \geq 1$ and every n × n matrix A on the line with $\det(A) = 0$: there is a unit vector v ($v\cdot v = 1$) with $Av = 0$. *Used for:* Zero modes and gauge modes: a singular Hamiltonian or stiffness matrix has a normalised static mode. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

681. A nonzero null vector forces a zero determinant. For every size n, every n × n matrix A on the line and every vector $v \neq 0$ with $Av = 0$: $\det(A) = 0$. *Used for:* A zero mode, such as a gauge direction or a static solution, makes a stiffness or kinetic matrix singular. *Price:* **free**; exact.

682. Secular equation for eigenvalues. For every size $n \geq 1$, every n × n matrix A on the line and every real number l: A has a unit eigenvector for l, a vector v with $v\cdot v = 1$ and $Av = l\cdot v$, exactly when $\det(l\cdot 1 - A) = 0$. *Used for:* Finding energy levels and squared normal-mode frequencies as roots of a secular determinant, with a normalised state for each root. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

683. Eigenvalues are roots of the characteristic polynomial. For every size n, every n × n matrix A on the line, every real number l and every unit vector v ($v\cdot v = 1$) with $Av = l\cdot v$: $\det(l\cdot 1 - A) = 0$. *Used for:* Every energy found with a normalised eigenvector is a root of the secular equation $\det(E\cdot 1 - H) = 0$, as in Hückel and tight-binding calculations. *Price:* **free**; exact.

684. A singular matrix has a nonzero null vector. For every size $n \geq 1$ and every n × n matrix A on the line with $\det(A) = 0$: there is a vector $v \neq 0$ with $Av = 0$. *Used for:* Existence of a zero mode of a singular Hamiltonian, stiffness or constraint matrix. *Price:* **open**; at most LLPO + AC_ω (upper bound only).

685. Positive semidefinite: ⟨x, Ax⟩ is a nonnegative real. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: $A \succeq 0$ if and only if $A\dagger = A$ and, for every complex vector x, $\langle x, Ax\rangle$ is a nonnegative real number (its real part is ≥ 0 and its imaginary part is 0). *Used for:* Density matrices: a quantum state $\rho \succeq 0$ assigns every normalized state vector ψ a real, nonnegative weight $\langle\psi, \rho\psi\rangle$, the probability of finding the system in ψ. *Price:* **free**; exact.

686. A† A is positive semidefinite. For all sizes m and n and every complex m × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the n × n matrix $A\dagger A$ is positive semidefinite, where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Gram matrices and measurement operators: the overlap matrix of any list of state vectors ($A\dagger A$, with the vectors as columns of A) and every operator $M\dagger M$ of a quantum measurement are positive semidefinite. *Price:* **free**; exact.

687. A A† is positive semidefinite. For all sizes m and n and every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the n × n matrix $AA\dagger$ is positive semidefinite, where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Building density matrices: $\rho = \psi_1\psi_1\dagger + \ldots + \psi_m\psi_m\dagger$ for any list of (unnormalized) state vectors, and the reduced state $AA\dagger$ of a bipartite pure state with coefficient matrix A, are positive semidefinite. *Price:* **free**; exact.

688. Sum of positive semidefinite matrices. For every size n and all complex n × n matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$ and $B \succeq 0$, then $A + B \succeq 0$. *Used for:* Mixing quantum states: a sum of positive operators is positive, so a mixture $p_1\rho_1 + p_2\rho_2$ of density matrices, whose weighted terms $p_1\rho_1$ and $p_2\rho_2$ are positive, is again positive. *Price:* **free**; exact.

689. Nonnegative multiple of a positive semidefinite matrix. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number a

that is a nonnegative real (real part ≥ 0, imaginary part 0), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$, then $a \cdot A \succeq 0$. *Used for:* Weighting and normalizing states: $p\rho$ for a probability $p \geq 0$, and ρ divided by its trace, stay positive. *Price:* **free**; exact.

690. Congruence B† A B preserves positive semidefiniteness. For all sizes n and m, every complex $n \times n$ matrix A and every complex $n \times m$ matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$, then the $m \times m$ matrix $B\dagger AB \succeq 0$. *Used for:* Transforming positive operators: a positive observable or state stays positive under a change of basis or in the Heisenberg picture ($U\dagger AU$), and under restriction to a subspace (B an isometry onto it). *Price:* **free**; exact.

691. Congruence B A B† preserves positive semidefiniteness. For all sizes n and m, every complex $n \times n$ matrix A and every complex $m \times n$ matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$, then the $m \times m$ matrix $BAB\dagger \succeq 0$. *Used for:* Quantum operations: each Kraus term $K\rho K\dagger$ of a quantum channel applied to a density matrix ρ is positive, including channels between spaces of different dimension. *Price:* **free**; exact.

692. Diagonal entries of a positive semidefinite matrix are nonnegative. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$, then every diagonal entry A_{ii} is a nonnegative real number (real part ≥ 0, imaginary part 0). *Used for:* Populations: in any basis, the diagonal entries of a density matrix are the occupation probabilities of the basis states, and they are nonnegative. *Price:* **free**; exact.

693. Trace of a positive semidefinite matrix is nonnegative. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: if $A \succeq 0$, then its trace $A_{11} + \ldots + A_{nn}$ is a nonnegative real number (real part ≥ 0, imaginary part 0). *Used for:* Total probability: the trace of a density matrix or of any positive operator is a nonnegative real number; a normalized state has trace 1. *Price:* **free**; exact.

694. Identity matrix is positive semidefinite. For every size n, the complex $n \times n$ identity matrix 1 is positive semidefinite, where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* The identity is positive, so the maximally mixed state $1/n$, a positive multiple of it, is a valid density matrix; the identity is also the measurement operator of a certain outcome. *Price:* **free**; exact.

695. Zero matrix is positive semidefinite. For every size n, the complex $n \times n$ zero matrix is positive semidefinite, where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* The measurement operator of an impossible outcome, and the starting point when positive operators are summed. *Price:* **free**; exact.

696. Diagonal matrix is positive semidefinite exactly when its entries are. For every size n and all complex numbers $d_1, \ldots, d_n$ (in the complex numbers over the line), where a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$: the diagonal matrix $\mathrm{diag}(d_1, \ldots, d_n) \succeq 0$ if and only if every d_k is a nonnegative real number (real part ≥ 0, imaginary part 0). *Used for:* Classical probability inside quantum mechanics: a state diagonal in a known basis, such as a thermal state in the energy basis, $\mathrm{diag}(p_1, \ldots, p_n)$, is valid exactly when every $p_k \geq 0$. *Price:* **free**; exact.

697. Positive semidefinite exactly when the eigenvalues are nonnegative. For every size n, every complex $n \times n$ matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian ($H\dagger = H$), every complex $n \times n$ matrix V with $V\dagger V = 1$, and all real numbers $\lambda_1, \ldots, \lambda_n$ on the line with $HV = V\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ (the columns of V are an orthonormal eigenbasis of H with these eigenvalues): H is positive semidefinite if and only if $\lambda_1, \ldots, \lambda_n$ are all ≥ 0. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. The eigenbasis is given, not constructed. *Used for:* Reading positivity off the spectrum: once an orthonormal eigenbasis is known, a candidate density matrix or a Hamiltonian bounded below by zero is

checked through its eigenvalues. *Price:* **free**; exact.

698. Eigenvalues of a positive semidefinite matrix are nonnegative. For every size n, every positive semidefinite complex n × n matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), every complex n × n matrix V with V†V = 1, and all real numbers $\lambda_1, \ldots, \lambda_n$ on the line with HV = V diag($\lambda_1, \ldots, \lambda_n$): every $\lambda_k \geq 0$. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. The eigenbasis is given, not constructed. *Used for:* The eigenvalues of a density matrix are the probabilities of its pure-state components, and positive observables such as $H - E_0$ (E_0 the ground energy) have nonnegative spectrum. *Price:* **free**; exact.

699. Square root of a positive semidefinite matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$, there is a complex n × n matrix B with B† = B, $B \succeq 0$ and BB = A. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Square roots of density matrices: the fidelity $F(\rho, \sigma) = (\mathrm{tr}\, \sqrt{(\sqrt{\rho}\, \sigma\, \sqrt{\rho})})^2$ between quantum states, the purification of a mixed state, and polar decompositions of operators. *Price:* **free**; exact.

700. Positive semidefinite exactly when A = B† B. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): $A \succeq 0$ if and only if A = B†B for some complex n × n matrix B. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Every density matrix can be written ρ = B†B, the starting point of purifications and of Gram-matrix descriptions of sets of quantum states. *Price:* **free**; exact.

701. Spectral theorem for positive semidefinite matrices. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$, there are a complex n × n matrix V with V†V = 1 and real numbers $\lambda_1, \ldots, \lambda_n$ on the line, all ≥ 0, with AV = V diag($\lambda_1, \ldots, \lambda_n$). Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Spectral decomposition of a density matrix, $\rho = p_1 v_1 v_1^\dagger + \ldots + p_n v_n v_n^\dagger$ with orthonormal pure states v_k and probabilities $p_k \geq 0$, used for the von Neumann entropy $-\Sigma\, p_k \ln p_k$. *Price:* **LLPO + AC$_\omega$**; exact.

702. Zero expectation in a positive semidefinite matrix means kernel. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$ and every complex vector x: $\langle x, Ax\rangle = 0$ if and only if Ax = 0. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* A state with zero weight under a density matrix ρ ($\langle \psi, \rho\psi\rangle = 0$) is annihilated by ρ, so it is orthogonal to the support of ρ; this holds for every positive observable as well. *Price:* **free**; exact.

703. Positive semidefinite matrix of zero trace is zero. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$: the trace $A_{11} + \ldots + A_{nn}$ is 0 if and only if A = 0. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* A positive operator with zero total weight vanishes: a block of a density matrix or a measurement operator of trace zero is zero. *Price:* **free**; exact.

704. Powers of a positive semidefinite matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$: every power A^k, k = 0, 1, 2, ..., is positive semidefinite ($A^0 = 1$). Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Powers of density matrices: the purity $\mathrm{tr}\, \rho^2$ and the Rényi entropies built from $\mathrm{tr}\, \rho^k$, and powers of positive Hamiltonians. *Price:* **free**; exact.

705. Transpose of a positive semidefinite matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if $A \succeq 0$, then its transpose $A^T \succeq 0$. Here a square complex matrix A is positive semidefinite, $A \succeq 0$, when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* The transpose (equivalently the complex conjugate) of a density matrix is a valid state; the partial transpose of a two-part state need not be, which is the Peres test for entanglement. *Price:* **free**; exact.

706. Conjugate transpose of a positive semidefinite matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if $A \succeq 0$, then $A^\dagger \succeq 0$. Here a

square complex matrix A is positive semidefinite, $A \succeq 0$, when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Density matrices and positive observables are self-adjoint, so passing to the adjoint, as in the Heisenberg-picture dual, leaves them valid. *Price:* **free**; exact.

707. Trace of a sum. For every size n and all n × n matrices A and B with entries on the line: tr(A + B) = tr A + tr B, where tr is the sum of the diagonal entries. *Used for:* Expectation values add over the terms of an observable (with the distributive law): $\langle A + B\rangle = \langle A\rangle + \langle B\rangle$, such as energy as kinetic plus potential energy. *Price:* **free**; exact.

708. Trace of a difference. For every size n and all n × n matrices A and B with entries on the line: tr(A − B) = tr A − tr B, where tr is the sum of the diagonal entries. *Used for:* The expectation value of a perturbation $H_2 - H_1$ is the difference of the two expectation values, as in energy shifts between two Hamiltonians. *Price:* **free**; exact.

709. Trace of a scalar multiple. For every set of scalars S acting on the line additively (r·(a + b) = r·a + r·b and r·0 = 0; for example, multiplication by a number on the line), every r in S, every size n and every n × n matrix A with entries on the line: tr(r·A) = r·tr A, where r·A multiplies every entry by r and tr is the sum of the diagonal entries. *Used for:* Scaling an observable by a number scales its expectation value, as with a coupling constant in front of an interaction term. *Price:* **free**; exact.

710. Trace of the zero matrix. For every size n, the trace (the sum of the diagonal entries) of the n × n zero matrix with entries on the line is 0. *Used for:* The zero operator has expectation value zero in every state. *Price:* **free**; exact.

711. Trace of the transpose. For every size n and every n × n matrix A with entries on the line: $\mathrm{tr}(A^{\mathrm{T}}) = \mathrm{tr}\, A$, where A^{T} is the transpose and tr is the sum of the diagonal entries. *Used for:* Real observables and states: transposing (the adjoint over the reals) leaves traces and expectation values unchanged. *Price:* **free**; exact.

712. Trace of the identity is the dimension. For every size n, the trace (the sum of the diagonal entries) of the n × n identity matrix with entries on the line is n, the natural number n read on the line. *Used for:* The dimension of a state space as $\mathrm{tr}\, 1 = n$; the normalisation of the maximally mixed state 1/n. *Price:* **free**; exact.

713. Trace of a product is cyclic. For all natural numbers m and n, every m × n matrix A and every n × m matrix B with entries on the line: tr(AB) = tr(BA), where tr is the sum of the diagonal entries (of the m × m matrix AB and of the n × n matrix BA). *Used for:* Expectation values can be written either way round, $\mathrm{tr}(\rho A) = \mathrm{tr}(A\rho)$; traces and partition functions do not depend on the basis. *Price:* **free**; exact.

714. Trace of three factors is cyclic. For all natural numbers m, n and p and matrices with entries on the line, A of shape m × n, B of shape n × p and C of shape p × m: tr((AB)C) = tr((CA)B), where tr is the sum of the diagonal entries. *Used for:* Moving an operator around a trace of three factors, as in showing that the Schrödinger and Heisenberg pictures give equal expectation values. *Price:* **free**; exact.

715. Matrix multiplication is associative. For all natural numbers m and n, any index sets I and K, and matrices with entries on the line, L of shape I × m, M of shape m × n and N of shape n × K: (LM)N = L(MN). *Used for:* Successive linear transformations, such as evolution steps, rotations and transfer matrices, compose without regard to grouping. *Price:* **free**; exact.

716. Identity matrix is a left unit. For every natural number m, any index set K and every m × K matrix M with entries on the line: 1M = M, where 1 is the m × m identity matrix. *Used for:* The identity transformation composed with any linear map changes nothing, as with a zero-time evolution applied after an operation. *Price:* **free**; exact.

717. Identity matrix is a right unit. For every natural number n, any index set I and every I × n matrix M with entries on the line: M1 = M, where 1 is the n × n identity matrix. *Used for:* Inserting the identity, written as a complete set of basis states, between two factors leaves a product unchanged. *Price:* **free**; exact.

718. Matrix product distributes over a sum on the right. For every natural number n, any index sets I and K, every I × n matrix L and all n × K matrices M and N with entries on the line: L(M + N) = LM + LN. *Used for:* Linearity of operators: an operator applied to a sum of operators, or to a superposition of state columns, acts term by term. *Price:* **free**; exact.

719. Matrix product distributes over a sum on the left. For every natural number m, any index sets I and K, all I × m matrices L and M and every m × K matrix N with entries on the line: (L + M)N = LN + MN. *Used for:* A sum of operators acts term by term, as for a Hamiltonian H = T + V applied to states. *Price:* **free**; exact.

720. Scalar factor in the left matrix of a product. For every monoid of scalars S acting on the line as a distributive action (1·a = a, (st)·a = s·(t·a), s·(a + b) = s·a + s·b and s·0 = 0) with (s·a)b = s·(ab) for all a, b on the line (for example, multiplication by a number on the line), every s in S, every natural number m, any

index sets I and K, every I × m matrix M and every m × K matrix N with entries on the line: (s·M)N = s·(MN), where s·M multiplies every entry by s. *Used for:* Coupling constants and time steps factor out of products of operators, as in (τH)N = τ(HN) in a time-stepping scheme. *Price:* **free**; exact.

721. Scalar factor in the right matrix of a product. For every monoid of scalars S acting on the line as a distributive action (1·a = a, (st)·a = s·(t·a), s·(a + b) = s·a + s·b and s·0 = 0) with a(s·b) = s·(ab) for all a, b on the line (for example, multiplication by a number on the line), every s in S, every natural number m, any index sets I and K, every I × m matrix M and every m × K matrix N with entries on the line: M(s·N) = s·(MN), where s·N multiplies every entry by s. *Used for:* Moving numerical factors, such as normalisations, through a product of operators. *Price:* **free**; exact.

722. Transpose of a product reverses the order. For every natural number m, any index sets I and K, every I × m matrix M and every m × K matrix N with entries on the line: $(MN)^{T} = N^{T}M^{T}$, where ^{T} is the transpose. *Used for:* Undoing a sequence of rotations: for orthogonal matrices, $(R_1R_2)^{T} = R_2{}^{T}R_1{}^{T}$ applies the inverse steps in reverse order; adjoints of products of real operators. *Price:* **free**; exact.

723. Successive matrix-vector products. For all natural numbers n and o, any index set I, every I × n matrix M, every n × o matrix N and every vector v of length o with entries on the line: M(Nv) = (MN)v, where Xv is the matrix-vector product, $(Xv)(i) = \Sigma_j X(i, j)v(j)$. *Used for:* Applying two transformations to a state vector in turn equals applying their product once, as for successive evolution steps or rotations. *Price:* **free**; exact.

724. Matrix times a sum of vectors. For every natural number n, any index set I, every I × n matrix A and all vectors x and y of length n with entries on the line: A(x + y) = Ax + Ay, where $(Av)(i) = \Sigma_j A(i, j)v(j)$. *Used for:* Superposition: a linear evolution or a linear response acts on a sum of states or of inputs term by term. *Price:* **free**; exact.

725. Dot product is symmetric. For every length n and all vectors v and w of length n with entries on the line: v · w = w · v, where $v \cdot w = \Sigma_i v(i)w(i)$. *Used for:* Symmetry of the Euclidean inner product: the work F · d of a force along a displacement, or a projection, can be computed either way round. *Price:* **free**; exact.

726. Trace of three factors is cyclic, right-grouped. For all natural numbers m, n and p and matrices with entries on the line, A of shape m × n, B of shape n × p and C of shape p × m: tr(A(BC)) = tr(C(AB)), where tr is the sum of the diagonal entries. *Used for:* Rearranging traces of operator products in perturbation theory and linear response, such as tr(ρ(HA)) = tr(A(ρH)). *Price:* **free**; exact.

727. Trace of a product of transposes. For all natural numbers m and n, every m × n matrix A and every n × m matrix B with entries on the line: $\mathrm{tr}(A^{T}B^{T}) = \mathrm{tr}(AB)$, where ^{T} is the transpose and tr is the sum of the diagonal entries. *Used for:* Trace formulas for real operators do not depend on the row-vector or column-vector convention: transposing every factor keeps the trace. *Price:* **free**; exact.

728. Product with a zero matrix on the right. For every natural number n, any index sets I and K and every I × n matrix M with entries on the line: M0 = 0, where the first 0 is the zero n × K matrix and the second is the zero I × K matrix. *Used for:* Any operator applied after the zero operator still gives zero: a switched-off coupling contributes nothing to a product of operators. *Price:* **free**; exact.

729. Product with a zero matrix on the left. For every natural number m, any index sets I and K and every m × K matrix M with entries on the line: 0M = 0, where the first 0 is the zero I × m matrix and the second is the zero I × K matrix. *Used for:* An operator that sends every state to zero composes with any operator to the zero operator. *Price:* **free**; exact.

730. Matrix times the zero vector. For every natural number n, any index set I and every I × n matrix A with entries on the line: A0 = 0, that is, A sends the zero vector of length n to the zero vector indexed by I, where $(Av)(i) = \Sigma_j A(i, j)v(j)$. *Used for:* Linear response: no input (zero displacement, zero applied field) gives no output. *Price:* **free**; exact.

731. Sum of matrices times a vector. For every natural number n, any index set I, all I × n matrices A and B and every vector x of length n with entries on the line: (A + B)x = Ax + Bx, where $(Av)(i) = \Sigma_j A(i, j)v(j)$. *Used for:* Several couplings acting on one state add their effects: $(H_1 + H_2)\psi = H_1\psi + H_2\psi$. *Price:* **free**; exact.

732. Matrix times a scaled vector. For every set of scalars S acting on the line additively (s·(a + b) = s·a + s·b and s·0 = 0) with a(s·b) = s·(ab) for all a, b on the line (for example, multiplication by a number on the line), every s in S, every natural number n, any index set I, every I × n matrix M and every vector v of length n with entries on the line: M(sv) = s(Mv), where sv multiplies every entry by s and $(Mv)(i) = \Sigma_j M(i, j)v(j)$. *Used for:* Linear response: scaling the input, such as an applied field or a displacement, scales the output; the overall normalisation of a state passes through an operator. *Price:* **free**; exact.

733. Identity matrix times a vector. For every length m and every vector v of length m with entries on the line: 1v = v, where 1 is the m × m identity matrix and $(Av)(i) = \Sigma_j A(i, j)v(j)$. *Used for:* The

identity evolution leaves every state vector unchanged. *Price:* **free**; exact.

734. Dot product is additive in the first vector. For every length n and all vectors u, v and w of length n with entries on the line: $(u + v) \cdot w = u \cdot w + v \cdot w$, where $x \cdot y = \Sigma_i\, x(i)y(i)$. *Used for:* The work done by a sum of forces is the sum of their works, $(F_1 + F_2) \cdot d$. *Price:* **free**; exact.

735. Dot product is additive in the second vector. For every length n and all vectors u, v and w of length n with entries on the line: $u \cdot (v + w) = u \cdot v + u \cdot w$, where $x \cdot y = \Sigma_i\, x(i)y(i)$. *Used for:* The work done by a force along two successive displacements adds, $F \cdot (d_1 + d_2) = F \cdot d_1 + F \cdot d_2$. *Price:* **free**; exact.

736. Scalar factor in the first vector of a dot product. For every set of scalars S acting on the line additively ($s\cdot(a + b) = s\cdot a + s\cdot b$ and $s\cdot 0 = 0$) with $(s\cdot a)b = s\cdot(ab)$ for all a, b on the line (for example, multiplication by a number on the line), every s in S, every length n and all vectors v and w of length n with entries on the line: $(sv) \cdot w = s(v \cdot w)$, where sv multiplies every entry by s and $x \cdot y = \Sigma_i\, x(i)y(i)$. *Used for:* Scaling a force scales the work it does and the power it delivers. *Price:* **free**; exact.

737. Scalar factor in the second vector of a dot product. For every set of scalars S acting on the line additively ($s\cdot(a + b) = s\cdot a + s\cdot b$ and $s\cdot 0 = 0$) with $a(s\cdot b) = s\cdot(ab)$ for all a, b on the line (for example, multiplication by a number on the line), every s in S, every length n and all vectors v and w of length n with entries on the line: $v \cdot (sw) = s(v \cdot w)$, where sw multiplies every entry by s and $x \cdot y = \Sigma_i\, x(i)y(i)$. *Used for:* Scaling a displacement or a velocity scales the projection of a force onto it, as in the power $F \cdot v$. *Price:* **free**; exact.

738. Transpose of the identity matrix. For every size n, the transpose of the n × n identity matrix with entries on the line is the identity matrix: $1^{\mathrm{T}} = 1$. *Used for:* The identity operator is its own adjoint over the reals, so it is a valid real observable; the Euclidean metric, the identity matrix, is symmetric. *Price:* **free**; exact.

739. Trace of the negative of a matrix. For every size n and every n × n matrix A with entries on the line: $\mathrm{tr}(-A) = -\mathrm{tr}\, A$, where tr is the sum of the diagonal entries. *Used for:* Reversing the sign of an observable, such as the coupling to an applied field, reverses its expectation value. *Price:* **free**; exact.

740. Trace of an outer product is a dot product. For every length n and all vectors a and b of length n with entries on the line: the trace of the outer product ab^{T} (the n × n matrix with entry a(i)b(j)) equals $a \cdot b = \Sigma_i\, a(i)b(i)$. *Used for:* Pure-state normalisation: for a real state vector v, $\mathrm{tr}(vv^{\mathrm{T}}) = v \cdot v$, so a unit vector gives a density matrix of trace 1. *Price:* **free**; exact.

741. Trace is unchanged by a change of basis. For every size m, all m × m matrices M and W with entries on the line such that MW = 1 and WM = 1 (W is the inverse of M), and every m × m matrix N: $\mathrm{tr}((MN)W) = \mathrm{tr}\, N$, where tr is the sum of the diagonal entries. *Used for:* Basis independence: a change of basis $N \mapsto MNM^{-1}$ keeps traces, partition functions and expectation values. *Price:* **free**; exact.

742. Trace is unchanged by the inverse change of basis. For every size m, all m × m matrices M and W with entries on the line such that MW = 1 and WM = 1 (W is the inverse of M), and every m × m matrix N: $\mathrm{tr}((WN)M) = \mathrm{tr}\, N$, where tr is the sum of the diagonal entries. *Used for:* Traces agree whichever way a basis change is written, $N \mapsto M^{-1}NM$; diagonalising a matrix keeps its trace, so the trace is the sum of the eigenvalues on the diagonal. *Price:* **free**; exact.

743. Only the zero vector has zero squared length. For every length n and every vector v of length n with entries on the line: $v \cdot v = 0$ exactly when v = 0, where $v \cdot v = \Sigma_i\, v(i)^2$. *Used for:* Positive definiteness of the Euclidean inner product: a displacement or a field configuration with zero squared length is zero. *Price:* **free**; exact.

744. Trace of a commutator is zero. For every size n and all n × n matrices X and Y with entries on the line: $\mathrm{tr}(XY - YX) = 0$, where tr is the sum of the diagonal entries. *Used for:* No matrices of any finite size $n \geq 1$ satisfy a canonical commutation relation $XP - PX = c\cdot 1$ with $c \neq 0$ (the left side has trace 0, the right side $n\cdot c$), which is why position and momentum need an infinite-dimensional state space. *Price:* **free**; exact.

745. Moving a matrix across a dot product. For all natural numbers m and n, every vector v of length m, every m × n matrix A and every vector w of length n with entries on the line: $v \cdot (Aw) = (v\, A) \cdot w$, where $(Aw)(i) = \Sigma_j\, A(i, j)w(j)$, $(v\, A)(j) = \Sigma_i\, v(i)A(i, j)$ and $x \cdot y = \Sigma_k\, x(k)y(k)$. *Used for:* Matrix elements $v^{\mathrm{T}}Aw$ of an operator or of a bilinear form can be evaluated from either side, as for the energy $x^{\mathrm{T}}Ky$ of a network of springs. *Price:* **free**; exact.

746. Successive row-vector products. For all natural numbers m and n, any index set K, every vector v of length m, every m × n matrix M and every n × K matrix N with entries on the line: $(v\, M)\, N = v\, (M\, N)$, where the row-vector product is $(v\, X)(j) = \Sigma_i\, v(i)X(i, j)$. *Used for:* Probability row vectors of a Markov chain: two steps taken in turn equal one step with the product of the transition matrices, $(p\, P)\, Q = p\, (P\, Q)$. *Price:* **free**; exact.

747. Transposed matrix times a vector. For every natural number m, any index set K, every m × K matrix A and every vector x of length m with entries

on the line: $A^{T}x = x\,A$, where $(A^{T}x)(j) = \Sigma_i\, A(i, j)x(i)$ is the matrix-vector product with the transpose and $(x\,A)(j) = \Sigma_i\, x(i)A(i, j)$ is the row-vector product. *Used for:* Switching between row and column conventions: a probability distribution p evolved by a stochastic matrix P as the row product p P equals the column product $P^{T}p$. *Price:* **free**; exact.

748. Vector times a transposed matrix. For every natural number n, any index set I, every I × n matrix A and every vector x of length n with entries on the line: $xA^{T} = Ax$, where $(xA^{T})(i) = \Sigma_j\, x(j)A(i, j)$ is the row-vector product with the transpose and $(Ax)(i) = \Sigma_j\, A(i, j)x(j)$. *Used for:* Writing the action of a real linear map on row vectors (covectors such as gradients) through the transpose, and converting back to the column product. *Price:* **free**; exact.

749. Row vector times a sum of matrices. For every natural number m, any index set K, all m × K matrices A and B and every vector x of length m with entries on the line: $x(A + B) = x\,A + x\,B$, where $(x\,A)(j) = \Sigma_i\, x(i)A(i, j)$. *Used for:* A distribution driven by a generator that is a sum of processes changes term by term: the rate matrices of a master equation add. *Price:* **free**; exact.

750. Sum of row vectors times a matrix. For every natural number m, any index set K, every m × K matrix A and all vectors x and y of length m with entries on the line: $(x + y)A = x\,A + y\,A$, where $(x\,A)(j) = \Sigma_i\, x(i)A(i, j)$. *Used for:* Mixtures evolve term by term: a weighted combination of probability distributions under one Markov step. *Price:* **free**; exact.

751. Row vector times the identity matrix. For every length m and every vector v of length m with entries on the line: $v1 = v$, where 1 is the m × m identity matrix and $(v\,A)(j) = \Sigma_i\, v(i)A(i, j)$. *Used for:* A trivial Markov step (the identity transition matrix) leaves every distribution unchanged. *Price:* **free**; exact.

752. Row vector times the zero matrix. For every natural number m, any index set K and every vector v of length m with entries on the line: $v0 = 0$, where 0 is the zero m × K matrix, the result is the zero vector indexed by K, and $(v\,A)(j) = \Sigma_i\, v(i)A(i, j)$. *Used for:* A zero coupling or transfer matrix sends every row vector, such as a source distribution, to zero. *Price:* **free**; exact.

753. Row vector times an outer product. For every natural number m, any index set K, all vectors u and v of length m and every vector w indexed by K with entries on the line: $u(vw^{T}) = (u \cdot v)w$, where vw^{T} is the outer product with entry $v(i)w(j)$, $(u\,X)(j) = \Sigma_i\, u(i)X(i, j)$ and $u \cdot v = \Sigma_i\, u(i)v(i)$. *Used for:* A rank-one operator vw^{T} acting on a vector gives the overlap times the target vector, as for a real projector vv^{T} onto a unit vector v. *Price:* **free**; exact.

754. Outer product of a sum of vectors. For any index sets I and K, all vectors w_1 and w_2 indexed by I and every vector v indexed by K with entries on the line: $(w_1 + w_2)v^{T} = w_1v^{T} + w_2v^{T}$, where xv^{T} is the outer product with entry $x(i)v(j)$. *Used for:* Expanding dyadic products of sums, as for inertia and stress tensors or real density matrices built from superposed vectors. *Price:* **free**; exact.

755. Trace of a 1 × 1 matrix. For every 1 × 1 matrix A with its entry on the line, the trace of A is its single entry. *Used for:* A one-level system: an operator is a single number, such as one energy level, and its trace is that number. *Price:* **free**; exact.

756. Trace of a 2 × 2 matrix. For every 2 × 2 matrix A with entries on the line: $\operatorname{tr} A = A_{11} + A_{22}$, the sum of its two diagonal entries. *Used for:* Two-level systems (spin ½, qubits): the trace of a 2 × 2 operator, such as the normalisation $p_1 + p_2 = 1$ of the two diagonal populations of a qubit. *Price:* **free**; exact.

757. Trace of a 3 × 3 matrix. For every 3 × 3 matrix A with entries on the line: $\operatorname{tr} A = A_{11} + A_{22} + A_{33}$, the sum of its three diagonal entries. *Used for:* Three-dimensional space and three-level systems: the pressure is minus one third of the trace of the stress tensor, and a rotation by angle θ has trace $1 + 2\cos\theta$. *Price:* **free**; exact.

758. Conjugate transpose of a product reverses the order. For every natural number m, any index sets I and K, every I × m matrix A and every m × K matrix B with complex entries (on the complex numbers over the line): $(AB)\dagger = B\dagger A\dagger$, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* Undoing a sequence of quantum gates or evolutions: $(UV)\dagger = V\dagger U\dagger$ applies the inverses in reverse order. *Price:* **free**; exact.

759. Conjugate transpose taken twice. For any index sets I and K and every I × K matrix X with complex entries (on the complex numbers over the line): $(X\dagger)\dagger = X$, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* Taking the adjoint twice returns the operator; turning a column state into a row state and back returns the state. *Price:* **free**; exact.

760. Trace of the conjugate transpose. For every size n and every n × n matrix A with complex entries (on the complex numbers over the line): $\operatorname{tr}(A\dagger) = \overline{\operatorname{tr} A}$, where A† is the conjugate transpose (its entry at (i, j) is the complex conjugate of A(j, i)), conj is complex conjugation and tr is the sum of the diagonal entries. *Used for:* The trace of a Hermitian operator (A† = A) equals its own complex conjugate,

so it is real; complex conjugates of traces of operator products. *Price:* **free**; exact.

761. A matrix times its conjugate transpose is Hermitian. For every natural number m, any index set I and every I × m matrix A with complex entries (on the complex numbers over the line): (AA†)† = AA†, that is, AA† is Hermitian, where A† is the conjugate transpose (its entry at (i, j) is the complex conjugate of A(j, i)). *Used for:* The reduced density matrix ρ = AA† of a two-part pure state, with A the matrix of its amplitudes, is Hermitian. *Price:* **free**; exact.

762. Conjugate transpose times the matrix is Hermitian. For every natural number m, any index set K and every m × K matrix A with complex entries (on the complex numbers over the line): (A†A)† = A†A, that is, A†A is Hermitian, where A† is the conjugate transpose (its entry at (i, j) is the complex conjugate of A(j, i)). *Used for:* Operators of the form A†A are Hermitian: number operators a†a, measurement operators K†K built from Kraus operators, Gram matrices of state vectors. *Price:* **free**; exact.

763. Sum of Hermitian matrices is Hermitian. For any index set I and all I × I matrices A and B with complex entries (on the complex numbers over the line): if A† = A and B† = B, then (A + B)† = A + B, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* A Hamiltonian that is a sum of Hermitian terms (kinetic, potential and interaction energy) is Hermitian. *Price:* **free**; exact.

764. Identity matrix is Hermitian. For every size n, the n × n identity matrix with complex entries (on the complex numbers over the line) equals its own conjugate transpose: 1† = 1, where X† has entry at (i, j) the complex conjugate of X(j, i). *Used for:* The identity is an observable, reading 1 in every state; it is also the trivial unitary evolution. *Price:* **free**; exact.

765. Conjugate transpose of a sum. For any index sets I and K and all I × K matrices A and B with complex entries (on the complex numbers over the line): (A + B)† = A† + B†, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* Hermitian parts of operators: A + A† is Hermitian, which builds observables such as a + a† from non-Hermitian operators. *Price:* **free**; exact.

766. Conjugate transpose of the zero matrix. For any index sets I and K, the conjugate transpose of the zero I × K matrix with complex entries (on the complex numbers over the line) is the zero K × I matrix. *Used for:* The zero operator is Hermitian, the trivial observable; a switched-off coupling stays zero under adjoints. *Price:* **free**; exact.

767. Conjugate transpose of the negative of a matrix. For any index sets I and K and every I × K matrix A with complex entries (on the complex numbers over the line): (−A)† = −A†, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* Anti-Hermitian generators of unitary evolution (A† = −A): sign changes pass through the adjoint. *Price:* **free**; exact.

768. Conjugate transpose of a difference. For any index sets I and K and all I × K matrices A and B with complex entries (on the complex numbers over the line): (A − B)† = A† − B†, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* Commutators: the adjoint of AB − BA is B†A† − A†B†, so the commutator of two Hermitian observables is anti-Hermitian. *Price:* **free**; exact.

769. Conjugate transpose of a complex multiple. For every complex number c (on the complex numbers over the line), any index sets I and K and every I × K matrix A with complex entries: (c A)† = $\bar{c}$ A†, where c A multiplies every entry by c, $\bar{c}$ is the complex conjugate of c, and A† is the conjugate transpose (its entry at (i, j) is the complex conjugate of A(j, i)). *Used for:* The generator −i H of unitary time evolution is anti-Hermitian when H is Hermitian, since (−i H)† = i H. *Price:* **free**; exact.

770. Conjugating a Hermitian matrix as B†AB keeps it Hermitian. For every natural number m, any index set K, every m × m matrix A and every m × K matrix B with complex entries (on the complex numbers over the line): if A† = A, then (B†AB)† = B†AB, with the product grouped as (B†A)B, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* An observable stays Hermitian under a change of basis or a projection: U†HU in the Heisenberg picture, effective Hamiltonians P†HP on a subspace. *Price:* **free**; exact.

771. Conjugating a Hermitian matrix as BAB† keeps it Hermitian. For every natural number m, any index set I, every m × m matrix A and every I × m matrix B with complex entries (on the complex numbers over the line): if A† = A, then (BAB†)† = BAB†, with the product grouped as (BA)B†, where X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* A density matrix stays Hermitian under unitary evolution ρ ↦ UρU† and under a Kraus operation ρ ↦ KρK†. *Price:* **free**; exact.

772. Conjugate transpose of an outer product. For any index sets I and K, every vector w indexed by I and every vector v indexed by K with complex entries (on the complex numbers over the line): $(wv^{T})^{\dagger} = \bar{v}\bar{w}^{T}$, where xy^{T} is the outer product with entry x(i)y(j), the bar conjugates every entry, and X†

is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)). *Used for:* The adjoint of a transition operator from one state to another is the reverse transition (raising and lowering operators); a pure-state projector $\psi\psi\dagger$ is Hermitian. *Price:* **free**; exact.

773. A complex vector of zero norm is zero. For every length n and every vector v of length n with complex entries (on the complex numbers over the line): Σ_i conj(v(i))·v(i) = 0 exactly when v = 0, where conj is complex conjugation. *Used for:* Positive definiteness of the quantum inner product: the only state vector of zero norm is the zero vector. *Price:* **free**; exact.

774. Mixed product of Kronecker products. For all natural numbers k and k′, all index sets I, J, I′, J′ and all matrices with entries on the line, A of shape I × k, B of shape k × J, A′ of shape I′ × k′ and B′ of shape k′ × J′: (AB) ⊗ (A′B′) = (A ⊗ A′)(B ⊗ B′). Here X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′), and the product on the right sums over all k·k′ index pairs. *Used for:* Composite systems: independent operations on two subsystems compose part by part, $(U_1 \otimes V_1)(U_2 \otimes V_2) = U_1U_2 \otimes V_1V_2$, as in two-qubit gate circuits and in the transition matrices of two independent Markov chains. *Price:* **free**; exact.

775. Trace of a Kronecker product. For all sizes m and n, every m × m matrix A and every n × n matrix B with entries on the line: tr(A ⊗ B) = tr A · tr B. Here A ⊗ B is the Kronecker product, with entry A(i, j)·B(i′, j′) at row (i, i′) and column (j, j′), and each trace is the sum of the diagonal entries (m·n of them for A ⊗ B). *Used for:* Product observables in product states: with the mixed-product rule, tr((ρ ⊗ σ)(A ⊗ B)) = tr(ρA)·tr(σB); the dimension of a composite system is the product of the dimensions of its parts. *Price:* **free**; exact.

776. Kronecker product of two identity matrices. For any two index sets I and J (finite or not) on which equality of indices is decidable: the Kronecker product of the identity matrix on I with the identity matrix on J, entries on the line, is the identity matrix on the pairs (i, j), that is 1 ⊗ 1 = 1. The Kronecker product X ⊗ Y has entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Doing nothing to each part does nothing to the whole: the identity operator of a composite quantum system is the product of the identities of its parts. *Price:* **free**; exact.

777. Kronecker product of diagonal matrices. For any index sets I and J on which equality of indices is decidable, every family a of numbers on the line indexed by I and every family b indexed by J: diag(a) ⊗ diag(b) = diag(c), where c(i, j) = a(i)·b(j). Here diag(a) is the diagonal matrix with a(i) at position (i, i) and zeros elsewhere, and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Independent subsystems: diagonal density matrices with probabilities p(i) and q(j) combine to the diagonal matrix of joint probabilities p(i)q(j); the Boltzmann weights of non-interacting parts multiply. *Price:* **free**; exact.

778. Kronecker product with the identity on the right. For any index sets I and J, any index set O on which equality of indices is decidable, and every I × J matrix A with entries on the line: A ⊗ 1, with 1 the identity matrix on O, is the block-diagonal matrix with one copy of A in each diagonal block. Entry by entry: A ⊗ 1 has entry A(i, j) at row (i, o) and column (j, o′) when o = o′, and 0 when o ≠ o′. The Kronecker product X ⊗ Y has entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Local operators: an observable or gate acting on the first subsystem alone, A ⊗ 1, written as a block-diagonal matrix with one block per state of the second subsystem. *Price:* **free**; exact.

779. Kronecker product with the identity on the left. For any index sets I and J, any index set O on which equality of indices is decidable, and every I × J matrix B with entries on the line: 1 ⊗ B, with 1 the identity matrix on O, is the block-diagonal matrix with one copy of B in each diagonal block after the two parts of every row pair and every column pair are swapped. Entry by entry: 1 ⊗ B has at row (o, i) and column (o′, j) the entry of that block-diagonal matrix at row (i, o) and column (j, o′), which is B(i, j) when o = o′ and 0 when o ≠ o′. The Kronecker product X ⊗ Y has entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Local operators on the second subsystem, 1 ⊗ B, such as a single-qubit gate applied to the second qubit of a two-qubit register. *Price:* **free**; exact.

780. Transpose of a Kronecker product. For any index sets I, J, I′, J′, every I × J matrix A and every I′ × J′ matrix B with entries on the line: $A^T \otimes B^T = (A \otimes B)^T$, where ^{T} is the transpose and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Adjoints of real operators on composite systems (over the reals the adjoint is the transpose), for example undoing real two-qubit gates or local rotations factor by factor. *Price:* **free**; exact.

781. Kronecker product distributes over a sum on the left. For any index sets I, J, I′, J′, all I × J matrices A_1 and A_2 and every I′ × J′ matrix B with entries on the line: $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$. Here X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Linearity in the first factor: an operator on one subsystem that is a sum of terms gives a composite operator that splits term by term, $(H_1 + H_2) \otimes 1 = H_1 \otimes 1 + H_2 \otimes 1$. *Price:* **free**; exact.

782. Kronecker product distributes over a sum on the right. For any index sets I, J, I′, J′, every I × J matrix A and all I′ × J′ matrices B_1 and B_2 with entries on the line: $A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$. Here X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Linearity in the second factor: expanding an operator on a composite system into a sum of product terms, as in decompositions of two-qubit operators into Pauli products. *Price:* **free**; exact.

783. Kronecker product is associative. For any index sets I, J, I′, J′, I″, J″ and matrices with entries on the line, A of shape I × J, B of shape I′ × J′ and C of shape I″ × J″: (A ⊗ B) ⊗ C and A ⊗ (B ⊗ C) agree once every index triple is regrouped, that is, for all indices the entry of (A ⊗ B) ⊗ C at row ((i, i′), i″) and column ((j, j′), j″) equals the entry of A ⊗ (B ⊗ C) at row (i, (i′, i″)) and column (j, (j′, j″)). Here X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Systems of three or more parts: the operator on a multi-part system does not depend on how the parts are grouped, as for qubit registers and spin chains. *Price:* **free**; exact.

784. Kronecker product with a zero matrix on the left. For any index sets I, J, I′, J′ and every I′ × J′ matrix B with entries on the line: 0 ⊗ B = 0, where the first 0 is the zero I × J matrix, the result is the zero matrix on the index pairs, and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* A vanishing operator on the first subsystem makes every product term containing it vanish: zero coupling terms drop out of a composite Hamiltonian. *Price:* **free**; exact.

785. Kronecker product with a zero matrix on the right. For any index sets I, J, I′, J′ and every I × J matrix A with entries on the line: A ⊗ 0 = 0, where the first 0 is the zero I′ × J′ matrix, the result is the zero matrix on the index pairs, and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Switching off one side of a coupling A ⊗ B removes the whole product term from a composite operator. *Price:* **free**; exact.

786. Scalar factor in the first Kronecker factor. For every set of scalars S acting on the line with (r·a)·b = r·(a·b) for all r in S and all a, b on the line (for example, multiplication by a number on the line), every r in S, any index sets I, J, I′, J′, every I × J matrix A and every I′ × J′ matrix B with entries on the line: (r A) ⊗ B = r(A ⊗ B), where r A multiplies every entry by r and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Coupling constants in front of product operators: (J A) ⊗ B = J (A ⊗ B) for an exchange coupling of strength J between two spins. *Price:* **free**; exact.

787. Scalar factor in the second Kronecker factor. For every set of scalars S acting on the line with a·(r·b) = r·(a·b) for all r in S and all a, b on the line (for example, multiplication by a number on the line), every r in S, any index sets I, J, I′, J′, every I × J matrix A and every I′ × J′ matrix B with entries on the line: A ⊗ (r B) = r(A ⊗ B), where r B multiplies every entry by r and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Moving a coupling strength or a normalisation factor out of the second factor of a product operator on a composite system. *Price:* **free**; exact.

788. Conjugate transpose of a Kronecker product. For any index sets I, J, I′, J′, every I × J matrix A and every I′ × J′ matrix B with complex entries (on the complex numbers over the line): (A ⊗ B)† = A† ⊗ B†. Here X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)) and X ⊗ Y is the Kronecker product, with entry X(i, j)·Y(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Adjoints of operators on composite quantum systems; a product A ⊗ B of Hermitian observables on two subsystems is Hermitian. *Price:* **free**; exact.

789. Kronecker product of unitary matrices is unitary. For all sizes n and m, every n × n matrix U and every m × m matrix V with complex entries (on the complex numbers over the line): if U†U = 1 and UU† = 1, and V†V = 1 and VV† = 1, then (U ⊗ V)†(U ⊗ V) = 1 and (U ⊗ V)(U ⊗ V)† = 1, where 1 is the identity matrix on the n·m index pairs. Here X† is the conjugate transpose (its entry at (i, j) is the complex conjugate of X(j, i)) and U ⊗ V is the Kronecker product, with entry U(i, j)·V(i′, j′) at row (i, i′) and column (j, j′). *Used for:* Independent closed evolutions of two quantum systems form a closed evolution of the pair; gates applied in parallel to a quantum register are unitary. *Price:* **free**; exact.

790. Exponential series of a matrix converges, with a rate. For every size n and every complex n × n matrix X (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), the exponential series $\Sigma_m X^m/m!$ converges entrywise to a matrix E, with an explicit rate: there is a function M on the natural numbers such that for every j and every N ≥ M(j), each entry of the partial sum $X^0/0! + X^1/1! + \ldots + X^{N-1}/(N-1)!$ is within 1/(j+1) of the corresponding entry of E, in its real part and in its imaginary part. One rate M serves every entry. Convergence is read as the limit of the partial sums taken in order. *Used for:* Computing the propagator $e^{-i\,t\,H}$ of a finite quantum system, or the solution matrix $e^{t\,A}$ of linear equations dx/dt = Ax, by truncating the series with a guaranteed error at each stage. *Price:* **free**; exact.

791. Matrix exponential is well defined. For

every size n and every complex n × n matrix X (entries in the complex numbers over the line), there is exactly one matrix E such that the exponential series Σ_m $X^m/m!$ converges entrywise to E. So exp X is a single, definite matrix. *Used for:* Gives the propagator $e^{-i\,t\,H}$ and the Gibbs factor $e^{-\beta\,H}$ one definite value, so every identity about them refers to a single matrix. *Price:* **free**; exact.

792. Exponential of the zero matrix is the identity. For every size n and every complex n × n matrix E (entries in the complex numbers over the line): if the exponential series Σ_m $0^m/m!$ of the zero matrix converges entrywise to E, then E is the identity matrix. That is, exp 0 = 1. *Used for:* The initial condition of quantum time evolution: at time zero the propagator is the identity, $U(0) = e^0 = 1$. *Price:* **free**; exact.

793. Exponential of the conjugate transpose. For every size n and all complex n × n matrices X and E (entries in the complex numbers over the line): if the exponential series Σ_m $X^m/m!$ converges entrywise to E, then the exponential series of the conjugate transpose X† converges entrywise to E†. That is, exp(X†) = (exp X)†. *Used for:* Taking the adjoint of a propagator, $(e^{-i\,t\,H})^\dagger = e^{i\,t\,H}$ for Hermitian H, used to evolve bras and to write observables in the Heisenberg picture, A(t) = U† A U. *Price:* **free**; exact.

794. Exponential of a Hermitian matrix is Hermitian. For every size n and all complex n × n matrices H and E (entries in the complex numbers over the line): if H is Hermitian (H† = H) and the exponential series Σ_m $H^m/m!$ converges entrywise to E, then E is Hermitian (E† = E). *Used for:* Thermal states: for a Hamiltonian H and real β the Gibbs factor $e^{-\beta H}$ is Hermitian, so the density matrix $e^{-\beta H}/Z$ of a system in equilibrium is Hermitian. *Price:* **free**; exact.

