Contents

Preface 11

1 Chapter 1 14

The ZF Fragment and the von Neumann Natural Numbers 14
1.0.1 Roadmap . . . 15
1.0.2 Named moves . . . 16

1.1 Axioms for Sets 16
1.1.1 Starting again from sets . . . 16
1.1.2 The axioms . . . 17
1.1.3 The first consequences . . . 20
1.1.4 Exercises . . . 22
1.1.4.1 A. Check your understanding . . . 22
1.1.4.2 B. Practice . . . 22
1.1.4.3 C. Going deeper . . . 23

1.2 Ordered Pairs Built from Sets 24
1.2.1 Order from sets . . . 25
1.2.2 The characteristic property . . . 26
1.2.3 Products, and the working rules recovered . . . 28
1.2.4 Exercises . . . 30
1.2.4.1 A. Check your understanding . . . 30
1.2.4.2 B. Practice . . . 31
1.2.4.3 C. Going deeper . . . 31

1.3 Successors, Inductive Sets and the Natural Numbers 33
1.3.1 Going one step further . . . 34
1.3.2 Inductive sets and the axiom of infinity . . . 35
1.3.3 The least inductive set . . . 36
1.3.4 Induction on the natural numbers . . . 38
1.3.5 Exercises . . . 40
1.3.5.1 A. Check your understanding . . . 40
1.3.5.2 B. Practice . . . 40
1.3.5.3 C. Going deeper . . . 41

1.4 The Natural Numbers Satisfy Peano's Axioms 42
1.4.1 Transitive sets . . . 42
1.4.2 Peano's axioms for the natural numbers . . . 44
1.4.3 Peano systems . . . 45
1.4.4 Pairs are not natural numbers . . . 47
1.4.5 Exercises . . . 48

1.4.5.1 A. Check your understanding . . . 48
1.4.5.2 B. Practice . . . 49
1.4.5.3 C. Going deeper . . . 50

1.5 The Recursion Theorem; Addition and Multiplication 52
1.5.1 Recursion on ω . . . 52
1.5.2 Addition and multiplication . . . 56
1.5.3 The laws of arithmetic . . . 58
1.5.4 Exercises . . . 60
1.5.4.1 A. Check your understanding . . . 60
1.5.4.2 B. Practice . . . 61
1.5.4.3 C. Going deeper . . . 62

1.6 Order and Well-Ordering of the Natural Numbers 64
1.6.1 Less than is "element of" . . . 64
1.6.2 Well-ordering . . . 66
1.6.3 Order and arithmetic . . . 67
1.6.4 Powers and parity . . . 69
1.6.5 Tuples and lists as sets . . . 71
1.6.6 Exercises . . . 72
1.6.6.1 A. Check your understanding . . . 72
1.6.6.2 B. Practice . . . 73
1.6.6.3 C. Going deeper . . . 74

1.7 All Peano Systems Are Alike 75
1.7.1 Two Peano systems, and what "alike" should mean . . . 76
1.7.2 The categoricity theorem . . . 76
1.7.3 Arithmetic and order travel along the isomorphism . . . 79
1.7.4 Identifying the two sets of natural numbers . . . 80
1.7.5 How later chapters cite *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 . . . 82
1.7.6 Exercises . . . 84
1.7.6.1 A. Check your understanding . . . 84
1.7.6.2 B. Practice . . . 84
1.7.6.3 C. Going deeper . . . 85
1.7.6.4 R. Review . . . 87

2 Chapter 2 89

Building the Integers and the Rationals from Equivalence Classes 89
2.0.1 Roadmap . . . 90
2.0.2 Named moves . . . 91

2.1 The Integers as Classes of Pairs 91
2.1.1 What we use about the natural numbers . . . 91
2.1.2 A naive attempt . . . 92
2.1.3 Pairs that record a difference . . . 92
2.1.4 Pictures of classes . . . 94
2.1.5 Why this definition . . . 95
2.1.6 Exercises . . . 97
2.1.6.1 A. Check your understanding . . . 97
2.1.6.2 B. Practice . . . 98

2.1.6.3 C. Going deeper . . . 99

2.2 Well-Definedness, and the Arithmetic of the Integers 100
2.2.1 Definitions through representatives . . . 100
2.2.2 The lemma . . . 101
2.2.3 Sum, product and negative of integers . . . 103
2.2.4 Rules that fail . . . 105
2.2.5 The natural numbers inside the integers . . . 106
2.2.6 Exercises . . . 107
2.2.6.1 A. Check your understanding . . . 107
2.2.6.2 B. Practice . . . 108
2.2.6.3 C. Going deeper . . . 109