795. Exponential of minus a matrix is its inverse. For every size n and all complex n × n matrices X, E and F (entries in the complex numbers over the line): if the exponential series of X converges entrywise to E and the exponential series of –X converges entrywise to F, then EF = 1 and FE = 1. So exp(–X) is the two-sided inverse of exp X. *Used for:* Running quantum evolution backward: $e^{i\,t\,H}$ undoes $e^{-i\,t\,H}$, so the state at time t determines the state at time zero. *Price:* **free**; exact.

796. Matrix exponential is invertible. For every size n and all complex n × n matrices X and E (entries in the complex numbers over the line): if the exponential series Σ_m $X^m/m!$ converges entrywise to E, then E has a two-sided inverse: there is a matrix F with EF = 1 and FE = 1. *Used for:* Every propagator $e^{-i\,t\,H}$ and every fundamental matrix $e^{t\,A}$ of a linear system dx/dt = Ax is invertible, so solutions can be run forward and backward in time. *Price:* **free**; exact.

797. Exponential commutes with what commutes with its argument. For every size n and all complex n × n matrices X, Q and E (entries in the complex numbers over the line): if XQ = QX and the exponential series Σ_m $X^m/m!$ converges entrywise to E, then EQ = QE. So exp X commutes with every matrix that commutes with X. *Used for:* Conservation laws in quantum mechanics: an observable Q that commutes with the Hamiltonian H commutes with the evolution $e^{-i\,t\,H}$, so its expectation value stays constant in time. *Price:* **free**; exact.

798. Exponentials of commuting matrices commute. For every size n and all complex n × n matrices X, Y, E and F (entries in the complex numbers over the line): if XY = YX, the exponential series of X converges entrywise to E and the exponential series of Y converges entrywise to F, then EF = FE. *Used for:* Commuting symmetry operations: a rotation $e^{-i\,\theta\,Jz}$ about the z axis commutes with the time evolution $e^{-i\,t\,H}$ when H commutes with the angular momentum component Jz, and commuting gauge transformations can be applied in either order. *Price:* **free**; exact.

799. Exponential of –i t H is unitary for Hermitian H. For every size n, every Hermitian complex n × n matrix H (H† = H, entries in the complex numbers over the line) and every real number t on the line (the real numbers constructed from distinctions): if the exponential series of –i t H converges entrywise to U, then U†U = 1 and UU† = 1. That is, $e^{-i\,t\,H}$ is unitary. *Used for:* Conservation of probability: the time evolution $e^{-i\,t\,H}$ of a finite quantum system (qubits, spins, few-level atoms) preserves the norm of every state vector. *Price:* **free**; exact.

800. Unitary time evolution of a finite quantum system. For every size n and every Hermitian complex n × n matrix H (H† = H, entries in the complex numbers over the line) there is a function U from the real numbers on the line to complex n × n matrices such that: for every real t the exponential series of –i t H converges entrywise to U(t); U(0) = 1; U(t) U(u) = U(t + u) for all real t and u; and every U(t) is unitary, U(t)† U(t) = 1 and U(t) U(t)† = 1. *Used for:* The Schrödinger evolution of a finite quantum system as a one-parameter group of unitary operators U(t) = $e^{-i\,t\,H}$ in time. *Price:* **free**; exact.

801. Exponential of a sum of commuting matrices. For every size n and all complex n × n matrices A and B (entries in the complex numbers over the line) with AB = BA: if the exponential series of A, of B and of A + B converge entrywise to E, F and G respectively, then G = EF. That is, $e^{A+B} = e^A\,e^B$ for commuting matrices. *Used for:* Splitting a Hamiltonian into commuting parts, $e^{-i\,t\,(H_0 + V)}$ =

$e^{-i\,t\,H_0}\,e^{-i\,t\,V}$ when $H_0\,V = V\,H_0$, and factoring the Gibbs state of non-interacting subsystems into a product. *Price:* **free**; exact.

802. Orthogonal matrix: one product test, $A\,A^T = 1$. For every size n and every real n × n matrix A (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: A is orthogonal if and only if $AA^T = 1$. So orthonormal rows already give orthonormal columns. *Used for:* Building a rotation from an orthonormal set of row vectors, such as a rigid body's frame or the principal axes of an inertia tensor: one product check suffices and the columns are then orthonormal too. *Price:* **free**; exact.

803. Orthogonal matrix: one product test, $A^T\,A = 1$. For every size n and every real n × n matrix A (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: A is orthogonal if and only if $A^TA = 1$. So orthonormal columns already give orthonormal rows. *Used for:* Changing between orthonormal bases: a matrix whose columns are an orthonormal basis (normal modes, principal axes) is a rotation or reflection, and its rows are orthonormal as well. *Price:* **free**; exact.

804. Product of orthogonal matrices is orthogonal. For every size n and all real n × n matrices A and B (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: if A and B are orthogonal, then AB is orthogonal. *Used for:* Composing rotations and reflections of space, for example successive Euler-angle rotations of a rigid body or two symmetry operations of a crystal. *Price:* **free**; exact.

805. Identity matrix is orthogonal. For every size n, the n × n identity matrix 1 (entries on the line) is orthogonal: $1^T1 = 1$ and $11^T = 1$. *Used for:* The trivial symmetry: the identity is the neutral element of every rotation group and crystallographic point group. *Price:* **free**; exact.

806. Transpose of an orthogonal matrix is orthogonal. For every size n and every real n × n matrix A (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: if A is orthogonal, then its transpose A^T is orthogonal. *Used for:* Undoing a rotation: the inverse rotation A^T is again a rotation or reflection, used to switch between the body frame and the space frame of a rigid body. *Price:* **free**; exact.

807. Inverse of an orthogonal matrix is its transpose. For every size n and all real n × n matrices A and B (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: if A is orthogonal and AB = 1, then $B = A^T$. Every right inverse of an orthogonal matrix is its transpose. *Used for:* Inverting a rotation by transposing it, with no matrix inversion, when converting vectors and tensors between rotated frames. *Price:* **free**; exact.

808. Orthogonal matrix satisfies $A^T\,A = 1$. For every size n and every real n × n matrix A (entries on the line), where a square matrix A is orthogonal when $A^TA = 1$ and $AA^T = 1$: if A is orthogonal, then $A^TA = 1$. *Used for:* Orthonormal columns of a rotation matrix, used when expanding vectors in a rotated orthonormal basis. *Price:* **free**; exact.

809. Orthogonal matrices preserve the dot product. For every size n, every real n × n matrix A (entries on the line) that is orthogonal ($A^TA = 1$ and $AA^T = 1$), and all vectors x and y with n real entries: $(Ax) \cdot (Ay) = x \cdot y$, where $u \cdot v = u_1v_1 + \ldots + u_nv_n$. *Used for:* Rotations and reflections preserve angles and lengths: scalar products of positions, velocities and forces, and hence work and energy, are unchanged in every rotated frame. *Price:* **free**; exact.

810. Orthogonal matrices preserve length. For every size n, every real n × n matrix A (entries on the line) that is orthogonal ($A^TA = 1$ and $AA^T = 1$), and every vector x with n real entries: $(Ax) \cdot (Ax) = x \cdot x$, where $u \cdot v = u_1v_1 + \ldots + u_nv_n$. That is, $\|Ax\|^2 = \|x\|^2$. *Used for:* Rotations preserve lengths and distances: the kinetic energy $\tfrac{1}{2}m\|v\|^2$ and the distance between two points are unchanged in every rotated frame. *Price:* **free**; exact.

811. Product of determinant-one matrices has determinant one. For every size n and all real n × n matrices A and B (entries on the line): if $\det A = 1$ and $\det B = 1$, then $\det(AB) = 1$. So the special linear group SL(n), the matrices of determinant 1, is closed under products. *Used for:* Composing volume-preserving linear maps: phase-space maps of a Hamiltonian system with one degree of freedom (Liouville's theorem), ray-transfer (ABCD) matrices of optical systems between media of equal refractive index, deformations of an incompressible material. *Price:* **free**; exact.

812. Inverse in SL(n) is the adjugate. For every size n and every real n × n matrix A (entries on the line) with $\det A = 1$, the adjugate adj A (the transposed matrix of cofactors) satisfies $A\,\mathrm{adj}\,A = 1$ and $\mathrm{adj}\,A \cdot A = 1$, and $\det(\mathrm{adj}\,A) = 1$. So the inverse of a determinant-one matrix is its adjugate, again of determinant 1. *Used for:* Inverting a volume-preserving map exactly, with no division: for a 2 × 2 ray-transfer or phase-space matrix (a b; c d) of determinant 1 the inverse is (d −b; −c a). *Price:* **free**; exact.

813. Transpose of a determinant-one matrix has determinant one. For every size n and every real n × n matrix A (entries on the line): if $\det A = 1$, then

$\det A^T = 1$. So SL(n) is closed under transposition. *Used for:* Covectors such as momenta and gradients transform with the inverse transpose of a linear change of coordinates; for a volume-preserving change the transpose again has determinant one. *Price:* **free**; exact.

814. Product of rotations is a rotation (SO(n)). For every size n and all real n × n matrices A and B (entries on the line), where A is in SO(n) when it is orthogonal ($A^TA = 1$ and $AA^T = 1$) and $\det A = 1$: if A and B are in SO(n), then AB is in SO(n). *Used for:* Composing proper rotations of rigid bodies and molecules: two rotations give a rotation, never a reflection. *Price:* **free**; exact.

815. Inverse of a rotation is a rotation (SO(n)). For every size n and every real n × n matrix A (entries on the line), where A is in SO(n) when it is orthogonal ($A^TA = 1$ and $AA^T = 1$) and $\det A = 1$: if A is in SO(n), then A^T is in SO(n). Since A^T is the inverse of A, SO(n) is closed under inverses. *Used for:* Rotating a rigid body or a reference frame back: the reverse of a proper rotation is a proper rotation. *Price:* **free**; exact.

816. Product of invertible matrices is invertible. For every size n and all real n × n matrices A and B (entries on the line): if A has a two-sided inverse (some A' with $AA' = 1$ and $A'A = 1$) and B has a two-sided inverse, then AB has a two-sided inverse. *Used for:* Composing invertible changes of coordinates or of basis, such as transfer matrices of layered optical or acoustic media and conversions between crystal lattice bases. *Price:* **free**; exact.

817. Reflection of one axis is orthogonal with determinant −1. For every n, the real (n+1) × (n+1) diagonal matrix R = diag(−1, 1, ..., 1) (entries on the line), which reverses the first coordinate, is orthogonal ($R^TR = 1$ and $RR^T = 1$) and has determinant −1. *Used for:* Mirror reflections and the parity operation on one axis: an orthogonal map that reverses orientation, used to separate proper from improper symmetry operations of crystals and molecules. *Price:* **free**; exact.

818. Every orthogonal matrix is a rotation or a reflected rotation. For every n and every real (n+1) × (n+1) matrix A (entries on the line) with $A^TA = 1$ and $AA^T = 1$: either A is in SO(n+1) (orthogonal with $\det A = 1$), or RA is in SO(n+1) and A = R(RA), where R = diag(−1, 1, ..., 1) reflects the first axis. Which case holds is decided without any omniscience principle. *Used for:* Every rigid symmetry of a crystal or molecule is a proper rotation or a mirror reflection followed by a proper rotation, which sorts point-group operations into proper and improper. *Price:* **free**; exact.

819. Entries of a 2 × 2 rotation matrix. For every real 2 × 2 matrix M = (a b; c d) (entries on the line): M is in SO(2) ($M^TM = 1$, $MM^T = 1$ and $\det M = 1$) if and only if $a = d$, $b = -c$ and $a^2 + b^2 = 1$. *Used for:* Recognizing a plane rotation from its entries, in planar rigid-body motion and in the rotation of linear polarization by optical elements. *Price:* **free**; exact.

820. Rotation matrix (c −s; s c) and the unit circle. For all real numbers c and s on the line: the matrix R = (c −s; s c) is in SO(2) ($R^TR = 1$, $RR^T = 1$ and $\det R = 1$) if and only if $c^2 + s^2 = 1$. *Used for:* Describing a plane rotation by a cosine and sine pair (c, s) on the unit circle, without choosing an angle, as in planar mechanics and polarization optics. *Price:* **free**; exact.

821. Plane rotations compose by adding angles. For all real numbers c_1, s_1, c_2, s_2 on the line with ${c_1}^2 + {s_1}^2 = 1$ and ${c_2}^2 + {s_2}^2 = 1$, writing R(c, s) for the matrix (c −s; s c): $R(c_1, s_1)\,R(c_2, s_2) = R(c_1c_2 - s_1s_2, s_1c_2 + c_1s_2)$, and $(c_1c_2 - s_1s_2)^2 + (s_1c_2 + c_1s_2)^2 = 1$. *Used for:* Composing planar rotations by adding angles (the cosine and sine addition formulas), as in planar rigid-body kinematics and rotations of the polarization plane. *Price:* **free**; exact.

822. Lorentz condition is two-sided. For every n and every real (n+1) × (n+1) matrix Λ (entries on the line), where η = diag(−1, 1, ..., 1) is the Minkowski metric in 1 + n dimensions and a matrix L is a Lorentz transformation when $L^T\eta L = \eta$: $\Lambda^T\eta\Lambda = \eta$ if and only if $\Lambda\eta\Lambda^T = \eta$. (n = 3 is spacetime.). *Used for:* Special relativity: a Lorentz transformation can be recognized from either form of the defining condition, and its transpose is again one. *Price:* **free**; exact.

823. Product of Lorentz transformations is a Lorentz transformation. For every n and all real (n+1) × (n+1) matrices Λ and M (entries on the line), where η = diag(−1, 1, ..., 1) is the Minkowski metric in 1 + n dimensions and a matrix L is a Lorentz transformation when $L^T\eta L = \eta$: if Λ and M are Lorentz transformations, then ΛM is a Lorentz transformation. *Used for:* Composing changes of inertial frame in special relativity: successive boosts and rotations give a Lorentz transformation. *Price:* **free**; exact.

824. Identity is a Lorentz transformation. For every n, the (n+1) × (n+1) identity matrix 1 (entries on the line) satisfies $1^T\eta 1 = \eta$, where η = diag(−1, 1, ..., 1) is the Minkowski metric in 1 + n dimensions. So the identity is a Lorentz transformation. *Used for:* Staying in the original inertial frame: the identity is the neutral element of the Lorentz group of special relativity. *Price:* **free**; exact.

825. Inverse of a Lorentz transformation is $\eta\,\Lambda^T\,\eta$. For every n and every real (n+1) × (n+1) matrix Λ (entries on the line), where η = diag(−1, 1, ..., 1) is the

Minkowski metric in 1 + n dimensions and a matrix L is a Lorentz transformation when $L^T\eta L = \eta$: if Λ is a Lorentz transformation, then the matrix $\eta\Lambda^T\eta$ satisfies $\Lambda(\eta\Lambda^T\eta) = 1$ and $(\eta\Lambda^T\eta)\Lambda = 1$, and $\eta\Lambda^T\eta$ is again a Lorentz transformation. *Used for:* Changing back between inertial frames in special relativity: the inverse transformation is $\eta\Lambda^T\eta$, obtained with no matrix inversion (for a boost it reverses the velocity). *Price:* **free**; exact.

826. Transpose of a Lorentz transformation is a Lorentz transformation. For every n and every real (n+1) × (n+1) matrix Λ (entries on the line), where η = diag(−1, 1, ..., 1) is the Minkowski metric in 1 + n dimensions and a matrix L is a Lorentz transformation when $L^T\eta L = \eta$: if Λ is a Lorentz transformation, then Λ^T is a Lorentz transformation. *Used for:* Covectors such as gradients of fields and four-momenta with a lowered index transform with the inverse transpose of a Lorentz transformation; the transpose itself stays in the Lorentz group. *Price:* **free**; exact.

827. Lorentz transformations have determinant 1 or −1. For every n and every real (n+1) × (n+1) matrix Λ (entries on the line), where η = diag(−1, 1, ..., 1) is the Minkowski metric in 1 + n dimensions and a matrix L is a Lorentz transformation when $L^T\eta L = \eta$: if Λ is a Lorentz transformation, then $\det \Lambda = 1$ or $\det \Lambda = -1$. Which of the two holds is decided without any omniscience principle. *Used for:* Separating proper Lorentz transformations (determinant 1: boosts and rotations) from improper ones (determinant −1: those including a parity or time reversal). *Price:* **free**; exact.

828. Boost matrix in 1 + 1 dimensions. For all real numbers c and s on the line: the matrix B = (c s; s c) satisfies $B^T\eta B = \eta$, with η = diag(−1, 1) the Minkowski metric in 1 + 1 dimensions, if and only if $c^2 - s^2 = 1$. *Used for:* Boosts along one axis in special relativity: with $c = \gamma$ and $s = \gamma v$ (units with the speed of light 1) the matrix is a Lorentz transformation exactly when $\gamma^2(1 - v^2) = 1$. *Price:* **free**; exact.

829. Collinear boosts compose by adding rapidities. For all real numbers c_1, s_1, c_2, s_2 on the line with ${c_1}^2 - {s_1}^2 = 1$ and ${c_2}^2 - {s_2}^2 = 1$, writing B(c, s) for the matrix (c s; s c): $B(c_1, s_1)\,B(c_2, s_2) = B(c_1c_2 + s_1s_2, s_1c_2 + c_1s_2)$, and $(c_1c_2 + s_1s_2)^2 - (s_1c_2 + c_1s_2)^2 = 1$. *Used for:* Relativistic velocity addition: two boosts along one axis make one boost, with rapidities adding (the cosh and sinh addition formulas), which gives $w = (u + v)/(1 + uv)$. *Price:* **free**; exact.

830. Unitary matrix: one product test, U U† = 1. For every size n and every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): U is unitary if and only if UU† = 1. *Used for:* Checking that a quantum gate or a scattering matrix is unitary with a single product: orthonormal rows already give orthonormal columns. *Price:* **free**; exact.

831. Unitary matrix: one product test, U† U = 1. For every size n and every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): U is unitary if and only if U†U = 1. *Used for:* A matrix whose columns are orthonormal quantum states (the images of a basis under an evolution or a beam splitter) is unitary, so its rows are orthonormal too. *Price:* **free**; exact.

832. Product of unitary matrices is unitary. For every size n and all complex n × n matrices U and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): if U and V are unitary, then UV is unitary. *Used for:* Quantum circuits and time steps: a sequence of unitary gates or of short evolutions is again a unitary evolution. *Price:* **free**; exact.

833. Identity matrix is unitary. For every size n, the complex n × n identity matrix 1 (entries in the complex numbers over the line) is unitary: 1†1 = 1 and 11† = 1. *Used for:* The identity gate, doing nothing to a quantum state, is a unitary evolution: the neutral element of quantum gate sets and gauge groups. *Price:* **free**; exact.

834. Conjugate transpose of a unitary matrix is unitary. For every size n and every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): if U is unitary, then U† is unitary. *Used for:* Reversing a quantum evolution: the adjoint U† of a unitary gate is the inverse gate, used to uncompute in quantum algorithms and to move observables to the Heisenberg picture. *Price:* **free**; exact.

835. Inverse of a unitary matrix is its conjugate transpose. For every size n and all complex n × n matrices U and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): if U is unitary and UV = 1, then V = U†. Every right inverse of a unitary matrix is its conjugate transpose. *Used for:* Inverting a quantum gate or a scattering matrix by taking its conjugate transpose, with no matrix inversion. *Price:* **free**; exact.

836. Unitary matrix satisfies U† U = 1. For every size n and every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where a square complex matrix U is unitary when U†U = 1 and UU† = 1 (U† is the conjugate transpose): if U is unitary, then U†U = 1. *Used for:* Conservation of total probability: for a unitary evolution or S-matrix the outcome probabilities from any normalized input state sum to one. *Price:* **free**; exact.

837. Unitary matrices preserve the inner product. For every size n, every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is unitary (U†U = 1 and UU† = 1), and all vectors u and v with n complex entries: $\langle Uu, Uv\rangle = \langle u, v\rangle$, where $\langle u, v\rangle = \bar{u}_1 v_1 + \ldots + \bar{u}_n v_n$ (conjugate-linear in the first vector). *Used for:* Transition amplitudes $\langle \varphi, \psi\rangle$ and the probabilities $\|\langle \varphi, \psi\rangle\|^2$ are unchanged when both states undergo one common time evolution or one common change of basis. *Price:* **free**; exact.

838. Product of special unitary matrices is special unitary. For every size n and all complex n × n matrices U and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where U is in SU(n) when it is unitary (U†U = 1 and UU† = 1) and det U = 1: if U and V are in SU(n), then UV is in SU(n). *Used for:* Composing gauge transformations of the weak interaction (SU(2)) and the strong interaction (SU(3), color), and composing spin rotations. *Price:* **free**; exact.

839. Conjugate transpose of a special unitary matrix is special unitary. For every size n and every complex n × n matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), where U is in SU(n) when it is unitary (U†U = 1 and UU† = 1) and det U = 1: if U is in SU(n), then U† is in SU(n). Since U† is the inverse of U, SU(n) is closed under inverses. *Used for:* The inverse of an SU(2) or SU(3) gauge transformation, or of a spin rotation, is its adjoint and lies in that group too. *Price:* **free**; exact.

840. Special unitary 2 × 2 matrices as unit quaternions. For every complex 2 × 2 matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): U is in SU(2) (U†U = 1, UU† = 1 and det U = 1) if and only if there are complex numbers a and b with U = (a −b̄; b ā) and āa + b̄b = 1, that is $\|a\|^2 + \|b\|^2 = 1$. *Used for:* Spin-½ rotations and single-qubit gates of determinant 1: each is described by two Cayley–Klein parameters (a, b) with $\|a\|^2 + \|b\|^2 = 1$, a unit quaternion. *Price:* **free**; exact.

841. Product law of 2 × 2 special unitary matrices. For all complex numbers a, b, c, d (in the complex numbers over the line) with āa + b̄b = 1 and c̄c + đd = 1, writing S(a, b) for the matrix (a −b̄; b ā): S(a, b) S(c, d) = S(ac − b̄d, bc + ād), and $\|ac - \bar{b}d\|^2 + \|bc + \bar{a}d\|^2 = 1$. *Used for:* Composing spin-½ rotations and single-qubit gates through their Cayley–Klein parameters, the matrix form of quaternion multiplication. *Price:* **free**; exact.

842. Commuting square root of a positive semidefinite matrix. For every size n and every n × n matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line) with $A \succeq 0$, there is an n × n complex matrix B with B† = B, $B \succeq 0$ and BB = A such that every n × n complex matrix Q that commutes with A (QA = AQ) also commutes with B (QB = BQ). Here a square complex matrix G is positive semidefinite, $G \succeq 0$, when G† = G and the real part of $\langle x, Gx\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Square roots of density matrices in the fidelity $F(\rho, \sigma) = (\mathrm{tr}\,\sqrt{(\sqrt{\rho}\,\sigma\,\sqrt{\rho})})^2$ and in the Petz recovery map; the root $\sqrt{\rho}$ commutes with every conserved charge that commutes with ρ, so it keeps the state's symmetries. *Price:* **free**; exact.

843. Polar decomposition of a square complex matrix. For every size n and every n × n matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line), there are n × n complex matrices W and P with W†W = 1 (W unitary), $P \succeq 0$ and A = WP. A is not assumed to be invertible. Here a square complex matrix P is positive semidefinite, $P \succeq 0$, when P† = P and the real part of $\langle x, Px\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Splitting a quantum measurement operator into a unitary and a positive part, $M = W\cdot\sqrt{(M^\dagger M)}$, including rank-deficient measurement operators such as projectors, to describe the back-action of a generalized measurement. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

844. Orthogonal vectors as a scaled orthonormal basis. For every size n, every list of n complex vectors $m_1, \ldots, m_n$ with n entries each (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every list of real numbers $\sigma_1, \ldots, \sigma_n \geq 0$ such that $\langle m_k, m_k\rangle = \sigma_k^2$ for every k and $\langle m_j, m_k\rangle = 0$ whenever j ≠ k: there are vectors $u_1, \ldots, u_n$ with $\langle u_j, u_k\rangle = 1$ when j = k and 0 when j ≠ k (an orthonormal basis) and $\sigma_k \cdot u_k = m_k$ for every k. Here $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. Where $\sigma_k = 0$ the vector u_k is not fixed by m_k and has to be found orthogonal to all the others; the statement does not say which σ_k vanish. *Used for:* Completing the left singular vectors of a rank-deficient matrix, and extending a set of mutually orthogonal states or optical modes of known norms, some possibly zero, to a full orthonormal basis. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

845. Polar decomposition of an invertible complex matrix. For every size n, every n × n matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number e with det(A)·e = 1 (an explicit inverse of the determinant): there are n × n complex matrices W and P with W†W = 1 (W unitary), P ⪰ 0 and A = WP. Here a square complex matrix P is positive semidefinite, P ⪰ 0, when P† = P and the real part of ⟨x, Px⟩ is ≥ 0 for every complex vector x, with ⟨x, y⟩ = $\bar{x}_1y_1$ + … + $\bar{x}_ny_n$. *Used for:* Löwdin symmetric orthogonalisation: the unitary polar factor W = A(A†A)$^{-1/2}$ of a matrix of overlapping orbitals is the orthonormal set closest to them, used in quantum chemistry and in building Wannier functions. *Price:* **free**; exact.

846. Schmidt decomposition of a bipartite pure state. For every size n and every n × n matrix C with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line), there are n × n complex matrices U and V with U†U = 1 and V†V = 1 and real numbers s_1, …, s_n ≥ 0 with C = U·diag(s_1, …, s_n)·V†. Read C as the coefficient matrix of a pure state ψ = Σ C_{ij} e_i⊗e_j of two n-level systems (e_1, …, e_n the standard basis): then ψ = s_1 a_1⊗b_1 + … + s_n a_n⊗b_n with orthonormal vectors a_k (the columns of U) and orthonormal vectors b_k (the complex conjugates of the columns of V), and the s_k are the Schmidt coefficients of ψ. *Used for:* Entanglement of a pure state of two subsystems: the Schmidt coefficients s_k give the spectrum s_k^2 of each reduced density matrix, the entanglement entropy −Σ s_k^2 ln s_k^2 and the Schmidt rank. *Price:* **LLPO + AC**$_\omega$; exact.

847. Polar factor of a complex vector. For every size n ≥ 1 and every complex vector m with n entries (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), there are a complex vector u with ⟨u, u⟩ = 1 and a real number r ≥ 0 with ⟨m, m⟩ = r^2 and r·u = m, where ⟨x, y⟩ = $\bar{x}_1y_1$ + … + $\bar{x}_ny_n$. So m = ‖m‖·u with u a unit vector. The vector m is not assumed to be nonzero. *Used for:* Renormalising a quantum state after a projective measurement or an imaginary-time step, ψ = ‖ψ‖·u, when the unnormalised result may vanish. *Price:* **LLPO + AC**$_\omega$; exact.

848. Determinant of the adjugate matrix. For every size n and every n × n matrix A on the line (real entries): det(adj(A)) = det(A)$^{n-1}$, where adj(A) is the adjugate, whose (i, j) entry is $(-1)^{i+j}$ times the determinant of A with row j and column i deleted. At size 0 both sides are 1. A is not assumed to be invertible. *Used for:* Nanson's formula in continuum mechanics: the cofactor (area-change) matrix of a 3 × 3 deformation gradient F has determinant det(F)^2, so the area and volume changes of a deformed body stay consistent. *Price:* **free**; exact.

849. Adjugate of the adjugate matrix. For every size n other than 1 and every n × n matrix A on the line (real entries): adj(adj(A)) = det(A)$^{n-2}$·A, where adj(A) is the adjugate, whose (i, j) entry is $(-1)^{i+j}$ times the determinant of A with row j and column i deleted. At size 0 the exponent n − 2 is read as 0 and both sides are the empty matrix. A is not assumed to be invertible. *Used for:* Recovering a 3 × 3 matrix from its cofactor matrix: for a deformation gradient F in nonlinear elasticity, adj(adj(F)) = det(F)·F, so the area-change map fixes F up to its volume change. *Price:* **free**; exact.

850. Chebyshev series of a Hermitian matrix, coefficients with representatives. For every size n, every n × n Hermitian matrix H with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line; H† = H) with −1 ⪯ H ⪯ 1 (both 1 − H and 1 + H positive semidefinite), and every sequence of real coefficients a_0, a_1, a_2, … on the line that is given together with representatives (for each k, a sequence of rationals q_0, q_1, … with $|q_i - q_l| \le 1/(i+1) + 1/(l+1)$ for all i and l, whose limit is a_k) and has an explicit tail modulus for Σ $|a_k|$ (a function K such that $|a_p| + |a_{p+1}| + \dots + |a_{p+m-1}| \le 1/(j+1)$ for every j, every p ≥ K(j) and every m): the partial sums of $a_0T_0(H) + a_1T_1(H) + a_2T_2(H) + \dots$ converge entry by entry to a matrix F with F† = F and FH = HF; and for every n × n matrix V with V†V = 1 and real numbers λ_1, …, λ_n and μ_1, …, μ_n such that HV = V·diag(λ_1, …, λ_n) and, for each k, the scalar series $a_0T_0(\lambda_k) + a_1T_1(\lambda_k) + \dots$ converges to μ_k, FV = V·diag(μ_1, …, μ_n). So F = f(H) for the function $f(t) = \Sigma\, a_kT_k(t)$. Here $T_0(H) = 1$, $T_1(H) = H$ and $T_{j+2}(H) = 2H\cdot T_{j+1}(H) - T_j(H)$, and $T_j(t)$ for a number t follows this recursion with t in place of H; and a matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, with ⟨x, y⟩ = $\bar{x}_1y_1$ + … + $\bar{x}_ny_n$. *Used for:* Chebyshev expansions of the thermal operator $e^{-\beta H}$ (modified Bessel-function coefficients) and of cos(Ht) and sin(Ht), the real and imaginary parts of the Chebyshev propagator, for a Hamiltonian rescaled to spectrum in [−1, 1]; explicit coefficient formulas come with representatives. *Price:* **free**; exact.

851. Unitary dilation of a Hermitian contraction. For every size n and every n × n Hermitian matrix H with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line; H† = H) with −1 ⪯ H ⪯ 1 (both 1 − H and 1 + H positive semidefinite), there is an n × n complex matrix U with U†U = 1 and UU† = 1 that commutes with H (UH = HU), with H = (U + U†)/2 and $T_k(H) = (U^k + (U\dagger)^k)/2$ for every k ≥ 0. Here $T_0(H) = 1$, $T_1(H) = H$ and $T_{k+2}(H) = 2H\cdot T_{k+1}(H) - T_k(H)$ are the

Chebyshev polynomials, and a matrix G is positive semidefinite when $G\dagger = G$ and the real part of $\langle x, Gx\rangle$ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Qubitization and quantum walks: a Hamiltonian rescaled into $[-1, 1]$ is the average of a unitary walk operator and its inverse, so k steps of the walk produce the Chebyshev polynomial $T_k(H)$, the matrix form of $T_k(\cos\theta) = \cos k\theta$. *Price:* **free**; exact.

852. Chebyshev polynomials of a Hermitian contraction are contractions. For every size n, every $n \times n$ Hermitian matrix H with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line; $H\dagger = H$) with $-1 \preceq H \preceq 1$ (both $1 - H$ and $1 + H$ positive semidefinite), and every $k \geq 0$: $\|T_k(H)x\|^2 \leq \|x\|^2$ for every complex vector x, where $\|x\|^2 = \langle x, x\rangle = \bar{x}_1 x_1 + \ldots + \bar{x}_n x_n$. So the operator norm of $T_k(H)$ is at most 1. Here $T_0(H) = 1$, $T_1(H) = H$ and $T_{k+2}(H) = 2H\cdot T_{k+1}(H) - T_k(H)$ are the Chebyshev polynomials, and a matrix G is positive semidefinite when $G\dagger = G$ and the real part of $\langle x, Gx\rangle$ is ≥ 0 for every complex vector x. *Used for:* Error control in Chebyshev expansions: each term $T_k(H)$ of a rescaled Hamiltonian has norm at most 1, so cutting $\Sigma\, a_k T_k(H)$ after N terms errs by at most Σ over $k \geq N$ of $|a_k|$, in the kernel polynomial method and Chebyshev propagators. *Price:* **free**; exact.

853. Spectral projection below a gap of a Hermitian matrix. For every size n, every $n \times n$ Hermitian matrix H with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line; $H\dagger = H$), and all real numbers c and $r > 0$ such that H has a gap of half-width r around c, that is $\|(H - c)x\|^2 \geq r^2\|x\|^2$ for every complex vector x, so that no eigenvalue of H lies strictly between $c - r$ and $c + r$: there is an $n \times n$ matrix P with $P\dagger = P$, $P^2 = P$ and $PH = HP$ such that, for every $n \times n$ matrix V with $V\dagger V = 1$ and real numbers $\lambda_1, \ldots, \lambda_n$ with $HV = V\cdot\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and every k, $Pv_k = v_k$ when $\lambda_k \leq c - r$ and $Pv_k = 0$ when $\lambda_k \geq c + r$, where v_k is the k-th column of V. Here $\|x\|^2 = \langle x, x\rangle = \bar{x}_1 x_1 + \ldots + \bar{x}_n x_n$. P is obtained without naming any eigenbasis. *Used for:* Fermi-sea projector of a gapped insulator or gapped ground state: the projection onto all states below the gap, used for ground-state density matrices, Berry phases, Chern numbers and Wannier functions. *Price:* **free**; exact.

854. Positive definite matrices are uniformly positive. For every size n and every $n \times n$ matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line) that is literally positive definite, meaning $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$: there is a real number $\varepsilon > 0$ such that the real part of $\langle x, Ax\rangle$ is at least $\varepsilon\cdot\langle x, x\rangle$ for every complex vector x. Here $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$. *Used for:* Stability margins: a positive definite Hessian, stiffness matrix or mass-weighted force-constant matrix has a uniform positive curvature ε, which gives a minimum restoring force and a strictly positive lowest normal-mode frequency. *Price:* **free**; exact.

855. Singular value decomposition of square complex matrices. For every size n and every $n \times n$ matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line), there are $n \times n$ complex matrices U and V with $U\dagger U = 1$ and $V\dagger V = 1$ and real numbers $s_1, \ldots, s_n \geq 0$ with $AV = U\cdot\mathrm{diag}(s_1, \ldots, s_n)$. So $A = U\cdot\mathrm{diag}(s_1, \ldots, s_n)\cdot V\dagger$ with U and V unitary; the s_k are the singular values of A, in no prescribed order. *Used for:* Truncating matrix-product states in DMRG and other tensor-network simulations by keeping the largest singular values, and low-rank approximation of response and correlation matrices. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

856. Determinant of a matrix exponential (Jacobi's formula). For every size n and every $n \times n$ matrix A with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line): if the exponential series $1 + A + A^2/2! + A^3/3! + \ldots$ converges entrywise to a matrix E, and the exponential series of the complex number $\mathrm{tr}(A) = A_{11} + \ldots + A_{nn}$ converges to a number e, then $\det(E) = e$. So $\det(\exp A) = \exp(\mathrm{tr}\, A)$. *Used for:* Liouville's theorem for linear flows: a phase-space volume carried by $\dot{x} = A\cdot x$ scales by $\det(e^{t\cdot A}) = e^{t\cdot \mathrm{tr}\, A}$; for a quantum evolution, $\det(e^{-i\cdot H\cdot t}) = e^{-i\cdot t\cdot \mathrm{tr}\, H}$, so traceless generators give special unitary evolutions. *Price:* **free**; exact.

857. Simultaneous diagonalization of commuting Hermitian matrices. For every size n and every two $n \times n$ Hermitian matrices H and K with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line; $H\dagger = H$ and $K\dagger = K$) that commute ($HK = KH$), there are an $n \times n$ complex matrix V with $V\dagger V = 1$ and real numbers $\lambda_1, \ldots, \lambda_n$ and $\kappa_1, \ldots, \kappa_n$ with $HV = V\cdot\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $KV = V\cdot\mathrm{diag}(\kappa_1, \ldots, \kappa_n)$. So the columns of V form one orthonormal basis of joint eigenvectors of H and K. *Used for:* Labelling energy eigenstates by a commuting symmetry: a common eigenbasis of a Hamiltonian and a conserved quantity such as parity, momentum or angular momentum, as in a complete set of commuting observables. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

858. Spectral theorem for normal matrices. For every size n and every $n \times n$ matrix N with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line) that is normal, $NN\dagger = N\dagger N$: there are an $n \times n$ complex matrix V with $V\dagger V = 1$ and complex numbers $\mu_1, \ldots,$

μ_n with $NV = V\cdot\mathrm{diag}(\mu_1, \ldots, \mu_n)$. So $N = V\cdot\mathrm{diag}(\mu_1, \ldots, \mu_n)\cdot V\dagger$ with V unitary: an orthonormal basis of eigenvectors with complex eigenvalues. *Used for:* Normal non-Hermitian operators with complex eigenvalues: circulant couplings such as a non-reciprocal (Hatano–Nelson) ring with periodic boundary conditions, diagonalised by plane waves, and operators $A + i\cdot B$ built from commuting observables A and B. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

859. Spectral theorem for unitary matrices. For every size n and every $n \times n$ matrix U with complex entries (on the complex numbers over the line, whose real and imaginary parts lie on the line) with $U\dagger U = 1$: there are an $n \times n$ complex matrix V with $V\dagger V = 1$ and complex numbers $\mu_1, \ldots, \mu_n$ with $\bar{\mu}_k\mu_k = 1$ for every k and $UV = V\cdot\mathrm{diag}(\mu_1, \ldots, \mu_n)$. So $U = V\cdot\mathrm{diag}(\mu_1, \ldots, \mu_n)\cdot V\dagger$ with every eigenvalue on the unit circle. *Used for:* Floquet quasi-energies and quantum gates: the eigenphases of a one-period evolution operator or of a quantum circuit, with an orthonormal basis of Floquet states or gate eigenstates. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

860. Positive definite implies positive semidefinite. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then A is positive semidefinite. Here a square complex matrix A is positive semidefinite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, and positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Strict positivity includes positivity: a full-rank density matrix, a nondegenerate covariance matrix or a mass matrix is in particular positive semidefinite, so every result proved for semidefinite matrices applies to it. *Price:* **free**; exact.

861. Positive definite plus positive semidefinite. For every size n and all complex $n \times n$ matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite and B is positive semidefinite, then $A + B$ is positive definite. Here a square complex matrix A is positive semidefinite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, and positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Regularisation and added couplings: adding a positive semidefinite term (springs with zero modes, a penalty, a coupling) to a positive definite mass, stiffness or covariance matrix keeps it positive definite; ridge regularisation $K + \lambda\cdot 1$ with $\lambda > 0$ is this statement with the roles of the two terms exchanged. *Price:* **free**; exact.

862. Positive multiple of a positive definite matrix. For every size n, every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number c that is a positive real (real part > 0, imaginary part 0): if A is positive definite, then $c\cdot A$ is positive definite. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Changing units or temperature: a positive multiple of a positive definite mass, inertia or metric matrix, such as a unit conversion or the factor $1/(k\cdot T)$ in a Boltzmann weight $\exp(-x^{\mathrm{T}}Ax/(k\cdot T))$, with Boltzmann's constant k and the temperature T„ stays positive definite. *Price:* **free**; exact.

863. Diagonal entries of a positive definite matrix are positive. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then every diagonal entry A_{ii} is a positive real number (real part > 0, imaginary part 0). Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Variances, masses and self-overlaps: each diagonal entry of a nondegenerate covariance matrix (a variance), of a mass matrix (a mass) and of the overlap matrix of independent states ($\langle\varphi_i, \varphi_i\rangle$) is positive. *Price:* **free**; exact.

864. Trace of a positive definite matrix is positive. For every size $n \geq 1$ and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then its trace $A_{11} + \ldots + A_{nn}$ is a positive real number (real part > 0, imaginary part 0). Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Total variance and normalisation: the trace of a nondegenerate covariance matrix (the total variance), of a positive definite inertia tensor (the sum of the principal moments) and of an unnormalised full-rank density matrix is a positive real, so dividing by it is allowed. *Price:* **free**; exact.

865. Congruence $B\dagger AB$ keeps positive definiteness. For all sizes n and m, every complex n

× n matrix A and every complex n × m matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite and B is injective (Bx = By implies x = y, for complex vectors x and y of length m), then the m × m matrix B†AB is positive definite. Here a square complex matrix A is positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries of x and y); x ≠ 0 means only that x is not the zero vector (the negation of x = 0). *Used for:* Restricting a positive energy to coordinates or a subspace: with positions x = Bq given by independent generalised coordinates q, the kinetic-energy matrix $B^T M B$ is positive definite; Rayleigh–Ritz projection of a positive definite operator onto an independent trial basis. *Price:* **free**; exact.

866. Inverse of a positive definite matrix is positive definite. For every size n and all complex n × n matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite and B is a two-sided inverse of A (AB = 1 and BA = 1), then B is positive definite. Here a square complex matrix A is positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). *Used for:* Inverse metric and inverse mass matrix: the inverse metric g^{-1} used to raise indices, the inverse mass matrix M^{-1} that gives velocities from momenta in Hamilton's equations, and a Gaussian's covariance as the inverse of its precision matrix are positive definite. *Price:* **free**; exact.

867. Eigenvalues of a positive definite matrix are positive. For every size n, every complex n × n matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), every complex n × n matrix V with V†V = 1, and all real numbers $\lambda_1, \ldots, \lambda_n$ on the line with HV = V diag($\lambda_1, \ldots, \lambda_n$) (the columns of V are an orthonormal eigenbasis of H with these eigenvalues): if H is positive definite, then every $\lambda_k > 0$. Here a square complex matrix A is positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The eigenbasis is given, not constructed. *Used for:* Stable normal modes: the eigenvalues ω^2 of a positive definite stiffness matrix in mass-weighted coordinates are positive, so every mode oscillates with a real frequency; a positive definite quadratic Hamiltonian has positive mode energies. *Price:* **free**; exact.

868. Identity matrix is positive definite. For every size n, the complex n × n identity matrix 1 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) is positive definite: for every complex vector x ≠ 0, $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2 > 0$. Here a square complex matrix A is positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The hypothesis says only that x is not the zero vector; turning that into a positive length is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already at size 1 the statement asks $r^2 > 0$ for every real r ≠ 0. *Used for:* The Euclidean metric δ_{ij} and the overlap matrix of an orthonormal basis are positive definite: every vector that is not zero has positive length, the property that makes the dot product an inner product. *Price:* **MP**; exact.

869. Positive semidefinite with trivial kernel is positive definite. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive semidefinite and Ax = 0 only for x = 0 (for every complex vector x, Ax = 0 implies x = 0), then A is positive definite. Here a square complex matrix A is positive semidefinite when A† = A and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x, and positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). For x ≠ 0 the hypotheses give only that the real part of $\langle x, Ax\rangle$ is ≥ 0 and is not zero; concluding that it is > 0 is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Ruling out zero modes: a positive semidefinite Hessian, stiffness or overlap matrix whose only null vector is zero is positive definite, so an equilibrium with no zero modes is a strict minimum. *Price:* **MP**; exact.

870. A†A is positive definite for injective A. For all sizes m and n and every complex m × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is injective (Ax = Ay implies x = y, for complex vectors x and y of length n), then the n × n matrix A†A is positive definite. Here a square complex matrix A is positive definite when A† = A and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector x ≠ 0, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). For x ≠ 0 injectivity gives only that Ax is not the zero vector; concluding $\langle x, A^\dagger Ax\rangle = \|Ax\|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly

that a real number that is not zero has an inverse. *Used for:* Least squares and induced metrics: the normal-equations matrix $A^{T}A$ of a fit whose design matrix A has independent columns is positive definite, so the fit has a unique best solution; and the metric $J^{T}J$ induced by an embedding with injective Jacobian J is positive definite for the reason just given. *Price:* **MP**; exact.

871. Determinant of a positive definite matrix is positive. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then det(A) is a positive real number (real part > 0, imaginary part 0). Here a square complex matrix A is positive definite when A† = A and the real part of ⟨x, Ax⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). *Used for:* Gaussian integrals and volume elements: the normalisation $\sqrt{((2\pi)^n/\det A)}$ of $\exp(-\tfrac{1}{2}\, x^{T}Ax)$, Gaussian prefactors in partition functions and path integrals, and the Riemannian volume element √(det g) all need det > 0. *Price:* **free**; exact.

872. A positive definite matrix is invertible. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then A has a two-sided inverse, a matrix B with AB = 1 and BA = 1. Here a square complex matrix A is positive definite when A† = A and the real part of ⟨x, Ax⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). *Used for:* Solving with a positive definite matrix: accelerations from Mẍ = F with a mass matrix M, raising indices with the inverse metric, and inverting a covariance matrix to write a Gaussian density. *Price:* **free**; exact.

873. Positive definite exactly when A = B†B with B invertible. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): A is positive definite if and only if A = B†B for some complex n × n matrix B that has a two-sided inverse (BC = 1 and CB = 1 for some C). Here a square complex matrix A is positive definite when A† = A and the real part of ⟨x, Ax⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). From left to right no principle is needed (B can be taken as the positive square root of A). From right to left, x ≠ 0 gives only that Bx is not the zero vector; concluding ⟨x, Ax⟩ = $\|Bx\|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Factoring a positive energy or metric: writing a mass matrix, metric or covariance as B†B (a Cholesky factor or a square root) to pass to mass-weighted or whitened coordinates y = Bx, in which the quadratic form becomes $\|y\|^2$. *Price:* **MP**; exact.

874. Positive semidefinite matrix: positive definite exactly when invertible. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is positive semidefinite: A is positive definite if and only if A has a two-sided inverse (AB = 1 and BA = 1 for some B). Here a square complex matrix A is positive semidefinite when A† = A and the real part of ⟨x, Ax⟩ is ≥ 0 for every complex vector x, and positive definite when A† = A and the real part of ⟨x, Ax⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). From left to right no principle is needed. From right to left, x ≠ 0 gives only that the real part of ⟨x, Ax⟩ is ≥ 0 and is not zero; concluding that it is > 0 is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Nondegeneracy tests: a positive semidefinite Fisher information, overlap or covariance matrix is strictly positive exactly when it can be inverted, which decides whether parameters are identifiable or a basis is free of redundancy. *Price:* **MP**; exact.

875. Positive definite exactly when all eigenvalues are positive. For every size n, every complex n × n matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with H† = H, every complex n × n matrix V with V†V = 1, and all real numbers $\lambda_1, \ldots, \lambda_n$ on the line with HV = V diag($\lambda_1, \ldots, \lambda_n$) (the columns of V are an orthonormal eigenbasis of H with these eigenvalues): H is positive definite if and only if every $\lambda_k > 0$. Here a square complex matrix A is positive definite when A† = A and the real part of ⟨x, Ax⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\bar{x}_1y_1 + \ldots + \bar{x}_ny_n$; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The eigenbasis is given, not constructed. From left to right no principle is needed. From right to left, x ≠ 0 gives only that the coefficient vector c = V†x is not zero; concluding ⟨x, Hx⟩ = $\lambda_1|c_1|^2 + \ldots + \lambda_n|c_n|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Stability read from the spectrum: an equilibrium of a potential is a strict minimum exactly when every eigenvalue of its Hessian is positive, and a quadratic Hamiltonian is strictly

positive exactly when every mode energy is, once the eigenbasis is known. *Price:* **MP**; exact.

876. Diagonal matrix positive definite exactly when entries are positive. For every size n and all complex numbers $d_1, \ldots, d_n$ (in the complex numbers over the line, whose real and imaginary parts lie on the line): the diagonal matrix $\mathrm{diag}(d_1, \ldots, d_n)$ is positive definite if and only if every d_i is a positive real number (real part > 0, imaginary part 0). Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). From left to right no principle is needed. From right to left, $x \neq 0$ gives only that x is not the zero vector; concluding $d_1|x_1|^2 + \ldots + d_n|x_n|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Principal axes: a diagonal inertia tensor, mass matrix, covariance of independent variables or diagonal metric $\mathrm{diag}(a^2, b^2, c^2)$ is positive definite exactly when every principal moment, mass, variance or diagonal entry is positive. *Price:* **MP**; exact.

877. LDL factorisation of a positive definite matrix. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then $A = L \cdot \mathrm{diag}(d_1, \ldots, d_n) \cdot L\dagger$ for some complex $n \times n$ matrix L that is lower triangular with ones on the diagonal ($L_{ii} = 1$, and $L_{ij} = 0$ whenever $i < j$) and some real numbers $d_1, \ldots, d_n$ on the line, all > 0. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Cholesky and LDL solves: factoring positive definite stiffness, mass and normal-equation matrices in finite-element and normal-mode computations, and generating correlated Gaussian noise from a covariance matrix. *Price:* **free**; exact.

878. Transpose of a positive definite matrix. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then its transpose A^{T} (entries swapped, $(A^{\mathrm{T}})_{ij} = A_{ji}$, with no complex conjugation) is positive definite. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Transposing a whole quantum state: the full transpose ρ^{T} of a full-rank density matrix is again positive definite (only the partial transpose can fail to be positive, which is what the Peres–Horodecki entanglement test detects); also switching row and column conventions. *Price:* **free**; exact.

879. Conjugate transpose of a positive definite matrix. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite, then its conjugate transpose $A\dagger$ is positive definite. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). *Used for:* Moving a positive operator across an inner product: the adjoint of a positive definite metric, Hamiltonian or density matrix is again positive definite (it is the operator itself), so the bra and ket forms of a quadratic expression agree. *Price:* **free**; exact.

880. Literal and uniform positive definiteness agree. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $A\dagger = A$: A is positive definite if and only if A is uniformly positive: there is a real number $\varepsilon > 0$ on the line such that the real part of $\langle x, Ax\rangle$ is $\geq \varepsilon\|x\|^2$ for every complex vector x, where $\|x\|^2 = \langle x, x\rangle = |x_1|^2 + \ldots + |x_n|^2$. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). From left to right no principle is needed: the margin ε is constructed. From right to left, $x \neq 0$ gives only that x is not the zero vector; concluding $\varepsilon\|x\|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Coercivity: a positive definite Hessian, stiffness or kinetic-energy matrix comes with a uniform margin $\langle x, Ax\rangle \geq \varepsilon\|x\|^2$, the bound behind energy estimates, stability margins, a positive lower bound on the spectrum of a quadratic Hamiltonian, and convergence rates of iterative solvers. *Price:* **MP**; exact.

881. Invertible change of basis preserves positive definiteness. For every size n and all complex $n \times n$ matrices U, W and X (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $UW = 1$ and $WU = 1$ (U is invertible with inverse W): $U\dagger XU$ is positive definite if and only if X is positive definite. Here a square complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the

negation of x = 0). *Used for:* Changing coordinates by any invertible, not necessarily unitary, matrix: positive definiteness of a metric, inertia tensor, kinetic-energy matrix or Hamiltonian does not depend on the basis (the positive definite case of Sylvester's law of inertia). *Price:* **free**; exact.

882. Inverse is positive definite exactly when the matrix is. For every size n and all complex n × n matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with AB = 1 and BA = 1 (B is the inverse of A): B is positive definite if and only if A is positive definite. Here a square complex matrix A is positive definite when $A^\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$; $x \neq 0$ means only that x is not the zero vector (the negation of x = 0). *Used for:* Passing between a matrix and its inverse in both directions: a Lagrangian mass matrix and the Hamiltonian inverse mass matrix, a metric and its inverse, and a Gaussian's covariance and precision matrix are positive definite together. *Price:* **free**; exact.

883. Congruence BAB† keeps positive definiteness. For all sizes n and m, every complex n × n matrix A and every complex m × n matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if A is positive definite and the row-vector map $v \mapsto v\cdot B$ is injective ($v\cdot B = w\cdot B$ implies v = w, for complex row vectors v and w of length m, where $(v\cdot B)_j = v_1 B_{1j} + \ldots + v_m B_{mj}$), then the m × m matrix $BAB^\dagger$ is positive definite. Here a square complex matrix A is positive definite when $A^\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries of x and y); $x \neq 0$ means only that x is not the zero vector (the negation of x = 0). *Used for:* Pushing a nondegenerate covariance or positive operator through a linear map: the covariance $B\Sigma B^\dagger$ of the image of a nondegenerate Gaussian under a map B with independent rows is nondegenerate, so the transformed variables still have a proper joint density. *Price:* **free**; exact.

884. Gram–Schmidt and QR for uniformly independent vectors. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is uniformly independent: there is a real number $\varepsilon > 0$ on the line with $\|c_1 v_1 + \ldots + c_k v_k\|^2 \geq \varepsilon(|c_1|^2 + \ldots + |c_k|^2)$ for all complex numbers $c_1, \ldots, c_k$: there are orthonormal vectors $u_1, \ldots, u_k$ of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and k × k upper triangular matrices R and S (entries below the diagonal are 0) with RS = SR = 1 and every diagonal entry R_{jj} a positive real number, such that $v_j = R_{1j}u_1 + \ldots + R_{kj}u_k$ and $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j. Since R and S are upper triangular, $u_1, \ldots, u_j$ and $v_1, \ldots, v_j$ span one common space for every j: this is Gram–Schmidt orthonormalisation, and the QR factorisation of the n × k matrix whose columns are $v_1, \ldots, v_k$. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. *Used for:* Orthonormalising a basis that comes with a certified overlap bound: Gram–Schmidt and QR on Krylov (Lanczos, Arnoldi) vectors, trial wavefunctions or atomic orbitals whose overlap matrix has its smallest eigenvalue bounded away from zero. *Price:* **free**; exact.

885. Gram–Schmidt for linearly independent vectors. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is linearly independent in the literal sense: whenever $c_1 v_1 + \ldots + c_k v_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is): there are orthonormal vectors $u_1, \ldots, u_k$ of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and k × k upper triangular matrices R and S (entries below the diagonal are 0) with RS = SR = 1 and every diagonal entry R_{jj} a positive real number, such that $v_j = R_{1j}u_1 + \ldots + R_{kj}u_k$ and $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j. Since R and S are upper triangular, $u_1, \ldots, u_j$ and $v_1, \ldots, v_j$ span one common space for every j: this is Gram–Schmidt orthonormalisation, and the QR factorisation of the n × k matrix whose columns are $v_1, \ldots, v_k$. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. The hypothesis is of the 'only zero' kind and gives no positive length; normalising the residual vectors needs a positive length, and getting it is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already one vector (r) with a real $r \neq 0$ is literally independent, and normalising it inverts r. *Used for:* Textbook Gram–Schmidt: orthonormalising states known only to be linearly independent, such as a basis of a degenerate eigenspace, a set of trial wavefunctions or non-orthogonal orbitals, before computing matrix elements. *Price:* **MP**; exact.

886. Löwdin orthonormalisation for uniformly independent vectors. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is uniformly independent: there is a real number $\varepsilon > 0$ on the line with $\|c_1 v_1 + \ldots + c_k v_k\|^2 \geq \varepsilon(|c_1|^2 + \ldots + |c_k|^2)$ for all complex numbers $c_1, \ldots, c_k$: there are orthonormal vectors $u_1, \ldots, u_k$ of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and a k × k matrix S with $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j, S positive semidefinite ($S^\dagger = S$ and

the real part of $\langle x, Sx\rangle$ is ≥ 0 for every complex vector x), SGS = 1 for the Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$), and S invertible (ST = TS = 1 for some T). Then S^2 is the inverse of G, so S is a positive semidefinite square root of G^{-1}: the matrix written $G^{-1/2}$ in Löwdin's symmetric orthonormalisation u = VS. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. *Used for:* Löwdin symmetric orthonormalisation of non-orthogonal atomic orbitals (the orbitals multiplied by the inverse square root of their overlap matrix), used in tight-binding and Hückel models and in Löwdin population analysis, for orbital sets with a certified overlap bound. *Price:* **free**; exact.

887. Löwdin orthonormalisation for linearly independent vectors. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is): there are orthonormal vectors $u_1, \ldots, u_k$ of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and a k × k matrix S with $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j, S positive semidefinite ($S^\dagger = S$ and the real part of $\langle x, Sx\rangle$ is ≥ 0 for every complex vector x), SGS = 1 for the Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$), and S invertible (ST = TS = 1 for some T). Then S^2 is the inverse of G, so S is a positive semidefinite square root of G^{-1}: the matrix written $G^{-1/2}$ in Löwdin's symmetric orthonormalisation u = VS. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. The hypothesis is of the 'only zero' kind and gives no positive lower bound on the overlaps; producing one is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already one vector (r) with a real $r \neq 0$ is literally independent, and normalising it inverts r. *Used for:* Löwdin symmetric orthonormalisation of atomic orbitals under the textbook hypothesis that the orbitals are merely linearly independent. *Price:* **MP**; exact.

888. Gram matrix is positive semidefinite. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the k × k Gram matrix G with $G_{ij} = \langle v_i, v_j\rangle$ is positive semidefinite, where a square complex matrix A is positive semidefinite when $A^\dagger = A$ and the real part of $\langle x, Ax\rangle$ is ≥ 0 for every complex vector x. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. *Used for:* Overlap and correlation matrices: the overlap matrix $\langle \varphi_i, \varphi_j\rangle$ of any set of states, orbitals or coherent states, and the correlation matrix of a set of signals, are positive semidefinite. *Price:* **free**; exact.

889. Positive definite Gram matrix gives linear independence. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if the k × k Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$) is positive definite, then $v_1, \ldots, v_k$ are linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is). Here a square complex matrix A is positive definite when $A^\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries); $x \neq 0$ means only that x is not the zero vector (the negation of x = 0). *Used for:* Certifying a basis: a positive definite overlap matrix shows that a set of basis functions, coherent states or trial wavefunctions has no linear dependence. *Price:* **free**; exact.

890. Gram matrix of independent vectors is positive definite. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if $v_1, \ldots, v_k$ are linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is), then the k × k Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$) is positive definite. Here a square complex matrix A is positive definite when $A^\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries); $x \neq 0$ means only that x is not the zero vector (the negation of x = 0). For a coefficient vector $c \neq 0$ the hypothesis gives only that $c_1v_1 + \ldots + c_kv_k$ is not the zero vector; concluding $\langle c, Gc\rangle = \|c_1v_1 + \ldots + c_kv_k\|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Non-orthogonal bases: the overlap matrix of linearly independent atomic orbitals or tight-binding basis functions is positive definite, so the generalized eigenvalue problem Hc = ESc is well posed. *Price:* **MP**; exact.

891. Gram matrix positive definite exactly when vectors are independent. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the k × k Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$) is positive definite if and only if $v_1, \ldots, v_k$ are linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is). Here a square

complex matrix A is positive definite when $A\dagger = A$ and the real part of $\langle x, Ax\rangle$ is > 0 for every complex vector $x \neq 0$, with $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries); $x \neq 0$ means only that x is not the zero vector (the negation of $x = 0$). From left to right no principle is needed. From right to left, for a coefficient vector c $\neq 0$ the hypothesis gives only that $c_1v_1 + \ldots + c_kv_k$ is not the zero vector; concluding $\langle c, Gc\rangle = \|c_1v_1 + \ldots + c_kv_k\|^2 > 0$ is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* Testing a basis for overcompleteness: a set of Gaussian basis functions or coherent states is linearly independent exactly when its overlap matrix is positive definite, the standard check for near-linear dependence in quantum chemistry codes. *Price:* **MP**; exact.

892. Literal and uniform linear independence agree. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): $v_1, \ldots, v_k$ are linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers $c_1, \ldots, c_k$, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is) if and only if they are uniformly independent: there is a real number $\varepsilon > 0$ on the line with $\|c_1v_1 + \ldots + c_kv_k\|^2 \geq \varepsilon(|c_1|^2 + \ldots + |c_k|^2)$ for all complex numbers $c_1, \ldots, c_k$. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. From right to left no principle is needed. From left to right the hypothesis gives no positive length for any combination; producing the uniform bound ε is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* The two meanings of an independent basis in numerical physics: no nontrivial combination vanishes, versus the smallest eigenvalue of the overlap matrix is bounded away from zero (a finite condition number), which is what solvers and orthonormalisation routines check. *Price:* **MP**; exact.

893. QR decomposition of an invertible matrix. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that has a two-sided inverse ($AB = 1$ and $BA = 1$ for some B): there are complex $n \times n$ matrices Q, R and S with $Q\dagger Q = 1$ and $QQ\dagger = 1$ (Q is unitary), $A = QR$, R and S upper triangular (entries below the diagonal are 0), every diagonal entry R_{jj} a positive real number, and $RS = SR = 1$. *Used for:* The QR factorisation in numerical physics: QR iterations for the eigenvalues of Hamiltonians, stable least-squares solves, and the repeated re-orthonormalisation used to compute Lyapunov exponents of dynamical systems. *Price:* **free**; exact.

894. Columns independent exactly when the matrix is invertible. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the columns of A are linearly independent in the literal sense ($Ac = 0$ for a complex vector c implies $c = 0$) if and only if A has a two-sided inverse ($AB = 1$ and $BA = 1$ for some B). From right to left no principle is needed. From left to right the hypothesis gives no positive lower bound on $\|Ac\|$; building the inverse needs one, and getting it is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already at size 1 this is $A = (r)$ with a real $r \neq 0$. *Used for:* Square linear systems: n equations in n unknowns (normal-mode, circuit or equilibrium equations) have a unique solution for every right-hand side exactly when the columns of the coefficient matrix are independent. *Price:* **MP**; exact.

895. Gram–Schmidt on an arbitrary list of vectors. For all sizes n and k and every list of k complex vectors $f_1, \ldots, f_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), with no independence hypothesis: there are vectors $w_1, \ldots, w_k$ that are pairwise orthogonal ($\langle w_i, w_j\rangle = 0$ for $i \neq j$) and $k \times k$ upper triangular matrices R and S, each with every diagonal entry equal to 1, such that $f_j = R_{1j}w_1 + \ldots + R_{kj}w_k$ and $w_j = S_{1j}f_1 + \ldots + S_{kj}f_k$ for every j. So w_j is f_j minus a combination of $f_1, \ldots, f_{j-1}$, and a vector that is a combination of the earlier ones comes out as 0. Here $\langle x, y\rangle = \sum_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. The construction has to decide at every step whether the remaining residual is zero or not, and that decision is exactly what costs the limited principle of omniscience (LPO); already the two vectors $f_1 = (a)$ and $f_2 = (1)$ of length 1 decide whether the real number a is 0 or apart from 0. *Used for:* Gram–Schmidt on an overcomplete or possibly dependent set: orthogonalising a redundant list of states, functions or measurement vectors, where dependent vectors come out as zero and are pruned. *Price:* **LPO**; exact.

896. Gram–Schmidt output exists exactly for uniform independence. For all sizes n and k and every list of k complex vectors $v_1, \ldots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), the following are equivalent: (a) there are orthonormal vectors $u_1, \ldots, u_k$ of length n ($\langle u_i, u_j\rangle = 1$ if $i = j$ and 0 otherwise) and $k \times k$ upper triangular matrices R and S (entries below the diagonal are 0) with $RS = SR = 1$ and every diagonal entry R_{jj} a positive real number, such that $v_j = R_{1j}u_1 + \ldots + R_{kj}u_k$ and $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for

every j; (b) v_1, ..., v_k are uniformly independent: there is a real number $\varepsilon > 0$ on the line with $\|c_1v_1 + \ldots + c_kv_k\|^2 \geq \varepsilon(|c_1|^2 + \ldots + |c_k|^2)$ for all complex numbers c_1, ..., c_k. Since R and S are upper triangular, u_1, ..., u_j and v_1, ..., v_j span one common space for every j: this is Gram–Schmidt orthonormalisation, and the QR factorisation of the n × k matrix whose columns are v_1, ..., v_k. Here $\langle x, y\rangle = \Sigma_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. *Used for:* What orthonormalisation actually needs: a Gram–Schmidt or QR output exists exactly when the vectors have a uniform overlap bound, the quantity numerical codes monitor as the smallest overlap eigenvalue or the condition number. *Price:* **free**; exact.

897. Löwdin output exists exactly for uniform independence. For all sizes n and k and every list of k complex vectors v_1, ..., v_k of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), the following are equivalent: (a) there are orthonormal vectors u_1, ..., u_k of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and a k × k matrix S with $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j, S positive semidefinite ($S\dagger = S$ and the real part of $\langle x, Sx\rangle$ is ≥ 0 for every complex vector x), SGS = 1 for the Gram matrix G ($G_{ij} = \langle v_i, v_j\rangle$), and S invertible (ST = TS = 1 for some T); (b) v_1, ..., v_k are uniformly independent: there is a real number $\varepsilon > 0$ on the line with $\|c_1v_1 + \ldots + c_kv_k\|^2 \geq \varepsilon(|c_1|^2 + \ldots + |c_k|^2)$ for all complex numbers c_1, ..., c_k. In (a), S^2 is the inverse of G, so S is a positive semidefinite square root of G^{-1}, the matrix written $G^{-1/2}$ in Löwdin's symmetric orthonormalisation u = VS. Here $\langle x, y\rangle = \Sigma_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. *Used for:* What Löwdin orthonormalisation of atomic orbitals needs: exactly a uniform positive lower bound on the overlap matrix of the orbitals, no more and no less. *Price:* **free**; exact.

898. Gram–Schmidt output exists exactly for independent vectors. For all sizes n and k and every list of k complex vectors v_1, ..., v_k of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), the following are equivalent: (a) there are orthonormal vectors u_1, ..., u_k of length n ($\langle u_i, u_j\rangle = 1$ if i = j and 0 otherwise) and k × k upper triangular matrices R and S (entries below the diagonal are 0) with RS = SR = 1 and every diagonal entry R_{jj} a positive real number, such that $v_j = R_{1j}u_1 + \ldots + R_{kj}u_k$ and $u_j = S_{1j}v_1 + \ldots + S_{kj}v_k$ for every j; (b) v_1, ..., v_k are linearly independent in the literal sense: whenever $c_1v_1 + \ldots + c_kv_k = 0$ for complex numbers c_1, ..., c_k, every $c_i = 0$ (only the zero combination vanishes; this says nothing about how long a nonzero combination is). Since R and S are upper triangular, u_1, ..., u_j and v_1, ..., v_j span one common space for every j: this is Gram–Schmidt orthonormalisation, and the QR factorisation of the n × k matrix whose columns are v_1, ..., v_k. Here $\langle x, y\rangle = \Sigma_i \bar{x}_i y_i$ (summed over all entries) and $\|x\|^2 = \langle x, x\rangle$. From (a) to (b) no principle is needed. From (b) to (a) the hypothesis gives no positive length for the residual vectors; normalising them is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. *Used for:* The textbook equivalence: a set of states can be orthonormalised by Gram–Schmidt with a triangular change of basis exactly when the states are linearly independent. *Price:* **MP**; exact.

899. Rows independent exactly when the matrix is invertible. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the rows of A are linearly independent in the literal sense (if c_1 times row 1 plus ... plus c_n times row n is the zero row for complex numbers c_1, ..., c_n, then every $c_i = 0$) if and only if A has a two-sided inverse (AB = 1 and BA = 1 for some B). From right to left no principle is needed. From left to right the hypothesis gives no positive lower bound on the length of a nonzero combination of rows; building the inverse needs one, and getting it is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already at size 1 this is A = (r) with a real $r \neq 0$. *Used for:* Independent equations: n linear constraints or measurement rows on n unknowns determine the unknowns uniquely, through an inverse matrix, exactly when the rows are independent; also invertibility of a change-of-basis matrix read from its rows. *Price:* **MP**; exact.

900. Hermitian block matrix and its Schur complement. For all sizes m and k, every complex m × m matrix A with $A\dagger = A$, every complex m × k matrix B, every complex k × k matrix D and every complex m × m matrix W that is either a two-sided inverse of A (AW = 1 and WA = 1) or the zero matrix (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the block matrix M with rows (A B) and (B† D), of size m + k, is Hermitian ($M\dagger = M$) if and only if the Schur complement D − B†WB is Hermitian. Here X† is the conjugate transpose of X. W stands for A^{-1} in the convention that gives a matrix with no inverse the inverse 0; the statement holds for either value. *Used for:* Löwdin partitioning (Feshbach projection): the effective Hamiltonian of the retained states is a Schur complement D − B†A^{-1}B, with A the eliminated block shifted by a real energy, and it is Hermitian exactly when the full Hamiltonian is. *Price:* **free**; exact.

901. Hermitian block matrix, Schur complement of the second block. For all sizes m

and k, every complex m × m matrix A, every complex m × k matrix B, every complex k × k matrix D with D† = D and every complex k × k matrix W that is either a two-sided inverse of D (DW = 1 and WD = 1) or the zero matrix (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the block matrix M with rows (A B) and (B† D), of size m + k, is Hermitian (M† = M) if and only if the Schur complement A − BWB† is Hermitian. Here X† is the conjugate transpose of X. W stands for D^{-1} in the convention that gives a matrix with no inverse the inverse 0; the statement holds for either value. *Used for:* Low-energy effective Hamiltonians: eliminating a high-energy block D (shifted by a real energy) leaves A − $BD^{-1}B$† acting on the low-energy states, and that operator is Hermitian exactly when the full Hamiltonian is. *Price:* **free**; exact.

902. Positive semidefinite block matrix and its Schur complement. For all sizes m and k, every positive definite complex m × m matrix A, every complex m × k matrix B, every complex k × k matrix D and every complex m × m matrix W with AW = 1 and WA = 1 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the block matrix M with rows (A B) and (B† D), of size m + k, is positive semidefinite if and only if the Schur complement D − B†WB is positive semidefinite. Here X† is the conjugate transpose of X; a square complex matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, and positive definite when G† = G and the real part of ⟨x, Gx⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = Σ_i $\bar{x}_i y_i$ summed over all entries; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The inverse W of A is given. *Used for:* Gaussian conditioning in statistical physics and estimation: a joint covariance matrix with blocks A, B, B†, D is a valid covariance (positive semidefinite) exactly when the conditional covariance D − B†A^{-1}B is. *Price:* **free**; exact.

903. Positive semidefinite block matrix, Schur complement of the second block. For all sizes m and k, every complex m × m matrix A, every complex m × k matrix B, every positive definite complex k × k matrix D and every complex k × k matrix W with DW = 1 and WD = 1 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the block matrix M with rows (A B) and (B† D), of size m + k, is positive semidefinite if and only if the Schur complement A − BWB† is positive semidefinite. Here X† is the conjugate transpose of X; a square complex matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, and positive definite when G† = G and the real part of ⟨x, Gx⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = Σ_i $\bar{x}_i y_i$ summed over all entries; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The inverse W of D is given. *Used for:* Eliminating stiff or fast coordinates from a quadratic energy (static condensation, adiabatic elimination): with a positive definite block D, the full stiffness or Hessian matrix is positive semidefinite exactly when the effective stiffness A − $BD^{-1}B$† of the remaining coordinates is. *Price:* **free**; exact.

904. Positive definite block matrix and its Schur complement. For all sizes m and k, every positive definite complex m × m matrix A, every complex m × k matrix B, every complex k × k matrix D and every complex m × m matrix W with AW = 1 and WA = 1 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the block matrix M with rows (A B) and (B† D), of size m + k, is positive definite if and only if the Schur complement D − B†WB is positive definite. Here X† is the conjugate transpose of X; a square complex matrix G is positive definite when G† = G and the real part of ⟨x, Gx⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = Σ_i $\bar{x}_i y_i$ summed over all entries; x ≠ 0 means only that x is not the zero vector (the negation of x = 0). The inverse W of A is given. No principle beyond the hypotheses is needed: a positive definite matrix with at least one row already supplies the step from 'not zero' to 'positive' (Markov's principle), and when m = k = 0 there is nothing to prove. *Used for:* Stability of coupled systems: the Hessian of a two-block energy, such as strain coupled to an order parameter or two coupled sets of oscillators, is positive definite (a strict minimum, a stable equilibrium) exactly when one block and its Schur complement are; this test also decides convergence of a Gaussian integral over a block-partitioned quadratic form. *Price:* **free**; exact.

905. Finite sum of positive semidefinite matrices. For every size n, every number of terms N and all complex n × n matrices X(1), X(2), ..., X(N) (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that are each positive semidefinite: the sum X(1) + X(2) + ... + X(N) is positive semidefinite (for N = 0 the sum is the zero matrix). Here a square complex matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, with ⟨x, y⟩ = $\bar{x}_1 y_1$ + ... + $\bar{x}_n y_n$. *Used for:* Mixtures and noise sums: a density matrix written as a weighted sum of pure-state projectors, a Kraus sum Σ $K_i \rho K_i$† whose terms are each positive semidefinite, and a noise covariance summed over independent sources are positive semidefinite. *Price:* **free**; exact.

906. Resolvent of a Hermitian matrix off the real axis. For every size n, every complex n × n matrix H with H† = H (entries in the complex

numbers over the line, whose real and imaginary parts lie on the line) and every complex number z with $(\mathrm{Im}\ z)^2 > 0$, which holds whenever $\mathrm{Im}\ z > 0$ or $\mathrm{Im}\ z < 0$: the matrix $z - H$ (that is, $z \cdot 1 - H$) has a two-sided inverse, a complex $n \times n$ matrix R with $(z - H)R = 1$ and $R(z - H) = 1$. R is the resolvent $(z - H)^{-1}$. The hypothesis is a positive bound on Im z, not merely that Im z is not zero. *Used for:* Retarded and advanced Green's functions $G(E \pm i\eta) = (E \pm i\eta - H)^{-1}$ of a finite Hamiltonian exist for every real energy E and every broadening $\eta > 0$: spectral functions, local densities of states and Green's-function transport calculations at finite η. *Price:* **free**; exact.

907. Resolvent when the imaginary part is not zero. For every size n, every complex $n \times n$ matrix H with $H\dagger = H$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number z with $\mathrm{Im}\ z \neq 0$: the matrix $z - H$ (that is, $z \cdot 1 - H$) has a two-sided inverse, a complex $n \times n$ matrix R with $(z - H)R = 1$ and $R(z - H) = 1$. The hypothesis $\mathrm{Im}\ z \neq 0$ says only that $\mathrm{Im}\ z = 0$ is false, the negation of an equation; obtaining an inverse from it is the step that costs Markov's principle, which on the line says exactly that a real number that is not zero has an inverse. Already at size 1 with $H = 0$ the statement says that $i \cdot r$ has an inverse for every real $r \neq 0$. *Used for:* The textbook form of 'a Hermitian Hamiltonian has a Green's function $(z - H)^{-1}$ at every complex energy off the real axis', with off the axis stated as $\mathrm{Im}\ z \neq 0$, as in complex-energy contour integrals and the analytic continuation of Green's functions. *Price:* **MP**; exact.

908. Points of the spectrum of a Hermitian matrix are real. For every size n, every complex $n \times n$ matrix H with $H\dagger = H$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number z: if $z - H$ (that is, $z \cdot 1 - H$) has no two-sided inverse, then $\mathrm{Im}\ z = 0$. So every point of the spectrum of H, where $z - H$ fails to be invertible, lies on the real axis. *Used for:* Energies of a closed quantum system are real: the poles of the Green's function $(z - H)^{-1}$ of a Hermitian Hamiltonian lie on the real energy axis, and complex poles (resonances with a finite lifetime) need a non-Hermitian effective Hamiltonian. *Price:* **free**; exact.

909. Resolvent formula with a real positive denominator. For every size n, every complex $n \times n$ matrix H with $H\dagger = H$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), every complex number z and every complex $n \times n$ matrix W that is a two-sided inverse of $Q = (H - \mathrm{Re}\ z)^2 + (\mathrm{Im}\ z)^2 \cdot 1$ ($QW = 1$ and $WQ = 1$): the matrix $(\bar{z} - H)W$ is a two-sided inverse of $z - H$, where $\bar{z}$ is the complex conjugate of z and $z - H$ means $z \cdot 1 - H$. So $(z - H)^{-1} = (\bar{z} - H)((H - \mathrm{Re}\ z)^2 + (\mathrm{Im}\ z)^2)^{-1}$; the inverse W is given. *Used for:* Lorentzian broadening: with $z = E + i\eta$ the formula gives the anti-Hermitian part of $G(E + i\eta)$ as $-i\eta((H - E)^2 + \eta^2)^{-1}$, the matrix Lorentzian behind spectral functions and broadened densities of states. *Price:* **free**; exact.

910. First resolvent identity. For every size n, every complex $n \times n$ matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line; H need not be Hermitian), all complex numbers z and w, and all complex $n \times n$ matrices R(z) and R(w) such that R(z) is a two-sided inverse of $z - H$ and R(w) is a two-sided inverse of $w - H$ ($z - H$ meaning $z \cdot 1 - H$): $R(z) - R(w) = (w - z) \cdot R(z)R(w)$. Both resolvents are given. *Used for:* Green's functions at two complex energies: $G(z) - G(w) = (w - z)G(z)G(w)$ shows that the derivative of G with respect to z is $-G^2$, and gives the analytic structure of G used in spectral sum rules and many-body perturbation theory. *Price:* **free**; exact.

911. Resolvent norm bound by the imaginary part. For every size n, every complex $n \times n$ matrix H with $H\dagger = H$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), every complex number z and every complex $n \times n$ matrix R with $(z - H)R = 1$ and $R(z - H) = 1$ ($z - H$ meaning $z \cdot 1 - H$): $(\mathrm{Im}\ z)^2 \cdot \|Rx\|^2 \le \|x\|^2$ for every complex vector x, where $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. So the resolvent $(z - H)^{-1}$ has operator norm at most $1/|\mathrm{Im}\ z|$, here written without division. *Used for:* A broadening η bounds the response of a finite system: $\|G(E + i\eta)\| \le 1/\eta$ at every real energy E, the basic a priori estimate in linear-response theory and in Green's-function numerics at finite broadening. *Price:* **free**; exact.

912. Resolvent at a real energy inside a spectral gap. For every size n, every complex $n \times n$ matrix H with $H\dagger = H$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and all real numbers c and r on the line such that H has a spectral gap of half-width r around c, meaning $r > 0$ and $r^2\|x\|^2 \le \|(H - c)x\|^2$ for every complex vector x (so no eigenvalue lies strictly between $c - r$ and $c + r$): the matrix $c - H$ (that is, $c \cdot 1 - H$) has a two-sided inverse R, and $r^2\|Rx\|^2 \le \|x\|^2$ for every complex vector x. Here $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. So the resolvent $(c - H)^{-1}$ at the real number c exists and has operator norm at most $1/r$. *Used for:* Green's function at an energy inside a band gap: it exists and is bounded by one over the distance to the spectrum, which controls gap-energy Green's functions and evanescent states in insulators, perturbation theory around a gapped ground state and Riesz projections onto isolated bands. *Price:* **free**; exact.

913. Difference of two inverse matrices. For every size n and all complex $n \times n$ matrices A, B, W

and W′ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) such that either W is a two-sided inverse of A and W′ is a two-sided inverse of B (AW = WA = 1 and BW′ = W′B = 1), or W = 0 and W′ = 0: W − W′ = W(B − A)W′. When A and B are invertible this is $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$; the second case covers the convention that assigns the inverse 0 to a matrix with none, as when A and B are both singular. *Used for:* Dyson's equation in difference form: with A = E − H_0 − V and B = E − H_0 the identity reads G − G_0 = G V G_0, relating full and unperturbed propagators; it also gives the first-order change of any inverted response matrix. *Price:* **free**; exact.

914. Kronecker product of positive semidefinite matrices. For all sizes n and m, every positive semidefinite complex n × n matrix A and every positive semidefinite complex m × m matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the Kronecker product A ⊗ B is positive semidefinite. Here A ⊗ B is indexed by the n·m pairs (i, i′) and has entry A(i, j)·B(i′, j′) at row (i, i′) and column (j, j′); a square complex matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, with ⟨x, y⟩ = $\Sigma_p \bar{x}_p y_p$ summed over all pair indices p (respectively all indices of A or B). *Used for:* Uncorrelated composite systems: for density matrices ρ and σ of two independently prepared subsystems, ρ ⊗ σ is positive semidefinite, a valid joint state; separable (Kronecker) covariance models of space-time noise are valid covariances by this theorem. *Price:* **free**; exact.

915. Kronecker product of positive definite matrices. For all sizes n and m, every positive definite complex n × n matrix A and every positive definite complex m × m matrix B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the Kronecker product A ⊗ B is positive definite. Here A ⊗ B is indexed by the n·m pairs (i, i′) and has entry A(i, j)·B(i′, j′) at row (i, i′) and column (j, j′); a square complex matrix G is positive definite when G† = G and the real part of ⟨x, Gx⟩ is > 0 for every complex vector x ≠ 0, with ⟨x, y⟩ = $\Sigma_p \bar{x}_p y_p$ summed over all pair indices p (respectively all indices of A or B); x ≠ 0 means only that x is not the zero vector (the negation of x = 0). No principle beyond the hypotheses is needed: a positive definite A with at least one row already supplies the step from 'not zero' to 'positive' (Markov's principle), and when n = 0 there is nothing to prove. *Used for:* Thermal states of non-interacting subsystems: the product $e^{-\beta H_1}/Z_1 \otimes e^{-\beta H_2}/Z_2$ of two full-rank Gibbs states is positive definite, so the joint state is faithful and its logarithm and relative entropies are defined; and a tensor product of positive definite metrics is a positive definite metric on the tensor-product space. *Price:* **free**; exact.

916. Woodbury identity for a low-rank update. For all sizes n and k, every complex n × n matrix A with a given two-sided inverse A^{-1}, every complex n × k matrix U, every complex k × k matrix C with a given two-sided inverse C^{-1}, every complex k × n matrix V, and every complex k × k matrix S that is a two-sided inverse of $C^{-1} + VA^{-1}U$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the matrix $Y = A^{-1} - A^{-1}USVA^{-1}$ is a two-sided inverse of A + UCV, and every left or right inverse of A + UCV equals Y. So $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, with each inverse on the right given as a two-sided inverse. *Used for:* Dyson equation for a rank-k perturbation: the Green's function of H_0 plus a rank-k coupling follows from that of H_0 and one k × k inversion (T-matrix of a separable potential, impurity Green's functions, Lippmann–Schwinger equations with finite-rank interactions); also the covariance update of the Kalman filter. *Price:* **free**; exact.

917. Binomial inverse theorem for a low-rank update. For all sizes n and k, every complex n × n matrix A with a given two-sided inverse A^{-1}, every complex n × k matrix U, every complex k × k matrix C (not assumed invertible), every complex k × n matrix V, and every complex k × k matrix T that is a two-sided inverse of $C + CVA^{-1}UC$ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): the matrix $Y = A^{-1} - A^{-1}UCTCVA^{-1}$ is a two-sided inverse of A + UCV, and every left or right inverse of A + UCV equals Y. So $(A + UCV)^{-1} = A^{-1} - A^{-1}UC(C + CVA^{-1}UC)^{-1}CVA^{-1}$, with each inverse on the right given as a two-sided inverse. *Used for:* Rank-k Green's-function and covariance updates, as with the Woodbury identity (impurity and separable-potential T-matrices, Kalman-type updates), arranged so that only the k × k matrix $C + C{\cdot}V{\cdot}A^{-1}{\cdot}U{\cdot}C$ is inverted and the inverse of C never appears. *Price:* **free**; exact.

918. Matrix Chebyshev series with coefficients given as numbers. For every size n, every complex n × n matrix H with H† = H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and $-1 \preceq H \preceq 1$ (both 1 − H and 1 + H positive semidefinite), and every sequence of real coefficients $a_0, a_1, a_2, \ldots$ on the line, given only as numbers, with an explicit tail modulus for $\Sigma\,|a_k|$ (a function K such that $|a_p| + |a_{p+1}| + \ldots + |a_{p+m-1}| \le 1/(j+1)$ for every j, every $p \ge K(j)$ and every m): there is a complex n × n matrix F such that the partial sums of $a_0T_0(H) + a_1T_1(H) + a_2T_2(H) + \ldots$ converge entry by entry to F (real and imaginary parts), F† = F and FH =

HF; and for every n × n matrix V with V†V = 1 and real numbers $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$ such that HV = V·diag($\lambda_1, \ldots, \lambda_n$) and, for each k, the scalar series $a_0T_0(\lambda_k) + a_1T_1(\lambda_k) + \ldots$ converges to μ_k, FV = V·diag($\mu_1, \ldots, \mu_n$). So F = f(H) for f(t) = $\Sigma\, a_kT_k(t)$ on every orthonormal eigenbasis. Here $T_0(H) = 1$, $T_1(H) = H$ and $T_{j+2}(H) = 2H\cdot T_{j+1}(H) - T_j(H)$, and $T_j(t)$ for a number t follows this recursion with t in place of H; a matrix G is positive semidefinite when G† = G and the real part of ⟨x, Gx⟩ is ≥ 0 for every complex vector x, with $\langle x, y\rangle = \bar{x}_1y_1 + \ldots + \bar{x}_ny_n$. Because the coefficients are given only as numbers, forming the limit uses the completeness of the line (a Cauchy sequence of points with a modulus has a limit), and the statement implies that completeness: at size 1 with H = 1 it says that every real series with an explicit tail modulus for $\Sigma\, |a_k|$ has a sum. *Used for:* Kernel polynomial method and Chebyshev propagators: $e^{-\beta H}$, cos(Ht), sin(Ht) or a Fermi function of a Hamiltonian rescaled into [−1, 1] is computed as the limit of $\Sigma\, a_kT_k(H)$, with no decision about the spectrum of H, here for coefficient sequences specified only as numbers. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

919. Moore–Penrose pseudoinverse of any square matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), with no rank hypothesis: there is an n × n matrix B, the Moore–Penrose pseudoinverse A^+, with ABA = A, BAB = B, (AB)† = AB and (BA)† = BA, where M† is the conjugate transpose of M. The construction runs Gram–Schmidt on the columns of A and of A† and must decide, for each column, whether its residual length (the pivot) is zero or apart from zero; that decision is exactly what costs the limited principle of omniscience (LPO). Already the 1 × 1 matrix A = (a) with a real forces it: ABA = A gives a(ar − 1) = 0 for the real part r of B, and that decides whether a is 0 or apart from 0. No countable choice is used. *Used for:* Inverting a matrix that may be singular: least-squares fits to a redundant or degenerate model, effective operators and response matrices with zero modes, and the pseudoinverse of a singular Fisher or covariance matrix in data analysis. *Price:* **LPO**; exact.

920. Generalized inverse of any square matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), with no rank hypothesis: there is an n × n matrix B with ABA = A (a generalized inverse, also called an inner inverse of A). Building B decides, for each column of A, whether its Gram–Schmidt residual length (the pivot) is zero or apart from zero, and that decision is exactly what costs the limited principle of omniscience (LPO). The lower bound is the 1 × 1 matrix A = (a) with a real: ABA = A gives a(ar − 1) = 0 for the real part r of B, which decides whether a is 0 or apart from 0. No countable choice is used. *Used for:* Solving a consistent linear system Ax = y whose matrix may be singular: if y = Ax_0 for some x_0, then x = By solves it, since A(By) = $ABAx_0 = Ax_0$ = y; used for constraint equations with gauge freedom and for linear response in redundant coordinates. *Price:* **LPO**; exact.

921. Normal equations of least squares are solvable. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex vector b of length n, with no rank hypothesis on A: the normal equations A†Ax = A†b have a solution x, where A† is the conjugate transpose of A. These are the equations of the least-squares problem of making ‖Ax − b‖ as small as possible; the row asserts only that they can be solved. Solving them decides, for each column of A, whether its Gram–Schmidt residual length (the pivot) is zero or apart from zero, and that decision is exactly what costs the limited principle of omniscience (LPO): already A = (a) with a real and b = (1) give $a^2x = a$, so a(ar − 1) = 0 for the real part r of the solution, which decides whether a is 0 or apart from 0. No countable choice is used. *Used for:* Linear least-squares fitting of measurements to a model that may be degenerate or over-parametrised: the fitted parameters solve the normal equations, as in detector calibration, spectral unmixing and linearised parameter estimation. *Price:* **LPO**; exact.

922. Orthogonal projector onto the range of a matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), with no rank hypothesis: there is an n × n matrix P that is Hermitian (P† = P) and idempotent ($P^2 = P$), with PA = A and Px = 0 for every complex vector x with A†x = 0. So P is the orthogonal projector onto the range (column space) of A: it fixes every column of A and kills every vector orthogonal to all the columns. The price is the weak limited principle of omniscience (WLPO) together with countable choice. The upper bound takes a singular value decomposition A = U diag($s_1, \ldots, s_n$) V† (which costs LLPO with countable choice, and WLPO gives LLPO) and decides only whether each singular value is zero or not zero, never an inverse of it, setting P = U diag($e_1, \ldots, e_n$) U† with each e_j equal to 0 or 1. The lower bound, with no choice, is the 1 × 1 matrix A = (a) with a real: P = (p) has p real, $p^2 = p$, pa = a and p = 0 when a = 0, and comparing p with 0 and 1 decides whether a = 0 or a ≠ 0. *Used for:* The projector onto

the column space of a design or measurement matrix (the hat matrix of least squares), the signal subspace in subspace methods of spectroscopy and array processing, and restriction to the states reachable through a given coupling, all without knowing the rank in advance. *Price:* **WLPO + AC$_\omega$**; exact.

923. Orthogonal projector onto the kernel of a matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), with no rank hypothesis: there is an n × n matrix K that is Hermitian (K† = K) and idempotent (K² = K), with AK = 0 and Kx = x for every complex vector x with Ax = 0. So K is the orthogonal projector onto the kernel (null space) of A: its range is exactly the set of zero modes of A. The price is the weak limited principle of omniscience (WLPO) together with countable choice. The upper bound is K = 1 − Q, with Q the orthogonal projector onto the range of A†, which comes from a singular value decomposition (LLPO with countable choice; WLPO gives LLPO) by deciding only whether each singular value is zero or not zero. The lower bound, with no choice, is the 1 × 1 matrix A = (a) with a real: K = (k) has k real, $k^2 = k$, ak = 0 and k = 1 when a = 0, and comparing k with 0 and 1 decides whether a = 0 or a ≠ 0. *Used for:* Isolating zero modes without knowing how many there are: gauge directions of a constraint matrix, translational and rotational zero modes of a Hessian or stiffness matrix, and the zero-energy subspace of a Hamiltonian. *Price:* **WLPO + AC$_\omega$**; exact.

924. Pseudoinverse of a matrix with invertible determinant. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex number e with det(A)·e = 1 (an inverse of the determinant is given): there is an n × n matrix B with AB = 1 and BA = 1 that also satisfies the four Penrose equations ABA = A, BAB = B, (AB)† = AB and (BA)† = BA, where M† is the conjugate transpose of M. So for a matrix of full rank, given through an invertible determinant, the Moore–Penrose pseudoinverse is the ordinary inverse; the proof takes B = e·adj(A), with adj(A) the adjugate of A, and needs no decision. *Used for:* Matrices certified nonsingular by an explicit inverse of the determinant, such as a nondegenerate metric, a normal-mode matrix with no zero frequency or the Jacobian of a change of variables: inverse and pseudoinverse coincide and cost nothing. *Price:* **free**; exact.

925. Pseudoinverse with uniformly independent columns. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) whose columns are uniformly independent, meaning there is a real number ε > 0 on the line with $\|Ac\|^2 \geq \varepsilon\|c\|^2$ for every complex vector c, where $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$: there is an n × n matrix B with AB = 1 and BA = 1 that satisfies the four Penrose equations ABA = A, BAB = B, (AB)† = AB and (BA)† = BA, where M† is the conjugate transpose of M. So for a square matrix of full rank, given as a certified lower bound on $\|Ac\|$, the Moore–Penrose pseudoinverse is the two-sided inverse, and finding it needs no decision. *Used for:* Inverting a well-conditioned square system whose smallest singular value has a known positive lower bound, as for a calibrated detector response or a transfer matrix with a certified gap above zero: the pseudoinverse is the inverse and costs nothing. *Price:* **free**; exact.

926. Least squares with uniformly independent vectors. For all sizes n and k, every list of k complex vectors $v_1, \dots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is uniformly independent, meaning there is a real number ε > 0 on the line with $\|c_1v_1 + \dots + c_kv_k\|^2 \geq \varepsilon(|c_1|^2 + \dots + |c_k|^2)$ for all complex numbers $c_1, \dots, c_k$, and every complex vector b of length n: the equations Gc = $(\langle v_1, b\rangle, \dots, \langle v_k, b\rangle)$ have exactly one solution c = $(c_1, \dots, c_k)$, where G is the k × k Gram matrix with entries $G_{ij} = \langle v_i, v_j\rangle$. For the n × k matrix A whose columns are $v_1, \dots, v_k$, these are the normal equations A†Ac = A†b of fitting b by $c_1v_1 + \dots + c_kv_k$ in least squares. Here $\langle x, y\rangle = \bar{x}_1y_1 + \dots + \bar{x}_ny_n$ and $\|x\|^2 = \langle x, x\rangle$. No decision is needed: G has a two-sided inverse. *Used for:* Fitting data to k known templates or basis functions with a certified non-degenerate overlap: the coefficients of a spectral fit with known line shapes, or of a state expanded in a non-orthogonal basis, exist and are unique. *Price:* **free**; exact.

927. Projection onto the span of independent vectors. For all sizes n and k and every list of k complex vectors $v_1, \dots, v_k$ of length n (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is uniformly independent, meaning there is a real number ε > 0 on the line with $\|c_1v_1 + \dots + c_kv_k\|^2 \geq \varepsilon(|c_1|^2 + \dots + |c_k|^2)$ for all complex numbers $c_1, \dots, c_k$: there is an n × n matrix P that is Hermitian (P† = P) and idempotent ($P^2 = P$), fixes every v_i ($Pv_i = v_i$) and kills every vector orthogonal to all of them (if $\langle v_i, x\rangle = 0$ for every i, then Px = 0). So P is the orthogonal projection onto the span of $v_1, \dots, v_k$. Here $\langle x, y\rangle = \bar{x}_1y_1 + \dots + \bar{x}_ny_n$ and $\|x\|^2 = \langle x, x\rangle$. No decision is needed: the proof orthonormalises the vectors and adds up the projectors onto the results. *Used for:* Projecting onto a model space spanned by known independent states: a set of non-orthogonal atomic orbitals in quantum

chemistry, the model space of an effective Hamiltonian, or the span of independent signal templates. *Price:* **free**; exact.

928. Logarithm of a uniformly positive matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian ($A\dagger = A$) and uniformly positive, meaning there is a real number $s > 0$ on the line with $\mathrm{Re}\langle x, Ax\rangle \geq s\|x\|^2$ for every complex vector x: there is an n × n matrix L, the logarithm $\log A$, that is Hermitian ($L\dagger = L$), commutes with A ($LA = AL$) and has $\exp(L) = A$, where $\exp(X) = E$ means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E. L acts as the ordinary logarithm on every eigenbasis: whenever V is an n × n matrix with $V\dagger V = 1$ and $AV = V\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ for real numbers $\lambda_1, \dots, \lambda_n$, and $\mu_1, \dots, \mu_n$ are real numbers with $\exp(\mu_k) = \lambda_k$, then $LV = V\,\mathrm{diag}(\mu_1, \dots, \mu_n)$. No eigenbasis is used to build L: it comes from the series $\log(1 - Y) = -(Y + Y^2/2 + Y^3/3 + \dots)$ applied to $Y = 1 - A/K$ for a suitable real number $K > 0$. Here $\langle x, y\rangle = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$ and $\|x\|^2 = \langle x, x\rangle$. *Used for:* The logarithm of a positive operator: the modular (entanglement) Hamiltonian $-\log \rho$ of a faithful state, the relative entropy $\mathrm{tr}\,\rho(\log \rho - \log \sigma)$, and an effective Hamiltonian read off as the logarithm of a positive transfer matrix. *Price:* **free**; exact.

929. Real powers of a uniformly positive matrix. For every size n and every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian ($A\dagger = A$) and uniformly positive, meaning there is a real number $s > 0$ on the line with $\mathrm{Re}\langle x, Ax\rangle \geq s\|x\|^2$ for every complex vector x: there are a Hermitian n × n matrix L that commutes with A and has $\exp(L) = A$, and a matrix P(t) for every real number t on the line with $P(t) = \exp(t \cdot L)$, such that every P(t) is Hermitian and commutes with A, $P(0) = 1$, $P(1) = A$, $P(t)P(u) = P(t + u)$ for all real numbers t and u, and $P(1/2)$ is positive semidefinite (Hermitian with $\mathrm{Re}\langle x, P(1/2)x\rangle \geq 0$ for every x) with $P(1/2)^2 = A$. So $P(t) = A^t = \exp(t \log A)$ are the real powers of A, negative powers included, and $P(1/2)$ is a positive square root of A. Here $\exp(X) = E$ means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\langle x, y\rangle = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$ and $\|x\|^2 = \langle x, x\rangle$. *Used for:* Real powers of a positive operator: $\rho^{1/2}$ in the fidelity and in purifications, ρ^{α} in Rényi entropies, $S^{-1/2}$ for an overlap matrix S in Löwdin orthogonalisation, and interpolations between two positive metrics. *Price:* **free**; exact.