2.3 The Integers as an Ordered Ring 110
2.3.1 The laws of a ring . . . 111
2.3.2 The integers form a commutative ring . . . 112
2.3.3 Ordered rings . . . 114
2.3.4 The order of the integers . . . 116
2.3.5 Rules for inequalities . . . 118
2.3.6 Least elements, and a ring that cannot be ordered . . . 120
2.3.7 Exercises . . . 122
2.3.7.1 A. Check your understanding . . . 122
2.3.7.2 B. Practice . . . 122
2.3.7.3 C. Going deeper . . . 123

2.4 The Rationals as Classes of Fractions 124
2.4.1 Fractions as pairs . . . 125
2.4.2 Adding and multiplying classes . . . 127
2.4.3 The rationals form a field . . . 129
2.4.4 The integers inside the rationals . . . 131
2.4.5 Other denominators . . . 132
2.4.6 Exercises . . . 133
2.4.6.1 A. Check your understanding . . . 133
2.4.6.2 B. Practice . . . 133
2.4.6.3 C. Going deeper . . . 135

2.5 The Rationals as an Ordered Field; the Archimedean Property 137
2.5.1 Which fractions are positive . . . 137
2.5.2 The rationals form an ordered field . . . 139
2.5.3 Between two rationals . . . 141
2.5.4 The Archimedean property . . . 141
2.5.5 The integer part . . . 143
2.5.6 Exercises . . . 144
2.5.6.1 A. Check your understanding . . . 144
2.5.6.2 B. Practice . . . 145
2.5.6.3 C. Going deeper . . . 146

2.6 The Gaps in the Rationals 148
2.6.1 No rational squares to two . . . 148
2.6.2 The set below the gap . . . 150
2.6.3 A rational sequence aimed at the gap . . . 152
2.6.4 Exercises . . . 154

2.6.4.1 A. Check your understanding . . . 154
2.6.4.2 B. Practice . . . 155
2.6.4.3 C. Going deeper . . . 156
2.6.4.4 R. Review . . . 158

3 Chapter 3 162

Divisibility, Division with Remainder, and Modular Arithmetic 162
3.0.1 Roadmap . . . 164
3.0.2 Named moves . . . 164

3.1 Divisibility and Division with Remainder 165
3.1.1 Divisibility on the constructed integers . . . 165
3.1.2 Division with remainder . . . 167
3.1.3 Even and odd . . . 169
3.1.4 Small primes . . . 170
3.1.5 Exercises . . . 171
3.1.5.1 A. Check your understanding . . . 171
3.1.5.2 B. Practice . . . 171
3.1.5.3 C. Going deeper . . . 172

3.2 The Greatest Common Divisor, Bézout, and the Euclidean Algorithm 172
3.2.1 Integer combinations . . . 173
3.2.2 The greatest common divisor . . . 173
3.2.3 Bézout's theorem . . . 174
3.2.4 The Euclidean algorithm . . . 176
3.2.5 Exercises . . . 180
3.2.5.1 A. Check your understanding . . . 180
3.2.5.2 B. Practice . . . 180
3.2.5.3 C. Going deeper . . . 181

3.3 Euclid's Lemma and the Fundamental Theorem of Arithmetic 183
3.3.1 Euclid's lemma . . . 183
3.3.2 Lists, products and prime factorizations . . . 184
3.3.3 Primes dividing products . . . 185
3.3.4 The least divisor is prime . . . 187
3.3.5 The fundamental theorem . . . 187
3.3.6 Lowest terms . . . 189
3.3.7 Exercises . . . 190
3.3.7.1 A. Check your understanding . . . 190
3.3.7.2 B. Practice . . . 191
3.3.7.3 C. Going deeper . . . 193

3.4 Ideals, Quotient Rings, and the Integers Modulo n 195
3.4.1 What a quotient needs . . . 196
3.4.2 The quotient ring . . . 197
3.4.3 The ideals of the integers . . . 199
3.4.4 The ring of integers modulo n . . . 200
3.4.5 Exercises . . . 202
3.4.5.1 A. Check your understanding . . . 202
3.4.5.2 B. Practice . . . 203
3.4.5.3 C. Going deeper . . . 204