930. Entropy sign for a state bounded by the identity. For every size n and every complex n × n matrix ρ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian, uniformly positive (there is a real number $s > 0$ on the line with $\mathrm{Re}\langle x, \rho x\rangle \geq s\|x\|^2$ for every complex vector x) and below the identity, $\rho \preceq 1$ ($\mathrm{Re}\langle x, \rho x\rangle \leq \|x\|^2$ for every x): there is an n × n matrix L, the logarithm of ρ, that is Hermitian, commutes with ρ, has $\exp(L) = \rho$, and satisfies $LV = V\,\mathrm{diag}(\mu_1, \dots, \mu_n)$ whenever $V\dagger V = 1$, $\rho V = V\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ with real numbers λ_k, and μ_k are real numbers with $\exp(\mu_k) = \lambda_k$; and the real part of $\mathrm{tr}(\rho L)$ is ≤ 0. So the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho \log \rho)$ is nonnegative (stated for its real part). Here $\exp(X) = E$ means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\langle x, y\rangle = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$ and $\|x\|^2 = \langle x, x\rangle$. *Used for:* Nonnegativity of the von Neumann entropy of a full-support state given with the operator bound $\rho \preceq 1$, as for a reduced density matrix of a subsystem: the base case of entropy inequalities for finite quantum systems. *Price:* **free**; exact.

931. Entropy sign for a full-rank density matrix. For every size n and every complex n × n matrix ρ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian, uniformly positive (there is a real number $s > 0$ on the line with $\mathrm{Re}\langle x, \rho x\rangle \geq s\|x\|^2$ for every complex vector x) and of trace 1 (the real part of $\rho_{11} + \dots + \rho_{nn}$ is 1, which for a Hermitian matrix is the whole trace): there is an n × n matrix L, the logarithm of ρ, that is Hermitian, commutes with ρ, has $\exp(L) = \rho$, and satisfies $LV = V\,\mathrm{diag}(\mu_1, \dots, \mu_n)$ whenever $V\dagger V = 1$, $\rho V = V\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ with real numbers λ_k, and μ_k are real numbers with $\exp(\mu_k) = \lambda_k$; and the real part of $\mathrm{tr}(\rho L)$ is ≤ 0. So the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho \log \rho)$ of a density matrix of full rank is nonnegative (stated for its real part). The row covers only density matrices bounded below by a positive multiple of the identity: a density matrix with a zero eigenvalue, such as a pure state, is outside it. Here $\exp(X) = E$ means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\langle x, y\rangle = \bar{x}_1 y_1 + \dots + \bar{x}_n y_n$ and $\|x\|^2 = \langle x, x\rangle$. *Used for:* The von Neumann entropy of a mixed state with full support is nonnegative, for example a thermal state $\exp(-\beta H)/Z$ at finite temperature: the entropy S in the free energy $F = E - TS$ and in the second law. *Price:* **free**; exact.

932. Lie–Trotter product formula with its error bound. For every size n, all complex n × n matrices H_1 and H_2 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that are Hermitian ($H_1\dagger = H_1$, $H_2\dagger = H_2$), every real number t on the line, every real number $c \geq 0$ that bounds the commutator in operator norm ($\|(H_1H_2 - H_2H_1)x\| \leq c\|x\|$ for every complex vector x) and every

number of steps $M = m + 1 \geq 1$: if $E = \exp(-it(H_1 + H_2))$, $E_1 = \exp(-itH_1/M)$ and $E_2 = \exp(-itH_2/M)$, then $\|(E - (E_1E_2)^M)x\| \leq (t^2c/(2M))\|x\|$ for every complex vector x. So alternating the two evolutions in M short steps reproduces the full evolution with operator-norm error at most $t^2\|[H_1, H_2]\|/(2M)$. Here $\exp(X) = E$ means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$. *Used for:* Trotterised time evolution: split-operator integration of the Schrödinger equation, Suzuki–Trotter steps in digital quantum simulation and time slicing in path integrals, with the number of steps for a target accuracy read off from $t^2\|[H_1, H_2]\|/(2M)$. *Price:* **free**; exact.

933. Every complex matrix has an eigenvector. For every size $n \geq 1$ and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): there are a complex vector v with $\langle v, v\rangle = 1$ and a complex number μ with $Av = \mu v$, where $\langle x, y\rangle = \bar{x}_1y_1 + \dots + \bar{x}_ny_n$. So every complex square matrix has a unit eigenvector. Already 2×2 Hermitian matrices force LLPO, with no choice. With no principle at all, the statement gives an exact root of every monic complex polynomial of degree at least 1, through its companion matrix. *Used for:* Existence of a normal mode or stationary state for any finite linear dynamics, Hermitian or not: damped oscillators, non-Hermitian effective Hamiltonians of open systems, transfer and Floquet matrices, and rate matrices of Markov processes. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact: the statement implies LLPO with no choice; conversely, under LLPO and countable choice the statement holds exactly when the approximate fundamental theorem of algebra holds (every monic complex polynomial p of degree at least 1 has, for every k, a point z with $|p(z)|^2 \leq 1/(k+1)$), and the exact fundamental theorem of algebra, which is free (row 482), gives the approximate one.

934. Schur triangularization of a complex matrix. For every size n and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): there are $n \times n$ matrices U and T with $U\dagger U = 1$ (U is unitary), T upper triangular (every entry below the diagonal is 0) and $AU = UT$, that is $A = UTU\dagger$. With no principle, this is equivalent to the statement that every complex $n \times n$ matrix with $n \geq 1$ has a unit eigenvector (row 933): the first column of U is such an eigenvector, and conversely unit eigenvectors triangularise A one column at a time. *Used for:* Triangularising a non-normal matrix by a unitary change of basis: the Schur form behind numerically stable eigenvalue algorithms, the analysis of non-Hermitian Hamiltonians and Liouvillians of open quantum systems, and transient growth in hydrodynamic stability. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact: through the equivalence above it has the price of row 933, which rests on the free fundamental theorem of algebra (row 482); the statement implies LLPO with no choice.

935. Operator norm of a complex matrix exists. For every size $n \geq 1$ and every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): there is a real number c on the line that is the operator norm of A and also the operator norm of $A\dagger$, and whose square c^2 is the operator norm of $A\dagger A$, where $A\dagger$ is the conjugate transpose of A. So $\|A\|$, the supremum of $\|Ax\|$ over unit vectors x, exists as a number, with $\|A\dagger\| = \|A\|$ and $\|A\dagger A\| = \|A\|^2$ (the C*-identity). Here a real number c is the operator norm of a matrix M when $c \geq 0$, $\|Mx\|^2 \leq c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j + 1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \geq c^2 - 1/(j + 1)$; $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$. The construction finds c^2 as the top value of $\|Ax\|^2$ over unit vectors without deciding any comparison. The size must be at least one: the empty 0×0 matrix has no unit vector, so no number meets the last clause. *Used for:* The largest factor by which an operator can amplify a state: the norm in every perturbation and error bound of quantum dynamics ($\|H_1 - H_2\|$, $\|[A, B]\|$), the spectral norm behind condition numbers and numerical stability, and the size of a coupling or of a gate error. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

936. Operator norm is unique. For every size n, every complex $n \times n$ matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and all real numbers a and b on the line: if a and b are both operator norms of A, then $a = b$. Here a real number c is the operator norm of a matrix M when $c \geq 0$, $\|Mx\|^2 \leq c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j + 1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \geq c^2 - 1/(j + 1)$; $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$. So the operator norm, defined by its properties, names a single number. *Used for:* Lets the operator norm be used as a definite quantity: two computations of the largest amplification of one operator, for example by power iteration and from an eigenvalue bound, must agree on its value. *Price:* **free**; exact.

937. Submultiplicativity of the operator norm. For every size n, all complex $n \times n$ matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and all real numbers a, b and c on the line: if a is the operator norm of A, b is the operator norm of B and c is the operator norm of the product AB, then $c \leq ab$. So $\|AB\|$

$\le \|A\|\|B\|$. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Error growth through a sequence of operations: a product of time steps, gates or transfer matrices is bounded by the product of their norms, as in stability estimates for time-stepping schemes and in the growth of perturbations along a chain of transfer matrices. *Price:* **free**; exact.

938. Triangle inequality for the operator norm. For every size n, all complex n × n matrices A and B (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and all real numbers a, b and c on the line: if a is the operator norm of A, b is the operator norm of B and c is the operator norm of A + B, then $c \le a + b$. So $\|A + B\| \le \|A\| + \|B\|$. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Adding up error sources: a Hamiltonian $H_0 + V_1 + V_2$ with two perturbations has $\|V_1 + V_2\| \le \|V_1\| + \|V_2\|$, the step that combines separate error terms in perturbation bounds and in Trotter error budgets. *Price:* **free**; exact.

939. Operator norm is unchanged by unitary factors. For every size n, all complex n × n matrices U and A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $U\dagger U = 1$ and $UU\dagger = 1$ (U is unitary; U† is its conjugate transpose), and every real number a on the line: if a is the operator norm of A, then a is also the operator norm of UA and of AU. So $\|UA\| = \|AU\| = \|A\|$. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Norms do not depend on the basis or the picture: an operator keeps its norm after any change of orthonormal basis and in the interaction picture, so error bounds survive unitary frame changes such as the rotating frames of NMR and qubit control. *Price:* **free**; exact.

940. Norm of a diagonalized Hermitian matrix from its eigenvalues. For every size n, all complex n × n matrices H and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $V\dagger V = 1$ and $VV\dagger = 1$ (V is unitary; V† is its conjugate transpose), all real numbers $\lambda_1, \ldots, \lambda_n$ on the line with $HV = V\cdot\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ (a given orthonormal diagonalization of H, which makes H Hermitian), and every real number c that is the operator norm of H: $|\lambda_k| \le c$ for every k, and for every tolerance $1/(j+1)$ some index k has $\lambda_k^2 \ge c^2 - 1/(j+1)$. So $\|H\| = \max|\lambda_k|$, stated without naming an index where the maximum is reached. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* The norm of a Hamiltonian or an observable is its largest eigenvalue in absolute value: $\|H\| = \max|E_k|$ sets the stable time step of explicit integrators, the spectral window of Chebyshev and Krylov propagators, and the strength of a perturbation in energy units. *Price:* **free**; exact.

941. Operator norm compared with the Frobenius norm. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every real number c on the line that is the operator norm of A: $c^2 \le \mathrm{Re}\,\mathrm{tr}(A\dagger A) \le n\cdot c^2$, where A† is the conjugate transpose of A and $\mathrm{tr}(A\dagger A) = \Sigma\,|A_{il}|^2$, summed over all entries, is the squared Frobenius norm $F(A)^2$. So $\|A\| \le F(A) \le \sqrt{n}\cdot\|A\|$. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Cheap, computable bounds on the operator norm: the Frobenius norm, the square root of the sum of all squared matrix elements, brackets $\|A\|$ within a factor $\sqrt{n}$, used to bound coupling strengths and errors in large numerical problems where the largest eigenvalue is too expensive to find. *Price:* **free**; exact.

942. Matrix entries and the operator norm bound each other. For every size n, every complex n × n matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every real number c on the line that is the operator norm of A: $|A_{il}|^2 \le c^2$ for every entry A_{il}; and $c \le n\cdot K$ for every real number $K \ge 0$ with $|A_{il}|^2 \le K^2$ for every entry. So no entry exceeds $\|A\|$ in absolute value, and $\|A\| \le n\cdot\max|A_{il}|$. Here a real number c is the operator norm of a matrix M when $c \ge 0$, $\|Mx\|^2 \le c^2\|x\|^2$ for every complex vector x, and for every tolerance $1/(j+1)$ some unit vector x ($\|x\|^2 = 1$) has $\|Mx\|^2 \ge c^2 - 1/(j+1)$; $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Estimating the strength of a Hamiltonian from its matrix elements: a tight-binding or spin Hamiltonian on n states whose couplings are all at most K in size has $\|H\| \le n\cdot K$, and one large matrix element forces a large norm; used for quick bounds on energy ranges and time steps. *Price:* **free**; exact.

943. Moore–Penrose pseudoinverse of any rectangular matrix. For all sizes n and m and every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary

parts lie on the line), with no rank hypothesis: there is a complex m × n matrix B, the Moore–Penrose pseudoinverse A^+, with ABA = A, BAB = B, (AB)† = AB and (BA)† = BA, where M† is the conjugate transpose of M. The price is exactly the limited principle of omniscience (LPO), with no countable choice. The upper bound places A in the top-left corner of a square matrix of size n + m, filled with zeros elsewhere, takes the square pseudoinverse (which costs LPO) and keeps its m × n corner block. The lower bound is already the 1 × 1 case A = (a) with a real: ABA = A gives a(ar − 1) = 0 for the real part r of B, and that decides whether a is 0 or apart from 0. *Used for:* Fitting m model parameters to n measurements when the design matrix may be rank-deficient: the pseudoinverse gives the minimum-norm least-squares estimate in detector unfolding, tomographic reconstruction and the calibration of over- or under-determined linear models. *Price:* **LPO**; exact.

944. Normal equations of a rectangular least-squares problem are solvable. For all sizes n and m, every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every complex vector b of length n, with no rank hypothesis on A: the normal equations A†Ax = A†b have a solution x, a complex vector of length m, where A† is the conjugate transpose of A. These are the equations of the least-squares problem of making $\|Ax - b\|$ as small as possible (n measurements, m parameters); the row asserts only that they can be solved. The price is exactly the limited principle of omniscience (LPO), with no countable choice. The upper bound takes x = A^+b with the Moore–Penrose pseudoinverse A^+, since $A†AA^+ = A†(AA^+)† = (AA^+A)† = A†$. The lower bound is already A = (a) with a real and b = (1): $a^2x = a$ gives a(ar − 1) = 0 for the real part r of x, which decides whether a is 0 or apart from 0. *Used for:* Linear least-squares fitting with more measurements than parameters or a degenerate model: the fitted parameters of a calibration curve, a spectral decomposition or a linearised orbit determination solve the normal equations, whatever the rank of the design matrix. *Price:* **LPO**; exact.

945. Pseudoinverse of a matrix with uniformly independent columns. For all sizes n and m and every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) whose columns are uniformly independent, meaning there is a real number ε > 0 on the line with $\|Ax\|^2 \geq \varepsilon\|x\|^2$ for every complex vector x of length m, where $\|x\|^2 = |x_1|^2 + \ldots + |x_m|^2$: there is a complex m × n matrix B with BA = 1 (the m × m identity) that also satisfies the four Penrose equations ABA = A, BAB = B, (AB)† = AB and (BA)† = BA, where M† is the conjugate transpose of M. So B is both the Moore–Penrose pseudoinverse and a left inverse of A. It is B = $(A†A)^{-1}A†$, where the inverse of A†A is a convergent series with an explicit rate, so no decision is needed. *Used for:* Overdetermined fits with a certified non-degenerate design, where every parameter direction moves the predictions by a definite amount, as in a well-conditioned calibration or a fit to known line shapes: the estimator $(A†A)^{-1}A†b$ then costs nothing. *Price:* **free**; exact.

946. Pseudoinverse of a matrix with uniformly independent rows. For all sizes n and m and every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) whose rows are uniformly independent, meaning there is a real number ε > 0 on the line with $\|A†y\|^2 \geq \varepsilon\|y\|^2$ for every complex vector y of length n, where A† is the conjugate transpose of A and $\|y\|^2 = |y_1|^2 + \ldots + |y_n|^2$: there is a complex m × n matrix B with AB = 1 (the n × n identity) that also satisfies the four Penrose equations ABA = A, BAB = B, (AB)† = AB and (BA)† = BA. So B is both the Moore–Penrose pseudoinverse and a right inverse of A. It is B = $A†(AA†)^{-1}$, built with no decision. *Used for:* Underdetermined linear systems with independent constraints, more unknowns than equations: distributing a required field or force over redundant coils or actuators, or reconstructing a source from fewer sensors than unknowns; the right inverse $A†(AA†)^{-1}$ meets every target exactly. *Price:* **free**; exact.

947. Minimum-norm solution of an underdetermined system. For all sizes n and m and every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) whose rows are uniformly independent (there is a real number ε > 0 on the line with $\|A†y\|^2 \geq \varepsilon\|y\|^2$ for every complex vector y of length n, where A† is the conjugate transpose of A): the n × n matrix AA† has a two-sided inverse W (AA†W = 1 and WAA† = 1), and for every complex vector b of length n the vector x = A†Wb has three properties: Ax = b; x is the only solution of Ax = b of the form A†w for a complex vector w (the only solution in the range of A†); and every solution x′ of Ax′ = b satisfies $\|x'\|^2 = \|x\|^2 + \|x' - x\|^2$, so x is the solution of least norm. Here $\|x\|^2 = |x_1|^2 + \ldots + |x_m|^2$. *Used for:* Choosing the least-effort solution when the constraints leave freedom: minimum-energy control inputs, minimum-norm current estimates in magnetoencephalography and EEG source imaging, and the smallest correction that meets a set of linear constraints in beam steering. *Price:* **free**; exact.

948. Least-squares Pythagorean identity. For all

sizes n and m, every complex n × m matrix A (entries in the complex numbers over the line, whose real and imaginary parts lie on the line), every complex vector b of length n and every complex vector x_0 of length m that solves the normal equations $A\dagger Ax_0 = A\dagger b$, where $A\dagger$ is the conjugate transpose of A, with no rank hypothesis: for every complex vector x of length m, $\|Ax - b\|^2 = \|Ax_0 - b\|^2 + \|A(x - x_0)\|^2$, where $\|y\|^2 = |y_1|^2 + \ldots + |y_k|^2$ for a vector of length k. So every solution of the normal equations makes $\|Ax - b\|$ as small as possible, and any other x is worse by exactly $\|A(x - x_0)\|^2$. *Used for:* The geometry of a least-squares fit: the residual of the best fit is orthogonal to everything the model can predict, which splits the total squared error into the best-fit residual plus the model-space part, the decomposition behind χ^2 goodness-of-fit tests and the analysis of variance. *Price:* **free**; exact.

949. Truncation error of a given singular value decomposition. For every size n, all complex n × n matrices A, U and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with $U\dagger U = 1$, $V\dagger V = 1$ and $AV = U\cdot\mathrm{diag}(s_1, \ldots, s_n)$ for real numbers $s_1, \ldots, s_n$ on the line (a given singular value decomposition $A = U\cdot\mathrm{diag}(s_1, \ldots, s_n)\cdot V\dagger$; the s_k need not be ordered or nonnegative), and every set S of indices among 1, ..., n (given by a yes-or-no value for each index): let $E = A - \sum_k s_k u_k v_k\dagger$, the sum running over the indices k not in S, where u_k and v_k are the k-th columns of U and V. Then the real part of $\mathrm{tr}(E\dagger E)$ equals $\sum_k s_k{}^2$, the sum running over the indices k in S. So dropping a set of singular triples leaves a squared Frobenius error equal to the sum of the squares of the dropped singular values. Here $M\dagger$ is the conjugate transpose of M and tr(M) the sum of its diagonal entries. *Used for:* Low-rank compression with a known error: truncating a Schmidt decomposition or a matrix product state in DMRG and tensor-network simulations, where the discarded weight $\Sigma\, s_k{}^2$ is exactly the squared error, and principal component analysis of data. *Price:* **free**; exact.

950. Evolution depends Lipschitz-continuously on the Hamiltonian. For every size n, all complex n × n matrices H_1 and H_2 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that are Hermitian ($H_1\dagger = H_1$ and $H_2\dagger = H_2$), all real numbers t and c ≥ 0 on the line such that c bounds the difference in operator norm ($\|(H_1 - H_2)x\|^2 \le c^2\|x\|^2$ for every complex vector x): if $E_1 = \exp(-itH_1)$ and $E_2 = \exp(-itH_2)$, then $\|(E_1 - E_2)x\|^2 \le (|t|c)^2\|x\|^2$ for every complex vector x. So $\|e^{-itH_1} - e^{-itH_2}\| \le |t|\|H_1 - H_2\|$ in operator norm. The proof replaces Duhamel's integral formula by many short time steps and needs no integral. Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \ldots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Comparing two models of one system: the true Hamiltonian and an approximation of it (a truncated basis, an effective or mean-field Hamiltonian) give evolutions that differ by at most $|t|\|H_1 - H_2\|$ at time t, the bound used for error budgets in Hamiltonian simulation and quantum control. *Price:* **free**; exact.

951. Evolved state under a perturbed Hamiltonian. For every size n, all complex n × n matrices H and V (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that are Hermitian ($H\dagger = H$ and $V\dagger = V$), all real numbers t and v ≥ 0 on the line such that v bounds V in operator norm ($\|Vx\|^2 \le v^2\|x\|^2$ for every complex vector x), and every complex vector ψ of length n: if $E_1 = \exp(-i\cdot t\cdot(H + V))$ and $E_2 = \exp(-i\cdot t\cdot H)$, then $\|E_1\psi - E_2\psi\|^2 \le (|t|v)^2\|\psi\|^2$. So adding a perturbation V to the Hamiltonian moves the state evolved for a time t by at most $|t|\|V\|\|\psi\|$. Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \ldots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* Robustness of quantum dynamics to an unknown or neglected term: a stray field, a residual coupling or a calibration error V shifts the evolved state by at most $|t|\|V\|\|\psi\|$, which sets how long a quantum simulation or a gate sequence stays faithful to the intended Hamiltonian. *Price:* **free**; exact.

952. Short-time evolution stays close to the identity. For every size n, every complex n × n matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian ($H\dagger = H$), every real number τ on the line and every real number K ≥ 0 that bounds H in operator norm ($\|Hx\|^2 \le K^2\|x\|^2$ for every complex vector x): if $E = \exp(-i\tau H)$, then $\|(E - 1)x\|^2 \le (|\tau|K)^2\|x\|^2$ for every complex vector x. So the evolution over a time τ differs from the identity by at most $|\tau|\|H\|$ in operator norm. Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \ldots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \ldots + |x_n|^2$. *Used for:* A Hamiltonian of strength $\|H\|$ acting for a time τ moves any state ψ by at most $|\tau|\|H\|\|\psi\|$: the short-time estimate behind step-size control in numerical quantum dynamics, and a crude limit on how fast a quantum state can change under a bounded Hamiltonian. *Price:* **free**; exact.

953. Unitary evolution is Lipschitz in time. For every size n, every complex n × n matrix H (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is Hermitian ($H\dagger = H$), all real numbers t and s on the line and every real number K ≥ 0 that bounds H in operator norm ($\|Hx\|^2 \le K^2\|x\|^2$ for every complex vector x): if $E_t = \exp(-i\cdot t\cdot H)$ and $E_s = \exp(-i\cdot s\cdot H)$, then $\|(E_t - E_s)x\|^2 \le (|t$

$- s|K)^2\|x\|^2$ for every complex vector x. So $\|e^{-i\cdot t\cdot H} - e^{-i\cdot s\cdot H}\| \le |t - s|\|H\|$: the evolution is Lipschitz in time with constant $\|H\|$. Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$. *Used for:* Timing errors in quantum dynamics: a state read out or stored at time s instead of t differs by at most $|t - s|\|H\|\|\psi\|$, which bounds the effect of timing jitter in pulse sequences and the error of interpolating between saved time steps of a simulation. *Price:* **free**; exact.

954. Strang splitting with its error bound. For every size n, all complex n × n matrices H_1 and H_2 (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that are Hermitian ($H_1\dagger = H_1$ and $H_2\dagger = H_2$), every real number t on the line, all real numbers $c_1 \ge 0$ and $c_2 \ge 0$ that bound the double commutators in operator norm ($\|[H_1, [H_1, H_2]]x\|^2 \le {c_1}^2\|x\|^2$ and $\|[H_2, [H_1, H_2]]x\|^2 \le {c_2}^2\|x\|^2$ for every complex vector x, where [X, Y] = XY − YX), and every number of steps M = m + 1 ≥ 1: put $\tau = t/M$, $E = \exp(-it(H_1 + H_2))$, $A = \exp(-i\tau H_1/2)$ and $B = \exp(-i\tau H_2)$. Then $\|(E - (ABA)^M)x\|^2 \le r^2\|x\|^2$ for every complex vector x, with $r = |t|^3(c_1/8 + c_2/12)/M^2$. So the Strang step (half a step of H_1, a full step of H_2, another half step of H_1), repeated M times, reproduces the full evolution with operator-norm error at most $|t|^3(\|[H_1, [H_1, H_2]]\|/8 + \|[H_2, [H_1, H_2]]\|/12)/M^2$. Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E, and $\|x\|^2 = |x_1|^2 + \dots + |x_n|^2$. *Used for:* Second-order splitting of quantum dynamics: the split-step Fourier method for the Schrödinger equation (half a kinetic step, a full potential step, half a kinetic step), second-order Trotter–Suzuki circuits in digital quantum simulation and symmetric time slicing of path integrals, with the number of steps for a target accuracy read off from the $1/M^2$ error. *Price:* **free**; exact.

955. Von Neumann entropy of every density matrix. For every size n and every complex n × n matrix ρ (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is a density matrix, meaning positive semidefinite ($\rho\dagger = \rho$ and the real part of $\langle x, \rho x\rangle$ is ≥ 0 for every complex vector x) with trace 1 (the real part of $\rho_{11} + \dots + \rho_{nn}$ is 1), pure states and other singular ρ included: there is a real number $S \ge 0$ on the line, the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho \log \rho)$, such that $S = -(\lambda_1\ell_1 + \dots + \lambda_n\ell_n)$ whenever V is a complex n × n matrix with $V\dagger V = 1$, $\lambda_1, \dots, \lambda_n$ are real numbers with $\rho V = V\cdot\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ (an orthonormal eigenbasis of ρ with its eigenvalues), and $\ell_1, \dots, \ell_n$ are real numbers such that, for each k, either $\lambda_k = 0$ or $\exp(\ell_k) = \lambda_k$. So $S = -\sum_k \lambda_k \log \lambda_k$ on every eigenbasis, with 0 log 0 = 0: a zero eigenvalue contributes nothing, whatever ℓ_k is. Building S needs no eigenbasis and no decision about which eigenvalues vanish: with Y = 1 − ρ, S is the limit of the increasing numbers $S(N) = \mathrm{Re}\,\mathrm{tr}(\rho(Y + Y^2/2 + \dots + Y^N/N))$, and the terms from N on add at most n/(N + 1). Here $\langle x, y\rangle = \bar{x}_1y_1 + \dots + \bar{x}_ny_n$, and $\exp(\ell) = \lambda$ means that the exponential series of ℓ converges to λ. *Used for:* The entropy of any quantum state of a finite system, mixed or pure, full rank or not: the entanglement entropy of a subsystem, the thermodynamic entropy of a Gibbs state, and the optimal rate of quantum data compression. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

956. Exponential of a traceless anti-Hermitian matrix is special unitary. For every size n and all complex n × n matrices X and E (entries in the complex numbers over the line, whose real and imaginary parts lie on the line): if X is anti-Hermitian ($X\dagger = -X$, where X† is the conjugate transpose), traceless ($X_{11} + \dots + X_{nn} = 0$) and E = exp(X), then $E\dagger E = 1$, $EE\dagger = 1$ and $\det E = 1$. So the exponential maps the Lie algebra su(n) into the special unitary group SU(n). Here exp(X) = E means that the partial sums of the series $1 + X + X^2/2! + X^3/3! + \dots$ converge entrywise to E. *Used for:* The exponential map from generators to symmetry transformations: every exp(X) with X in su(n) is special unitary, as used for SU(2) spin rotations, SU(3) colour transformations in quantum chromodynamics and quantum gates written as $\exp(-i\cdot H)$ with a traceless H. *Price:* **free**; exact.

957. One-parameter group of a traceless anti-Hermitian matrix is special unitary. For every size n, every complex n × n matrix X (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) that is anti-Hermitian ($X\dagger = -X$, where X† is the conjugate transpose) and traceless ($X_{11} + \dots + X_{nn} = 0$), every real number t on the line and every complex n × n matrix E: if $E = \exp(t\cdot X)$, then $E\dagger E = 1$, $EE\dagger = 1$ and $\det E = 1$. So the whole one-parameter group $t \mapsto \exp(t\cdot X)$ lies in the special unitary group SU(n). Here $\exp(t\cdot X) = E$ means that the partial sums of the series $1 + t\cdot X + t^2X^2/2! + t^3X^3/3! + \dots$ converge entrywise to E. *Used for:* Evolution under a traceless Hamiltonian $H = i\cdot X$: $\exp(-i\cdot t\cdot H) = \exp(t\cdot X)$ is special unitary at every time t, so spin precession (SU(2)) and flavour or colour rotations (SU(3)) stay in the special unitary group, with unit determinant at every time. *Price:* **free**; exact.

958. Exponential of a real skew-symmetric matrix is a rotation. For every size n, every n × n matrix A with real entries on the line (the real numbers constructed from distinctions) that is skew-symmetric ($A^T = -A$), and every complex n × n

matrix E (entries in the complex numbers over the line) with E = exp(A), meaning that the partial sums of the series 1 + A + $A^2/2!$ + $A^3/3!$ + … converge entrywise to E: every entry of E is real, so E = R for a real n × n matrix R, and R is a proper rotation: R^TR = 1, RR^T = 1 and det R = 1. So the exponential maps the real skew-symmetric matrices, the Lie algebra so(n), into the rotation group SO(n). *Used for:* Rotations generated by angular velocities: a constant angular velocity, written as a skew-symmetric matrix, integrates to a proper rotation with no reflection, as in rigid-body kinematics and the precession of a classical magnetic moment in a constant field. *Price:* **free**; exact.

959. Exponential of a matrix whose square is a scalar. For every size n, all complex n × n matrices X and E (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) and every real number s on the line: if $X^2 = -s\cdot 1$ and E = exp(X), then E = C(s)·1 + S(s)·X, where $C(s) = \sum_j (-s)^j/(2j)!$ and $S(s) = \sum_j (-s)^j/(2j+1)!$, both sums over j = 0, 1, 2, and so on. For $s = \omega^2 \geq 0$ these are cos ω and (sin ω)/ω; for $s = -\kappa^2 \leq 0$ they are cosh κ and (sinh κ)/κ. No square root of s and no decision about the sign of s is needed. Here exp(X) = E means that the partial sums of the series 1 + X + $X^2/2!$ + $X^3/3!$ + … converge entrywise to E. *Used for:* Closed-form exponentials of 2 × 2 generators: every traceless 2 × 2 matrix squares to −det(X)·1, so the spin-1/2 rotations and Rabi oscillations of a two-level system (cos and sin) and boosts in one space dimension (cosh and sinh) come out in closed form. *Price:* **free**; exact.

960. Spin-1/2 rotation formula with Pauli matrices. For all real numbers a, b, c and w on the line (the real numbers constructed from distinctions) with $a^2 + b^2 + c^2 = 1$, and every complex 2 × 2 matrix E (entries in the complex numbers over the line): if E = exp(−iw(aσ₁ + bσ₂ + cσ₃)), then E = cos w·1 − i sin w·(aσ₁ + bσ₂ + cσ₃), where σ_1 = (0 1; 1 0), σ_2 = (0 −i; i 0) and σ_3 = (1 0; 0 −1) are the Pauli matrices, cos w = $\sum_j (-1)^j w^{2j}/(2j)!$ and sin w = $\sum_j (-1)^j w^{2j+1}/(2j+1)!$. Here exp(X) = E means that the partial sums of the series 1 + X + $X^2/2!$ + $X^3/3!$ + … converge entrywise to E. *Used for:* Rotating a spin-1/2: the rotation by angle θ about the unit axis n̂ = (a, b, c) acts on spinors as exp(−i(θ/2)n̂·σ) = cos(θ/2) − i sin(θ/2)n̂·σ, the formula behind single-qubit gates, NMR and ESR pulses, and the sign change of a spinor under a full 2π turn. *Price:* **free**; exact.

961. Rodrigues rotation formula. For all real numbers a, b, c and θ on the line (the real numbers constructed from distinctions) with $a^2 + b^2 + c^2 = 1$, and every complex 3 × 3 matrix E (entries in the complex numbers over the line): if E = exp(θK), where K = (0 −c b; c 0 −a; −b a 0) is the cross-product matrix of the unit axis n̂ = (a, b, c) (Kv = n̂ × v for every vector v), then E = 1 + sin θ·K + (1 − cos θ)·K^2, with sin θ = $\sum_j (-1)^j \theta^{2j+1}/(2j+1)!$ and cos θ = $\sum_j (-1)^j \theta^{2j}/(2j)!$. So exp(θK) sends v to v cos θ + (n̂ × v) sin θ + n̂(n̂·v)(1 − cos θ). Here exp(X) = E means that the partial sums of the series 1 + X + $X^2/2!$ + $X^3/3!$ + … converge entrywise to E. *Used for:* The rotation by angle θ about a unit axis in three-dimensional space, in closed form: attitude of rigid bodies and spacecraft, rotation of molecular and crystal coordinates, and the rotation operators of classical mechanics. *Price:* **free**; exact.

962. Rotation generated by a cross-product matrix lies in SO(3). For all real numbers a, b, c and θ on the line (the real numbers constructed from distinctions), the axis (a, b, c) not necessarily of length one, and every complex 3 × 3 matrix E (entries in the complex numbers over the line): if E = exp(θK), where K = (0 −c b; c 0 −a; −b a 0) is the cross-product matrix of (a, b, c), then every entry of E is real, so E = R for a real 3 × 3 matrix R, and R^TR = 1, RR^T = 1 and det R = 1. So exp(θK) is a proper rotation of space. Here exp(X) = E means that the partial sums of the series 1 + X + $X^2/2!$ + $X^3/3!$ + … converge entrywise to E. *Used for:* Every constant angular velocity generates a proper rotation: the attitude update of a gyroscope or spacecraft over one time step, and rotations of molecules and crystals, stay in SO(3) with no reflection. *Price:* **free**; exact.

963. Every unit complex number is a phase $e^{i\theta}$. For every complex number u (in the complex numbers over the line, whose real and imaginary parts lie on the line) with $(\mathrm{Re}\ u)^2 + (\mathrm{Im}\ u)^2 = 1$: there is a real number θ on the line with exp(iθ) = u, the exponential taken of the 1 × 1 matrix (iθ), meaning that the partial sums of 1 + iθ + $(i\theta)^2/2!$ + $(i\theta)^3/3!$ + … converge to u. So every element of the circle group U(1) is the exponential of a purely imaginary number, an element of its Lie algebra. No window is required for θ: the row does not ask for the principal value in (−π, π]. *Used for:* Phases of quantum amplitudes: every unit complex number is a phase factor $e^{i\theta}$, as in U(1) gauge transformations of electromagnetism, the global phase of a wave function and the phase of a plane wave or a coherent state. *Price:* **free**; exact.

964. Every SU(2) matrix is the exponential of an su(2) matrix. For every complex 2 × 2 matrix U (entries in the complex numbers over the line, whose real and imaginary parts lie on the line) with U†U = 1 and det U = 1 (U is in SU(2); U† is its conjugate transpose): there is a complex 2 × 2 matrix X with X† = −X (anti-Hermitian) and $X_{11} + X_{22} = 0$ (traceless) such that U = exp(X), meaning that the partial sums of 1 + X + $X^2/2!$ + $X^3/3!$ + … converge entrywise to U. The price is exactly the lesser limited principle of omniscience (LLPO) together with countable choice.

The upper bound diagonalizes U by a unitary V (the spectral theorem for unitary matrices, which costs LLPO with countable choice), writes the eigenvalues as $e^{i\theta}$ and $e^{-i\theta}$ with an angle θ that costs nothing, and takes $X = V \cdot \mathrm{diag}(i\theta, -i\theta) \cdot V\dagger$. The lower bound, with no choice, uses test matrices near -1, the 2π rotation: for a real number y, a logarithm of an SU(2) matrix built from y decides whether $y \geq 0$ or $y \leq 0$, and that sign decision is LLPO. *Used for:* Every spin-1/2 rotation or single-qubit gate in SU(2) is generated by some traceless Hamiltonian $H = i \cdot X$ acting for unit time, $U = \exp(-i \cdot H)$; finding that generator, for instance to implement a gate as a single pulse or to interpolate between two rotations, costs a sign decision near the 2π rotation $U = -1$. *Price:* **LLPO + $\mathbf{AC}_{\omega}$**; exact.

Q.7 Spectra, lattice models and thermodynamic limits

965. Fekete's subadditive lemma. If $u(m + n) \leq u(m) + u(n)$ for all m, n and u(n)/n is bounded below for $n \geq 1$, then u(n)/n converges to a limit no larger than any u(n)/n. *Used for:* thermodynamic limit of free energy or ground energy per site. *Price:* **LPO**; exact (the converse is monotone convergence, row 1302). Proved in Proposition 28.63(1).

966. Fekete's lemma, bounded-defect case. If $|u(m + n) - u(m) - u(n)| \leq C$ for all m, n, there is a point L with $|u(n)/n - L| \leq C/n$ for every $n \geq 1$. *Used for:* energy per site of a chain whose halves interact through a bounded boundary term. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Proposition 28.63(2).

967. Top level value of a symmetric matrix. Every symmetric $n \times n$ matrix A on the line, $n \geq 1$, has a value λ with $x \cdot Ax \leq \lambda(x \cdot x)$ for every x, and unit vectors x with $x \cdot Ax$ within every tolerance of λ. *Used for:* highest energy of a finite real Hamiltonian; highest normal-mode frequency. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

968. Ground level value of a symmetric matrix. Every symmetric $n \times n$ matrix A on the line, $n \geq 1$, has a value λ with $\lambda(x \cdot x) \leq x \cdot Ax$ for every x, and unit vectors x with $x \cdot Ax$ within every tolerance of λ. *Used for:* ground-state energy of a finite real Hamiltonian by the variational principle. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

969. Spectral theorem at every finite precision. Every symmetric $n \times n$ matrix A on the line, $n \geq 1$, has at every precision ε an exactly orthonormal basis of vectors v with values μ such that every coordinate of $Av - \mu v$ is at most ε in size. *Used for:* diagonalizing finite Hamiltonians and normal-mode problems to any finite precision. *Price:* **free**; exact.

970. Thermodynamic limit over boxes in any dimension. A translation-invariant function F of boxes in the n-dimensional lattice that changes by at most C times the area of the cut when any side is split has a value f with $|F/\text{volume} - f| \leq C(1/L_1 + \ldots + 1/L_n)$ for every box with sides $L_1, \ldots, L_n \geq 1$. *Used for:* free energy per site of a system with surface interactions in any dimension, with its finite-size correction. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Theorem 28.62.

971. Every energy level of a symmetric matrix. Every symmetric $n \times n$ matrix on the line has levels $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ such that, for a fixed constant K, the j-th largest value of every diagonalization accurate to ε lies within $K\varepsilon$ of λ_j. *Used for:* the full spectrum of a finite Hamiltonian: every energy level, each to any stated precision. *Price:* **free**; exact.

972. Ground energy of a lattice model splits over boxes. For a nearest-neighbour model on boxes of the n-dimensional lattice, with finitely many states per site and bond energies at most J_0 in size, splitting a box changes its ground energy by at most J_0 times the area of the cut. *Used for:* the ground energy per site of lattice magnets and lattice gases, with its finite-size correction. *Price:* **free**; exact.

973. Spectral theorem when the levels are known and apart. A symmetric $n \times n$ matrix on the line with n values that are pairwise apart, each of which has approximate unit eigenvectors at every precision, has an exactly orthonormal basis of eigenvectors with exactly those values. *Used for:* energy eigenstates of a nondegenerate finite Hamiltonian; Born probabilities for observables with distinct outcomes. *Price:* **free**; exact (no choice at all). Proved in Theorem 28.67.

974. Eigenvector of a simple energy level. For a symmetric $n \times n$ matrix on the line, $n \geq 1$, a level that is apart from every other level has an exact unit eigenvector with that level as its eigenvalue. *Used for:* the ground state of a gapped finite Hamiltonian; eigenstates of nondegenerate energy levels. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

975. Free energy of a lattice model splits over boxes. For a nearest-neighbour model on boxes of

the n-dimensional lattice with bond energies at most J_0 in size, at every inverse temperature $\beta \geq 0$ the free energy $-\log Z$ changes by at most βJ_0 times the area of the cut when a box is split. *Used for:* free energy per site of lattice magnets and lattice gases at every temperature, with its finite-size correction. *Price:* **free**; exact.

976. Determinant as the product of the energy levels. For every real symmetric $n \times n$ matrix A on the line, $n \geq 1$, $\det(t \cdot 1 - A)$ equals the product of $t - \lambda_j$ over its levels λ_j for every t, so $\det A$ is the product of the levels. *Used for:* characteristic polynomials of finite Hamiltonians; Gaussian integrals and partition functions as products over normal modes. *Price:* **free**; exact.

977. Spectral projector onto a gapped window of levels. For a symmetric $n \times n$ matrix on the line, $n \geq 1$, and a run of consecutive levels separated by gaps from the levels above and below it, there is an orthogonal projector P commuting with the matrix whose trace is the number of levels in the run and whose range has energies between the lowest and highest level of the run. *Used for:* the probability of finding the energy in a gapped window, even when levels inside the window coincide. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

978. Free energy of a finite-range lattice model splits. For a lattice model on boxes in n dimensions with finitely many pair interactions of range at most R, each at most J_0 in size, splitting a box changes the free energy $-\log Z$ by at most $\beta\, J_0\, r\, R$ times the area of the cut, where r is the number of interaction terms and $\beta \geq 0$. *Used for:* free energy per site of lattice systems with interactions beyond nearest neighbours, at every temperature. *Price:* **free**; exact.

979. Functions of a symmetric matrix. For a symmetric $n \times n$ matrix A on the line, $n \geq 1$, and a Lipschitz function f, the matrix f(A) exists: it is symmetric, commutes with A, is the limit of $\Sigma\, f(\mu)\, v\, v^{T}$ over diagonalizations, and its trace is the sum of f over the levels of A. *Used for:* Gibbs matrices $e^{-\beta H}$ and partition functions as sums over energy levels, with the exponential made constant outside an interval that contains every level (f(A) depends only on the values of f near the levels); time evolution of finite quantum systems. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

980. Weyl's bound for energy levels. If two symmetric $n \times n$ matrices on the line, $n \geq 1$, have Rayleigh quotients that differ by at most $e(x \cdot x)$ for every vector x, then each level of one is within e of the corresponding level of the other. *Used for:* stability of energy levels under a bounded perturbation; error bounds in perturbation theory. *Price:* **free**; exact.

981. Energies of non-interacting systems add. For symmetric matrices A and B on the line with level lists, the levels of the combined system $A \otimes 1 + 1 \otimes B$ are the sums of one level of A and one level of B, sorted from the largest; two level lists of one matrix always agree. *Used for:* energies of independent subsystems; partition functions of non-interacting systems multiply. *Price:* **free**; exact.

982. Partition function of two coupled blocks. For symmetric matrices A and B on the line and a symmetric coupling E that changes every Rayleigh quotient by at most e times the squared length, the partition function of $A \otimes 1 + 1 \otimes B + E$ at inverse temperature $\beta \geq 0$ lies within a factor $e^{\beta\, e}$ of the product of the partition functions of A and B. *Used for:* almost additivity of the free energy of quantum systems; surface terms in the thermodynamic limit. *Price:* **free**; exact.

983. Quantum lattice free energy is almost additive. For symmetric matrices indexed by boxes of the d-dimensional lattice that split across every cut into the two halves side by side plus a coupling of size at most J per unit of cut area, the free energy, minus the logarithm of the partition function at inverse temperature $\beta \geq 0$, changes by at most β J times the cut area when a box is split. *Used for:* thermodynamic limit of the free energy per site of quantum spin lattice models. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Theorem 28.64(1).

984. Functions of a symmetric matrix multiply. For a symmetric $n \times n$ matrix A on the line, $n \geq 1$, and Lipschitz functions f and g, the matrix f(A) times the matrix g(A) is the matrix of the product function f g at A. *Used for:* composing functions of a Hamiltonian, each made Lipschitz by a cut-off outside an interval that contains every level; products of observables that are functions of the energy. *Price:* **free**; exact.

985. Time evolution of a finite quantum system is unitary. For a symmetric $n \times n$ matrix A on the line, $n \geq 1$, and Lipschitz functions c and s with $c^2 + s^2 = 1$, the matrices $C = c(A)$ and $S = s(A)$ commute and $C^2 + S^2$ is the identity matrix; with c and s the cosine and sine of t times the energy, the evolution C minus i times S is unitary. *Used for:* unitary time evolution of finite quantum systems; conservation of probability. *Price:* **free**; exact.

986. Gibbs state of a finite quantum system. For a symmetric $n \times n$ matrix A on the line, $n \geq 1$, and inverse temperature $\beta \geq 0$, the Gibbs matrix $\rho = e^{-\beta A}/Z$ exists: it is symmetric, commutes with A, is positive, has trace 1, and $Z\, \rho$ is the limit of the sums

of $e^{-\beta\,\mu}$ v v^{T} over diagonalizations. *Used for:* thermal states and thermal expectation values of finite quantum systems. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

987. Ising chain in a transverse field splits at every cut. For real numbers J and g, the transverse-field Ising chain on n sites with open ends, with nearest-neighbour coupling −J between the spins along one axis and field −g flipping each spin, is on any cut into a left and a right part the two parts side by side plus one bond whose effect on every expectation value is at most |J| times the squared length. *Used for:* surface terms of the quantum Ising chain; locality of quantum spin chain Hamiltonians. *Price:* **free**; exact. Proved in Theorem 28.61.

988. Free energy per site of the quantum Ising chain. For real numbers J and g and inverse temperature $\beta \geq 0$, the free energy of the transverse-field Ising chain on n sites with open ends, minus the logarithm of its partition function (β times the thermodynamic free energy), divided by n, has a limit f, and on n sites it is within $\beta\,|J| / n$ of f. *Used for:* thermodynamic limit of the quantum Ising chain; quantum phase transitions in one dimension. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Corollary 28.65(1).

989. Quantum Ising model splits at every cut. For real numbers J and g and every dimension d, the transverse-field Ising model on a box of the d-dimensional lattice with open boundaries, with coupling −J between neighbouring spins along one axis and field −g flipping each spin, is on any cut of the box into two boxes the two parts side by side plus the bonds across the cut, whose effect on every expectation value is at most |J| times the area of the cut times the squared length. *Used for:* surface terms of quantum spin models; locality of lattice Hamiltonians in every dimension. *Price:* **free**; exact. Proved in Theorem 28.61.

990. Free energy per site of the quantum Ising model. For real numbers J and g, every dimension d and inverse temperature $\beta \geq 0$, the free energy per site of the transverse-field Ising model on boxes of the d-dimensional lattice with open boundaries, the free energy being minus the logarithm of the partition function (β times the thermodynamic free energy), has a limit f, and on a box with sides $L_1, \ldots, L_d$ it is within $\beta\,|J|\,(1/L_1 + \ldots + 1/L_d)$ of f. *Used for:* thermodynamic limit of the quantum Ising model; quantum phase transitions in two and three dimensions. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Corollary 28.65(1).

991. Ground energies of quantum lattice models are almost additive. For symmetric matrices indexed by boxes of the d-dimensional lattice that split across every cut into the two halves side by side plus a coupling of size at most J per unit of cut area, the lowest energy level changes by at most J times the cut area when a box is split. *Used for:* ground-state energy per site of quantum lattice models; surface energies. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Theorem 28.64(2).

992. Ground energy per site of the quantum Ising chain. For real numbers J and g, the lowest energy level of the transverse-field Ising chain on n sites with open ends, divided by n, has a limit, and on n sites it is within $|J| / n$ of that limit. *Used for:* ground-state energy density of the quantum Ising chain; quantum phase transitions at zero temperature. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Corollary 28.65(2).

993. Ground energy per site of the quantum Ising model. For real numbers J and g and every dimension d, the lowest energy level of the transverse-field Ising model on boxes of the d-dimensional lattice with open boundaries, divided by the number of sites, has a limit e_0, and on a box with sides $L_1, \ldots, L_d$, each at least 1, it is within $|J|\,(1/L_1 + \ldots + 1/L_d)$ of e_0. *Used for:* ground-state energy per site of quantum spin models in two and three dimensions; quantum phase transitions at zero temperature. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Corollary 28.65(2).

994. Collecting located points into a sequence. Choosing, for every natural number, the one point that a relation assigns to it, when rational approximations of that point within every tolerance are picked out by a relation that can be decided case by case, is exactly as strong as countable unique choice of yes-or-no answers. The free allowance contains that choice: countable choice at pairs of ratios, applied to the pair (b, b) for the unique bit b. *Used for:* collecting computed physical quantities, one per box or per length, into one function. *Price:* **free**; exact (equivalent, with no other choice, to countable unique choice of yes-or-no answers).

Q.8 Topology, compactness and fixed points

995. Brouwer's fixed point theorem on an interval. On the line, if a function is continuous at every point of an interval [a, b] ($a \le b$) and sends every point of [a, b] into [a, b], then it has a fixed point in [a, b]. *Used for:* equilibria of one-variable maps: a mixed equilibrium of a two-strategy game, a self-consistent solution with one order parameter. *Price:* **LLPO**; exact (the converse goes through the intermediate value theorem, row 1312; its choices, of yes-or-no answers and of representatives of the map's values, are inside the free allowance).