3.5 Units, Prime Moduli, and the Chinese Remainder Theorem 205
3.5.1 Units of a commutative ring . . . 206
3.5.2 Which classes are units . . . 207
3.5.3 The field of integers modulo a prime . . . 208
3.5.4 Two congruences at once . . . 210
3.5.5 Exercises . . . 212
3.5.5.1 A. Check your understanding . . . 212
3.5.5.2 B. Practice . . . 213
3.5.5.3 C. Going deeper . . . 215
3.5.5.4 R. Review . . . 217

4 Chapter 4 220

Ordered Fields and Sequences in an Ordered Field 220
4.0.1 Roadmap . . . 221
4.0.2 Named moves . . . 222

4.1 Working in an Ordered Field 222
4.1.1 The standing setup . . . 222
4.1.2 Fractions . . . 223
4.1.3 Halving and comparing . . . 224
4.1.4 Absolute value . . . 225
4.1.5 Exercises . . . 228
4.1.5.1 A. Check your understanding . . . 228
4.1.5.2 B. Practice . . . 228
4.1.5.3 C. Going deeper . . . 230

4.2 The Rationals inside an Ordered Field; the Archimedean Property 231
4.2.1 The natural numbers in an ordered field . . . 231
4.2.2 A copy of the rationals in every ordered field . . . 233
4.2.3 The Archimedean property . . . 236
4.2.4 Integer part and density . . . 237
4.2.5 A field that is not Archimedean . . . 239
4.2.6 Exercises . . . 240
4.2.6.1 A. Check your understanding . . . 240
4.2.6.2 B. Practice . . . 240
4.2.6.3 C. Going deeper . . . 241

4.3 Sequences and Limits in an Ordered Field 244
4.3.1 Sequences . . . 244
4.3.2 The limit of a sequence . . . 245
4.3.3 Archimedes and the sequence $1/(n+1)$ . . . 248
4.3.4 Changing finitely many terms . . . 249
4.3.5 A sequence has at most one limit . . . 250
4.3.6 Convergent sequences are bounded . . . 251
4.3.7 Exercises . . . 252
4.3.7.1 A. Check your understanding . . . 252
4.3.7.2 B. Practice . . . 252
4.3.7.3 C. Going deeper . . . 254

4.4 The Algebra and Order of Limits 255
4.4.1 Two tools . . . 255

4.4.2 The limit laws . . . 256
4.4.3 Limits and order . . . 259
4.4.4 Exercises . . . 260
4.4.4.1 A. Check your understanding . . . 260
4.4.4.2 B. Practice . . . 261
4.4.4.3 C. Going deeper . . . 263

4.5 Cauchy Sequences 265
4.5.1 Terms that agree with one another . . . 265
4.5.2 Convergent sequences are Cauchy, and Cauchy sequences are bounded . . . 267
4.5.3 Sums and products of Cauchy sequences . . . 268
4.5.4 The sign of a Cauchy sequence that is not null . . . 270
4.5.5 Sign stability and reciprocals . . . 271
4.5.6 Small steps do not make a Cauchy sequence . . . 273
4.5.7 Exercises . . . 274
4.5.7.1 A. Check your understanding . . . 274
4.5.7.2 B. Practice . . . 275
4.5.7.3 C. Going deeper . . . 276

4.6 A Cauchy Sequence with No Rational Limit 278
4.6.1 The sequence again . . . 278
4.6.2 The theorem . . . 279
4.6.3 What this chapter did not prove . . . 281
4.6.4 Exercises . . . 282
4.6.4.1 A. Check your understanding . . . 282
4.6.4.2 B. Practice . . . 282
4.6.4.3 C. Going deeper . . . 283
4.6.4.4 R. Review . . . 285

5 Chapter 5 289

Constructing the Real Numbers from Cauchy Sequences 289
5.0.1 Roadmap . . . 291
5.0.2 Named moves . . . 291

5.1 Sequences That Should Converge 292
5.1.1 Two sequences aimed at one hole . . . 292
5.1.2 Cauchy sequences of rationals . . . 293
5.1.3 When two sequences name the same number . . . 294
5.1.4 The truncations and the Newton sequence . . . 296
5.1.5 Finite changes, and shifts . . . 298
5.1.6 Exercises . . . 299
5.1.6.1 A. Check your understanding . . . 299
5.1.6.2 B. Practice . . . 300
5.1.6.3 C. Going deeper . . . 301

5.2 The Real Numbers and Their Arithmetic 302
5.2.1 The real numbers as classes . . . 302
5.2.2 Arithmetic on representatives . . . 303
5.2.3 The null sequences form an ideal . . . 305
5.2.4 Inverses . . . 306
5.2.5 Exercises . . . 308