996. Brouwer's fixed point theorem on the unit square. On the line, if a function of two variables is uniformly continuous (with a modulus of continuity) on the unit square $[0, 1] \times [0, 1]$ and sends the square into itself, then it has a fixed point in the square. *Used for:* equilibria of two-variable maps: a Nash equilibrium with two mixing probabilities, a self-consistent solution with two order parameters. *Price:* **LLPO**; exact for uniformly continuous maps: the converse uses weak König's lemma, whose choices are yes-or-no answers, and representatives of the moves of the map at the dyadic corners, all inside the free allowance; for maps that are only continuous at each point see row 1058.

997. Banach's fixed point theorem on a complete invariant set. On the line, let K be a set such that every Cauchy sequence of points of K converges to a point of K, let f send K into K, and let f shrink the distance between any two points of K by a factor q with $0 \le q < 1$. Then for every start x_0 in K the iterates of f converge to a fixed point p in K, and $(1 - q)\cdot|f^n(x_0) - p| \le q^n\cdot|f(x_0) - x_0|$ for every n. *Used for:* iterations confined to a closed region: self-consistent-field and Picard iterations inside a ball or a constraint set. *Price:* **free**; exact.

998. The continuous image of a connected set is connected. On the line, call a set connected when any two open sets that together cover it and each contain a point of it share a point of it. If a set is connected and f is continuous on it, then the set of values of f on it is connected. *Used for:* a continuous quantity on a connected parameter set takes an interval of values, which is the source of every intermediate-value statement. *Price:* **free**; exact.

999. Locally constant integer functions on connected sets are constant. On the line, let K be a connected set (any two open sets that together cover K and each contain a point of K share a point of K) and let c be an integer-valued function that is locally constant on K: near every point of K, c is constant on K. Then c is constant on K: it has one single value at all points of K. *Used for:* quantized invariants such as winding numbers, charges and Chern numbers cannot change along a connected set of parameters. *Price:* **free**; exact.

1000. An integer-valued locally constant function on an interval is constant. On the line, an integer-valued function that is locally constant on a closed interval [a, b] (near every point of [a, b] it is constant on [a, b]) is constant on [a, b]: it has one single value at all points of [a, b]. *Used for:* quantized invariants such as winding numbers, charges and Chern numbers cannot change along a parameter interval. *Price:* $\mathbf{AC}_\omega$; exact.

1001. The line has no nontrivial open and closed sets. On the line, if a set of points is open and its complement is open, then the set is empty or it contains every point. *Used for:* a connected state space cannot be split into two disjoint open parts: no nontrivial superselection sectors. *Price:* **EM** + $\mathbf{AC}_\omega$; exact.

1002. Limit points are limits of sequences of members. On the line, for a set K and a point x: for every tolerance K has a member within that tolerance of x if and only if some sequence of points of K converges to x. *Used for:* every argument that takes the limit of a sequence of allowed states. *Price:* $\mathbf{AC}_\omega$; exact.

1003. A set that contains its limit points is complete. On the line, if a set K contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x), then every Cauchy sequence of points of K converges to a point of K. *Used for:* closed constraint sets and closed subspaces are complete, so iterations that start in them converge inside them. *Price:* **free**; exact (the proof uses the completeness of the line and forms a limit point).

1004. A complete set contains its limit points. On the line, if every Cauchy sequence of points of a set K converges to a point of K, then K contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x). *Used for:* a complete subspace of states is closed under limits. *Price:* $\mathbf{AC}_\omega$; exact.

1005. Cantor's intersection theorem. On the line, let $S_0, S_1, S_2, \ldots$ be sets that each contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x), such that for every N some point lies in the first N + 1 sets at once, and such that for every tolerance, from some index on any two points of one set S_n are within that tolerance of each other. Then some point lies in every S_n. *Used for:* nested-region and bisection arguments: existence of the limiting point. *Price:* **free**; exact (the proof chooses rational approximations of points of the sets, which the free allowance contains, and forms a limit point).

1006. Baire's category theorem on the line. On the line, let U_0, U_1, U_2, ... be open sets (every point of U_n has a tolerance band of points of U_n around it) such that every point has, within every tolerance, a point of each U_n. Then every point has, within every tolerance, a point that lies in all the U_n at once. *Used for:* generic properties of systems; uniform boundedness of operator families in quantum theory. *Price:* **free**; exact (the proof chooses rational intervals inside the open sets, which the free allowance contains, and forms a limit point).

1007. The preimage of a set containing its limit points. On the line, if f is continuous and a set K contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x), then the set of points x with f(x) in K also contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x). *Used for:* level sets and constraint surfaces {F = c} contain their limit points. *Price:* **free**; exact.

1008. Where two continuous functions agree contains its limit points. On the line, if f and g are continuous, then the set of points where f(x) = g(x) contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x). *Used for:* a constraint surface {F = G}: limits of states that satisfy the constraint satisfy it. *Price:* **free**; exact.

1009. A ball contains its limit points. On the line, for any centre c and radius r, the set of points x with $|x - c| \le r$ contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x). *Used for:* closed neighbourhoods: limits of states within a radius stay within it. *Price:* **free**; exact.

1010. The intersection of two sets containing their limit points. On the line, if two sets K and L each contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x), then so does the set of points lying in both. *Used for:* combining two constraints on a state. *Price:* **free**; exact.

1011. The intersection of two open sets is open. On the line, if two sets are open (every point of the set has a tolerance band of points of the set around it), then so is the set of points lying in both. *Used for:* combining two open conditions or two neighbourhoods. *Price:* **free**; exact.

1012. Limit of a sequence in a limit-closed set. On the line, if a set K contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x), and a sequence is eventually in K and converges to L, then L is in K. *Used for:* the limit of states that satisfy a constraint satisfies it. *Price:* **free**; exact.

1013. A closed interval contains its limit points. On the line, for any a and b the interval of points x with $a \le x \le b$ contains every point that its members approach (every point x such that, for every tolerance, the set has a member within that tolerance of x). *Used for:* bounded parameter ranges: limits of parameters in range stay in range. *Price:* **free**; exact.

1014. A compact set has an open complement. On the line, if a set K is compact (every countable cover of K by open intervals with fraction endpoints has a finite subcover made of its first intervals), then every point outside K has a tolerance band around it that contains no point of K. *Used for:* a compact set of allowed states keeps a margin from every excluded state, so small perturbations cannot reach it. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the price is excluded middle).

1015. Where two continuous functions disagree is open. On the line, if f and g are continuous, then every point x with $f(x) \ne g(x)$ has a tolerance band around it in which $f(y) \ne g(y)$ at every point y. *Used for:* a constraint F = G that fails at a state fails at every nearby state: violations are stable. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the statement is free).

1016. The complement of a closed interval is open. On the line, for any a and b, every point x that does not satisfy $a \le x \le b$ has a tolerance band around it containing no point of the interval [a, b]. *Used for:* a parameter out of range stays out of range under small perturbations. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the statement is free).

1017. The complement of a closed ball is open. On the line, for any centre c and radius r, every point x with $|x - c| \le r$ false has a tolerance band around it containing no point of the ball of points y with $|y - c| \le r$. *Used for:* a state outside a ball stays outside under small perturbations. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the statement is free).

1018. The union of two sets with open complements. On the line, if the complements of two sets K and L are open (every point outside K has a tolerance band outside K, and every point outside L has a tolerance band outside L), then the complement of the set of points lying in K or in L is open. *Used for:* two constraints either of which may hold, such as alternative allowed regions. *Price:* **free**; exact.

1019. The intersection of two sets with open complements. On the line, if the complements of two sets K and L are open (every point outside K has a tolerance band outside K, and every point outside L has a tolerance band outside L), then the complement of the set of points lying in both K and L is open. *Used for:* combining two constraints on a state. *Price:* **open**; at most weak excluded middle, and more sharply at most this principle: whenever a property N of natural numbers is stable (if not-not N(k) then N(k)) and persists (N(k) gives N(k+1)), and for a statement A both "A fails" and "A does not fail" lead to some k with N(k), then some k has N(k). With Markov's principle available this statement is exactly equivalent to that principle, and then implies LPO. Without Markov's principle no lower bound is proved: the encodings tried need it, because requiring the single point 0 to be closed already costs Markov's principle. The principle is not one of the letters.

1020. The preimage of a set with open complement. On the line, if f is continuous and the complement of a set K is open, then the complement of the set of points x with f(x) in K is open. *Used for:* level sets of continuous quantities have open complements. *Price:* **free**; exact.

1021. Limit of a sequence in a set with open complement. On the line, if the complement of a set K is open, and a sequence is eventually in K and converges to L, then L is in K. *Used for:* the limit of states that satisfy a constraint satisfies it, for an arbitrary constraint set with open complement. *Price:* **EM**; exact.

1022. A set with open complement is complete. On the line, if the complement of a set K is open, then every Cauchy sequence of points of K converges to a point of K. *Used for:* constraint sets with open complement are complete, so iterations that start in them converge inside them. *Price:* **EM**; exact (the proof uses the completeness of the line).

1023. A complete set has an open complement. On the line, if every Cauchy sequence of points of a set K converges to a point of K, then every point outside K has a tolerance band around it containing no point of K. *Used for:* a complete subspace of states keeps a margin from every excluded state. *Price:* **open**; at least LPO and at most EM + AC_{ω}. With countable choice this statement is exactly equivalent to the conjunction of two principles: Markov's principle for stable persistent properties (if N is stable under double negation, N(k) gives N(k+1), and it is not refuted that some k has N(k), then some k has N(k)) and the double negation shift (if ¬¬Q(n) for every n, then ¬¬(Q(n) for every n)); the lower bound needs only that the line is complete, which the free allowance supplies. The first principle is not one of the letters, so the exact price in the letters is open.

1024. Finitely enumerable sets contain their limit points. On the line, let S be a set given by a list of points, repeated entries allowed. If for every tolerance S has a member within that tolerance of a point x, then x is a member of S. *Used for:* a discrete spectrum, such as the eigenvalues of a finite matrix, is closed: a limit of approximate eigenvalues is an eigenvalue; the price is the degeneracy problem of telling which of two nearly equal levels the limit is. *Price:* **LLPO**; exact.

1025. Finite sets contain their limit points, without repeated points. On the line, let S be a set given by a list of points with no repeated entries. If for every tolerance S has a member within that tolerance of a point x, then x is a member of S. *Used for:* a discrete spectrum of distinct eigenvalues is closed: a limit of approximate eigenvalues is an eigenvalue. *Price:* $\mathbf{MP}^{\vee}$; exact.

1026. A finite set has an open complement. On the line, let S be a set given by a list of points with no repeated entries. Then every point x outside S has a tolerance band around it that contains no point of S. *Used for:* a point outside the spectrum of a finite matrix has a margin free of eigenvalues, so small perturbations cannot reach the spectrum. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, only upper bounds are proved).

1027. A single point contains its limit points. On the line, for any point a and any point x: if for every tolerance the set containing only a has a member within that tolerance of x, that is, if a is within every tolerance of x, then x equals a. *Used for:* a limit of states that all equal a given state equals it. *Price:* **free**; exact.

1028. A single point has an open complement. On the line, for any point a, every point x different from a has a tolerance band around it in which every point differs from a. *Used for:* a state different from a reference state stays different under small perturbations. *Price:* **MP**; exact.

1029. A finite set is compact. On the line, every set given by a list of points is compact: whenever countably many open intervals with fraction endpoints cover the points of the list, finitely many of them, the first N for some N, already cover them. *Used for:* finitely many levels or measured values are covered by finitely many neighbourhoods. *Price:* **free**; exact.

1030. A single point is compact. On the line, for any point a, whenever countably many open intervals with fraction endpoints cover a, finitely many of them, the first N for some N, already cover it. *Used for:* a reference state is covered by finitely many

neighbourhoods of any cover. *Price:* **free**; exact.

1031. Path-connected subsets of the line are connected. On the line, let K be a set containing a point x_0 such that for every point y of K some path, continuous on [0, 1], runs inside K from x_0 to y. Then K is nonempty and connected: any two open sets that together cover K and each contain a point of K share a point of K. *Used for:* a set of parameters that can be moved through continuously cannot be split into two separated parts. *Price:* $\mathbf{AC}_\omega$; exact.

1032. Path-connected subsets of the plane are connected. On the plane of pairs of points of the line with the maximum distance, let K be a set containing a point p_0 such that for every point q of K some path, continuous on [0, 1], runs inside K from p_0 to q. Then K is nonempty and connected: any two open sets of the plane that together cover K and each contain a point of K share a point of K. *Used for:* a region of two parameters that can be moved through continuously cannot be split into two separated parts. *Price:* $\mathbf{AC}_\omega$; exact.

1033. Convex subsets of the line are path-connected. On the line, let K be a nonempty set such that for any points x and y of K and any a and b with $a \geq 0$, $b \geq 0$ and $a + b = 1$, the point $a \cdot x + b \cdot y$ lies in K. Then any two points of K are joined by a path, continuous on [0, 1], that stays in K. *Used for:* convex parameter regions, such as intervals of couplings, can be traversed continuously. *Price:* **free**; exact (the two-point conclusion is stronger than the base-point form of the statement).

1034. Convex subsets of the plane are path-connected. On the plane of pairs of points of the line with the maximum distance, let K be a nonempty set such that for any points p and q of K and any a and b with $a \geq 0$, $b \geq 0$ and $a + b = 1$, the point $a \cdot p + b \cdot q$ lies in K. Then any two points of K are joined by a path, continuous on [0, 1], that stays in K. *Used for:* convex state spaces of two parameters, such as a segment of probability distributions on three outcomes, can be traversed continuously. *Price:* **free**; exact (the two-point conclusion is stronger than the base-point form of the statement).

1035. Convex subsets of the line are connected. On the line, let K be a set such that for any points x and y of K and any a and b with $a \geq 0$, $b \geq 0$ and $a + b = 1$, the point $a \cdot x + b \cdot y$ lies in K. Then any two open sets that together cover K and each contain a point of K share a point of K (and K is nonempty when it contains a point). *Used for:* a convex parameter region cannot be split into two separated parts. *Price:* $\mathbf{AC}_\omega$; exact.

1036. Convex subsets of the plane are connected. On the plane of pairs of points of the line with the maximum distance, let K be a set such that for any points p and q of K and any a and b with $a \geq 0$, $b \geq 0$ and $a + b = 1$, the point $a \cdot p + b \cdot q$ lies in K. Then any two open sets of the plane that together cover K and each contain a point of K share a point of K (and K is nonempty when it contains a point). *Used for:* a convex state space of two parameters cannot be split into two separated parts. *Price:* $\mathbf{AC}_\omega$; exact.

1037. Open intervals are open. On the line, for any a and b, every point x with $a < x < b$ has a tolerance band around it in which every point y also satisfies $a < y < b$. Here $a < x$ means that $x - a$ is at a known positive distance from 0 (the apartness order). *Used for:* an open range of a parameter stays open under small perturbations. *Price:* **free**; exact.

1038. Open half-lines are open (above). On the line, for any a, every point x with $a < x$ has a tolerance band around it in which every point y also satisfies $a < y$. Here $a < x$ means that $x - a$ is at a known positive distance from 0 (the apartness order). *Used for:* a strict lower bound on a quantity is stable under small perturbations. *Price:* **free**; exact.

1039. Open half-lines are open (below). On the line, for any b, every point x with $x < b$ has a tolerance band around it in which every point y also satisfies $y < b$. Here $x < b$ means that $b - x$ is at a known positive distance from 0 (the apartness order). *Used for:* a strict upper bound on a quantity is stable under small perturbations. *Price:* **free**; exact.

1040. Open balls are open. On the line, for any centre c and radius r, every point x with $|x - c| < r$ has a tolerance band around it in which every point y also satisfies $|y - c| < r$. Here the strict inequality is the apartness order: $r - |x - c|$ is at a known positive distance from 0. *Used for:* states within a given distance of a reference state form a neighbourhood: small perturbations stay inside. *Price:* **free**; exact.

1041. Closed half-lines contain their limit points (above). On the line, for any a, the set of points x with $a \leq x$ contains its limit points: if for every tolerance the set has a member within that tolerance of a point x, then $a \leq x$. *Used for:* a limit of quantities that stay above a bound stays above it. *Price:* **free**; exact.

1042. Closed half-lines contain their limit points (below). On the line, for any a, the set of points x with $x \leq a$ contains its limit points: if for every tolerance the set has a member within that tolerance of a point x, then $x \leq a$. *Used for:* a limit of quantities that stay below a bound stays below it. *Price:* **free**; exact.

1043. Closed half-lines have an open complement (above). On the line, for any a, every point x for which $a \leq x$ fails has a tolerance band

around it in which a ≤ y fails for every point y. *Used for:* a quantity that violates a lower bound violates it under every small perturbation. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the statement is free).

1044. Closed half-lines have an open complement (below). On the line, for any a, every point x for which x ≤ a fails has a tolerance band around it in which y ≤ a fails for every point y. *Used for:* a quantity that violates an upper bound violates it under every small perturbation. *Price:* **MP**; exact (with the other reading of closed, containing its limit points, the statement is free).

1045. Open half-lines open under the derived strict order (above). On the line, write x > a for: a ≤ x holds and x ≤ a fails. For any a, every point x with x > a has a tolerance band around it in which every point y also satisfies y > a. *Used for:* a strict lower bound, read as not at most the bound, is stable under small perturbations. *Price:* **MP**; exact.

1046. Open half-lines open under the derived strict order (below). On the line, write x < b for: x ≤ b holds and b ≤ x fails. For any b, every point x with x < b has a tolerance band around it in which every point y also satisfies y < b. *Used for:* a strict upper bound, read as not at least the bound, is stable under small perturbations. *Price:* **MP**; exact.

1047. Closed intervals are totally bounded. On the line, for any a and b and any tolerance, there is a finite list of points such that every point x with a ≤ x ≤ b is within that tolerance of a listed point. *Used for:* a bounded region of a parameter can be discretized by a finite grid at every accuracy. *Price:* **free**; exact.

1048. The rationals are dense in the line. On the line, every point x has, within every tolerance, a fraction z/(d+1) with z an integer and d a natural number. *Used for:* every real quantity is approximated to any accuracy by a fraction: numerical approximation. *Price:* **free**; exact.

1049. The line has a countable dense sequence. On the line, there is a sequence of points such that every point has a member of the sequence within every tolerance. *Used for:* a countable list of trial values suffices to approximate every real value: separability of the line. *Price:* **free**; exact.

1050. Heine–Borel theorem on the unit square. On the line, if countably many open rectangles with fraction endpoints cover the unit square [0, 1] × [0, 1], then finitely many of them, the first N for some N, already cover it. *Used for:* compact parameter regions of two variables: the compactness step behind uniform estimates and fixed-point arguments. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω); the square statement implies the one-variable theorem.

1051. Compact sets are closed and bounded (open-complement reading). On the line, if every cover of a set K by countably many open intervals with fraction endpoints has a finite subcover (the first N intervals), then every point outside K has a tolerance band containing no point of K, and K lies inside some interval [−B, B]. *Used for:* a compact set of allowed states keeps a margin from every excluded state and is bounded. *Price:* **MP**; exact.

1052. Subsets of compact sets with open complement are compact. On the line, let K be compact (every cover of K by countably many open intervals with fraction endpoints has a finite subcover), and let L be a subset of K such that every point outside L has a tolerance band containing no point of L. Then L is compact. *Used for:* a constraint set inside a compact region is compact, so the minimum of a continuous cost over it is attained. *Price:* **open**; at most WEM + MP + AC_ω (and at most EM + AC_ω); no lower bound is proved.

1053. Heine–Borel theorem: compact exactly when closed and bounded. On the line, a set K is compact (every cover by countably many open intervals with fraction endpoints has a finite subcover) if and only if every point outside K has a tolerance band containing no point of K and K lies inside some interval [−B, B]. *Used for:* compactness of the closed bounded parameter regions that physics works with. *Price:* **open**; at least MP and at most WEM + MP + AC_ω.

1054. Closed bounded sets are compact (open-complement reading). On the line, if every point outside a set K has a tolerance band containing no point of K, and K lies inside some interval [−B, B], then every cover of K by countably many open intervals with fraction endpoints has a finite subcover. *Used for:* a closed bounded region of states is compact: the compactness step behind uniform bounds. *Price:* **open**; at most WEM + MP + AC_ω; no lower bound is proved; with MP it gives the one-variable Heine–Borel theorem.

1055. Heine–Borel theorem on the unit cube. On the line, if countably many open boxes with fraction endpoints cover the unit cube [0, 1] × [0, 1] × [0, 1], then finitely many of them, the first N for some N, already cover it. *Used for:* compact three-dimensional regions of positions or parameters: the compactness step behind uniform estimates. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω); the cube statement implies the square and interval statements.

1056. Heine–Borel theorem on the unit cube of every dimension. On the line, for every n, if

countably many open boxes with fraction endpoints cover the unit cube $[0, 1]^n$ (n coordinates, each in [0, 1]), then finitely many of them, the first N for some N, already cover it. *Used for:* compact regions of any number of parameters, such as the configuration space of a finite system: the compactness step behind uniform estimates. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem, in every dimension (so at most LLPO + AC_ω); implies the one-variable theorem.

1057. Heine–Cantor theorem on the unit square. On the line, if a function from the unit square to the plane is continuous at every point of the square, then it is uniformly continuous on the square: for every tolerance there is one distance such that any two points of the square closer than it have images within the tolerance. *Used for:* uniform error estimates for a continuous map of two parameters on a compact region. *Price:* **open**; at least the decidable fan theorem (it implies the interval statement and so the fan theorem); at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives), and at most the full fan theorem with no choice; equivalent, with no choice, to the decidable fan theorem together with a presentation of the local control data by countable covers; under LLPO and countable choice that presentation is equivalent to the statement itself, so whether LLPO with countable choice gives it is the open question.

1058. Brouwer on the unit square, pointwise continuous maps. On the line, if a function from the unit square into itself is continuous at every point of the square (not necessarily uniformly), then it has a fixed point in the square. *Used for:* equilibria of two-variable maps when only pointwise continuity is known. *Price:* **open**; at least LLPO and at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives); the proof reads the modulus of uniform continuity off as a function from yes-or-no questions; for uniformly continuous maps the price is exactly LLPO (row 996).

1059. Order-connected subsets of the line are connected. On the line, let K be a set such that every point between two points of K lies in K (if x and y are in K and $x \le z \le y$ then z is in K). Then any two open sets that together cover K and each contain a point of K share a point of K. *Used for:* a parameter region that is an interval, however it is described, cannot be split into separated parts. *Price:* $\mathbf{AC}_\omega$; exact (equivalent to the connectedness of closed intervals, with no principle).

1060. Connected subsets of the line are order-connected. On the line, let K be a set such that any two open sets that together cover K and each contain a point of K share a point of K. Then every point between two points of K lies in K. *Used for:* a connected set of parameters is an interval: every intermediate value is attained. *Price:* **open**; at most LPO with no choice; at least LPO given countable choice, so exactly LPO in the presence of countable choice; the lower bound without that hypothesis is open.

1061. Connected equals order-connected on the line. On the line, a set K is connected (any two open sets that together cover K and each contain a point of K share a point of K) if and only if every point between two points of K lies in K. *Used for:* connected parameter regions of the line are exactly the intervals. *Price:* **LPO + $\mathbf{AC}_\omega$**; exact.

1062. Intersections of closed sets contain their limit points. On the line, for any family of sets, each of which contains its limit points, the set of points lying in all of them contains its limit points. *Used for:* intersections of constraints: a limit of states that satisfy every constraint satisfies every constraint. *Price:* **free**; exact.

1063. Limit points of a set form a closed set. On the line, for any set S, the set of limit points of S (points with members of S within every tolerance) contains its own limit points. *Used for:* taking limits of limits gives nothing new: the closure of a region is closed. *Price:* **free**; exact.

1064. Limit points lie in every closed superset. On the line, if a set T contains its limit points and contains a set S, then T contains every limit point of S. *Used for:* the closure of a region lies inside every closed region that contains it. *Price:* **free**; exact.

1065. A function on a finite list has a minimizer. On the line, let S be given by a nonempty list of points with no repeated entries (for each pair of positions the entries are not equal), and let f be any function from points to points. Then some member x of S satisfies $f(x) \le f(y)$ for every member y of S ($a \le b$ here means that $b < a$ fails). *Used for:* the lowest-energy configuration among finitely many candidates: a ground state of a system with finitely many configurations. *Price:* **LLPO**; exact (the lower bound uses no choice).

1066. A function on a finite list has a maximizer. On the line, let S be given by a nonempty list of points with no repeated entries (for each pair of positions the entries are not equal), and let f be any function from points to points. Then some member x of S satisfies $f(y) \le f(x)$ for every member y of S ($a \le b$ here means that $b < a$ fails). *Used for:* the most probable outcome or the highest-energy configuration among finitely many candidates. *Price:* **LLPO**; exact (the lower bound uses no choice).

1067. Distinct finitely many reals have a least member. On the line, for a nonempty list of points with no repeated entries, some member x of the list satisfies x ≤ y for every member y (a ≤ b here means that b < a fails). *Used for:* the lowest of finitely many distinct energy levels of a discrete spectrum. *Price:* **MP**$^\vee$; exact (the lower bound uses no choice).

1068. Distinct finitely many reals have a greatest member. On the line, for a nonempty list of points with no repeated entries, some member x of the list satisfies y ≤ x for every member y (a ≤ b here means that b < a fails). *Used for:* the highest of finitely many distinct energy levels of a discrete spectrum. *Price:* **MP**$^\vee$; exact (the lower bound uses no choice).

1069. Finitely many reals with repeats have a least member. On the line, for a nonempty list of points in which entries may repeat, some member x of the list satisfies x ≤ y for every member y (a ≤ b here means that b < a fails). *Used for:* the lowest of finitely many energy levels when levels may coincide (degenerate spectra). *Price:* **LLPO**; exact (the lower bound uses no choice).

1070. Distinct finitely many reals can be sorted. On the line, for every list of points with no repeated entries, some rearrangement of the list has every entry ≤ every later entry (a ≤ b here means that b < a fails). *Used for:* ordering the levels of a finite spectrum from the lowest up, as when filling levels one by one. *Price:* **MP**$^\vee$; exact (the lower bound uses no choice).

1071. Finitely many reals with repeats can be sorted. On the line, for every list of points, entries may repeat, some rearrangement of the list has every entry ≤ every later entry (a ≤ b here means that b < a fails). *Used for:* ordering measured values or levels from the lowest up when values may coincide. *Price:* **LLPO**; exact (the lower bound uses no choice).

1072. Every real has a nearest point in a finite set. On the line, for a nonempty list of points with no repeated entries and any point x, some member y of the list satisfies |x − y| ≤ |x − z| for every member z of the list (a ≤ b here means that b < a fails). *Used for:* rounding a measured value to the nearest allowed level of a finite spectrum: quantization. *Price:* **LLPO**; exact (the lower bound uses no choice).

1073. Finite sets are discrete: every point is isolated. On the line, for every list of points with no repeated entries and every member a of the list there is a tolerance such that any member y of the list within that tolerance of a equals a. *Used for:* a discrete spectrum has a gap around every level: the levels are separated from each other. *Price:* **MP**; exact (the lower bound uses no choice).

1074. Compactness of the space of bit sequences. On the line, if countably many finite strings of bits are given such that every infinite sequence of bits begins with at least one of them, then finitely many of them, the first N for some N, already do: every infinite sequence of bits begins with one of the first N strings. *Used for:* the configuration space of a spin system with two states per site is compact: the compactness step behind limit states of lattice models (a minimizer of a cost function carries its own price). *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω); also equivalent to the one-variable Heine–Borel theorem.

1075. Compactness of the space of sequences with m values. On the line, let m ≥ 2. If countably many finite strings with letters from {0, ..., m − 1} are given such that every infinite sequence of such letters begins with at least one of them, then the first N of them, for some N, already do. *Used for:* the configuration space of a lattice system with m states per site is compact: the compactness step behind limit states of lattice models (a minimizer of a cost function carries its own price). *Price:* **FAN**; exact for every m ≥ 2: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω).

1076. Convex subsets of the line are order-connected. On the line, let K be a set such that for any points x and y of K and any a and b with a ≥ 0, b ≥ 0 and a + b = 1, the point a·x + b·y lies in K. Then every point between two points of K lies in K (if x and y are in K and x ≤ z ≤ y then z is in K). *Used for:* a convex region of one parameter is an interval: every intermediate value between two allowed states is allowed. *Price:* **LLPO**; exact (the converse chooses one yes-or-no answer for each row of a table of numbers, a choice inside the free allowance); also at most LPO with no choice at all.

1077. Order-connected subsets of the line are convex. On the line, let K be a set such that every point between two points of K lies in K (if x and y are in K and x ≤ z ≤ y then z is in K). Then for any points x and y of K and any a and b with a ≥ 0, b ≥ 0 and a + b = 1, the point a·x + b·y lies in K. *Used for:* an interval of allowed values is closed under mixing: any mixture of two allowed states is allowed. *Price:* **LLPO**; exact (no choice anywhere).

1078. Convex equals order-connected on the line. On the line, a set K is convex (for any points x and y of K and any a and b with a ≥ 0, b ≥ 0 and a + b = 1, the point a·x + b·y lies in K) if and only if every point between two points of K lies in K. *Used for:* convex regions of one parameter are exactly the intervals. *Price:* **LLPO**; exact; each direction is exactly LLPO, and the converse of each chooses yes-or-no answers, a choice inside the free allowance.

1079. Order-connected sets of comparable points are convex. On the line, let K be a set such

that every point between two points of K lies in K, and such that any two different points x and y of K satisfy $x \le y$ or $y \le x$. Then for any points x and y of K and any a and b with $a \ge 0$, $b \ge 0$ and $a + b = 1$, the point $a \cdot x + b \cdot y$ lies in K. *Used for:* intervals of allowed values whose points are known to be ordered are closed under mixing. *Price:* **open**; at most LLPO (an order-connected set is convex at LLPO whatever the extra hypothesis); with the disjunctive Markov principle available the statement is exactly LLPO, so it is not free; no lower bound is proved without that principle; a variant that assumes the comparison for every pair, equal points included, is free.

1080. Path-connected subsets of the line are order-connected. On the line, let K be a set containing a point x_0 such that for every point y of K some path, continuous on [0, 1], runs inside K from x_0 to y. Then every point between two points of K lies in K. *Used for:* a set of states connected by continuous deformations is an interval: every intermediate value is attained. *Price:* **LLPO**; exact (the converse chooses yes-or-no answers and representatives of the path's values, both inside the free allowance); with the stronger hypothesis that any two points of K are joined by a path, the price is unchanged.

1081. Path-connected subsets of the line are convex. On the line, let K be a set containing a point x_0 such that for every point y of K some path, continuous on [0, 1], runs inside K from x_0 to y. Then for any points x and y of K and any a and b with $a \ge 0$, $b \ge 0$ and $a + b = 1$, the point $a \cdot x + b \cdot y$ lies in K. *Used for:* a set of states connected by continuous deformations is closed under mixing. *Price:* **LLPO**; exact (the converse chooses yes-or-no answers and representatives of the path's values, both inside the free allowance); with the stronger hypothesis that any two points of K are joined by a path, the price is unchanged.

1082. The join of finitely many reals is one of them. On the line, let S be given by a nonempty list of indices with no repeated entries, and let f give a point for each index. Then the largest of the values (computed pairwise as the larger of two points) f(i), i in S, equals f(i) for some index i in S. *Used for:* the largest energy of a finite spectrum is one of its levels. *Price:* **LLPO**; exact (the lower bound uses no choice); also exact for lists in which entries may repeat.

1083. The meet of finitely many reals is one of them. On the line, let S be given by a nonempty list of indices with no repeated entries, and let f give a point for each index. Then the smallest of the values (computed pairwise as the smaller of two points) f(i), i in S, equals f(i) for some index i in S. *Used for:* the lowest energy of a finite spectrum is one of its levels. *Price:* **LLPO**; exact (the lower bound uses no choice); also exact for lists in which entries may repeat.

1084. Join of finitely many reals is the least upper bound. On the line, for a nonempty list of indices and a point f(i) for each index, each value f(i) is at most the largest of the values (computed pairwise as the larger of two points) f(i), i in the list, and that largest value is at most a point c exactly when every value f(i) is at most c. *Used for:* the largest of finitely many measured values bounds them all and is the smallest such bound. *Price:* **free**; exact.

1085. Meet of finitely many reals is the greatest lower bound. On the line, for a nonempty list of indices and a point f(i) for each index, the smallest of the values (computed pairwise as the smaller of two points) f(i), i in the list, is at most each value f(i), and a point c is at most that smallest value exactly when c is at most every value f(i). *Used for:* the smallest of finitely many measured values is below them all and is the largest such bound. *Price:* **free**; exact.

1086. A finite list has a least member for any function. On the line, for every nonempty list of items, entries may repeat, and every function f from items to points, some member m of the list satisfies $f(m) \le f(a)$ for every member a of the list ($a \le b$ here means that $b < a$ fails). *Used for:* the lowest-cost configuration in a finite enumeration of candidates. *Price:* **LLPO**; exact (the lower bound uses no choice); the mirror statement for a greatest member has an equal price.

1087. A finite list has a first member of least value. On the line, for every nonempty list of items, entries may repeat, and every function f from items to points, some member m of the list satisfies $f(m) \le f(a)$ for every member a and, for every member a with $f(a) \le f(m)$, the first occurrence of m comes no later in the list than the first occurrence of a ($a \le b$ here means that $b < a$ fails). *Used for:* naming the first of the lowest-cost configurations in a fixed enumeration: ties must be decided. *Price:* **WLPO**; exact (the lower bound uses no choice); the mirror statement for the first greatest member has an equal price.

1088. Heine–Cantor theorem on the unit cube. On the line, if a function from the unit cube $[0, 1] \times [0, 1] \times [0, 1]$ to the line is continuous at every point of the cube (a distance for every point and tolerance, with closeness meaning every coordinate within the distance), then it is uniformly continuous on the cube: for every tolerance there is one distance that serves all pairs of points of the cube. *Used for:* one mesh serves a whole compact three-dimensional region: uniform error estimates for a field that is only known to be continuous at each point. *Price:* **open**; at least the decidable fan theorem; at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives),

and at most the full fan theorem with no choice; equivalent, with no choice, to the decidable fan theorem together with a presentation of the local control data by countable covers.

1089. Heine–Cantor theorem on the unit cube of every dimension. On the line, for every $n \geq 1$, if a function from the unit cube $[0, 1]^n$ (n coordinates, each in $[0, 1]$) to the line is continuous at every point of the cube, then it is uniformly continuous on the cube: for every tolerance there is one distance that serves all pairs of points of the cube. *Used for:* one mesh for a compact region of any number of parameters, such as the configuration space of a finite system. *Price:* **open**; at least the decidable fan theorem in every dimension; at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives), and at most the full fan theorem with no choice; equivalent, with no choice, to the decidable fan theorem together with a presentation of the local control data by countable covers, in every dimension.

1090. Arzelà–Ascoli with a common modulus, uniformly Cauchy form. On the line, let $f_0, f_1, f_2, \ldots$ be functions on $[0, 1]$ that are all bounded by one natural number M, and suppose that for every tolerance there is one distance such that every function f_n changes by at most the tolerance between any two points of $[0, 1]$ within that distance. Then some strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ makes the functions $f(\varphi(0)), f(\varphi(1)), \ldots$ uniformly Cauchy on $[0, 1]$: for every tolerance, from some m on, any two of them are within the tolerance at every point of $[0, 1]$. *Used for:* a bounded family with one common modulus of continuity has a uniformly Cauchy subfamily: a step toward limiting profiles in variational and evolution problems. *Price:* **LPO**; exact: the lower bound LPO uses no choice (constant functions); the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance.

1091. Arzelà–Ascoli for sequences equicontinuous at each point. On the line, let $f_0, f_1, f_2, \ldots$ be functions on $[0, 1]$ that are all bounded by one natural number M, and suppose that at every point x of $[0, 1]$, for every tolerance there is a distance such that every function f_n changes by at most the tolerance between x and any point of $[0, 1]$ within that distance. Then some strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ makes the functions $f(\varphi(0)), f(\varphi(1)), \ldots$ uniformly Cauchy on $[0, 1]$: for every tolerance, from some m on, any two of them are within the tolerance at every point of $[0, 1]$. *Used for:* limiting profiles exist when continuity is only known at each point, to one common degree for every profile of the family. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance; it goes through row 1090 and the Heine–Cantor theorem for a sequence of functions.

1092. A tuple of reals can be sorted by a permutation. On the line, for every n and every list of n points $f(0), \ldots, f(n-1)$, repeats allowed, some permutation σ of the places $0, \ldots, n-1$ satisfies $f(\sigma(i)) \leq f(\sigma(j))$ whenever $i \leq j$ ($a \leq b$ here means that $b < a$ fails). *Used for:* ordering n measured values or energy levels from the lowest up, as when filling levels one by one. *Price:* **LLPO**; exact (the lower bound uses no choice).

1093. A tuple of reals with injective values can be sorted. On the line, for every n and every list of n points $f(0), \ldots, f(n-1)$ such that $f(i) = f(j)$ forces $i = j$ (no two places carry equal values), some permutation σ of the places satisfies $f(\sigma(i)) \leq f(\sigma(j))$ whenever $i \leq j$ ($a \leq b$ here means that $b < a$ fails). *Used for:* ordering the levels of a nondegenerate spectrum from the lowest up, when no two levels coincide. *Price:* $\mathbf{MP}^{\vee}$; exact (the lower bound uses no choice); values that are pairwise apart, with a visible gap between any two, would be comparable outright and are not priced here.

1094. Heine–Cantor for a sequence of functions on the interval. On the line, let $f_0, f_1, f_2, \ldots$ be functions on $[0, 1]$ such that at every point x of $[0, 1]$, for every tolerance there is a distance such that every function f_n changes by at most the tolerance between x and any point of $[0, 1]$ within that distance. Then the family is uniformly equicontinuous: for every tolerance there is one distance that serves every function f_n and every pair of points of $[0, 1]$. *Used for:* one mesh for a whole compact interval and a whole family of fields or profiles that are only known to be continuous at each point to one common degree. *Price:* **open**; at least the decidable fan theorem (it gives the Heine–Cantor theorem for one function); at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives), and at most the full fan theorem with no choice.

1095. Arzelà–Ascoli with a convergent subsequence, common modulus. On the line, let $f_0, f_1, f_2, \ldots$ be functions on $[0, 1]$ that are all bounded by one natural number M, and suppose that for every tolerance there is one distance such that every function f_n changes by at most the tolerance between any two points of $[0, 1]$ within that distance. Then some strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ has the following property: at every point x of $[0, 1]$ the values $f(\varphi(0))(x), f(\varphi(1))(x), \ldots$ converge to

some limit, and for every tolerance there is one m_0 such that for all $m \geq m_0$ and every point x of [0, 1], $f(\varphi(m))(x)$ is within the tolerance of the limit at x. *Used for:* a bounded family with one common modulus of continuity has a uniformly convergent subfamily: limiting profiles of variational and evolution problems, given point by point. *Price:* **LPO**; exact: the lower bound LPO uses no choice (a convergent subsequence is uniformly Cauchy); the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance; the limits are given point by point.

1096. Arzelà–Ascoli with a convergent subsequence, pointwise equicontinuity. On the line, let $f_0, f_1, f_2, \ldots$ be functions on [0, 1] that are all bounded by one natural number M, and suppose that at every point x of [0, 1], for every tolerance there is a distance such that every function f_n changes by at most the tolerance between x and any point of [0, 1] within that distance. Then some strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ has the following property: at every point x of [0, 1] the values $f(\varphi(0))(x), f(\varphi(1))(x), \ldots$ converge to some limit, and for every tolerance there is one m_0 such that for all $m \geq m_0$ and every point x of [0, 1], $f(\varphi(m))(x)$ is within the tolerance of the limit at x. *Used for:* limiting profiles exist when continuity is only known at each point, to one common degree for every profile of the family. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance; it goes through the Heine–Cantor theorem for a sequence of functions.

1097. Arzelà–Ascoli, common modulus, convergent subsequence with continuous limits. On the line, let $f_0, f_1, f_2, \ldots$ be functions on [0, 1] that are all bounded by one natural number M, and suppose that for every tolerance there is one distance such that every function f_n changes by at most the tolerance between any two points of [0, 1] within that distance. Then some strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ has the following properties: at every point x of [0, 1] the values $f(\varphi(0))(x), f(\varphi(1))(x), \ldots$ converge to some limit; for every tolerance there is one m_0 such that for all $m \geq m_0$ and every point x of [0, 1], $f(\varphi(m))(x)$ is within the tolerance of the limit at x; and for every tolerance there is a distance such that, for any two points x and y of [0, 1] within that distance, the limit at x and the limit at y are within the tolerance of each other. *Used for:* the limiting profile is itself continuous, with one modulus serving the whole family and its limit. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance; the continuity of the limits needs no further principle.

1098. Arzelà–Ascoli with a limit function, common modulus. On the line, let $f_0, f_1, f_2, \ldots$ be functions on [0, 1] that are all bounded by one natural number M, and suppose that for every tolerance there is one distance such that every function f_n changes by at most the tolerance between any two points of [0, 1] within that distance. Then there are a strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ and a function g (only its values on [0, 1] matter) such that for every tolerance there is one m_0 with $f(\varphi(m))(x)$ within the tolerance of $g(x)$ for all $m \geq m_0$ and every point x of [0, 1]. *Used for:* a bounded family with one common modulus of continuity has a subfamily converging uniformly to a limit profile g: limiting profiles of variational and evolution problems. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance; the function g is built by extension from the rational points of [0, 1], not by choosing a value at each point; compactness of the closure of the family in the sup metric is not stated.

1099. Arzelà–Ascoli with a limit function, pointwise equicontinuity. On the line, let $f_0, f_1, f_2, \ldots$ be functions on [0, 1] that are all bounded by one natural number M, and suppose that at every point x of [0, 1], for every tolerance there is a distance such that every function f_n changes by at most the tolerance between x and any point of [0, 1] within that distance. Then there are a strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ and a function g (only its values on [0, 1] matter) such that for every tolerance there is one m_0 with $f(\varphi(m))(x)$ within the tolerance of $g(x)$ for all $m \geq m_0$ and every point x of [0, 1]. *Used for:* limiting profiles exist when continuity is only known at each point, to one common degree for every profile of the family. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance.

1100. Arzelà–Ascoli, limit function that is continuous. On the line, let $f_0, f_1, f_2, \ldots$ be functions on [0, 1] that are all bounded by one natural number M, and suppose that for every tolerance there is one distance such that every function f_n changes by at most the tolerance between any two points of [0, 1] within that distance. Then there are a strictly increasing sequence of indices $\varphi(0) < \varphi(1) < \ldots$ and a function g (only its values on [0, 1] matter) such that for every tolerance there is one m_0 with $f(\varphi(m))(x)$

within the tolerance of g(x) for all $m \geq m_0$ and every point x of [0, 1], and such that, for every tolerance and every distance that works for all functions f_n at that tolerance, g(x) and g(y) are within the tolerance whenever x and y are points of [0, 1] within that distance. *Used for:* the limiting profile is itself continuous, with the modulus of the family. *Price:* **LPO**; exact: the lower bound LPO uses no choice; the converse uses LPO with countable unique choice of yes-or-no answers and choice of representatives, both inside the free allowance.

1101. Compactness of the Hilbert cube of unit intervals. On the line, let countably many open boxes be given, each restricting only finitely many coordinates: box k restricts the coordinates $i < m(k)$ to open intervals with fraction endpoints and leaves the other coordinates free. If every sequence $x_0, x_1, x_2, \ldots$ of points of [0, 1] lies in some box, then finitely many of the boxes, the first N for some N, already contain every such sequence. *Used for:* the state space of an infinite lattice with one bounded real variable per site is compact: limit states and uniform estimates for infinite systems. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω through weak König's lemma), a price equal to that of Heine–Borel on [0, 1] and on every finite cube; the lower bound is the one-coordinate case.

1102. Lebesgue number lemma on the unit interval. On the line, if countably many open intervals with fraction endpoints cover the unit interval [0, 1] (every point x of [0, 1] lies strictly inside one of them), then there is a natural number d such that for every point x of [0, 1] some interval of the cover contains every point y of the line with $|y - x| < 1/(d + 1)$. *Used for:* a uniform grid size for a compact interval: every window of one fixed width around a point of the interval lies inside a single member of a countable covering family of open intervals with fraction endpoints. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem (so at most LLPO + AC_ω through weak König's lemma), a price equal to that of Heine–Borel on [0, 1]; the lower bound uses a finite grid and a sum of finitely many indices.

1103. Lebesgue number lemma on the unit square. On the line, if countably many open rectangles with fraction corners cover the unit square $[0, 1] \times [0, 1]$ (every point of the square lies strictly inside one of them), then there is a natural number d such that for every point $p = (p_1, p_2)$ of the square some rectangle of the cover contains every point $q = (q_1, q_2)$ with $|q_1 - p_1| < 1/(d + 1)$ and $|q_2 - p_2| < 1/(d + 1)$. *Used for:* a uniform grid size for a compact two-parameter region: every window of one fixed size around a point of the region lies inside a single member of a countable covering family of open rectangles with fraction corners. *Price:* **FAN**; exact: equivalent, with no choice, to the decidable fan theorem and to the Lebesgue number lemma on the unit interval; upper bound at most LLPO + AC_ω through weak König's lemma.

Q.9 Probability

1104. Variance is never negative. For any nonnegative weights on finitely many outcomes, the variance of an observable is zero or positive. *Used for:* Fluctuations of energy and other observables; heat capacity read as an energy variance. *Price:* **free**; exact.

1105. Variance as mean square minus squared mean. For a probability law on finitely many outcomes, the variance of X equals the mean of X^2 minus the square of the mean of X. *Used for:* Energy fluctuations $\langle \Delta E^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2$ in statistical mechanics. *Price:* **free**; exact.

1106. Variance of a scaled observable. For any weights on finitely many outcomes, multiplying an observable by c multiplies its variance by c^2. *Used for:* Changing the units or the scale of a measured quantity. *Price:* **free**; exact.

1107. Variance ignores a constant shift. For a probability law on finitely many outcomes, adding a constant to an observable leaves its variance unchanged. *Used for:* Moving the zero of energy does not change energy fluctuations. *Price:* **free**; exact.

1108. Variance is at most the mean square. For a probability law on finitely many outcomes, the variance of X is at most the mean of X^2. *Used for:* Bounding fluctuations by second moments. *Price:* **free**; exact.

1109. Popoviciu's variance bound. For a probability law on finitely many outcomes, an observable with values between a and b has variance at most $((b - a)/2)^2$. *Used for:* Bounded observables such as spins and occupation numbers have bounded fluctuations. *Price:* **free**; exact.

1110. Mean of a product of independent observables. When the joint law of two finite observables is the product of their separate laws, the mean of their product is the product of their means. *Used for:* Uncoupled subsystems; partition functions

that factorize. *Price:* **free**; exact.

1111. Variances of independent observables add (Bienaymé). When the joint law of two finite observables is the product of their separate laws, the variance of their sum is the sum of their variances. *Used for:* Random-walk diffusion, where the mean square displacement grows linearly with the number of steps; additivity of fluctuations. *Price:* **free**; exact.

1112. Markov's inequality for the exact tail set. For a nonnegative observable on finitely many outcomes and every level c, the set of outcomes where the observable is at least c is an event, and c times its probability is at most the mean. *Used for:* Tail bounds on nonnegative quantities such as energies and counts. *Price:* **WLPO**; exact.

1113. Markov's inequality for recorded tail events. For a nonnegative observable on finitely many outcomes, every event on whose outcomes the observable is at least c has c times its probability at most the mean. *Used for:* Tail bounds on nonnegative quantities, read off any recorded event in the tail. *Price:* **free**; exact.

1114. Chebyshev's inequality for the exact tail set. For an observable on finitely many outcomes and every $c > 0$, the set of outcomes where the observable is at least c away from its mean is an event, and c^2 times its probability is at most the variance. *Used for:* Concentration of measure; the sharpness of thermodynamic limits. *Price:* **WLPO**; exact.

1115. Chebyshev's inequality for recorded tail events. For an observable on finitely many outcomes and $c \geq 0$, every event on whose outcomes the observable is at least c away from its mean has c^2 times its probability at most the variance. *Used for:* Concentration of measure read off any recorded deviation. *Price:* **free**; exact.

1116. Chernoff bound for the exact tail set. For an observable X on finitely many outcomes, a level ε and $t \geq 0$, the set where X is at least ε is an event and its probability is at most $e^{-t\varepsilon}$ times the mean of $e^{t \cdot X}$. *Used for:* Large deviations; exponential concentration of sums. *Price:* **WLPO**; exact.

1117. Chernoff bound for recorded tail events. For an observable X on finitely many outcomes, a level ε and $t \geq 0$, every event on whose outcomes X is at least ε has probability at most $e^{-t\varepsilon}$ times the mean of $e^{t \cdot X}$. *Used for:* Large deviations read off any recorded tail event. *Price:* **free**; exact.

1118. Jensen's inequality for finite averages. For a function f convex on the whole real line and an observable X on a finite probability space, f of the mean of X is at most the mean of f(X). *Used for:* The Gibbs–Bogoliubov free-energy bound; Jarzynski's inequality $\langle e^{-W} \rangle \geq e^{-\langle W \rangle}$. *Price:* **free**; exact.

1119. Concavity of the entropy function. The function $-x \log x$ is concave on the nonnegative numbers. *Used for:* Concavity of entropy; thermodynamic stability. *Price:* **free**; exact.

1120. Gibbs' inequality (relative entropy is nonnegative). For two probability laws on finitely many outcomes, the second giving every outcome positive weight, the relative entropy of the first to the second is zero or positive. *Used for:* The second law of thermodynamics; relative entropy as a one-way distance between laws. *Price:* **free**; exact.

1121. Relative entropy vanishes only at equal laws. For two probability laws on finitely many outcomes, the second giving every outcome positive weight, the relative entropy is zero exactly when the two laws are equal. *Used for:* Equilibrium is the unique law with zero relative entropy to itself. *Price:* **free**; exact.

1122. Binary entropy is at most log 2. For a probability p between 0 and 1, $-p \log p - (1 - p) \log(1 - p)$ is at most log 2. *Used for:* A two-outcome record holds at most one bit of entropy. *Price:* **free**; exact.

1123. Markov's inequality at a resolvable level. For a nonnegative observable on finitely many outcomes, a level c and any tolerance, there is a level c' between c and c plus the tolerance at which the set where the observable is at least c' is an event, and c' times its probability is at most the mean. *Used for:* Tail bounds read at a level a finite record can resolve. *Price:* **free**; exact.

1124. Chebyshev's inequality at a resolvable level. For an observable on finitely many outcomes, a level $c > 0$ and any tolerance, there is a level c' between c and c plus the tolerance at which the set where the observable is at least c' away from its mean is an event, and c'^2 times its probability is at most the variance. *Used for:* Concentration of measure at a level a finite record can resolve. *Price:* **free**; exact.

1125. Variance of a sum of uncorrelated observables. For finitely many observables on finitely many outcomes, pairwise uncorrelated, the variance of their sum is the sum of their variances (uncorrelated is weaker than independent, so this is the stronger statement). *Used for:* Diffusion; fluctuations of extensive sums over many independent parts. *Price:* **free**; exact.

1126. Weak law of large numbers with a rate. On a finite probability space (finitely many outcomes, nonnegative weights summing to 1), take k pairwise uncorrelated observables with variances at most σ^2 and a number $\varepsilon \geq 0$. Every event on which their sum is at least $k\varepsilon$ away from its mean has $(k\varepsilon)^2$ times its probability at most $k\sigma^2$, that is probability at most $\sigma^2/(k\varepsilon^2)$ when $k\varepsilon > 0$. *Used for:* Averages over many

parts concentrate on their mean; sharp thermodynamic limits. *Price:* **free**; exact.

1127. Stationary law of a two-state Markov chain. Every Markov chain on two states has a stationary law: a probability law that one step of the chain leaves unchanged. *Used for:* Equilibrium of a two-level system with given up and down rates. *Price:* **LLPO + AC**$_\omega$; exact.

1128. Stationary law of a moving two-state chain. A Markov chain on two states that leaves at least one state with positive probability, $a = K_{01}$ and $b = K_{10}$ with $a + b > 0$, has the stationary law $(b, a)/(a + b)$. *Used for:* Detailed balance and the equilibrium population of a two-level system. *Price:* **free**; exact.

1129. Stationary law of a three-state Markov chain. Every Markov chain on three states has a stationary law: a probability law on the three states that one step of the chain leaves unchanged. *Used for:* Equilibrium of three-level systems and other small finite Markov models. *Price:* **LLPO + AC**$_\omega$; exact.

1130. Unique stationary law of an ergodic chain (Doeblin). A Markov chain on finitely many states in which some state is entered from every state with probability at least $\delta > 0$ has exactly one stationary law: a probability law that one step of the chain leaves unchanged, and every other such law equals it. *Used for:* Relaxation to a unique equilibrium in ergodic finite systems; Monte Carlo sampling converges. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1131. Approximately stationary laws of finite chains. Every Markov chain on finitely many states (at least one) has, for every tolerance, a probability law that one step of the chain changes by at most that tolerance at every state. *Used for:* Time averages of a finite system's state approach equilibrium. *Price:* **free**; exact.

1132. Strong law of large numbers on finite windows. On a finite probability space (finitely many outcomes, nonnegative weights summing to 1), let X_0, X_1, ... be observables with $|X_k| \le 1$ and mean 0, each uncorrelated with the sum $S_k = X_0 + \dots + X_{k-1}$ of those before it in the four products $S_k X_k$, $S_k^2 X_k^2$, $S_k^3 X_k$ and $S_k X_k^3$. For $\varepsilon > 0$ and a window $N < k \le M$, every event on whose outcomes some k in the window has $|S_k| \ge k\varepsilon$ has $\varepsilon^4 \cdot N$ times its probability at most 3, that is probability at most $3/(\varepsilon^4 N)$ when $N \ge 1$. *Used for:* Running averages of repeated measurements settle on the mean, with an explicit rate for every finite record. *Price:* **free**; exact.

1133. Strong-law window bound from a fourth-moment bound. On a finite probability space (finitely many outcomes, nonnegative weights summing to 1), let S_k be the sum of the first k of a sequence of observables, with mean of S_k^4 at most $C \cdot k^2$ for every k. For $\varepsilon > 0$ and a window $N < k \le M$, every event on whose outcomes some k in the window has $|S_k| \ge k\varepsilon$ has $\varepsilon^4 \cdot N$ times its probability at most C, that is probability at most $C/(\varepsilon^4 N)$ when $N \ge 1$. *Used for:* Almost-sure convergence of averages, made quantitative from moment bounds. *Price:* **free**; exact.

1134. Stationary laws of all finite Markov chains. Every Markov chain on finitely many states, two or more, has a stationary law: a probability law that one step of the chain leaves unchanged. *Used for:* The existence of equilibrium for every finite Markov model, as Perron–Frobenius theory gives it. *Price:* **LLPO + AC**$_\omega$; exact.

1135. Unique stationary law of a primitive chain. A Markov chain on finitely many states whose allowed moves each have probability at least $\delta > 0$, and whose graph of allowed moves reaches one state from every state in exactly m steps, has exactly one stationary law. *Used for:* A single equilibrium for irreducible aperiodic systems; the target law of a Monte Carlo sampler is the only law its moves leave unchanged. *Price:* **free**; exact.

1136. Gaussian integral over the whole line. For every $p > 0$ and every q, the integral of $e^{-p\,y^2 + q\,y}$ over the whole line is $\sqrt{(\pi/p)} \cdot e^{q^2/(4p)}$, and the integral of $y \cdot e^{-p\,y^2 + q\,y}$ over the whole line is $(q/(2p)) \cdot \sqrt{(\pi/p)} \cdot e^{q^2/(4p)}$, with π twice the quarter period of the cosine. *Used for:* Oscillator ground states, heat kernels, Gaussian wave packets and partition functions of quadratic energies. *Price:* **free**; exact. Proved in Theorem 12.95.

1137. Minimum principle on Cantor space. Every uniformly continuous function of infinite sequences of binary digits attains its minimum at some sequence. *Used for:* Ground states of systems with infinitely many two-state sites, such as spin chains, exist. *Price:* **LLPO + AC**$_\omega$; exact.

1138. Minimum principle on the simplex of laws. On the set of probability laws over finitely many outcomes, two or more, every uniformly continuous function attains its minimum. *Used for:* Variational principles over probability laws: minimum free energy, maximum entropy at fixed constraints. *Price:* **LLPO + AC**$_\omega$; exact.

1139. Normal density integrates to one. For every mean μ and variance $v > 0$, the integral of $e^{-(y-\mu)^2/(2v)}$ over the whole line is $\sqrt{(2\pi v)}$, so the normal density integrates to one. *Used for:* Normalisation of Gaussian wave packets, thermal (Maxwell) velocity laws and Gaussian noise. *Price:* **free**; exact.

1140. Mean of the normal law. For every mean μ and variance $v > 0$, the integral of $y \cdot e^{-(y-\mu)^2/(2v)}$ over

the whole line is $\mu\sqrt{(2\pi v)}$: the normal law's mean is μ. *Used for:* The centre of a Gaussian distribution of positions, velocities or measurement errors. *Price:* **free**; exact.

1141. Variance of the normal law. For every mean μ and variance $v > 0$, the integral of $(y - \mu)^2 \cdot e^{-(y-\mu)^2/(2v)}$ over the whole line is $v\sqrt{(2\pi v)}$: the normal law's variance is v. *Used for:* Spread of Brownian displacement; equipartition, where the velocity variance is the temperature times Boltzmann's constant over the mass. *Price:* **free**; exact.

1142. Moment generating function of the normal law. For every mean μ, variance $v > 0$ and real t, the integral of $e^{t\cdot y} \cdot e^{-(y-\mu)^2/(2v)}$ over the whole line is $\sqrt{(2\pi \cdot v)} \cdot e^{\mu \cdot t + v \cdot t^2/2}$. *Used for:* Cumulants of Gaussian noise; large deviations; partition functions of quadratic energies with a linear field. *Price:* **free**; exact.

1143. Even moments of the standard normal law. For every k, the integral of $y^{2k} \cdot e^{-y^2/2}$ over the whole line is $1 \cdot 3 \cdot 5 \cdots (2k - 1)$ times $\sqrt{(2\pi)}$. *Used for:* Wick's theorem for Gaussian fields; higher fluctuations of thermal and quantum ground-state variables. *Price:* **free**; exact.

1144. Characteristic function of the standard normal law. For every real t, the integral of $\cos(t\cdot y) \cdot e^{-y^2/2}$ over the whole line is $\sqrt{(2\pi)} \cdot e^{-t^2/2}$. *Used for:* The Fourier transform of a Gaussian: heat kernels, Gaussian wave packets in momentum space, and the input of the central limit theorem. *Price:* **free**; exact.

1145. Steady state of a finite master equation. Every master equation on finitely many states, two or more, with nonnegative transition rates has a steady state: a probability law that the rates leave unchanged. *Used for:* Steady states of rate equations: chemical kinetics, population dynamics, open quantum systems in the rate limit. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

1146. Unique steady state of a connected master equation. A master equation on finitely many states whose rates are at least $\delta > 0$ along a graph of moves that reaches one state from every state in a fixed number of steps has exactly one steady state. *Used for:* Connected rate equations have exactly one equilibrium, the only candidate for the state they relax to. *Price:* **free**; exact.

1147. Characteristic function of an independent sum. For two observables on finitely many outcomes whose joint law is the product of their separate laws, the characteristic function of their sum (its cosine and sine parts) is the product of their characteristic functions. *Used for:* Sums of independent kicks: random walks, Brownian motion as a limit of steps, the central limit theorem. *Price:* **free**; exact.

1148. Odd moments of the standard normal law vanish. For every k, the integrals of $y^{2k+1} \cdot e^{-y^2/2}$ over the symmetric intervals $[-L, L]$ tend to 0 as the whole number L grows: taken as a symmetric limit, the integral over the whole line is 0. *Used for:* Symmetric Gaussian noise has no skew; odd correlators of free Gaussian fields vanish. *Price:* **free**; exact.

1149. Sine transform of the standard normal density. For every real t, the integrals of $\sin(t\cdot y) \cdot e^{-y^2/2}$ over the symmetric intervals $[-L, L]$ tend to 0 as the whole number L grows: taken as a symmetric limit, the sine part of the characteristic function of the standard normal law (mean 0, variance 1) is 0. *Used for:* A centred Gaussian wave packet carries no phase; centred noise has a real spectrum. *Price:* **free**; exact.

1150. Characteristic function of a shifted normal law. For every mean μ and real t, the integrals of $\cos(t\cdot y)$ and $\sin(t\cdot y)$ against $e^{-(y-\mu)^2/2}$ over the whole line are $\sqrt{(2\pi)} \cdot e^{-t^2/2} \cdot \cos(\mu t)$ and $\sqrt{(2\pi)} \cdot e^{-t^2/2} \cdot \sin(\mu t)$. *Used for:* Gaussian wave packets centred away from the origin: momentum-space phase $e^{i\mu t}$. *Price:* **free**; exact.

1151. Detailed balance gives a stationary law. If a probability law and a Markov chain on finitely many states satisfy detailed balance (the weight of i times the rate from i to j equals the weight of j times the rate from j to i, for every pair of states), the law is stationary for the chain. *Used for:* Reversible dynamics and equilibrium: Metropolis and heat-bath sampling, microscopic reversibility. *Price:* **free**; exact.

1152. Stationary law of a birth–death chain. A Markov chain on the states $0, 1, \ldots, n$ that moves only between neighbouring states, with every downward rate from $i + 1$ to i positive, has a stationary probability law in detailed balance with it: the weight of $i + 1$ times the down-rate from $i + 1$ equals the weight of i times the up-rate from i. *Used for:* The Ehrenfest urn, one-dimensional diffusion between sites, queues and population models. *Price:* **free**; exact.

1153. Law of independent copies. For a probability law on finitely many outcomes and every k, the weights of k independent copies (the product of the k individual weights) are nonnegative and sum to 1. *Used for:* Repeated independent trials; ideal gases of independent particles; the outcome space of many independent measurements. *Price:* **free**; exact.

1154. Characteristic function of a sum of independent copies. For a probability law on finitely many values and every k, the characteristic function of the sum of k independent copies is the k-th complex power of the law's characteristic function. *Used for:* Random walks after k steps; the characteristic function behind the central limit

theorem for sums. *Price:* **free**; exact.

1155. Kolmogorov's criterion for reversibility. A Markov chain on finitely many states with every transition probability positive has a probability law in detailed balance with it exactly when, for every two states, the product of transition probabilities around the triangle through a fixed state is equal in both directions. *Used for:* Telling equilibrium dynamics from driven ones by reading the rates; reversible Markov models in chemistry and sampling. *Price:* **free**; exact.

1156. Deciding whether a chain is reversible. Deciding, for every Markov chain on finitely many states with all transition probabilities positive, whether it has a law in detailed balance with it. *Used for:* Deciding whether a steady state is an equilibrium or a driven nonequilibrium state. *Price:* **WLPO**; exact.

1157. Integral of a product of functions. If k functions each have an integral over the whole line, the product of the functions, one per coordinate, has as its iterated integral over k dimensions the product of the k integrals. *Used for:* Integrals over independent coordinates factor: partition functions of uncoupled degrees of freedom. *Price:* **free**; exact.

1158. Gaussian integral in k dimensions. For positive a_1, ..., a_k, the iterated integral of $e^{-(a_1 y_1^2 + \cdots + a_k y_k^2)}$ over k dimensions, taken one coordinate at a time over the whole line, is the product of the square roots of π/a_i. *Used for:* Partition functions of independent oscillators; momentum integrals of the ideal gas; multivariate normal laws with independent coordinates. *Price:* **free**; exact.

1159. Gaussian integral of a positive-definite form. For a quadratic form in k variables that equals a weighted sum of squares $d_1(y_1)^2 + d_2(y_2 + m_{21}y_1)^2 + d_3(y_3 + m_{31}y_1 + m_{32}y_2)^2 + \cdots$ with every weight d_i positive (its triangular factorisation), the iterated integral of $e^{-\text{form}}$ over k dimensions, taken one coordinate at a time over the whole line, is the product of the square roots of π/d_i. *Used for:* Normalisation of multivariate normal laws; partition functions of coupled oscillators and of free Gaussian fields. *Price:* **free**; exact.

1160. Every Markov chain on a tree is reversible. A Markov chain on finitely many states whose moves go only along the edges of a tree, with both directions of every edge possible, has a law in detailed balance with it. *Used for:* Branching state spaces (polymer and hierarchical models): equilibrium is automatic when the moves form a tree. *Price:* **free**; exact.

1161. Factorising a uniformly positive quadratic form. Every symmetric quadratic form in k variables that is at least $c > 0$ times the squared length is a weighted sum of k squares, each square combining one variable with the earlier ones, with every weight at least c. *Used for:* Diagonalising covariance, stiffness and inertia matrices; Gaussian elimination when every pivot is kept away from zero. *Price:* **free**; exact.

1162. Gaussian integral of a uniformly positive form. For every symmetric quadratic form in k variables that is at least $c > 0$ times the squared length, the integral of $e^{-\text{form}}$ over k dimensions, taken one coordinate at a time over the whole line, exists, and its square times the product of the form's pivots is π^k. *Used for:* Normalisation of every multivariate normal law; partition functions of coupled oscillators and lattice Gaussian fields. *Price:* **free**; exact.

1163. Kolmogorov's criterion for sparse chains (Wegscheider). A Markov chain whose possible moves form a symmetric network containing a spanning tree is reversible exactly when, on every move of the network, the products of rates along the tree paths balance, so that the rate products around every cycle agree in both directions. *Used for:* Wegscheider's condition for chemical reaction networks; equilibrium of Markov models on sparse state graphs. *Price:* **free**; exact.

1164. Deciding whether a sparse chain is reversible. Deciding, for every Markov chain whose possible moves form a symmetric network containing a spanning tree, whether it has a law in detailed balance with it. *Used for:* Deciding whether a reaction network's steady state is an equilibrium or a driven nonequilibrium state. *Price:* **WLPO**; exact.

1165. Triangular factorisation of a semidefinite quadratic form. Every symmetric quadratic form in finitely many variables that is never negative is a weighted sum of squares, each square combining one variable with the earlier ones, with every weight zero or positive. *Used for:* Degenerate covariance matrices and zero modes: factorising them requires knowing which pivots vanish exactly. *Price:* **LPO**; exact.

1166. Generating function of an independent sum. For two observables on finitely many outcomes whose joint law is the product of their separate laws, the mean of $e^{t(X+Y)}$ is the product of the means of e^{tX} and e^{tY}, for every real t. *Used for:* Cumulants of independent parts add; partition functions of uncoupled systems multiply. *Price:* **free**; exact.

1167. Generating function of a sum of independent copies. For a probability law on finitely many values, every k and every real t, the mean of e^{tS} over k independent copies, S their sum, is the k-th power of the mean of e^{tX}. *Used for:* Every moment of a sum of independent copies is read off one function; large deviations of averages; the free energy of k independent particles. *Price:* **free**; exact.

1168. Principal component of a covariance matrix. For every probability law on finitely many outcomes of a vector observable, the covariance matrix has a unit eigenvector whose eigenvalue is the largest variance of the observable along any unit direction. *Used for:* Principal component analysis; the dominant normal mode of thermal fluctuations; the direction of largest spread of a measurement. *Price:* **LLPO + AC**$_\omega$; exact.

1169. Chernoff–Cramér bound for sums of independent copies. For k independent copies of a law on finitely many values, a level a and a tilt $t \geq 0$, every event on which the sum is at least $k \cdot a$ has probability at most $e^{-t \cdot k \cdot a}$ times the k-th power of the mean of $e^{t \cdot X}$. *Used for:* Exponential concentration of averages of many independent measurements; large deviations of extensive quantities. *Price:* **free**; exact.

1170. Chernoff–Cramér bound for the exact tail set of a sum. For k independent copies of a law on finitely many values, a level a and a tilt $t \geq 0$, the set of outcomes on which the sum is at least $k \cdot a$ is an event and its probability is at most $e^{-t \cdot k \cdot a}$ times the k-th power of the mean of $e^{t \cdot X}$. *Used for:* Large-deviation bounds stated for the exact tail set of a sum. *Price:* **WLPO**; exact.

1171. Chernoff's optimal constant for sums of independent copies. For a law on finitely many values that gives positive weight to some value above a level a, the greatest lower bound ρ over tilts $t \geq 0$ of $e^{-t \cdot a}$ times the mean of $e^{t \cdot X}$ exists as a number, approached at some tilts; and for k independent copies every event on which the sum is at least $k \cdot a$ has probability at most ρ^k. *Used for:* Cramér's rate function of large deviations (ρ is e to the minus the rate); optimal exponential bounds on rare fluctuations of extensive quantities. *Price:* **AC**$_\omega$; exact (no decision principle); the upper bound uses countable choice.

1172. Largest variance along a unit direction. For a law on finitely many values of a vector observable, the largest variance along a unit direction exists as a number: no unit direction has larger variance, and unit directions come within any tolerance of it. *Used for:* Largest normal-mode amplitude of thermal fluctuations; the leading variance in principal component analysis; the top eigenvalue of a covariance or stiffness matrix. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1173. Optimal Chernoff tilt as a point. For a law on finitely many values that gives positive weight to some value above a level a and to some value below it, there is a tilt τ at which the tilted mean equals a, that is the mean of $(X - a) \cdot e^{\tau \cdot (X - a)}$ is zero, and the function $t \mapsto e^{-t \cdot a}$ times the mean of $e^{t \cdot X}$ is least at τ over every real t. *Used for:* Choosing the inverse temperature that gives a prescribed mean energy; the saddle point of Cramér's large deviations. *Price:* **AC**$_\omega$; exact (no decision principle); the upper bound uses countable choice.

1174. Ordered variances along the principal axes. For a law on finitely many values of a vector observable with $d + 1$ components, there are $d + 1$ numbers in decreasing order and a constant K such that, at every precision $1/(m+1)$, the covariance has an orthonormal basis with values whose eigen-equations hold within $1/(m+1)$ in every coordinate, and for every such basis and values the j-th largest value lies within $K/(m+1)$ of the j-th number. *Used for:* Variances along the principal axes of fluctuations; energies of normal modes; the spectrum of a covariance in principal component analysis. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1175. Markov chain tree theorem. For every square matrix P, let w(r) be the sum, over all trees on the states directed to r, of the product of the entries $P(k, f(k))$ along the tree's edges. Then for every state j, the sum over i of $w(i) \cdot P(i, j)$ equals w(j) times the sum over k of $P(j, k)$. *Used for:* Kirchhoff's and Hill's diagram method for steady states of master equations and of networks of chemical and biochemical reactions. *Price:* **free**; exact.

1176. Stationary law from tree weights. For a Markov chain on finitely many states whose total tree weight is positive, the tree weights divided by their total form a stationary law. *Used for:* Explicit steady states of finite Markov chains and master equations; steady-state populations of reaction networks. *Price:* **free**; exact.

1177. Positive tree weight of a reachable chain. For a square matrix with nonnegative entries off the diagonal and a graph of allowed moves, each with entry at least $\delta > 0$, that reaches one state r from every state in exactly m steps, the tree weight of r (the sum over trees directed to r of the products of their entries) is positive. *Used for:* Kirchhoff's and Hill's diagram method: steady states of connected chains and reaction networks are given by positive tree weights. *Price:* **free**; exact.

1178. Cramér's rate function at every interior level. For a law on finitely many values that gives positive weight to some value above a level a and to some value below it, there is a number $I \geq 0$ such that $e^{t \cdot a - I}$ is at most the mean of $e^{t \cdot X}$ for every real t, with equality at some t; that is, I is the largest value of $t \cdot a$ minus the logarithm of the mean of $e^{t \cdot X}$, and it is attained. *Used for:* Large-deviation rate of averages of many independent measurements; entropy as a Legendre transform in statistical mechanics. *Price:* **AC**$_\omega$; exact (no decision principle); the upper bound

uses countable choice.

1179. Cramér's rate function is convex. For a law on finitely many values, levels a and b and a weight s between 0 and 1, if Ia, Ib and Ic are the rates at a, b and s·a + (1 − s)·b, each in the attained form ($e^{t\cdot\text{level} - \text{rate}}$ at most the mean of $e^{t\cdot X}$ for every t, with equality at some t), then Ic is at most s·Ia + (1 − s)·Ib. *Used for:* Convexity of large-deviation rate functions; concavity of the microcanonical entropy. *Price:* **free**; exact.

1180. Weyl's bound for ordered variances. If the variances of two laws of a vector observable along every direction v differ by at most e times the squared length of v, then the j-th largest variances along the principal axes of the two laws differ by at most e, for every j. *Used for:* Stability of normal-mode spectra under small changes; how close sample principal variances are to the true ones. *Price:* **free**; exact.

1181. Cramér's lower bound: the rate is exact. For a law on finitely many values that gives positive weight to some value above a level a and to some value below it, let I be Cramér's rate at a in its attained form: $e^{t\cdot a - I}$ is at most the mean of $e^{t\cdot X}$ for every real t, with equality at some t. Then for every tolerance $\varepsilon = 1/(m+1)$ there is k_0 such that for every $k \geq k_0$ some event on which the sum of k independent copies is at least k·a has probability at least $e^{-k\cdot(I+\varepsilon)}$. *Used for:* Large deviations of averages of many independent measurements: the probability of a rare average decays exactly at Cramér's rate. *Price:* $\mathbf{AC}_{\omega}$; exact (no decision principle); the upper bound uses countable choice.

1182. Geometric convergence to equilibrium. A Markov chain on finitely many states whose allowed moves each have probability at least $\delta > 0$, and whose graph of allowed moves reaches one state from every state in exactly m steps, has a stationary law π, and from every initial law the law after q·m steps is within $2\cdot(1 - \delta^m)^q$ of π in total variation. *Used for:* Relaxation of finite systems to equilibrium at an explicit rate; convergence of Monte Carlo samplers. *Price:* **free**; exact.

1183. Kirchhoff's matrix-tree theorem. For every square matrix P, let L be its Laplacian: off the diagonal L(i, j) = −P(i, j), and on the diagonal L(i, i) is the sum of P(i, k) over $k \neq i$. Then for every state r, the tree weight of r (the sum over trees directed to r of the products of their entries) equals the determinant of L with row r and column r removed. *Used for:* Kirchhoff's 1847 counting of spanning trees in electrical networks; determinant formulas for steady states of master equations and reaction networks. *Price:* **free**; exact.

1184. Method of types: probability of a type class. For k independent copies of a law p on n letters and letter counts $c_1, \ldots, c_n$ summing to k, let P be the probability that the copies show exactly these counts. Then P times the product of $c_a^{c_a}$ is at most k^k times the product of $p_a^{c_a}$, and k^k times the product of $p_a^{c_a}$ is at most $(k + 1)^n$ times P times the product of $c_a^{c_a}$. With logarithms: P lies between $(k + 1)^{-n}\cdot e^{-k\cdot D}$ and $e^{-k\cdot D}$, where D is the relative entropy of the frequencies c/k from p. *Used for:* Boltzmann's counting of microstates (k^k over the product of $c_a^{c_a}$ is e to the k times the entropy of the frequencies); Sanov's large deviations of empirical frequencies. *Price:* **free**; exact.

1185. Steady states are Lipschitz in the rates. For two square matrices of size n with nonnegative entries off the diagonal, all entries at most B in size ($B \geq 1$), entries differing by at most ε, and total tree weights at least $w_0 > 0$, the laws proportional to the tree weights differ at every state by at most $(n + 1)\cdot n^n\cdot n\cdot B^n\cdot\varepsilon / w_0$. *Used for:* Response of nonequilibrium steady states of master equations and reaction networks to changes in the rates. *Price:* **free**; exact.

1186. Continuous-time relaxation of a master equation. Let Q be a matrix of transition rates on finitely many states: rows summing to zero, off-diagonal rates nonnegative, every diagonal entry at least −c with $c > 0$. Suppose the moves of a graph, each with rate at least e·c with $e > 0$, reach one state from every state in exactly m steps. Then Q has a steady state π, and for every initial law μ, every time $t \geq 0$, every limit E of the partial sums of $t^j/j!$ times the j-th power of Q (the exponential of t·Q) and every q, the law $\mu\cdot E$ is within $2\cdot(1 - e^m)^q + 2\cdot e^{-c\cdot t}$ times the sum over $j < q\cdot m$ of $(c\cdot t)^j/j!$ of π in total variation. *Used for:* Relaxation of continuous-time master equations (chemical kinetics, population dynamics, open quantum systems in the rate limit) to their steady state at an explicit rate. *Price:* **free**; exact.

1187. Sanov's upper bound for a set of types. For k independent copies of a law p on n letters, a number D and a decided set A of letter counts such that every count vector c in A summing to k satisfies k^k times the product of $p_a^{c_a}$ at most $e^{-k\cdot D}$ times the product of $c_a^{c_a}$ (relative entropy of the frequencies from p at least D), the probability that the letter counts of the copies lie in A is at most $(k + 1)^n\cdot e^{-k\cdot D}$. *Used for:* Large deviations of empirical frequencies (Sanov's theorem); the improbability of macrostates of lower entropy in Boltzmann's counting. *Price:* **free**; exact.

1188. Linear response of steady states. For square matrices Q and E whose total tree weight for Q is at least $w_0 > 0$, there are numbers D(j), a radius $s_0 > 0$ and a constant C such that for every s between $-s_0$ and s_0 the law proportional to the tree weights of Q + s·E differs from the law proportional to the tree

weights of Q by $s \cdot D(j)$ at every state j, up to an error of at most $C \cdot s^2$. *Used for:* Linear response of nonequilibrium steady states of master equations and reaction networks to small changes of the rates. *Price:* **free**; exact.

1189. Exponential of a matrix exists. For every square matrix Q on the line and every number t, the partial sums of $t^j/j!$ times the j-th power of Q converge at every entry. *Used for:* Time evolution of finite linear systems: master equations and finite quantum systems evolve by the exponential of t times their generator. *Price:* **free**; exact.

1190. Exponential of a rate matrix is stochastic. For a matrix of transition rates on finitely many states (rows summing to zero, off-diagonal rates nonnegative) and every time $t \geq 0$, every limit of the partial sums of the exponential of t times the rate matrix has nonnegative entries and rows summing to one. *Used for:* The solution of a master equation at every time is a probability law; the exponential of the generator is a Markov transition matrix. *Price:* **free**; exact.

1191. Static fluctuation–dissipation theorem. For finitely many states with energies H, an observable B coupled with strength s, an inverse temperature β and an observable A, the mean of A under the Gibbs law proportional to $e^{-\beta \cdot (H - s \cdot B)}$ equals its mean at $s = 0$ plus $s \cdot \beta$ times the covariance of A and B at $s = 0$, up to an error of at most $C \cdot s^2$ for all s between $-s_0$ and s_0, for some $s_0 > 0$ and C. *Used for:* Static susceptibilities equal β times equilibrium fluctuations (magnetic susceptibility, compressibility, heat capacity). *Price:* **free**; exact.

1192. Gibbs conditioning principle for types. For k independent copies of a law p on n letters, numbers D and η, and decided sets A and B of letter counts: suppose some count vector c in A with $c_1 + \ldots + c_n = k$ satisfies $e^{-k \cdot D}$ times the product of $c_a^{\,c_a}$ is at most k^k times the product of $p_a^{\,c_a}$ (the relative entropy of its frequencies from p is at most D), and every count vector c in B with $c_1 + \ldots + c_n = k$ satisfies k^k times the product of $p_a^{\,c_a}$ is at most $e^{-k \cdot (D + \eta)}$ times the product of $c_a^{\,c_a}$ (relative entropy at least $D + \eta$). Then the probability that the counts lie in B is at most $(k + 1)^{2n} \cdot e^{-k \cdot \eta}$ times the probability that they lie in A. *Used for:* Equivalence of ensembles and the maximum-entropy principle: under a macroscopic constraint the empirical frequencies concentrate on the frequencies of least relative entropy. *Price:* **free**; exact.

Q.10 Functional and Fourier analysis

1193. Cauchy–Schwarz inequality for inner products. In every real inner product space, the square of the inner product of two vectors is at most the product of their squared lengths: $\langle x, y\rangle^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$. *Used for:* Uncertainty relations and every bound on a correlation or a transition amplitude between two states. *Price:* **free**; exact.

1194. Cauchy–Schwarz inequality with norms. In every real inner product space, the absolute value of the inner product of two vectors is at most the product of their lengths: $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. *Used for:* Bounding the overlap of two quantum states by their norms. *Price:* **free**; exact.

1195. Triangle inequality in an inner product space. In every real inner product space, the length of a sum of two vectors is at most the sum of their lengths: $\|x + y\| \leq \|x\| + \|y\|$. *Used for:* Distances between states and error bounds on approximate state preparation. *Price:* **free**; exact.

1196. Parallelogram law for inner products. In every real inner product space, $\langle x + y, x + y\rangle + \langle x - y, x - y\rangle = 2(\langle x, x\rangle + \langle y, y\rangle)$. *Used for:* Recovering inner products from norms (polarization) and recognizing which norms come from an inner product. *Price:* **free**; exact.

1197. Pythagorean theorem for orthogonal vectors. In every real inner product space, if two vectors are orthogonal then the squared length of their sum is the sum of their squared lengths. *Used for:* Probabilities of mutually exclusive (orthogonal) outcomes add. *Price:* **free**; exact.

1198. Bessel's inequality for orthonormal families. For an orthonormal family of vectors and any finite set of its indices, the sum of the squared inner products of a vector x with those family members is at most $\|x\|^2$. *Used for:* The measured weights of a state on a set of orthogonal outcomes sum to at most its total norm. *Price:* **free**; exact.

1199. Orthonormal vectors are linearly independent. For an orthonormal family indexed by any set, a finite linear combination of its members that equals zero has every coefficient zero. *Used for:* Distinct orthonormal eigenstates are independent, so expansions in them are unique. *Price:* **free**; exact.

1200. Expansion in a finite orthonormal basis. If finitely many orthonormal vectors span the space, every vector x equals the sum over the basis of $\langle b, x\rangle$ times b. *Used for:* Expanding a quantum state in a finite basis of states. *Price:* **free**; exact.

1201. Parseval's identity in a finite basis. If

finitely many orthonormal vectors span the space, the inner product of any two vectors x and y equals the sum over the basis of $\langle x, b\rangle \cdot \langle b, y\rangle$. *Used for:* Inserting a resolution of the identity between two states. *Price:* **free**; exact.

1202. Eigenvectors of a symmetric operator are orthogonal. For a symmetric linear operator, an eigenvector for one eigenvalue and an eigenvector for a different eigenvalue are orthogonal. *Used for:* Eigenstates of an observable at different measured values are orthogonal. *Price:* **free**; exact.

1203. Nearest point characterized by orthogonality. For a subspace K and a point v of K, v is the point of K nearest to u exactly when u − v is orthogonal to every vector of K. *Used for:* Least-squares fitting and the variational principle for approximate states. *Price:* **free**; exact.

1204. Riesz representation in finite dimension. If finitely many orthonormal vectors span the space, every linear functional f equals the inner product with exactly one vector a: $f(x) = \langle a, x\rangle$ for all x. *Used for:* Bras and kets in a finite-dimensional quantum system. *Price:* **free**; exact.

1205. Orthogonal projection onto a line. For any vectors v and u in a real inner product space, there is a number t such that u − t·v is orthogonal to every multiple of v. *Used for:* Applying the measurement projector onto a state ψ, written $\psi\psi^*/\langle\psi, \psi\rangle$, to another state. *Price:* **LPO**; exact.

1206. Splitting a vector along and across a line. For any vectors v and u in a real inner product space, u is a multiple of v plus a vector orthogonal to every multiple of v. *Used for:* Splitting a state into its component along a reference state and the part orthogonal to it. *Price:* **LPO**; exact.

1207. Nearest multiple of a vector. For any vectors v and u in a real inner product space, some multiple of v is at least as close to u as every other multiple of v. *Used for:* The best fit of a state by multiples of one trial state. *Price:* **LPO**; exact.

1208. Double orthogonal complement of a line. For any vectors v and x in a real inner product space, x is orthogonal to every vector that is orthogonal to v exactly when x is a multiple of v. *Used for:* Identifying the one-dimensional subspace of a pure state from its orthogonal complement. *Price:* **LPO**; exact.

1209. Gram–Schmidt orthonormalization of independent vectors. Every linearly independent finite list of vectors in a real inner product space has an orthonormal list of equal length whose first k members span the space spanned by the first k given vectors, for every k. *Used for:* Building an orthonormal basis of states from independent states. *Price:* **MP**; exact.

1210. Existence of an orthonormal basis. Every real inner product space spanned by finitely many vectors has a finite orthonormal basis. *Used for:* Choosing a basis of states for a finite quantum system given by spanning states. *Price:* **LPO**; exact.

1211. Riesz representation on square-summable sequences. Every bounded linear functional f on the space of square-summable sequences is given by one square-summable sequence a: f(x) is the limit, as N grows, of the sums $a_0x_0 + a_1x_1 + \dots$ up to the N-th term, for every x. *Used for:* Bras and kets in infinite dimension: every bounded linear readout of a state is an overlap with a fixed state. *Price:* **LPO**; exact (limit points use the completeness of the line: countable choice of representatives).

1212. Norm of a bounded linear functional. Every bounded linear functional f on the space of square-summable sequences has a norm $L \geq 0$: $|f(x)| \leq L\cdot\|x\|$ for every x, and for every margin some x of length at most 1 has |f(x)| within that margin of L. *Used for:* The largest response of a linear measurement over normalized states. *Price:* **LPO**; exact (limit points use the completeness of the line: countable choice of representatives).

1213. Orthogonal projection onto any subspace of the line. Every subspace K of the real line, described by any condition at all, contains a point p such that u − p is orthogonal to every point of K. *Used for:* Projecting onto a subspace described only by a membership condition, with no spanning vectors given; it shows such descriptions can encode any yes-or-no question. *Price:* **EM**; exact.

1214. Orthogonal projection onto any subspace in finite dimension. In every real inner product space spanned by finitely many vectors, every subspace K, described by any condition at all, contains a point p such that u − p is orthogonal to every point of K, for every vector u. *Used for:* Projecting a state onto a subspace of a finite quantum system that is described only by a membership condition; it shows such descriptions can encode any yes-or-no question. *Price:* **EM**; exact.

1215. Inner product of square-summable sequences exists. For any two square-summable sequences x and y, the sums $x_0y_0 + x_1y_1 + \dots$ up to the N-th term converge to a number as N grows. *Used for:* The overlap of two states given by their amplitudes in an infinite basis. *Price:* **free**; exact.

1216. Norm of a square-summable sequence exists. For any square-summable sequence x there is a number $L \geq 0$ such that the sums $x_0^2 + x_1^2 + \dots$ up to the N-th term converge to L^2 as N grows. *Used for:* The total probability (norm) of a state given by its amplitudes in an infinite basis. *Price:* **free**; exact.

1217. Cauchy–Schwarz inequality for square-summable sequences. For square-summable sequences x and y, if the sums of $x_n y_n$ converge to I, the sums of x_n^2 to A and the sums of y_n^2 to B, then $I^2 \le A \cdot B$. *Used for:* Bounding transition amplitudes between states of an infinite-dimensional system. *Price:* **free**; exact.

1218. Completeness of the space of square-summable sequences. Every sequence of square-summable sequences of real numbers that is Cauchy in the square-summable distance converges, in that distance, to a square-summable sequence of real numbers. *Used for:* Limits of approximating states (perturbation series, variational sequences) are themselves states. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of representatives).

1219. Parseval's identity for an orthonormal basis. If an orthonormal sequence of square-summable sequences of real numbers has dense span, then for every square-summable sequence x the sums of the squared inner products of x with the first N basis members converge, as N grows, to the squared norm of x. *Used for:* Probabilities of the outcomes of a complete measurement add up to one. *Price:* **free**; exact.

1220. Parseval's identity for inner products. If an orthonormal sequence $b_0, b_1, b_2, \ldots$ of square-summable sequences of real numbers has dense span, then for all square-summable sequences x and y the sums of $\langle x, b_i\rangle \cdot \langle b_i, y\rangle$ over the first N basis members converge, as N grows, to $\langle x, y\rangle$. *Used for:* Inserting a complete set of states between two states in infinite dimension. *Price:* **free**; exact.

1221. Parseval's identity from a trivial orthogonal complement. If an orthonormal sequence of square-summable sequences of real numbers has no nonzero square-summable sequence orthogonal to all its members, then for every square-summable sequence x the sums of the squared inner products of x with the first N members converge, as N grows, to the squared norm of x. *Used for:* Recognizing a complete set of states from the fact that no state is orthogonal to all of them. *Price:* **open**; at least MP, with no choice; with the choice of representatives that the free allowance contains, exactly MP together with the double negation of LPO for sequences, one instance at a time: for every sequence of binary sequences, it is not false that some yes-or-no sequence says which of them contain a 1 (a principle that is not one of the letters); at most LPO, and at most MP together with the double negation shift (if $\neg\neg Q(n)$ for every n, then $\neg\neg$(Q(n) for every n)).

1222. Inversion of the discrete Fourier transform. For every $N \ge 1$ and every list f of N complex numbers, transforming f with the phases $e^{-2\pi i m k / N}$, transforming back with the phases $e^{2\pi i j k / N}$ and dividing by N returns f exactly. *Used for:* Bloch waves on a ring of N lattice sites are complete: every lattice state is recovered from its lattice-momentum amplitudes. *Price:* **free**; exact.

1223. Plancherel identity for the discrete Fourier transform. For every $N \ge 1$ and every list f of N complex numbers, the sum of the squared magnitudes of the discrete Fourier transform of f is N times the sum of the squared magnitudes of f. *Used for:* Passing to lattice momentum space preserves total probability (up to the normalization N). *Price:* **free**; exact.

1224. Orthogonality of the characters of a cyclic group. For every $N \ge 1$ and all j and m below N, the sum over k below N of $e^{2\pi i j k / N} \cdot e^{-2\pi i m k / N}$ is N when j = m and 0 otherwise. *Used for:* Plane waves of different lattice momenta on a ring are orthogonal. *Price:* **free**; exact.

1225. Numerical-radius bound for symmetric operators. For a symmetric linear operator T on a real inner product space and a number $M \ge 0$: if $|\langle T\,x, x\rangle| \le M \cdot \langle x, x\rangle$ for every x, then $|\langle T\,x, y\rangle| \le M \cdot \|x\| \cdot \|y\|$ for all x and y. *Used for:* Bounding all transition amplitudes of an observable by its largest expectation value on normalized states. *Price:* **free**; exact.

1226. Variational principle for the largest value. If a unit vector x_0 maximizes $\langle T\,x, x\rangle$ over all unit vectors x, for a bounded symmetric linear operator T on a real inner product space, then $T\,x_0 = \langle T\,x_0, x_0\rangle \cdot x_0$: x_0 is an eigenvector with that eigenvalue. *Used for:* The state of highest energy (or highest expectation of any observable) found by maximizing is an eigenstate. *Price:* **free**; exact.

1227. Variational principle for the ground state. If a unit vector x_0 minimizes $\langle T\,x, x\rangle$ over all unit vectors x, for a bounded symmetric linear operator T on a real inner product space, then $T\,x_0 = \langle T\,x_0, x_0\rangle \cdot x_0$: x_0 is an eigenvector with that eigenvalue. *Used for:* A ground state found by minimizing the energy over normalized states is an eigenstate of the Hamiltonian (the Rayleigh–Ritz principle). *Price:* **free**; exact.

1228. Variational principle with a nonzero maximizer of any length. If a vector x_0 that is not zero maximizes $\langle T\,x, x\rangle$ over all vectors x whose length equals the length of x_0, for a bounded symmetric linear operator T on a real inner product space, then there is a number μ with $\mu \cdot \langle x_0, x_0\rangle = \langle T\,x_0, x_0\rangle$ and $T\,x_0 = \mu \cdot x_0$. *Used for:* The variational principle when the trial state is not normalized, only known to be nonzero: the maximizing state is an eigenstate. *Price:* $\mathbf{MP}^{\vee}$; exact (the converse uses countable choice

of rational pairs, which the free allowance contains).

1229. Variational principle for a local extremum. If a unit vector x_0 is a local maximum or a local minimum of $\langle T\,x, x\rangle$ among the unit vectors within a fixed positive distance of x_0, for a bounded symmetric operator T on a real inner product space, then $T\,x_0 = \langle T\,x_0, x_0\rangle \cdot x_0$. *Used for:* A normalized state at which the energy is locally lowest or locally highest, even when it is not the ground state, is an eigenstate of the Hamiltonian. *Price:* **free**; exact.

1230. The trace does not depend on the basis. For a linear operator T on a real inner product space with two finite orthonormal bases b and c, the sum of $\langle T\,b_i, b_i\rangle$ over b equals the sum of $\langle T\,c_j, c_j\rangle$ over c. *Used for:* The trace of an observable or a density matrix takes one value whichever orthonormal basis computes it, so expectation values computed by traces are basis independent. *Price:* **free**; exact.

1231. The trace is the sum of the eigenvalues. If a linear operator T on a finite-dimensional real inner product space has an orthonormal basis of eigenvectors with eigenvalues λ_i, then the sum of $\langle T\,c_j, c_j\rangle$ over any orthonormal basis c equals the sum of the λ_i. *Used for:* The sum of the energy levels of a finite-level system equals the trace of its Hamiltonian. *Price:* **free**; exact.

1232. Uniform boundedness principle for functionals. If a sequence of bounded linear functionals on the space of square-summable sequences is bounded at each point, then one bound M works for all of them: the square of each value at x is at most M times the squared norm of x. *Used for:* Pointwise bounded families of linear measurements are uniformly bounded, which controls limits of measurement procedures. *Price:* **BD-N**; exact: the statement implies BD-N with no choice, and the converse uses only the choice of representatives that the free allowance contains; LPO also gives the statement.

1233. Existence of a Hilbert basis for a separable subspace. For every sequence v of square-summable sequences there are flags and vectors b such that the flagged vectors are orthonormal, the unflagged ones are zero, every b lies in the closed span of v, and every member of v lies in the closed span of b. *Used for:* Building a complete orthonormal set of states for any separable state space, such as the bound states of an atom or the modes of a field. *Price:* **free**; exact.

1234. Gram–Schmidt for an infinite independent sequence. Every linearly independent sequence of square-summable sequences has an orthonormal sequence such that, for each n, the first n members of each sequence are finite combinations of the first n members of the other. *Used for:* Building an orthonormal basis of states from independent states in an infinite-dimensional system, such as orthogonal polynomials or atomic orbitals. *Price:* **MP**; exact.

1235. Orthogonal projection onto the closed span of a sequence. For every sequence v of square-summable sequences and every square-summable x there is a point p in the closed span of v with $\langle x, v_j\rangle = \langle p, v_j\rangle$ for every j. *Used for:* Projecting a state onto the subspace spanned by infinitely many states, such as a band of energy levels or a symmetry sector. *Price:* **LPO**; exact (limit points use the completeness of the line: countable choice of representatives).

1236. Orthogonal projection onto a located closed subspace. For every sequence v of square-summable sequences, every square-summable x and every number d that is the distance from x to the closed span of v, there is a point p in that closed span with $\langle x, v_j\rangle = \langle p, v_j\rangle$ for every j. *Used for:* Projecting a state onto a subspace when the distance to it (for example the residual of a best fit) is known: then the projection costs nothing, whereas without the distance it costs LPO. *Price:* **free**; exact.

1237. Polarization identity for real inner products. In every real inner product space, $4\langle x, y\rangle = \|x + y\|^2 - \|x - y\|^2$. *Used for:* Recovering transition amplitudes from measured norms (probabilities) of superpositions. *Price:* **free**; exact.

1238. Equality in Cauchy–Schwarz, forward direction. In a real inner product space, if x and y are not zero and $|\langle x, y\rangle| = \|x\|\cdot\|y\|$, then y is a multiple of x. *Used for:* Two states whose overlap is as large as their lengths allow differ only by a factor. *Price:* **MP**; exact.

1239. Pythagorean theorem for finitely many orthogonal vectors. In a real inner product space, if finitely many vectors are pairwise orthogonal, the squared length of their sum is the sum of their squared lengths. *Used for:* Probabilities of finitely many mutually exclusive outcomes add. *Price:* **free**; exact.

1240. A symmetric operator with vanishing expectation values is zero. If a symmetric linear operator T on a real inner product space has $\langle Tx, x\rangle = 0$ for every x, then $T\,x = 0$ for every x. *Used for:* An observable whose expectation value vanishes in every state is the zero observable. *Price:* **free**; exact.

1241. Isometries preserve inner products. A linear map between real inner product spaces that preserves the length of every vector preserves the inner product of every pair of vectors. *Used for:* A transformation that preserves all probabilities of single states preserves all transition amplitudes (the real-linear part of Wigner's theorem). *Price:* **free**; exact.

1242. The projector onto a state is idempotent and symmetric. For a unit vector e in a real inner product space, the map $x \mapsto \langle e, x\rangle \cdot e$ applied twice equals itself applied once, and $\langle Px, y\rangle = \langle x, Py\rangle$ for all x and y. *Used for:* Repeating a measurement returns its first outcome, and the measurement projector is an observable. *Price:* **free**; exact.

1243. Uniform boundedness principle for operators. If a sequence of bounded linear operators on the space of square-summable sequences of real numbers is bounded at each point, then one bound M works for all of them: the squared norm of each image of x is at most M times the squared norm of x. *Used for:* Pointwise bounded families of evolutions or measurement maps are uniformly bounded, which controls limits of approximating dynamics. *Price:* **BD-N**; exact: the statement implies BD-N with no choice, and the converse uses only the choice of representatives that the free allowance contains.

1244. Riesz representation for functionals with a known norm. Every bounded linear functional f on the space of square-summable sequences of real numbers whose norm L is given is the inner product with one square-summable sequence a: f(x) is the limit of the sums $a_0x_0 + a_1x_1 + \dots$ up to the N-th term, for every x. *Used for:* Bras and kets in infinite dimension when the size of the readout is known: then the reference state is computed, with no decision. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of representatives).

1245. Equality in Cauchy–Schwarz, both directions. In a real inner product space, for nonzero x and y, $|\langle x, y\rangle| = \|x\|\cdot\|y\|$ exactly when $y = r\cdot x$ for some nonzero number r. *Used for:* Two states have maximal overlap exactly when one is a nonzero multiple of the other. *Price:* **MP**; exact.

1246. Top expectation value of a compact symmetric operator. For every compact symmetric operator T on the space of square-summable sequences of real numbers (compact: for every margin, the truncation of T to some finite number of coordinates is within that margin of T in operator norm) there is a number λ such that $\langle T\,x, x\rangle \le \lambda\cdot\|x\|^2$ for every x, and for every margin some unit vector x has $\langle T\,x, x\rangle$ within that margin of λ. *Used for:* The top (or, for $-T$, the ground) energy of a compact observable such as a bounded potential's resolvent, as a supremum over normalized states. *Price:* **free**; exact (no decision principle; the upper bound collects the top levels of the finite truncations, each determined uniquely, and forms their limit point, choices that the free allowance contains).

1247. Top eigenvector of a compact symmetric operator, positive case. For every compact symmetric operator T on the space of square-summable sequences of real numbers (compact: for every margin, the truncation of T to some finite number of coordinates is within that margin of T in operator norm) whose top expectation value λ over unit vectors is positive, some unit vector x satisfies $T\,x = \lambda\cdot x$. Without the positivity the statement is false: the diagonal operator with entries $-1/(n+1)$ has top value 0 and no eigenvector for it. *Used for:* A top eigenstate of a compact observable, the state of largest expectation value when that value is positive, exists as an actual eigenstate. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact; the lower bound uses no choice, the upper bound uses countable choice (through the compactness of a Hilbert cube, row 1251).

1248. Norm of a bounded operator. Every bounded linear operator T on the space of square-summable sequences of real numbers has a norm $L \ge 0$: $\|T\,x\|^2 \le L^2\|x\|^2$ for every x, and for every margin some x of norm at most 1 has $\|T\,x\|^2$ within that margin of L^2. *Used for:* The largest amplification of a linear evolution or response over normalized states. *Price:* **LPO**; exact (limit points use the completeness of the line: countable choice of representatives).

1249. Adjoint of a bounded operator. For every bounded linear operator T on the space of square-summable sequences of real numbers and every y there is z with $\langle T\,x, y\rangle = \langle x, z\rangle$ for every x. *Used for:* Moving an observable or evolution from kets to bras (the Hermitian conjugate) in infinite dimension. *Price:* **LPO**; exact (limit points use the completeness of the line: countable choice of representatives).

1250. Norm of a compact symmetric operator. Every compact symmetric operator T on the space of square-summable sequences of real numbers (compact: for every margin, the truncation of T to some finite number of coordinates is within that margin of T in operator norm) has a norm $L \ge 0$: $\|T\,x\|^2 \le L^2\|x\|^2$ for every x, and for every margin some x of norm at most 1 has $\|T\,x\|^2$ within that margin of L^2. *Used for:* The largest response of a compact observable, the larger of its top and bottom expectation values in absolute size. *Price:* **free**; exact (no decision principle; the upper bound collects the top levels of the finite truncations, each determined uniquely, and forms their limit point, choices that the free allowance contains); for a general bounded operator the same statement costs LPO (row 1248).

1251. Zero of a nonnegative Lipschitz function on a Hilbert cube. Fix tail bounds: the cube is the set of square-summable sequences of real numbers of norm at most 1 whose sum of squares beyond a given index for each margin is within that margin. Let F be a nonnegative function on the cube with a constant L

such that $(F(x) - F(y))^2$ is at most L times the squared distance of x and y for all points x and y of the cube. If F comes arbitrarily close to 0 on the cube, then F takes the value 0 at some point of the cube. *Used for:* Existence of minimizers by compactness in infinite dimension: ground states and best approximations found as limits of near-minimizers. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact; the lower bound uses no choice (one coordinate already gives LLPO), the upper bound uses countable choice.

1252. Adjoint of a bounded operator with a known column modulus. For a bounded linear operator T on the space of square-summable sequences and a vector y, if the numbers $\langle T\, e_n, y\rangle$ come with a known rate at which their squares sum, then the vector z of these numbers satisfies $\langle Tx, y\rangle = \langle x, z\rangle$ for every x. *Used for:* Moving an observable to the other side of an inner product (the adjoint of a measurement) when its matrix column is known with a convergence rate. *Price:* **free**; exact (no choice); without the rate the same statement costs LPO (row 1249).

1253. Riesz representation from the values on the unit vectors. A bounded linear functional f on the space of square-summable sequences equals $\langle x, w\rangle$ for every x, for any presented square-summable w whose n-th coordinate is $f(e_n)$. *Used for:* A linear measurement is the overlap with one fixed state, read off from its values on the basis states. *Price:* **free**; exact (no choice); finding w without it costs LPO (row 1211).

1254. Spectral theorem for compact symmetric operators, approximation form. For every compact symmetric operator T on the space of square-summable sequences and every margin, finitely many orthonormal eigenvectors $u_1, \ldots, u_m$ with eigenvalues $\mu_1, \ldots, \mu_m$ make $T - \Sigma\, \mu_i\langle\cdot, u_i\rangle u_i$ smaller than the margin in norm. *Used for:* Expanding a compact observable in its eigenstates up to any accuracy: the finite-mode truncation of a spectral decomposition. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact; the lower bound uses no choice (a two-by-two test block), the upper bound uses countable choice.

1255. Spectral theorem for compact symmetric operators, sequence form. Every compact symmetric operator T on the space of square-summable sequences has a sequence of pairwise orthogonal eigenvectors $u_0, u_1, \ldots$, each a unit vector or zero, with eigenvalues $\mu_0, \mu_1, \ldots$, such that the partial sums $\Sigma\, \mu_j\langle x, u_j\rangle u_j$ approach Tx uniformly on the unit ball. *Used for:* The full eigenmode expansion of a compact observable, as used for Hilbert–Schmidt integral operators and Green's functions. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact; the lower bound uses no choice, the upper bound uses countable choice (an earlier proof used dependent choice; the margins are now chosen independently).

1256. Riesz representation with a known rate for the functional's values. A bounded linear functional f on the space of square-summable sequences whose values $f(e_n)$ come with a known rate at which their squares sum equals $\langle x, w\rangle$ for some square-summable w and every x. *Used for:* A linear measurement is the overlap with one fixed state, when the size of its response to far basis states is known. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of representatives).

Q.11 A first survey across the areas

1257. Preorder axioms for the order of the line. On the line, $\le$ and $<$ satisfy the preorder axioms: $\le$ is reflexive and transitive, and $x < y$ holds exactly when $x \le y$ and not $y \le x$. *Used for:* comparing physical quantities in every bound and inequality. *Price:* **MP**; exact.

1258. Partial order axioms for the order of the line. On the line, $\le$ and $<$ satisfy the partial order axioms: $\le$ is reflexive, transitive and antisymmetric, and $x < y$ holds exactly when $x \le y$ and not $y \le x$. *Used for:* comparing physical quantities; concluding equality from bounds in both directions. *Price:* **MP**; exact.

1259. No zero divisors on the line. On the line, the ring of numbers has no zero divisors: whenever $x \cdot y = 0$, $x = 0$ or $y = 0$. *Used for:* solving factored equations, such as finding roots of characteristic polynomials. *Price:* **LLPO**; exact.

1260. Inverses of nonzero numbers. On the line, there are two different numbers, multiplication is commutative, and every number that is not equal to 0 has a multiplicative inverse. *Used for:* all algebra with division: ratios, rates and solving linear equations. *Price:* **MP**; exact. Proved in Theorem 28.9(6).

1261. Trichotomy of the order. For any two numbers x and y on the line, $x < y$, $x = y$ or $y < x$. *Used for:* deciding whether a quantity is below, at or above a threshold. *Price:* **LPO**; exact.

1262. Totality of the order $\le$. For any two numbers x and y on the line, $x \le y$ or $y \le x$. *Used for:* ranking quantities, such as ordering energy levels. *Price:* **LLPO**; exact.

1263. Order dichotomy: $x < y$ or $y \le x$. For any two numbers x and y on the line, $x < y$ or $y \le x$. *Used for:* case splits at a threshold: below a critical value, or at or above it. *Price:* **LPO**; exact.

1264. Order dichotomy: $x \le y$ or $y < x$. For any two numbers x and y on the line, $x \le y$ or $y < x$. *Used for:* case splits at a threshold: at or below a critical value, or above it. *Price:* **LPO**; exact.

1265. Deciding equality of two numbers. Any two numbers x and y on the line are either equal or not equal. *Used for:* testing whether two quantities agree exactly, such as a resonance condition. *Price:* **WLPO**; exact. Proved in Corollary 12.18.

1266. Splitting $\le$ into $<$ or $=$. On the line, $x \le y$ implies $x < y$ or $x = y$. *Used for:* separating strict from borderline cases, such as marginal stability. *Price:* **LPO**; exact.

1267. Absolute value equals x or −x. For every number x on the line, the absolute value of x equals x or equals −x. *Used for:* removing absolute values in magnitude and distance calculations. *Price:* **LLPO**; exact.

1268. The maximum is one of its arguments. For any two numbers x and y on the line, max(x, y) equals x or equals y. *Used for:* picking the larger of two quantities in piecewise and threshold laws. *Price:* **LLPO**; exact.

1269. Zero product rule. On the line, if $x \cdot y = 0$ then $x = 0$ or $y = 0$. *Used for:* solving factored equations for equilibria and roots. *Price:* **LLPO**; exact.

1270. Integer floor of a number. Every number x on the line has an integer floor: an integer m with $m \le x < m + 1$. *Used for:* counting whole periods, cycles or lattice cells; binning measured values. *Price:* **LPO**; exact.

1271. Unequal numbers are strictly ordered. On the line, if $x \ne y$ then $x < y$ or $y < x$. *Used for:* turning a difference between two quantities into a definite sign. *Price:* **MP**; exact.

1272. Weak inequality between unequal numbers is strict. On the line, if $x \le y$ and $x \ne y$ then $x < y$. *Used for:* strict bounds from a weak bound plus an exclusion, such as a nonzero energy gap. *Price:* **MP**; exact.

1273. Failure of $x \le y$ gives $y < x$. On the line, if $x \le y$ fails then $y < x$. *Used for:* arguments by contradiction about inequalities between physical quantities. *Price:* **MP**; exact.

1274. Failure of $y \le x$ gives $x < y$. On the line, if $y \le x$ fails then $x < y$. *Used for:* arguments by contradiction about inequalities between physical quantities. *Price:* **MP**; exact.

1275. Nonzero numbers have positive absolute value. On the line, every number $x \ne 0$ has an absolute value strictly greater than 0. *Used for:* positive magnitudes of nonzero quantities, needed before dividing by them. *Price:* **MP**; exact.

1276. Inverse of a nonzero number. On the line, every number $x \ne 0$ has a number y with $x \cdot y = 1$. *Used for:* dividing by nonzero quantities: ratios, rates and unit conversions. *Price:* **MP**; exact.

1277. Negation of $<$ is the reverse $\le$. On the line, $x < y$ fails exactly when $y \le x$. *Used for:* comparing physical quantities in every bound and inequality. *Price:* **free**; exact.

1278. Failure of $y < x$ gives $x \le y$. On the line, if $y < x$ fails then $x \le y$. *Used for:* deriving bounds by ruling out the opposite strict inequality. *Price:* **free**; exact.

1279. Antisymmetry of $\le$ on the line. On the line, $x \le y$ and $y \le x$ together give $x = y$. *Used for:* proving two quantities equal from bounds in both directions. *Price:* **free**; exact.

1280. Transitivity of $\le$. On the line, $x \le y$ and $y \le z$ give $x \le z$. *Used for:* chaining bounds, as in every error estimate. *Price:* **free**; exact.

1281. Irreflexivity of $<$. No number x on the line satisfies $x < x$. *Used for:* consistency of strict comparisons: no quantity exceeds itself. *Price:* **free**; exact.

1282. Strict then weak inequality gives strict. On the line, $x < y$ and $y \le z$ give $x < z$. *Used for:* chaining strict and weak bounds in estimates. *Price:* **free**; exact.

1283. Absolute values are nonnegative. For every number x on the line, the absolute value of x is at least 0. *Used for:* magnitudes, distances and error sizes are never negative. *Price:* **free**; exact.

1284. Triangle inequality for absolute values. For all numbers x and y on the line, $|x + y| \le |x| + |y|$. *Used for:* combining error bounds and adding displacements. *Price:* **free**; exact.

1285. Reverse triangle inequality. For all numbers x and y on the line, $|x| - |y| \le |x - y|$. *Used for:* bounding how much a magnitude changes under a perturbation. *Price:* **free**; exact.

1286. Absolute value of a product. For all numbers x and y on the line, $|x \cdot y| = |x| \cdot |y|$. *Used for:* magnitudes of products: gains, scale factors and relative errors. *Price:* **free**; exact.

1287. Zero absolute value means zero. For every number x on the line, $|x| = 0$ exactly when $x = 0$. *Used for:* a quantity of zero magnitude vanishes; zero error means exact agreement. *Price:* **free**; exact.

1288. A number is at most the maximum. For all numbers x and y on the line, $x \le \max(x, y)$. *Used for:*

bounding a quantity by the larger of two bounds. *Price:* **free**; exact.

1289. Maximum below a common upper bound. On the line, if $x \le z$ and $y \le z$ then $\max(x, y) \le z$. *Used for:* worst-case bounds over several cases. *Price:* **free**; exact.

1290. The minimum is at most each number. For all numbers x and y on the line, $\min(x, y) \le x$. *Used for:* taking the tighter of two bounds. *Price:* **free**; exact.

1291. Squares are nonnegative. For every number x on the line, $0 \le x \cdot x$. *Used for:* kinetic energies, intensities and variances are never negative. *Price:* **free**; exact.

1292. Sum of positive numbers is positive. On the line, if $0 < x$ and $0 < y$ then $0 < x + y$. *Used for:* adding positive masses, energies or probabilities stays positive. *Price:* **free**; exact.

1293. Product of positive numbers is positive. On the line, if $0 < x$ and $0 < y$ then $0 < x \cdot y$. *Used for:* signs of products, such as density times volume. *Price:* **free**; exact.

1294. Inverse of a positive number is positive. On the line, for every number x with $0 < x$, the inverse $1/x$ also satisfies $0 < 1/x$. *Used for:* periods from frequencies and resistances from conductances stay positive. *Price:* **free**; exact.

1295. A rational number between any two numbers. On the line, if $x < y$ then some rational number q satisfies $x < q < y$. *Used for:* approximating any real quantity by finite-precision rational readings. *Price:* **free**; exact.

1296. Density of the order. On the line, if $x < y$ then some number z satisfies $x < z < y$. *Used for:* continuum models: between any two positions or times lies another. *Price:* **free**; exact.

1297. Archimedean property of the line. For every number x on the line, some natural number n satisfies $x < n$. *Used for:* exceeding any quantity by enough unit steps; choosing scales and counts. *Price:* **free**; exact.

1298. Nonnegative square roots. On the line, every number $x \ge 0$ has a square root y with $y \ge 0$ and $y \cdot y = x$. *Used for:* magnitudes from sums of squares: lengths, speeds and root-mean-square values. *Price:* **free**; exact.

1299. Square of a sum. For all numbers a and b on the line, $(a + b)^2 = a^2 + 2ab + b^2$. *Used for:* expanding energies with cross terms, such as interference terms. *Price:* **free**; exact.

1300. Cauchy–Schwarz inequality in the plane. For all numbers a, b, c, d on the line, $(ac + bd)^2 \le (a^2 + b^2)(c^2 + d^2)$: the two-dimensional case, in squared form. *Used for:* bounding dot products: work, projections and correlation coefficients. *Price:* **free**; exact.

1301. Least upper bounds of sets. On the line, every set of numbers that has a member and an upper bound has a least upper bound. *Used for:* defining suprema: peak values, best constants and spectral bounds. *Price:* **WEM**; exact (the converse uses countable choice once, at pairs of ratios). Proved in Theorem 12.50.

1302. Monotone convergence theorem. On the line, every nondecreasing sequence that is bounded above converges to a limit. *Used for:* sums of positive contributions, such as total probability or accumulated energy. *Price:* **LPO**; exact (the converse chooses representatives and uses countable choice once at pairs of ratios, both inside the free allowance). Proved in Theorem 12.45.

1303. Monotone convergence for decreasing sequences. On the line, every nonincreasing sequence that is bounded below converges to a limit. *Used for:* improving variational upper bounds, such as ground-state energy estimates. *Price:* **LPO**; exact (apply row 1302 to $-x$). Proved in Theorem 12.45, applied to $-x$.

1304. Bolzano–Weierstrass theorem for bounded sequences. On the line, every bounded sequence has a convergent subsequence. *Used for:* extracting limiting states from bounded trajectories or approximations. *Price:* **LPO**; exact (the converse chooses representatives, yes-or-no decisions and indices, all inside the free allowance). Proved in Theorem 12.47(i).

1305. Every sequence has a monotone subsequence. On the line, every sequence has a subsequence that is nondecreasing or nonincreasing. *Used for:* a step in proving that bounded motions have limiting states. *Price:* **LPO**; exact (the converse chooses representatives, yes-or-no decisions and indices, all inside the free allowance). Proved in Theorem 12.47(iii).

1306. Bounded sequences have a cluster point. On the line, read with a bounded set in place of the compact set: every bounded sequence has a cluster point, approached within every tolerance $1/(k+1)$ by terms of arbitrarily late index. *Used for:* long-time behavior of bounded motions: recurrence and limit cycles. *Price:* **LPO**; exact (the converse chooses representatives, yes-or-no decisions and indices, all inside the free allowance). Proved in Theorem 12.47(ii).

1307. Least upper bound of a sequence bounded above. On the line, every sequence that is bounded above has a least upper bound. *Used for:* peak values over infinitely many times or cases. *Price:* **WLPO**;

exact (the converse chooses representatives and uses countable choice once at pairs of ratios, both inside the free allowance). Proved in Theorem 12.48.

1308. Greatest lower bound of a sequence bounded below. On the line, every sequence that is bounded below has a greatest lower bound. *Used for:* lowest energy over a sequence of trial states or configurations. *Price:* **WLPO**; exact (apply row 1307 to −x). Proved in Theorem 12.48, applied to $-x$.

1309. Terms of a sequence approach its supremum. On the line, every sequence bounded above has a least upper bound L such that every rational number below L is exceeded by some term. *Used for:* approaching a peak value by actual terms, such as maximal energy along an orbit. *Price:* **LPO**; exact (the converse chooses representatives and uses countable choice once at pairs of ratios, both inside the free allowance). Proved in Theorem 12.49.

1310. Limit superior of a bounded sequence. On the line, every bounded sequence has a limit superior: each tail has a least upper bound, and these bounds have a greatest lower bound. *Used for:* long-run worst-case values: asymptotic amplitudes and upper growth rates. *Price:* **WLPO**; exact (the converse applies row 1307 to every tail and row 1308 to the tail bounds; collecting those bounds, each a single point, uses only the free allowance).

1311. Completeness: Cauchy sequences converge. On the line, every Cauchy sequence converges to a limit. *Used for:* convergence of iterative computations and series, such as perturbation series. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Theorem 12.26.

1312. Intermediate value theorem on [0, 1]. On the line, every uniformly continuous function f on [0, 1] with f(0) ≤ 0 ≤ f(1) has a zero: some x in [0, 1] with f(x) = 0. *Used for:* existence of roots and equilibria, such as a balance point or crossing time. *Price:* **LLPO**; exact (the converse tabulates one yes-or-no answer for every dyadic ratio, a choice inside the free allowance). Proved in Theorem 12.54.

1313. Extreme value theorem: the maximum is attained. On the line, every uniformly continuous function f on [0, 1] attains a maximum: some x in [0, 1] has f(y) ≤ f(x) for every y in [0, 1]. *Used for:* existence of extreme configurations, such as turning points and peak energies. *Price:* **LLPO**; exact (the converse tabulates yes-or-no answers for adjacent dyadic intervals and uses countable choice at pairs of ratios for the suprema, both inside the free allowance). Proved in Theorem 12.60.

1314. Extreme value theorem: the minimum is attained. On the line, every uniformly continuous function f on [0, 1] attains a minimum: some x in [0, 1] has f(x) ≤ f(y) for every y in [0, 1]. *Used for:* existence of ground states and stable equilibria at minimum energy. *Price:* **LLPO**; exact (apply row 1313 to −f). Proved in Theorem 12.60, applied to $-f$.

1315. Continuous functions on [0, 1] are bounded above. On the line, every uniformly continuous function f on [0, 1] is bounded above there: some b has f(y) ≤ b for every y in [0, 1]. *Used for:* a priori bounds on fields and signals over a finite range. *Price:* **free**; exact.

1316. Supremum of a uniformly continuous function on [0, 1]. On the line, every uniformly continuous f on [0, 1] has a supremum s: f(y) ≤ s for every y in [0, 1], and values of f come within any tolerance of s. *Used for:* peak values of signals, potentials or fields over a finite range. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs). Proved in Theorem 12.58.

1317. Riemann sums of a uniformly continuous function converge. On the line, for every uniformly continuous f on [0, 1], the left Riemann sums over 2^n equal pieces of [0, 1] converge to a number. *Used for:* work, charge, mass and probability as integrals of continuous densities. *Price:* **free**; exact (a modulus of uniform continuity is chosen by countable choice of natural numbers; the Riemann sums then form a Cauchy sequence with a known modulus, and its limit point uses the completeness of the line: countable choice of rational pairs, Corollary 12.35).

1318. Limits of sequences are unique. On the line, a sequence that converges to L and also to L′ has L = L′. *Used for:* well-defined limiting values, such as one equilibrium or one extrapolated measurement. *Price:* **free**; exact.

1319. Limit of a sum. On the line, if sequences x and y converge to L and M, then the termwise sum x + y converges to L + M. *Used for:* adding limits of approximations, such as a total energy from its parts. *Price:* **free**; exact.

1320. Limit of a constant multiple. On the line, if a sequence x converges to L, then for every number c the sequence c · x converges to c · L. *Used for:* rescaling limits: unit changes and coupling constants. *Price:* **free**; exact.

1321. Squeeze theorem for sequences. On the line, if x ≤ y ≤ z termwise and the sequences x and z both converge to L, then y converges to L. *Used for:* pinning a quantity between lower and upper estimates that agree in the limit. *Price:* **free**; exact.

1322. 1/(n+1) tends to zero. On the line, the sequence 1/(n+1) converges to 0. *Used for:* vanishing step sizes and errors in discretizations. *Price:* **free**;

exact.

1323. A nondecreasing sequence stays below its limit. On the line, if a nondecreasing sequence converges to L, then every term is at most L. *Used for:* monotone approximations bound their limit, such as partial sums of positive terms. *Price:* **free**; exact.

1324. The identity function is continuous. On the line, the function $x \mapsto x$ is continuous at every point. *Used for:* the starting point for continuity of every formula built from a variable. *Price:* **free**; exact.

1325. Constant functions are continuous. On the line, for every number c, the constant function with value c is continuous at every point. *Used for:* constant fields and parameters as continuous ingredients of physical laws. *Price:* **free**; exact.

1326. Composition of continuous functions is continuous. On the line, if f and g are continuous functions, then $x \mapsto f(g(x))$ is continuous. *Used for:* continuity of composite laws, such as a potential evaluated along a path. *Price:* **free**; exact.

1327. Sum of continuous functions is continuous. On the line, if f and g are continuous functions, then $x \mapsto f(x) + g(x)$ is continuous. *Used for:* continuity of superposed fields and summed potentials. *Price:* **free**; exact.

1328. Maximum of continuous functions is continuous. On the line, if f and g are continuous functions, then $x \mapsto \max(f(x), g(x))$ is continuous. *Used for:* continuity of piecewise laws, such as thresholds and rectified signals. *Price:* **free**; exact.

1329. Absolute value is continuous. On the line, the function $x \mapsto |x|$ is continuous at every point. *Used for:* magnitudes and distances that vary continuously with the state. *Price:* **free**; exact.

1330. Continuity on a set passes to subsets. On the line, if f is continuous on a set T and S is a subset of T, then f is continuous on S. *Used for:* restricting a law to a smaller region or time interval. *Price:* **free**; exact.

1331. Epsilon-delta definition of continuity. On the line, f is continuous exactly when, for every x and every tolerance $1/(k+1)$, some $1/(j+1)$ works: points within $1/(j+1)$ of x have values within $1/(k+1)$ of $f(x)$. *Used for:* turning continuity into explicit error tolerances in calculations. *Price:* **free**; exact.

1332. Composition of uniformly continuous functions. On the line, if f and g are uniformly continuous functions, then $x \mapsto f(g(x))$ is uniformly continuous. *Used for:* error control through chained calculations, with one tolerance valid everywhere. *Price:* **free**; exact.

1333. Lipschitz functions are uniformly continuous. On the line, if $|f(x) - f(y)| \le K \cdot |x - y|$ for all x and y, with K a natural number, then f is uniformly continuous. *Used for:* stable dependence on data: bounded sensitivity in models and numerical schemes. *Price:* **free**; exact.

1334. Epsilon-delta definition of uniform continuity. On the line, f is uniformly continuous exactly when, for every tolerance $1/(k+1)$, one $1/(j+1)$ works everywhere: points within $1/(j+1)$ have values within $1/(k+1)$. *Used for:* uniform error tolerances in discretizations and numerical simulations. *Price:* **free**; exact.

1335. Banach fixed point theorem on the line. On the line, a contraction f with factor q ($0 \le q < 1$) has a fixed point p; iterates from any x_0 converge to p, with $(1 - q) \cdot |f^n(x_0) - p| \le q^n \cdot |f(x_0) - x_0|$. *Used for:* self-consistent solutions found by iteration, such as Picard iteration for equations of motion. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1336. A contraction has at most one fixed point. On the line, a contraction f with factor q ($0 \le q < 1$) has at most one fixed point: $f(p) = p$ and $f(p') = p'$ give $p = p'$. *Used for:* uniqueness of equilibria and of self-consistent solutions. *Price:* **free**; exact.

1337. Iterates of a contraction converge to a given fixed point. On the line, if a contraction f with factor q ($0 \le q < 1$) has a fixed point p, the iterates $f^n(x_0)$ from any start x_0 converge to p. *Used for:* convergence of iterative solvers to an equilibrium or steady state. *Price:* **free**; exact.

1338. A priori error bound for contraction iterates. On the line, if a contraction f with factor q ($0 \le q < 1$) has a fixed point p, then $(1 - q) \cdot |f^n(x_0) - p| \le q^n \cdot |f(x_0) - x_0|$ for every start x_0 and every n. *Used for:* estimating in advance how many iterations a self-consistent calculation needs. *Price:* **free**; exact.

1339. A posteriori error bound for contraction iterates. On the line, if a contraction f with factor q ($0 \le q < 1$) has a fixed point p, then $(1 - q) \cdot |f^n(x) - p| \le |f^{n+1}(x) - f^n(x)|$ for every x and every n. *Used for:* stopping rules for iterative solvers, read from the size of the last step. *Price:* **free**; exact.

1340. Distance to the fixed point from one step. On the line, if a contraction f with factor q ($0 \le q < 1$) has a fixed point p, then $(1 - q) \cdot |x - p| \le |f(x) - x|$ for every x. *Used for:* error estimates for approximate equilibria from their residual. *Price:* **free**; exact.

1341. Powers of a number below one tend to zero. On the line, for every r with $0 \le r < 1$, the powers r^n converge to 0. *Used for:* exponential decay step by step: damping, attenuation and radioactive decay. *Price:* **free**; exact.

1342. Geometrically shrinking steps give a Cauchy sequence. On the line, if $|u_n - u_{n+1}| \le C \cdot r^n$

for every n, with $0 \le r < 1$, then the sequence u is Cauchy. *Used for:* convergence of iterations whose corrections shrink geometrically. *Price:* **free**; exact.

1343. Sum of the geometric series. On the line, for every r with $0 \le r < 1$, the partial sums of $1 + r + r^2 + \cdots$ converge to the number s with $(1 - r) \cdot s = 1$. *Used for:* summing multiple reflections and repeated fractional losses or returns. *Price:* **free**; exact.

1344. The geometric series converges. On the line, for every r with $0 \le r < 1$, the partial sums of $1 + r + r^2 + \cdots$ converge. *Used for:* bounding perturbation series and iterated corrections by a geometric series. *Price:* **free**; exact.

1345. Terms of a convergent series tend to zero. On the line, if the partial sums of a series converge, then its terms converge to 0. *Used for:* a quick divergence test for series expansions. *Price:* **free**; exact.

1346. Absolute convergence implies convergence. On the line, if the partial sums of $|a_n|$ converge, then the partial sums of a_n converge. *Used for:* convergence of Fourier and perturbation series whose term sizes are summable. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1347. Comparison test for series. On the line, if $0 \le a_n \le b_n$ for every n and the partial sums of b_n converge, then the partial sums of a_n converge. *Used for:* proving sums of positive terms converge by bounding them, such as partition functions. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1348. Sum rule for derivatives. On the line, if f has derivative d_1 and g has derivative d_2 at x, then f + g has derivative $d_1 + d_2$ at x. *Used for:* rates of change of summed quantities, such as total energy or net force. *Price:* **free**; exact.

1349. Derivative of the identity function. On the line, the function $x \mapsto x$ has derivative 1 at every point. *Used for:* the base case for differentiating every polynomial law of motion. *Price:* **free**; exact.

1350. Derivative of a constant is zero. On the line, every constant function has derivative 0 at every point. *Used for:* constant quantities have zero rate of change: static and conserved values. *Price:* **free**; exact.

1351. Differentiable functions are continuous. On the line, a function that has a derivative at x is continuous at x. *Used for:* a trajectory that has a velocity has no jumps. *Price:* **free**; exact.

1352. The derivative at a point is unique. On the line, if f has derivative d_1 and also derivative d_2 at one point x, then $d_1 = d_2$. *Used for:* well-defined velocities, rates and slopes. *Price:* **free**; exact.

1353. Fermat's theorem at a local maximum. On the line, if f has derivative d at x and a local maximum at x, then $d = 0$. *Used for:* locating turning points and peaks, such as the top of a projectile's flight. *Price:* **free**; exact.

1354. Fermat's theorem at a local minimum. On the line, if f has derivative d at x and a local minimum at x, then $d = 0$. *Used for:* locating equilibria at potential minima and turning points of motion. *Price:* **free**; exact.

1355. Mean value inequality on [0, 1]. If f' is a uniform derivative of f on [0, 1] with $|f'| \le C$ there, then $|f(y) - f(x)| \le C \cdot |y - x|$ for all x, y in [0, 1]. *Used for:* bounding a change by its largest rate, such as distance by top speed times time. *Price:* **free**; exact.

1356. Nonnegative derivative gives an increasing function, on [0, 1]. If f' is a uniform derivative of f on [0, 1] and $f' \ge 0$ there, then $f(x) \le f(y)$ whenever $x \le y$ in [0, 1]. *Used for:* a quantity with a nonnegative rate of change, such as entropy, never decreases. *Price:* **free**; exact.

1357. Zero derivative gives a constant, on [0, 1]. If f' is a uniform derivative of f on [0, 1] and $f' = 0$ there, then $f(x) = f(y)$ for all x and y in [0, 1]. *Used for:* conservation laws: zero time derivative means constant energy or momentum. *Price:* **free**; exact.

1358. Rolle's theorem on [0, 1] with a uniform derivative. If f' is a uniform derivative of f on [0, 1] and $f(0) = f(1)$, then $f'(c) = 0$ at some point c of the closed interval [0, 1]. *Used for:* turning points: zero vertical velocity at the top of a throw landing at launch height. *Price:* **LLPO**; exact (the converse uses the choices of row 1312 and row 1313, which the free allowance contains). Proved in Theorem 12.87.

1359. Mean value theorem on [0, 1] with a uniform derivative. If f' is a uniform derivative of f on [0, 1], then $f'(c) = f(1) - f(0)$ at some point c of the closed interval [0, 1]. *Used for:* average velocity over a trip equals the instantaneous velocity at some moment. *Price:* **LLPO**; exact (the converse uses the choices of row 1312 and row 1313, which the free allowance contains). Proved in Theorem 12.87.

1360. Cauchy mean value theorem on [0, 1], closed-interval point. If f' and g' are uniform derivatives of f and g on [0, 1], then $(g(1) - g(0)) \cdot f'(c) = (f(1) - f(0)) \cdot g'(c)$ at some c in [0, 1]. *Used for:* limits of ratios by l'Hôpital's rule; comparing two rates along one motion. *Price:* **LLPO**; exact (the converse uses the choices of row 1312 and row 1313, which the free allowance contains). Proved in Theorem 12.87.

1361. Determinant of a product, 2 × 2. For 2 × 2 matrices A and B with entries on the line, the determinant of AB is $\det(A) \cdot \det(B)$. *Used for:* area factors of successive changes of variables multiply, as

Jacobians do. *Price:* **free**; exact.

1362. Determinant of the transpose, 2 × 2. Every 2 × 2 matrix with entries on the line and its transpose have equal determinants. *Used for:* Jacobians and linear systems give equal determinants by rows or by columns. *Price:* **free**; exact.

1363. Determinant of the identity, 2 × 2. The 2 × 2 identity matrix on the line has determinant 1. *Used for:* the identity map preserves areas and phase-space volume. *Price:* **free**; exact.

1364. Determinant of a scalar multiple, 2 × 2. For every point s of the line and every 2 × 2 matrix A, the determinant of s times A is $s^2 \cdot \det(A)$. *Used for:* scaling lengths by s scales areas by s^2. *Price:* **free**; exact.

1365. Cyclic property of the trace, 2 × 2. For 2 × 2 matrices A and B with entries on the line, the trace of AB equals the trace of BA. *Used for:* quantum expectation values as traces, and basis independence of the trace. *Price:* **free**; exact.

1366. Invertible exactly when the determinant is invertible, 2 × 2. A 2 × 2 matrix with entries on the line has a two-sided inverse exactly when its determinant has an inverse. *Used for:* deciding when two coupled linear equations have a unique solution. *Price:* **free**; exact.

1367. Left inverse gives an invertible determinant, 2 × 2. If BA is the identity for 2 × 2 matrices A and B with entries on the line, then det(A) has an inverse. *Used for:* reversible linear transformations, such as changes of basis, have nonzero volume factor. *Price:* **free**; exact.

1368. Inverse matrix from an invertible determinant, 2 × 2. If the determinant of a 2 × 2 matrix A on the line has an inverse, some matrix B gives AB equal to the identity. *Used for:* solving two coupled linear equations, such as a two-loop circuit, by matrix inversion. *Price:* **free**; exact.

1369. Inverse matrix from a nonzero determinant, 2 × 2. If the determinant of a 2 × 2 matrix A on the line is not equal to 0, then A has a two-sided inverse. *Used for:* solving linear equations when the determinant is known only to be nonzero. *Price:* **MP**; exact.

1370. Nonzero kernel vector forces zero determinant, 2 × 2. If a 2 × 2 matrix A on the line sends some vector not equal to zero to the zero vector, then det(A) = 0. *Used for:* normal modes: a nonzero mode vector requires the secular determinant to vanish. *Price:* **free**; exact.

1371. Injective implies surjective, 2 × 2. A 2 × 2 matrix on the line that sends only the zero vector to the zero vector reaches every vector of the plane. *Used for:* for linear equations, uniqueness of solutions gives existence for every source. *Price:* **MP**; exact.

1372. Surjective implies injective, 2 × 2. A 2 × 2 matrix on the line that reaches every vector of the plane sends only the zero vector to the zero vector. *Used for:* for linear equations, existence of solutions for every source gives uniqueness. *Price:* **free**; exact.

1373. Powers are monotone on nonnegative numbers. For points $0 \le a \le b$ of the line and every whole number n, $a^n \le b^n$. *Used for:* comparing energies such as $\frac{1}{2}mv^2$ or radiated power σT^4 as speed or temperature grows. *Price:* **free**; exact.

1374. Powers of a number below one become small. If $0 \le b < 1$ and $a > 0$ on the line, then some power b^n is less than a. *Used for:* geometric decay, such as attenuation through layers, falls below any threshold. *Price:* **free**; exact.

1375. Archimedean property through reciprocals 1/(n + 1). For every positive point x of the line, some whole number n has $1/(n + 1) < x$. *Used for:* refining a grid or time step below any required resolution. *Price:* **free**; exact.

1376. Limit of a product of sequences. If x(n) converges to L and y(n) converges to M on the line, then $x(n) \cdot y(n)$ converges to $L \cdot M$. *Used for:* limits of products, such as power as force times velocity. *Price:* **free**; exact.

1377. Cauchy sequences are bounded. For every Cauchy sequence x on the line there is a positive R with $|x(m) - x(n)| < R$ for all m and n. *Used for:* keeping iterates of a converging approximation scheme within fixed bounds. *Price:* **free**; exact.

1378. Product of continuous functions is continuous. If f and g are continuous at every point of the line, then so is the product $f \cdot g$. *Used for:* continuity of products such as pressure times volume or charge density times potential. *Price:* **free**; exact.

1379. Powers of continuous functions are continuous. If f is continuous at every point of the line, then so is $f(x)^n$ for every whole number n. *Used for:* continuity of power laws such as spring energy $\frac{1}{2}kx^2$ and T^4 radiation. *Price:* **free**; exact.

1380. Polynomials are continuous. Every polynomial, given by a finite list of coefficients on the line, is continuous at every point of the line. *Used for:* polynomial potentials, such as the anharmonic oscillator, vary continuously. *Price:* **free**; exact.

1381. Integral of a sum on [0, 1]. If I and J are the integrals of f and g over [0, 1], as limits of Riemann sums, then I + J is the integral of f + g. *Used for:* adding the work of several forces, or charge from several densities. *Price:* **free**; exact.

1382. Integral of a constant on [0, 1]. For every point c of the line, the Riemann sums of the constant function c over [0, 1] converge to c. *Used for:* work

done by a constant force; distance covered at constant speed. *Price:* **free**; exact.

1383. Monotonicity of the integral on [0, 1]. If f ≤ g on [0, 1] and I and J are their integrals over [0, 1], as limits of Riemann sums, then I ≤ J. *Used for:* bounding total energy or charge by integrating a larger density. *Price:* **free**; exact.

1384. Absolute value bound, non-strict form. For points a and b of the line, |a| ≤ b exactly when −b ≤ a ≤ b. *Used for:* tolerance bands: a measurement within plus or minus b of its target. *Price:* **free**; exact.

1385. Absolute value bound, strict form. For points a and b of the line, |a| < b exactly when −b < a < b. *Used for:* strict error bounds, such as an oscillation amplitude kept below a limit. *Price:* **free**; exact.

1386. The sequence 1/n tends to zero. The sequence 1/n on the line, with its term at n = 0 read as 0, converges to 0. *Used for:* vanishing of finite-size corrections of order 1/N. *Price:* **free**; exact.

1387. Cauchy criterion through distance bounds. A sequence x on the line is Cauchy exactly when, for every k, some N gives |x(m) − x(n)| ≤ 1/(k + 1) for all m, n ≥ N. *Used for:* testing convergence of iterative numerical solutions without knowing the limit. *Price:* **free**; exact.

1388. Convergent subsequence of a frequently bounded sequence. If |x(n)| ≤ b for infinitely many n, that is, for every N some n ≥ N has |x(n)| ≤ b, then some subsequence of x converges to a point of the line. *Used for:* extracting limiting states from bounded trajectories or minimizing sequences. *Price:* **LPO**; exact: a bounded sequence is frequently bounded, so the statement gives the Bolzano–Weierstrass theorem (row 1304) and LPO with no choice; conversely, choosing for every N an index n ≥ N with |x(n)| ≤ b, a choice of natural numbers inside the free allowance, gives a bounded subsequence, and the Bolzano–Weierstrass theorem at LPO finishes.

1389. Root of a linear equation. For points a and b of the line with a not equal to 0, some point x satisfies a·x + b = 0. *Used for:* solving linear balance equations, such as a spring's equilibrium kx = F. *Price:* **MP**; exact.

1390. Determinant formula for 2 × 2 matrices. A 2 × 2 matrix on the line with first row a, b and second row c, d has determinant ad − bc. *Used for:* area of a parallelogram; solving two linear equations by Cramer's rule. *Price:* **free**; exact.

1391. Inverse of the inverse, 2 × 2. If B is a two-sided inverse of the 2 × 2 matrix A on the line, and C is a two-sided inverse of B, then C = A. *Used for:* undoing the inverse of a change of basis returns the original transformation. *Price:* **free**; exact.

1392. Cayley–Hamilton theorem, 2 × 2. Every 2 × 2 matrix A on the line satisfies $A^2 - \mathrm{tr}(A)\cdot A + \det(A)\cdot I = 0$, where I is the identity. *Used for:* closed forms for powers and exponentials of 2 × 2 matrices, such as spin rotations. *Price:* **free**; exact.

1393. Real eigenvalue of a symmetric 2 × 2 matrix. For every symmetric 2 × 2 matrix on the line with rows a, b and b, d, some point μ satisfies (a − μ)(d − μ) − b^2 = 0. *Used for:* energy levels of a two-level system with real couplings; two coupled oscillators. *Price:* **free**; exact.

1394. Product rule for derivatives. If f has derivative d_1 and g has derivative d_2 at x on the line, then f·g has derivative $d_1\cdot g(x) + f(x)\cdot d_2$ at x. *Used for:* rates of change of products, such as momentum when both mass and velocity vary. *Price:* **free**; exact.

1395. Power rule for derivatives. If f has derivative d at x on the line, then f to the power n + 1 has derivative $(n + 1)\cdot f(x)^n\cdot d$ at x. *Used for:* forces from power-law potentials, such as kx from ½kx^2. *Price:* **free**; exact.

1396. Derivative of a polynomial. Every polynomial, given by a finite list of coefficients on the line, has at every point the derivative computed term by term from its coefficients. *Used for:* forces from polynomial potentials; velocities from polynomial trajectories. *Price:* **free**; exact.

1397. Monotone sequence converges to its least upper bound. If x(n) is nondecreasing on the line and L is its least upper bound, then x(n) converges to L. *Used for:* convergence of increasing lower bounds, such as partial sums of positive terms. *Price:* **MP + $\mathbf{AC}_\omega$**; exact.

1398. Cauchy sequences of fractions converge. Every Cauchy sequence of fractions, read as points of the line, converges to a point of the line. *Used for:* real quantities as limits of finite-precision rational measurements. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1399. A supremum is a limit of members. If L is an upper bound of a set S on the line and members of S come within every tolerance of L, then some sequence of members converges to L. *Used for:* approaching a best value, such as maximal work or efficiency, by actual processes. *Price:* $\mathbf{AC}_\omega$; exact.

1400. Image of [0, 1] is the interval [min, max]. For f uniformly continuous on [0, 1], there are m and M with m ≤ f ≤ M on [0, 1], and f takes every value between m and M on [0, 1]. *Used for:* a continuous profile on a segment, such as temperature, takes every value between its extremes. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

1401. Bound on an integral over [0, 1]. If $|f| \le C$ on [0, 1] and I is the integral of f over [0, 1], as a limit of Riemann sums, then $|I| \le C$. *Used for:* estimating impulse or work from a bound on the force. *Price:* **free**; exact.

1402. Lipschitz condition from a distance bound. If $|f(x) - f(y)| \le K \cdot |x - y|$ for all points x, y of the line, with K a whole number, then f is Lipschitz with constant K. *Used for:* bounded sensitivity of force laws, which makes solutions of equations of motion unique. *Price:* **free**; exact.

1403. Approximation from below by an increasing sequence. Every point x of the line is the limit of a strictly increasing sequence of points, each less than x. *Used for:* approaching a threshold value from below, such as a critical temperature. *Price:* **free**; exact.

1404. Bernoulli's inequality for a ≥ −2. For every point $a \ge -2$ of the line and every whole number n, $1 + n \cdot a \le (1 + a)^n$. *Used for:* bounds on compound growth and on first-order approximations of $(1 + x)^n$. *Price:* **free**; exact.

1405. Adding fractions with nonzero denominators. If b and d on the line are not equal to 0, then b, d and bd have inverses and $a/b + c/d = (ad + bc)/(bd)$. *Used for:* combining rates and resistances, such as parallel resistors $1/R_1 + 1/R_2$. *Price:* **MP**; exact.

1406. Chain rule for derivatives. If f has derivative a at x and g has derivative b at f(x) on the line, then g applied after f has derivative $b \cdot a$ at x. *Used for:* rate of change of potential energy along a path: slope times velocity. *Price:* **free**; exact.

1407. The derivative as a linear map. If f has derivative d at x on the line, then the linear map $h \mapsto h \cdot d$ is the derivative of f at x in the linear-map sense. *Used for:* linearizing dynamics near a point, as in small-oscillation and stability analysis. *Price:* **free**; exact.

1408. Comparison of functions through derivatives, on [0, 1]. If f′ and B′ are uniform derivatives of f and B on [0, 1], with $f(0) \le B(0)$ and $f' \le B'$ there, then $f \le B$ on [0, 1]. *Used for:* differential inequalities, such as Grönwall-type bounds on the growth of errors. *Price:* **free**; exact.

1409. The harmonic series diverges. For every point M of the line, the partial sums $1 + 1/2 + \ldots + 1/n$ are at least M from some n on. *Used for:* logarithmic divergences, such as mode sums and the potential of a long wire. *Price:* **free**; exact.

1410. Fundamental theorem of calculus, evaluation form, on [0, 1]. If f′ is a uniform derivative of f on [0, 1], then the Riemann sums of f′ over [0, 1] converge to $f(1) - f(0)$. *Used for:* work as a change in potential energy; displacement as the integral of velocity. *Price:* **free**; exact.

1411. Integral of x over [0, 1]. The Riemann sums of the function x over [0, 1] converge to 1/2. *Used for:* work done stretching a spring: integrating kx gives $\frac{1}{2}kx^2$. *Price:* **free**; exact.

1412. Integral of x^n over [0, 1]. For every whole number n, the Riemann sums of x^n over [0, 1] converge to $1/(n + 1)$. *Used for:* moments of inertia and fields of power-law densities. *Price:* **free**; exact.

1413. Monotone sequence diverges or converges. Every nondecreasing sequence on the line either exceeds every point from some index on, or converges to a point of the line. *Used for:* an increasing quantity, such as accumulated energy, either grows without bound or settles. *Price:* **LPO + $\mathbf{AC}_\omega$**; exact.

1414. Intermediate value theorem, point inside (0, 1). If f is uniformly continuous on [0, 1] and $f(0) < v < f(1)$, then $f(c) = v$ at some point c with $0 < c < 1$. *Used for:* existence of equilibria and zero crossings, such as where a force changes sign. *Price:* **LLPO**; exact (the zero that row 1312 gives for $f - v$ lies at a positive distance from both ends, since $f(0) < v < f(1)$ and f is uniformly continuous).

1415. Ratio bound gives a convergent series. If $0 \le r < 1$ and $|a(n + 1)| \le r \cdot |a(n)|$ from some n on, then the partial sums of a converge on the line. *Used for:* convergence of perturbation and power series with geometric decay. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1416. Ratio test for series. If $|a(n + 1)| = q(n) \cdot |a(n)|$ for every n and q(n) converges to a limit below 1, then the partial sums of a converge. *Used for:* radius of convergence of series solutions, such as those for the hydrogen atom. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1417. Convergence of the p-series. For a whole number p, the sum of $1/(n + 1)^p$ over all n converges on the line exactly when $p > 1$. *Used for:* convergence of lattice sums and of power-law tails. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1418. Heine–Borel theorem on [0, 1]. Every countable cover of [0, 1] by open intervals with fraction endpoints has a finite subcover: the first N intervals already cover [0, 1]. *Used for:* finite coverings of a bounded region, behind extreme values and uniform estimates. *Price:* **FAN**; exact (equivalent to the decidable fan theorem). Proved in Theorem 13.6.

1419. Positive derivative gives a strictly increasing function, on [0, 1]. If f′ is a uniform

derivative of f on [0, 1] and $f' > 0$ there, then $f(x) < f(y)$ whenever $x < y$ in [0, 1]. *Used for:* a quantity with a strictly positive rate, such as a heated body's temperature, keeps rising. *Price:* **free**; exact.

1420. Monotonicity of the limit superior. If $x(n) \le y(n)$ from some index on, and L and M are the limits superior of x and y (lowest tail least upper bounds), then $L \le M$. *Used for:* comparing long-run upper envelopes of fluctuating signals or trajectories. *Price:* **free**; exact.

1421. Square roots preserve strict order. For points $a \ge 0$ and $b \ge 0$ of the line, if $a^2 < b^2$ then $a < b$. *Used for:* comparing magnitudes, such as speeds, from their squares, such as kinetic energies. *Price:* **free**; exact.

1422. Continuity of the square root, read on roots. For each $a_0 \ge 0$ on the line and each tolerance, every $a \ge 0$ whose square is close enough to a_0^2 lies within that tolerance of a_0. *Used for:* stable magnitudes and speeds computed from squared quantities. *Price:* **free**; exact.

1423. Integration by parts on [0, 1]. If u', v' are uniform derivatives of u, v on [0, 1] and the Riemann sums of $u'v$ converge to J, those of uv' converge to $u(1)v(1) - u(0)v(0) - J$. *Used for:* deriving equations of motion from an action; boundary terms in field theory. *Price:* **free**; exact.

1424. Substitution rule for integrals on [0, 1]. If $F' = f$ and g' are uniform derivatives on [0, 1] and g maps [0, 1] into itself, the Riemann sums of $f(g(x))\cdot g'(x)$ converge to $F(g(1)) - F(g(0))$. *Used for:* changing variables in integrals, such as from time to position in energy integrals. *Price:* **free**; exact.

1425. Derivative of the square root. Given a function s with $s(y) \ge 0$ and $s(y)^2 = y$ for every $y \ge 0$, at each $x > 0$ its derivative is the point c with $c\cdot 2s(x) = 1$. *Used for:* rates of quantities under a root, such as speed computed from kinetic energy. *Price:* **free**; exact.

1426. Rank of a 2 × 2 matrix. Every 2 × 2 matrix on the line has a rank $r \le 2$: 0 for the zero matrix, 2 if its determinant is not 0, and 1 otherwise. *Used for:* counting independent equations or degrees of freedom in a linear system. *Price:* **WLPO**; exact.

1427. Quotient rule for derivatives. If c, d have derivatives c', d' at x on the line and e is a function with $d(y)\cdot e(y) = 1$ for every y, then $c\cdot e$ has derivative $(c'\cdot d(x) - c(x)\cdot d')\cdot e(x)^2$ at x. *Used for:* rates of ratios, such as the derivative of a potential falling off as $1/r$. *Price:* **free**; exact.

1428. Positive function has a positive integral, on [0, 1]. If f is uniformly continuous on [0, 1], positive at every point strictly between 0 and 1, and I is its integral over [0, 1], then $I > 0$. *Used for:* a positive density gives a positive total mass, charge or probability. *Price:* **free**; exact.

1429. Field structure on the line. the line, with its own sum and product, is a field in the standard sense: a total inverse with $0^{-1} = 0$ that inverts every point not equal to 0. *Used for:* the algebra of physical quantities: dividing by any nonzero quantity. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorem 28.15.

1430. Decidable linear order on the line. the order of the line is a linear order in the standard sense: any two points compare, and a procedure decides $x \le y$ and $x < y$ for every pair. *Used for:* ordering measured values, such as energies, times or temperatures. *Price:* **LPO function**; exact (equivalent to LPO as a function). Proved in Theorems 28.15 and 28.3(3).

1431. Unconditional summability equals absolute summability. A sequence on the line is summable, meaning its sums over finite sets of indices converge, exactly when its absolute values are summable. *Used for:* rearranging series freely, such as lattice sums and perturbation expansions. *Price:* **free**; exact (limit points use the completeness of the line: countable choice of rational pairs).

1432. The exponential grows without bound. For every point M of the line there is a point N with $\exp(x) \ge M$ for every $x \ge N$. *Used for:* runaway exponential growth, such as chain reactions and unstable modes. *Price:* **free**; exact.

1433. Floor of x divided by x tends to 1. For each k there is M such that, for every $x \ge M$, every integer floor m of x and every y with $y\cdot x = m$ satisfy $|y - 1| \le 1/(k + 1)$. *Used for:* counting whole periods or lattice sites in a long length: count over length tends to 1. *Price:* **free**; exact.

1434. Cauchy for fractions exactly when Cauchy on the line. A sequence of fractions is Cauchy for every positive fraction tolerance exactly when, read as points of the line, it is Cauchy for every positive tolerance on the line. *Used for:* trusting finite-precision rational computations to converge to real values. *Price:* **free**; exact.

1435. Continuity of the exponential function. The exponential function on the line is continuous at every point. *Used for:* small changes in a rate or exponent give small changes in decay factors. *Price:* **free**; exact.

1436. A closed interval is connected. If two open sets of the line together cover [a, b] and each contains a point of [a, b], then they share a point of [a, b]. *Used for:* intermediate-value arguments along a path, such as crossing a phase boundary. *Price:* $\mathbf{AC}_{\omega}$; exact.

1437. Continuous function unbounded both ways takes every value. A function continuous at every point of the line that tends to $+\infty$ at $+\infty$ and to $-\infty$ at $-\infty$ takes every value. *Used for:* existence of solutions of equations of state, such as a temperature for every energy. *Price:* **LLPO + $\mathbf{AC}_\omega$**; exact.

1438. Injective continuous function is strictly monotone. A function f continuous at every point of the line, with f(x) = f(y) only when x = y, is strictly increasing or strictly decreasing. *Used for:* inverting monotone calibration curves, such as a thermometer reading to a temperature. *Price:* **MP + $\mathbf{AC}_\omega$**; exact.

1439. Preimages of open sets are open. If f is continuous on the line and a set U contains a neighborhood of each of its points, then so does the set of x with f(x) in U. *Used for:* robust conditions: a strict inequality on a measured value survives small perturbations. *Price:* **free**; exact.

1440. Heine–Cantor theorem on a compact set. If K on the line is compact (every countable cover by open fraction-endpoint intervals has a finite subcover) and f is continuous on K, then f is uniformly continuous on K. *Used for:* one grid spacing that controls errors over a whole bounded region. *Price:* **open**; at most EM + AC_ω, and at most WEM.

1441. Heine–Cantor theorem on [0, 1]. A function continuous at every point of [0, 1], with points of [0, 1] as inputs, is uniformly continuous on [0, 1]. *Used for:* one grid spacing that controls numerical errors over a whole interval. *Price:* **open**; at least FAN; at most EM + AC_ω; at most WLPO (one yes-or-no answer for every row of a table of bits, collected by the free allowance, with representatives), and at most the full fan theorem with no choice; equivalent, with no choice, to the decidable fan theorem together with a presentation of the local control data by countable covers.

1442. Compact sets are bounded. If a set of the line is compact (every countable cover by open fraction-endpoint intervals has a finite subcover), it lies inside some interval [−B, B] with B a whole number. *Used for:* confined configurations, such as bound orbits, occupy a bounded region. *Price:* **free**; exact.

1443. Compact sets are closed. If a set of the line is compact (every countable cover by open fraction-endpoint intervals has a finite subcover), it contains every one of its limit points. *Used for:* limits of allowed configurations remain allowed configurations. *Price:* **EM**; exact.

1444. Sequentially compact implies compact. If every sequence in a set K of the line has a subsequence converging in K, then every countable cover of K by open fraction-endpoint intervals has a finite subcover. *Used for:* passing from limiting arguments to finite coverings in existence proofs. *Price:* **open**; at least LPO; at most EM + AC_ω.

1445. Compact exactly when sequentially compact. A set K of the line is compact (every countable cover by open fraction-endpoint intervals has a finite subcover) exactly when every sequence in K has a subsequence converging in K. *Used for:* existence of minimizers and limiting states in variational problems. *Price:* **EM + $\mathbf{AC}_\omega$**; exact.

1446. A compact set has a least element. A nonempty set of the line that is compact (every countable cover by open fraction-endpoint intervals has a finite subcover) has a least member. *Used for:* existence of a lowest energy or shortest time among allowed values. *Price:* **EM + $\mathbf{AC}_\omega$**; exact.

1447. Continuous image of a compact set is compact. If K is compact (every countable cover by open fraction-endpoint intervals has a finite subcover) and f is continuous on K, then the set of values of f on K is compact. *Used for:* a continuous observable on a closed bounded set of states has bounded, attained values. *Price:* **open**; at most EM + AC_ω, and at most WEM; equivalent to the margin property: for every countable cover of the values of f on K by open intervals with fraction endpoints, there are N and m such that every value lies in one of the first N intervals at distance at least 1/(m+1) from its ends.

1448. Compact exactly when totally bounded and complete. A set K of the line is compact (every countable cover by open fraction-endpoint intervals has a finite subcover) exactly when K has finite nets at every tolerance and is complete. *Used for:* approximating a bounded set of states by finitely many sample states. *Price:* **EM + $\mathbf{AC}_\omega$**; exact.

1449. Total boundedness by finite nets. A set of the line has a finite net for every relation containing a band $|x - y| \le 1/(k + 1)$ exactly when it has a finite net within each $1/(k + 1)$. *Used for:* discretizing a bounded region by a finite grid at any resolution. *Price:* **free**; exact.

1450. Closed subset of a compact set is compact. If C contains its limit points and lies inside a compact set K of the line (every countable cover by open fraction-endpoint intervals has a finite subcover), then C is compact too. *Used for:* restricting to a closed constraint set inside a bounded region. *Price:* **open**; at least LPO; at most EM + AC_ω.

1451. Compact space [0, 1] is a compact set. The interval [0, 1] is compact: every countable cover of [0, 1] by open intervals with fraction endpoints has a finite subcover, its first N intervals. *Used for:* finite coverings of a bounded region, behind extreme values and uniform estimates. *Price:* **FAN**; exact (equivalent

to the decidable fan theorem).

1452. The exponential is its own derivative. The exponential function on the line has derivative exp(x) at every point x. *Used for:* exponential decay and growth, where the rate is proportional to the amount. *Price:* **free**; exact.

1453. Fundamental theorem of calculus, derivative form. If F(x) is the integral of f from a to x for every x on the line, and f is continuous at b, then F has derivative f(b) at b. *Used for:* recovering a rate from an accumulated quantity, such as power from energy delivered. *Price:* **free**; exact.

1454. Integrals over adjacent intervals add. Where they exist, integrals of f on the line add: from a to b plus from b to c gives from a to c, for a, b and c in any order. *Used for:* splitting work or travel time over the segments of a path. *Price:* **free**; exact.

1455. Laplace expansion along the first row. For every square matrix on the line, the determinant, defined by cofactor expansion down the first column, equals the cofactor expansion along the first row. *Used for:* computing determinants of secular equations and Jacobians by cofactors. *Price:* **free**; exact.

1456. Rank–nullity theorem for matrices. Every matrix on the line has columns forming a basis of its column space, and a basis of its kernel indexed by the remaining columns: rank plus nullity is the column count. *Used for:* counting independent solutions of linear constraints and degrees of freedom. *Price:* **WLPO**; exact (the proof also collects yes-or-no answers by countable unique choice, which the free allowance contains).

1457. Rank of the transpose equals the rank. Every matrix on the line has a rank r, the size of a largest square minor with determinant not equal to 0, and its transpose also has rank r. *Used for:* row and column counts of independent equations agree in linear systems. *Price:* **WLPO**; exact.

1458. Injective linear map cannot lower dimension. If an $n \times m$ matrix on the line sends only the zero vector to the zero vector, then $m \le n$. *Used for:* a linear system with unique solutions needs at least as many equations as unknowns. *Price:* **free**; exact.

1459. A spanning family contains a basis. Every finite family of vectors in n-space on the line has a linearly independent subfamily that spans every vector of the family. *Used for:* removing redundant constraints or generators, such as dependent conservation laws. *Price:* **WLPO**; exact (the proof also collects yes-or-no answers by countable unique choice, which the free allowance contains).

1460. Basis of a finitely spanned space. The space spanned by finitely many vectors in n-space on the line has a basis drawn from those vectors. *Used for:* choosing independent coordinates or modes from an overcomplete set. *Price:* **WLPO**; exact (the proof also collects yes-or-no answers by countable unique choice, which the free allowance contains).

1461. Spectral theorem for symmetric matrices. For every n, every symmetric $n \times n$ matrix on the line has an orthonormal basis of eigenvectors, each with its eigenvalue. *Used for:* normal modes of coupled oscillators; diagonalizing real Hamiltonians and inertia tensors. *Price:* **LLPO**; exact (the converse uses weak König's lemma, whose choice collects a finite table for every depth, and rational approximations, both inside the free allowance). Proved in Theorem 28.66.

1462. Characteristic polynomial is monic. For every $n \times n$ matrix M on the line, the function $t \mapsto \det(t \cdot I - M)$ equals t^n plus a polynomial of degree below n. *Used for:* secular equations for energy levels and normal-mode frequencies. *Price:* **free**; exact.

1463. Determinant is the product of the eigenvalues. If the characteristic polynomial of an $n \times n$ matrix M on the line is given split as the product of $t - \lambda_i$, then det(M) is the product of the λ_i. *Used for:* volume change under a linear flow; products over modes in partition functions. *Price:* **free**; exact.

1464. Independent vectors number at most the dimension. A linearly independent family of k vectors in n-space on the line has $k \le n$. *Used for:* counting degrees of freedom, such as at most three independent directions in space. *Price:* **free**; exact.

1465. Existence of the exponential function. There is a function E on the line with $E(0) = 1$ whose derivative at every point x is E(x). *Used for:* radioactive decay, growth and damping, where the rate is proportional to the amount. *Price:* **free**; exact. Proved in Proposition 12.88.

1466. Exponential turns sums into products. For all points x and y of the line, $\exp(x + y) = \exp(x) \cdot \exp(y)$. *Used for:* evolution over successive time intervals; Boltzmann factors of independent systems multiply. *Price:* **free**; exact. Proved in Proposition 12.88.

1467. Integrals of uniformly continuous functions exist. If f is uniformly continuous on all of the line, then for all points a and b the integral of f from a to b exists as a point. *Used for:* total work, charge or probability from continuous densities. *Price:* **free**; exact (a modulus of uniform continuity is chosen by countable choice of natural numbers; the Riemann sums then form a Cauchy sequence with a known modulus, and its limit point uses the completeness of the line: countable choice of rational pairs, Corollary 12.35).

Bibliography

[1] Jonathan Washburn and Milan Lj. Zlatanović. The δ-calculus: From distinction to arithmetic. *Advances in Pure Mathematics* **16**(9) (2026), 705–739. doi:10.4236/apm.2026.169035.

[2] Richard Dedekind. *Was sind und was sollen die Zahlen?* Friedrich Vieweg und Sohn, Braunschweig, 1888.

[3] Thoralf Skolem. Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Mathematicae* **23** (1934), 150–161. doi:10.4064/fm-23-1-150-161.

[4] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction*, Volume I. Studies in Logic and the Foundations of Mathematics, Volume 121. North-Holland, Amsterdam, 1988.

[5] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction*, Volume II. Studies in Logic and the Foundations of Mathematics, Volume 123. North-Holland, Amsterdam, 1988.

[6] A. S. Troelstra (ed.). *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics, Volume 344. Springer, Berlin and Heidelberg, 1973. doi:10.1007/BFb0066739.

[7] Georg Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting (ed.), *Constructivity in Mathematics*, Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1959, 101–128.

[8] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* **38**(1) (1931), 173–198. doi:10.1007/BF01700692.

[9] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*, Volume II. Grundlehren der mathematischen Wissenschaften, Volume 50. Julius Springer, Berlin, 1939.

[10] M. H. Löb. Solution of a problem of Leon Henkin. *Journal of Symbolic Logic* **20**(2) (1955), 115–118. doi:10.2307/2266895.

[11] Leon Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic* **14**(3) (1949), 159–166. doi:10.2307/2267044.

[12] Wilhelm Ackermann. Die Widerspruchsfreiheit der allgemeinen Mengenlehre. *Mathematische Annalen* **114**(1) (1937), 305–315. doi:10.1007/BF01594179.

[13] Richard Kaye and Tin Lok Wong. On interpretations of arithmetic and set theory. *Notre Dame Journal of Formal Logic* **48**(4) (2007), 497–510. doi:10.1305/ndjfl/1193667707.

[14] Errett Bishop and Douglas Bridges. *Constructive Analysis*. Grundlehren der mathematischen Wissenschaften, Volume 279. Springer, Berlin and Heidelberg, 1985. doi:10.1007/978-3-642-61667-9.

[15] Dénes König. Über eine Schlußweise aus dem Endlichen ins Unendliche. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum* (Szeged) **3** (1927), 121–130.

[16] F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, Series 2, **30**(1) (1930), 264–286. doi:10.1112/plms/s2-30.1.264.

[17] P. Erdős and R. Rado. A partition calculus in set theory. *Bulletin of the American Mathematical Society* **62**(5) (1956), 427–489. doi:10.1090/S0002-9904-1956-10036-0.

[18] Russell C. Walker. *The Stone–Čech Compactification*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Volume 83. Springer, Berlin and Heidelberg, 1974. doi:10.1007/978-3-642-61935-9.

[19] Stephen Cole Kleene. *Introduction to Metamathematics*. Bibliotheca Mathematica, Volume 1. North-Holland, Amsterdam; P. Noordhoff, Groningen; D. Van Nostrand, New York and Toronto, 1952.

[20] Per Martin-Löf. *Intuitionistic Type Theory*. Notes by Giovanni Sambin of lectures given in Padua, June 1980. Bibliopolis, Naples, 1984.

[21] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, Series 2, **42**(1) (1937), 230–265. doi:10.1112/plms/s2-42.1.230.

[22] W. Forrest Stinespring. Positive functions on C^*-algebras. *Proceedings of the American Mathematical Society* **6**(2) (1955), 211–216. doi:10.1090/S0002-9939-1955-0069403-4.

[23] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications* **10**(3) (1975), 285–290. doi:10.1016/0024-3795(75)90075-0.

[24] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, Cambridge, 2018. doi:10.1017/9781316848142.

[25] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society* **10**(4) (1959), 545–551. doi:10.1090/S0002-9939-1959-0108732-6.

[26] E. P. Wigner. Die Messung quantenmechanischer Operatoren. *Zeitschrift für Physik* **133** (1952), 101–108. doi:10.1007/BF01948686. English translation by Paul Busch (2010), arXiv:1012.4372.

[27] Huzihiro Araki and Mutsuo M. Yanase. Measurement of quantum mechanical operators. *Physical Review* **120**(2) (1960), 622–626. doi:10.1103/PhysRev.120.622.

[28] Masanao Ozawa. Conservation laws, uncertainty relations, and quantum limits of measurements. *Physical Review Letters* **88**(5) (2002), 050402. doi:10.1103/PhysRevLett.88.050402.

[29] L. E. J. Brouwer. Über Definitionsbereiche von Funktionen. *Mathematische Annalen* **97**(1) (1927), 60–75. doi:10.1007/BF01447860.

[30] Josef Berger. The fan theorem and uniform continuity. In S. B. Cooper, B. Löwe and L. Torenvliet (eds.), *New Computational Paradigms. CiE 2005*, Lecture Notes in Computer Science, Volume 3526, Springer, Berlin and Heidelberg, 2005, 18–22. doi:10.1007/11494645_3.

[31] Josef Berger. Brouwer's fan theorem. Preprint, version 2, 2021. arXiv:2001.00064.

[32] Hajime Ishihara. Constructive reverse mathematics: compactness properties. In L. Crosilla and P. Schuster (eds.), *From Sets and Types to Topology and Analysis: Towards Practicable Foundations for Constructive Mathematics*, Oxford Logic Guides, Volume 48, Oxford University Press, 2005, 245–267. doi:10.1093/acprof:oso/9780198566519.003.0016.

[33] Hajime Ishihara. Weak König's lemma implies Brouwer's fan theorem: a direct proof. *Notre Dame Journal of Formal Logic* **47**(2) (2006), 249–252. doi:10.1305/ndjfl/1153858649.

[34] Robert S. Lubarsky and Hannes Diener. Separating the fan theorem and its weakenings. *Journal of Symbolic Logic* **79**(3) (2014), 792–813. doi:10.1017/jsl.2014.9.

[35] Tatsuji Kawai. A continuity principle equivalent to the monotone Π^0_1 fan theorem. *Archive for Mathematical Logic* **58**(3–4) (2019), 443–456. doi:10.1007/s00153-018-0644-1. Preprint arXiv:1808.08076.

[36] Hannes Diener and Iris Loeb. Sequences of real functions on $[0,1]$ in constructive reverse mathematics. *Annals of Pure and Applied Logic* **157**(1) (2009), 50–61. doi:10.1016/j.apal.2008.09.018.

[37] William H. Julian and Fred Richman. A uniformly continuous function on $[0,1]$ that is everywhere different from its infimum. *Pacific Journal of Mathematics* **111**(2) (1984), 333–340. doi:10.2140/pjm.1984.111.333.

[38] Errett Bishop. *Foundations of Constructive Analysis*. McGraw-Hill, New York, 1967.

[39] Ray Mines, Fred Richman and Wim Ruitenburg. *A Course in Constructive Algebra*. Universitext. Springer, New York, 1988. doi:10.1007/978-1-4419-8640-5.

[40] Douglas Bridges and Fred Richman. *Varieties of Constructive Mathematics*. London Mathematical Society Lecture Note Series, Volume 97. Cambridge University Press, Cambridge, 1987. doi:10.1017/CBO9780511565663.

[41] Douglas S. Bridges, Fred Richman and Peter Schuster. Linear independence without choice. *Annals of Pure and Applied Logic* **101**(1) (1999), 95–102. doi:10.1016/S0168-0072(99)00030-5.

[42] Mark Mandelkern. Constructively complete finite sets. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **34**(2) (1988), 97–103. doi:10.1002/malq.19880340202.

[43] Hajime Ishihara. Reverse mathematics in Bishop's constructive mathematics. *Philosophia Scientiæ*, Cahier spécial **6** (2006), 43–59. doi:10.4000/philosophiascientiae.406.

[44] Matthew Hendtlass. Equivalents of disjunctive Markov's principle. Preprint, 2016. arXiv:1611.02543.

[45] Amaury Pouly and Daniel S. Graça. Computational complexity of solving polynomial differential equations over unbounded domains. *Theoretical Computer Science* **626** (2016), 67–82. doi:10.1016/j.tcs.2016.02.002.

[46] Olivier Bournez, Daniel S. Graça and Amaury Pouly. Polynomial time corresponds to solutions of polynomial ordinary differential equations of polynomial length. *Journal of the ACM* **64**(6) (2017), Article 38. doi:10.1145/3127496.

[47] Ramon E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

[48] Wilhelm Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Zeitschrift für Mathematik und Physik* **46** (1901), 435–453.

[49] E. H. Kerner. Universal formats for nonlinear ordinary differential systems. *Journal of Mathematical Physics* **22**(7) (1981), 1366–1371. doi:10.1063/1.525074.

[50] S. C. Kleene. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic* **10**(4) (1945), 109–124. doi:10.2307/2269016.

[51] S. C. Kleene. Recursive functions and intuitionistic mathematics. In *Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts, 1950*, Volume I. American Mathematical Society, Providence, 1952, 679–685.

[52] Leon Henkin. Completeness in the theory of types. *Journal of Symbolic Logic* **15**(2) (1950), 81–91. doi:10.2307/2266967.

[53] Saul A. Kripke. Semantical analysis of intuitionistic logic I. In J. N. Crossley and M. A. E. Dummett (eds.), *Formal Systems and Recursive Functions*, Studies in Logic and the Foundations of Mathematics, Volume 40. North-Holland, Amsterdam, 1965, 92–130. doi:10.1016/S0049-237X(08)71685-9.

[54] Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae* **3** (1922), 133–181. doi:10.4064/fm-3-1-133-181.

[55] G. Kreisel. On the interpretation of non-finitist proofs, Part I. *Journal of Symbolic Logic* **16**(4) (1951), 241–267. doi:10.2307/2267908.

[56] Ulrich Kohlenbach. *Applied Proof Theory: Proof Interpretations and Their Use in Mathematics*. Springer Monographs in Mathematics. Springer, Berlin and Heidelberg, 2008. doi:10.1007/978-3-540-77533-1.

[57] Vladimir Lifschitz. CT_0 is stronger than CT_0! *Proceedings of the American Mathematical Society* **73**(1) (1979), 101–106. doi:10.1090/S0002-9939-1979-0512067-X.

[58] Jaap van Oosten. Lifschitz' realizability. *Journal of Symbolic Logic* **55**(2) (1990), 805–821. doi:10.2307/2274666.

[59] M. P. Fourman and J. M. E. Hyland. Sheaf models for analysis. In M. P. Fourman, C. J. Mulvey and D. S. Scott (eds.), *Applications of Sheaves*, Lecture Notes in Mathematics, Volume 753. Springer, Berlin and Heidelberg, 1979, 280–301. doi:10.1007/BFb0061823.

[60] Hajime Ishihara. Continuity properties in constructive mathematics. *Journal of Symbolic Logic* **57**(2) (1992), 557–565. doi:10.2307/2275292.

[61] E. Mark Gold. Limiting recursion. *Journal of Symbolic Logic* **30**(1) (1965), 28–48. doi:10.2307/2270580.

[62] Hilary Putnam. Trial and error predicates and the solution to a problem of Mostowski. *Journal of Symbolic Logic* **30**(1) (1965), 49–57. doi:10.2307/2270581.

[63] E. Mark Gold. Language identification in the limit. *Information and Control* **10**(5) (1967), 447–474. doi:10.1016/S0019-9958(67)91165-5.

[64] Kevin T. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, New York, 1996. doi:10.1093/oso/9780195091953.001.0001.

[65] Kurt Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums* **4** (1933), 34–38.

[66] Gerhard Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* **112**(1) (1936), 493–565. doi:10.1007/BF01565428.

[67] Gerhard Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik. *Archiv für mathematische Logik und Grundlagenforschung* **16**(3–4) (1974), 119–132. doi:10.1007/BF02015371.

[68] Harvey Friedman. Classically and intuitionistically provably recursive functions. In G. H. Müller and D. S. Scott (eds.), *Higher Set Theory*, Lecture Notes in Mathematics, Volume 669, Springer, Berlin and Heidelberg, 1978, 21–27. doi:10.1007/BFb0103100.

[69] G. Kreisel. On the interpretation of non-finitist proofs, Part II. Interpretation of number theory. Applications. *Journal of Symbolic Logic* **17**(1) (1952), 43–58. doi:10.2307/2267457.

[70] Kurt Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica* **12**(3–4) (1958), 280–287. doi:10.1111/j.1746-8361.1958.tb01464.x.

[71] A. M. Turing. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society*, Series 2, **45**(1) (1939), 161–228. doi:10.1112/plms/s2-45.1.161.

[72] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **14** (1968), 97–142. doi:10.1002/malq.19680140702.

[73] Hannes Diener. *Constructive Reverse Mathematics*. Habilitationsschrift, Universität Siegen, 2018; version 1.3, 2020. arXiv:1804.05495.

[74] Matt Hendtlass and Robert Lubarsky. Separating fragments of WLEM, LPO, and MP. *The Journal of Symbolic Logic* **81**(4) (2016), 1315–1343. doi:10.1017/jsl.2016.38.

[75] Ulrich Kohlenbach. On weak Markov's principle. *Mathematical Logic Quarterly* **48**, Suppl. 1 (2002), 59–65. doi:10.1002/1521-3870(200210)48:1+<59::AID-MALQ59>3.0.CO;2-I.

[76] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Second edition. Perspectives in Logic. Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511581007.

[77] I. D. Zaslavskiĭ and G. S. Ceĭtin. On singular coverings and properties of constructive functions connected with them. *Trudy Matematicheskogo Instituta imeni V. A. Steklova* **67** (1962), 458–502. In Russian.

[78] William W. Adams. The best two-dimensional Diophantine approximation constant for cubic irrationals. *Pacific Journal of Mathematics* **91** (1980), 29–30. doi:10.2140/pjm.1980.91.29.

[79] T. W. Cusick. The two-dimensional Diophantine approximation constant. II. *Pacific Journal of Mathematics* **105** (1983), 53–67. doi:10.2140/pjm.1983.105.53.

[80] J.-M. Dumont and A. Thomas. Systèmes de numération et fonctions fractales relatifs aux substitutions. *Theoretical Computer Science* **65** (1989), 153–169. doi:10.1016/0304-3975(89)90041-8.

[81] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*, sixth edition, revised by D. R. Heath-Brown and J. H. Silverman. Oxford University Press, Oxford, 2008.

[82] A. Hurwitz. Ueber die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche. *Mathematische Annalen* **39** (1891), 279–284. doi:10.1007/BF01206656.

[83] A. Ya. Khinchin. *Continued Fractions*. University of Chicago Press, Chicago, 1964.

[84] Wolfgang M. Schmidt. *Diophantine Approximation*. Lecture Notes in Mathematics, Volume 785. Springer, Berlin, 1980. doi:10.1007/BFb0087376.

[85] M. Takahashi. Parallel reductions in λ-calculus. *Information and Computation* **118** (1995), 120–127. doi:10.1006/inco.1995.1057.

[86] Bryce S. DeWitt. Quantum theory of gravity. I. The canonical theory. *Physical Review* **160** (1967), 1113–1148. doi:10.1103/PhysRev.160.1113.

[87] M. Fierz and W. Pauli. On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. *Proceedings of the Royal Society of London A* **173** (1939), 211–232. doi:10.1098/rspa.1939.0140.

[88] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, second edition. Graduate Texts in Mathematics 222, Springer, Cham, 2015.

[89] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*, second edition. Cambridge University Press, Cambridge, 2013.

[90] H. W. Kuhn. Some combinatorial lemmas in topology. *IBM Journal of Research and Development* **4** (1960), 518–524. doi:10.1147/rd.45.0518.

[91] Gerhart Lüders. Über die Zustandsänderung durch den Meßprozeß. *Annalen der Physik* (6) **8** (1951), 322–328. doi:10.1002/andp.19504430510.

[92] T. Regge. General relativity without coordinates. *Il Nuovo Cimento* **19** (1961), 558–571. doi:10.1007/BF02733251.

[93] L. Schläfli. On the multiple integral $\int^n dx\,dy\ldots dz$, whose limits are $p_1 = a_1x + b_1y + \cdots + h_1z > 0$, $p_2 > 0$, ..., $p_n > 0$, and $x^2 + y^2 + \cdots + z^2 < 1$. *Quarterly Journal of Pure and Applied Mathematics* **2** (1858), 269–301; **3** (1860), 54–68 and 97–108.

[94] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics* **21** (1950), 124–127. doi:10.1214/aoms/1177729893.

[95] Jonathan Washburn. The long-wave Regge graviton on the $\kappa = 1$ Kuhn lattice and the recognition cost's quadratic form. Preprint, version 1.1, Zenodo, 2026. doi:10.5281/zenodo.24027576.

[96] Jonathan Washburn. Regge's slope and the long-wave Einstein–Hilbert vertices from flat-space stationarity on the recognition lattice. Preprint, Zenodo, 2026. doi:10.5281/zenodo.23965762.

[97] J. Aczél. *Lectures on Functional Equations and Their Applications*. Mathematics in Science and Engineering, Volume 19. Academic Press, New York, 1966.

[98] J. Aczél and J. Dhombres. *Functional Equations in Several Variables*. Encyclopedia of Mathematics and its Applications, Volume 31. Cambridge University Press, Cambridge, 1989. doi:10.1017/CBO9781139086578.

[99] Ronald L. Graham, Donald E. Knuth and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*, second edition. Addison-Wesley, Reading, MA, 1994.

[100] M. Lothaire. *Algebraic Combinatorics on Words*. Encyclopedia of Mathematics and its Applications, Volume 90. Cambridge University Press, Cambridge, 2002. doi:10.1017/CBO9781107326019.

[101] H. Inose, Y. Yasuda and J. Murakami. A telemetering system by code modulation: Δ-Σ modulation. *IRE Transactions on Space Electronics and Telemetry* **SET-8** (1962), 204–209. doi:10.1109/iret-set.1962.5008839.

[102] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Cambridge, 2006.

[103] L. Mandelstam and Ig. Tamm. The uncertainty relation between energy and time in non-relativistic quantum mechanics. *Journal of Physics (USSR)* **9** (1945), 249–254.

[104] Norman Margolus and Lev B. Levitin. The maximum speed of dynamical evolution. *Physica D* **120** (1998), 188–195. doi:10.1016/S0167-2789(98)00054-2.

[105] Ulrich Berger, Wilfried Buchholz and Helmut Schwichtenberg. Refined program extraction from classical proofs. *Annals of Pure and Applied Logic* **114** (2002), 3–25. doi:10.1016/S0168-0072(01)00073-2.

[106] Sabine Koppelberg. *General Theory of Boolean Algebras*. Volume 1 of J. D. Monk and R. Bonnet (eds.), *Handbook of Boolean Algebras*. North-Holland, Amsterdam, 1989.

[107] Jean-Louis Krivine. Anneaux préordonnés. *Journal d'Analyse Mathématique* **12** (1964), 307–326. doi:10.1007/BF02807438.

[108] I. J. Schoenberg. On the Peano curve of Lebesgue. *Bulletin of the American Mathematical Society* **44** (1938), 519. doi:10.1090/S0002-9904-1938-06792-4.

[109] Gilbert Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen* **207** (1974), 87–97. doi:10.1007/BF01362149.

[110] M. H. Stone. The theory of representations for Boolean algebras. *Transactions of the American Mathematical Society* **40** (1936), 37–111. doi:10.1090/S0002-9947-1936-1501865-8.

[111] Eric K. van Douwen. The integers and topology. In K. Kunen and J. E. Vaughan (eds.), *Handbook of Set-Theoretic Topology*, North-Holland, Amsterdam, 1984, 111–167. doi:10.1016/B978-0-444-86580-9.50006-9.

[112] Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control* **45**(2) (1980), 117–135. doi:10.1016/S0019-9958(80)90285-5.

[113] Cyril Banderier and Sylviane Schwer. Why Delannoy numbers? *Journal of Statistical Planning and Inference* **135**(1) (2005), 40–54. doi:10.1016/j.jspi.2005.02.004.

[114] Jānis Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs*, Volume 1, Latvian State University, Riga, 1974, 82–88 (in Russian).

[115] Lenore Blum and Manuel Blum. Toward a mathematical theory of inductive inference. *Information and Control* **28**(2) (1975), 125–155. doi:10.1016/S0019-9958(75)90261-2.

[116] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler and Manfred K. Warmuth. Occam's razor. *Information Processing Letters* **24**(6) (1987), 377–380. doi:10.1016/0020-0190(87)90114-1.

[117] L. E. J. Brouwer. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen* **71** (1911), 97–115. doi:10.1007/BF01456931.

[118] John Case and Carl Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science* **25**(2) (1983), 193–220. doi:10.1016/0304-3975(83)90061-0.

[119] N. N. Čencov. *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs, Volume 53. American Mathematical Society, Providence, RI, 1982.

[120] Gregory J. Chaitin. Information-theoretic limitations of formal systems. *Journal of the ACM* **21**(3) (1974), 403–424. doi:10.1145/321832.321839.

[121] Ethan M. Coven and Gustav A. Hedlund. Sequences with minimal block growth. *Mathematical Systems Theory* **7**(2) (1973), 138–153. doi:10.1007/BF01762232.

[122] Bruno de Finetti. Funzione caratteristica di un fenomeno aleatorio. *Atti della Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali*, Serie 6, **4** (1931), 251–299.

[123] Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* **7**(1) (1937), 1–68.

[124] Persi Diaconis and David Freedman. Finite exchangeable sequences. *Annals of Probability* **8**(4) (1980), 745–764. doi:10.1214/aop/1176994663.

[125] Michael Dummett. A propositional calculus with denumerable matrix. *Journal of Symbolic Logic* **24**(2) (1959), 97–106. doi:10.2307/2964753.

[126] William Feller. *An Introduction to Probability Theory and Its Applications*, Volume II, second edition. Wiley, New York, 1971.

[127] Rūsiņš Freivalds and Carl H. Smith. On the role of procrastination in machine learning. *Information and Computation* **107**(2) (1993), 237–271. doi:10.1006/inco.1993.1068.

[128] Kurt Gödel. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien* **69** (1932), 65–66.

[129] Elad Hazan, Amit Agarwal and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning* **69**(2–3) (2007), 169–192. doi:10.1007/s10994-007-5016-8.

[130] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission* **1**(1) (1965), 1–7.

[131] Nick Littlestone. Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. *Machine Learning* **2**(4) (1988), 285–318. doi:10.1023/A:1022869011914.

[132] Marston Morse and Gustav A. Hedlund. Symbolic dynamics. *American Journal of Mathematics* **60**(4) (1938), 815–866. doi:10.2307/2371264.

[133] Marston Morse and Gustav A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics* **62**(1) (1940), 1–42. doi:10.2307/2371431.

[134] Jerzy Neyman. Contribution to the theory of the χ^2 test. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1949, 239–273.

[135] Alexander Ostrowski. Über einige Lösungen der Funktionalgleichung $\varphi(x)\cdot\varphi(y)=\varphi(xy)$. *Acta Mathematica* **41** (1916), 271–284. doi:10.1007/BF02422947.

[136] Emanuel Sperner. Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* **6** (1928), 265–272. doi:10.1007/BF02940617.

[137] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM* **27**(11) (1984), 1134–1142. doi:10.1145/1968.1972.

[138] Hermann Bondi. *Relativity and Common Sense: A New Approach to Einstein*. Doubleday, Garden City, New York, 1964.

[139] Max Born. Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik* **37** (1926), 863–867. doi:10.1007/BF01397477.

[140] William Feller. *An Introduction to Probability Theory and Its Applications*, Volume I, third edition. Wiley, New York, 1968.

[141] Andrew M. Gleason. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics* **6** (1957), 885–893. doi:10.1512/iumj.1957.6.56050.

[142] Lucien Hardy. Quantum theory from five reasonable axioms. Preprint, 2001. arXiv:quant-ph/0101012.

[143] P. Jordan and J. von Neumann. On inner products in linear, metric spaces. *Annals of Mathematics* **36** (1935), 719–723. doi:10.2307/1968653.

[144] Emmy Noether. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918), 235–257.

[145] Roger Penrose and Wolfgang Rindler. *Spinors and Space-Time*, Volume 1: Two-Spinor Calculus and Relativistic Fields. Cambridge University Press, Cambridge, 1984.

[146] Dale Rolfsen. *Knots and Links*. Mathematics Lecture Series, Volume 7. Publish or Perish, Berkeley, 1976.

[147] Edwin F. Taylor and John Archibald Wheeler. *Spacetime Physics: Introduction to Special Relativity*, second edition. W. H. Freeman, New York, 1992.

[148] Lawrence C. Washington. *Introduction to Cyclotomic Fields*, second edition. Graduate Texts in Mathematics, Volume 83. Springer, New York, 1997.

[149] William K. Wootters. Local accessibility of quantum states. In W. H. Zurek (ed.), *Complexity, Entropy and the Physics of Information*. Addison-Wesley, Redwood City, 1990.

[150] E. C. Zeeman. Causality implies the Lorentz group. *Journal of Mathematical Physics* **5** (1964), 490–493. doi:10.1063/1.1704140.

[151] Mark S. Ashbaugh, Carmen C. Chicone and Richard H. Cushman. The twisting tennis racket. *Journal of Dynamics and Differential Equations* **3** (1991), 67–85. doi:10.1007/BF01049489.

[152] Joseph Bertrand. Théorème relatif au mouvement d'un point attiré vers un centre fixe. *Comptes Rendus de l'Académie des Sciences de Paris* **77** (1873), 849–853.

[153] N. G. Chetaev. *The Stability of Motion*. Pergamon Press, Oxford, 1961.

[154] Thibault Damour and Gerhard Schäfer. Higher-order relativistic periastron advances and binary pulsars. *Il Nuovo Cimento B* **101** (1988), 127–176. doi:10.1007/BF02828697.

[155] Albert Einstein. Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin* (1915), 831–839.

[156] Leonhard Euler. *Theoria Motus Corporum Solidorum seu Rigidorum*. Rostock and Greifswald, 1765.

[157] D. M. Fradkin. Three-dimensional isotropic harmonic oscillator and SU_3. *American Journal of Physics* **33** (1965), 207–211. doi:10.1119/1.1971373.

[158] Herbert Goldstein. Prehistory of the "Runge–Lenz" vector. *American Journal of Physics* **43** (1975), 737–738. doi:10.1119/1.9745.

[159] T. H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics* (2) **20** (1919), 292–296. doi:10.2307/1967124.

[160] Johannes Kepler. *Astronomia Nova*. Heidelberg, 1609.

[161] Johannes Kepler. *Harmonices Mundi Libri V*. Linz, 1619.

[162] Joseph-Louis Lagrange. Essai sur le problème des trois corps (1772). In *Œuvres de Lagrange*, Volume 6. Gauthier-Villars, Paris, 1873.

[163] L. D. Landau and E. M. Lifshitz. *Mechanics*, third edition. Course of Theoretical Physics, Volume 1. Pergamon Press, Oxford, 1976.

[164] Pierre-Simon Laplace. *Traité de Mécanique Céleste*, Volume 1. Paris, 1799.

[165] W. Lenz. Über den Bewegungsverlauf und die Quantenzustände der gestörten Keplerbewegung. *Zeitschrift für Physik* **24** (1924), 197–207. doi:10.1007/BF01327245.

[166] Isaac Newton. *Philosophiæ Naturalis Principia Mathematica*. London, 1687.

[167] Carl Runge. *Vektoranalysis*, Volume 1. S. Hirzel, Leipzig, 1919.

[168] Karl Schwarzschild. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin* (1916), 189–196.

[169] Clifford M. Will. *Theory and Experiment in Gravitational Physics*, second edition. Cambridge University Press, Cambridge, 2018.

[170] Aurel Wintner. *The Analytical Foundations of Celestial Mechanics*. Princeton University Press, Princeton, 1941.

[171] Augustin-Louis Cauchy. *Cours d'analyse de l'École royale polytechnique. Première partie: Analyse algébrique*. Debure frères, Paris, 1821.

[172] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* **2** (1967), 299–318.

[173] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Series 5, **50** (1900), 157–175. doi:10.1080/14786440009463897.

[174] George Pólya. Über die Analogie der Kristallsymmetrie in der Ebene. *Zeitschrift für Kristallographie* **60** (1924), 278–282. doi:10.1524/zkri.1924.60.1.278.

[175] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly* **93**(5) (1986), 333–348. doi:10.1080/00029890.1986.11971821.

[176] Ludwig Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Classe* **66** (1872), 275–370.

[177] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Second edition. Wiley-Interscience, Hoboken, 2006. doi:10.1002/047174882X.

[178] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences of the United States of America* **93**(4) (1996), 1659–1664. doi:10.1073/pnas.93.4.1659.

[179] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability* **1**(1) (1991), 36–61. doi:10.1214/aoap/1177005980.

[180] Wolfgang Doeblin. Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples. *Bulletin mathématique de la Société roumaine des sciences* **39** (1937), no. 1, 57–115, and no. 2, 3–61.

[181] J. Willard Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, New York, 1902.

[182] Roy J. Glauber. Time-dependent statistics of the Ising model. *Journal of Mathematical Physics* **4**(2) (1963), 294–307. doi:10.1063/1.1703954.

[183] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing* **18**(6) (1989), 1149–1178. doi:10.1137/0218077.

[184] David A. Levin, Yuval Peres and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, 2009. doi:10.1090/mbk/058.

[185] Elliott H. Lieb and Derek W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics* **28** (1972), 251–257. doi:10.1007/BF01645779.

[186] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**(6) (1953), 1087–1092. doi:10.1063/1.1699114.

[187] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics* **105**(1) (1986), 123–131. doi:10.1007/BF01212345.

[188] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Translated by Amiel Feinstein. Holden-Day, San Francisco, 1964.

[189] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development* **4**(1) (1960), 66–82. doi:10.1147/rd.41.0066.

[190] Kane S. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation* **14**(3) (1966), 302–307. doi:10.1109/tap.1966.1138693.

[191] J. W. Alexander. A proof and extension of the Jordan–Brouwer separation theorem. *Transactions of the American Mathematical Society* **23** (1922), 333–349. doi:10.1090/S0002-9947-1922-1501206-6.

[192] Karol Borsuk. Drei Sätze über die n-dimensionale euklidische Sphäre. *Fundamenta Mathematicae* **20** (1933), 177–190. doi:10.4064/fm-20-1-177-190.

[193] Samuel Eilenberg and Norman Steenrod. *Foundations of Algebraic Topology*. Princeton Mathematical Series, Volume 15. Princeton University Press, Princeton, 1952. doi:10.1515/9781400877492.

[194] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[195] Giuseppe Peano. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen* **36** (1890), 157–160. doi:10.1007/BF01199438.

[196] Jonathan Washburn. On the Kuhn lattice, the long-wavelength vertices of Regge's action are one half of Einstein's through quartic order. Preprint, Zenodo, 2026. doi:10.5281/zenodo.24053932; Pith doi:10.67347/pith.preprint.2026.000050.

[197] Jonathan Washburn. Gravity from the recognition ledger: a finite sum over geometries with Einstein's equations at every order. Preprint, version 1.0, Zenodo, 2026. doi:10.5281/zenodo.24064849.

[198] Jonathan Washburn. The moral code from the physics of the record: least cost over acts with consent. Preprint, version 1.1, Zenodo, 2026. doi:10.5281/zenodo.24046120; Pith doi:10.67347/pith.preprint.2026.000049.

[199] Jeff Cheeger, Werner Müller and Robert Schrader. On the curvature of piecewise flat spaces. *Communications in Mathematical Physics* **92** (1984), 405–454. doi:10.1007/BF01210729.

[200] L. D. Landau and E. M. Lifshitz. *The Classical Theory of Fields*, fourth edition. Course of Theoretical Physics, Volume 2. Pergamon Press, Oxford, 1975.

[201] Stefan Felsner, Peter C. Fishburn and William T. Trotter. Finite three dimensional partial orders which are not sphere orders. *Discrete Mathematics* **201** (1999), 101–132. doi:10.1016/S0012-365X(98)00314-8.

[202] G. Hamel. Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x+y)=f(x)+f(y)$. *Mathematische Annalen* **60** (1905), 459–462.

[203] Pl. Kannappan. The functional equation $f(xy)+f(xy^{-1})=2f(x)f(y)$ for groups. *Proceedings of the American Mathematical Society* **19** (1968), 69–74. doi:10.2307/2036140.

[204] Serge Lang. *Introduction to Transcendental Numbers*. Addison-Wesley, Reading, MA, 1966.

[205] K. Ramachandra. Contributions to the theory of transcendental numbers (I) and (II). *Acta Arithmetica* **14** (1968), 65–72 and 73–88. doi:10.4064/aa-14-1-65-72 and doi:10.4064/aa-14-1-73-88.

[206] Michel Waldschmidt. *Diophantine Approximation on Linear Algebraic Groups*. Grundlehren der mathematischen Wissenschaften, Volume 326. Springer, Berlin, 2000.

[207] W. A. Howard. Hereditarily majorizable functionals of finite type. In A. S. Troelstra (ed.), *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, Volume 344, 454–461. Springer, Berlin, 1973.

[208] Marc Bezem. Strongly majorizable functionals of finite type: a model for bar recursion containing discontinuous functionals. *Journal of Symbolic Logic* **50** (1985), 652–660. doi:10.2307/2274319.

[209] Josef Hofbauer. A simple proof of $1+1/2^2+1/3^2+\cdots=\pi^2/6$ and related identities. *American Mathematical Monthly* **109** (2002), 196–200. doi:10.2307/2695334.

[210] Jonathan Washburn. Supplement to "On the Kuhn lattice, the long-wavelength vertices of Regge's action are one half of Einstein's through quartic order": the tables T_2, T_3, T_4 as rational functions of κ and the exact checking programs, in eight parts S1–S8. Version 1.0, the supplement file QuarticReggeEinstein_supplement_v1.0.zip of the Zenodo record of [196] (its checksum list SHA256SUMS has sha256 9f80897365ce5f8e03ed3f9d250f55e2c7975cf312f670c2ead28808318d83b5), Zenodo, 2026. doi:10.5281/zenodo.24053932.