5.2.5.1 A. Check your understanding . . . 308
5.2.5.2 B. Practice . . . 309
5.2.5.3 C. Going deeper . . . 310

5.3 Positive Classes and the Order 312
5.3.1 Comparing terms does not compare classes . . . 312
5.3.2 Positive sequences . . . 313
5.3.3 The positive classes . . . 315
5.3.4 The order of the reals . . . 316
5.3.5 Comparing classes through their terms . . . 317
5.3.6 The absolute value through representatives . . . 318
5.3.7 Exercises . . . 319
5.3.7.1 A. Check your understanding . . . 319
5.3.7.2 B. Practice . . . 320
5.3.7.3 C. Going deeper . . . 321

5.4 The Rationals Inside the Reals 323
5.4.1 A copy of the rationals . . . 324
5.4.2 Archimedes in the reals . . . 325
5.4.3 Between two reals there is a rational . . . 326
5.4.4 Rational tolerances suffice . . . 328
5.4.5 Each real number is the limit of its terms . . . 328
5.4.6 The square root of two . . . 329
5.4.7 Exercises . . . 331
5.4.7.1 A. Check your understanding . . . 331
5.4.7.2 B. Practice . . . 332
5.4.7.3 C. Going deeper . . . 333

5.5 Every Cauchy Sequence of Reals Converges 334
5.5.1 The question . . . 335
5.5.2 The theorem . . . 335
5.5.3 The approximant, by a rule . . . 337
5.5.4 Choice, and why it was not needed . . . 338
5.5.5 Exercises . . . 339
5.5.5.1 A. Check your understanding . . . 339
5.5.5.2 B. Practice . . . 339
5.5.5.3 C. Going deeper . . . 340

5.6 The Real Numbers Form a Complete Ordered Field 342
5.6.1 Complete ordered fields . . . 343
5.6.2 Powers of two in an ordered field . . . 344
5.6.3 From Cauchy-complete to complete . . . 345
5.6.4 The constructed reals . . . 347
5.6.5 What we built, and what comes next . . . 348
5.6.6 Exercises . . . 349
5.6.6.1 A. Check your understanding . . . 349
5.6.6.2 B. Practice . . . 349
5.6.6.3 C. Going deeper . . . 350
5.6.6.4 R. Review . . . 352

6 Chapter 6 **356**

The Complete Ordered Field Is Unique **356**
6.0.1 Roadmap . . . 357
6.0.2 Named moves . . . 358

6.1 What Completeness Gives at Once **358**
6.1.1 Maps that respect everything . . . 359
6.1.2 A complete ordered field is Archimedean . . . 361
6.1.3 The rationals below an element . . . 362
6.1.4 Square roots . . . 364
6.1.5 Exercises . . . 366
6.1.5.1 A. Check your understanding . . . 366
6.1.5.2 B. Practice . . . 367
6.1.5.3 C. Going deeper . . . 367

6.2 Building the Map from the Rationals Below **368**
6.2.1 What the map must be . . . 369
6.2.2 Sums and products, measured by rationals . . . 369
6.2.3 The map into a complete field . . . 371
6.2.4 Exercises . . . 374
6.2.4.1 A. Check your understanding . . . 374
6.2.4.2 B. Practice . . . 374
6.2.4.3 C. Going deeper . . . 376

6.3 Two Complete Ordered Fields, One Isomorphism **378**
6.3.1 The uniqueness theorem . . . 378
6.3.2 One real line . . . 380
6.3.3 Automorphisms . . . 382
6.3.4 Exercises . . . 384
6.3.4.1 A. Check your understanding . . . 384
6.3.4.2 B. Practice . . . 385
6.3.4.3 C. Going deeper . . . 386

6.4 Two Non-Examples **387**
6.4.1 Polynomial functions . . . 388
6.4.2 The field of rational functions . . . 391
6.4.3 The field $\mathbb{Q}(\sqrt{2})$ . . . 396
6.4.4 A symmetry that keeps the order . . . 399
6.4.5 What each hypothesis buys . . . 400
6.4.6 Exercises . . . 401
6.4.6.1 A. Check your understanding . . . 401
6.4.6.2 B. Practice . . . 402
6.4.6.3 C. Going deeper . . . 403

6.5 Assume the Axioms or Build a Model **405**
6.5.1 Two ways to begin analysis . . . 405
6.5.2 What uniqueness adds . . . 406
6.5.3 What is not carried across . . . 407
6.5.4 What *Analysis on the Real Line* takes from here . . . 408
6.5.5 Exercises . . . 409
6.5.5.1 A. Check your understanding . . . 409
6.5.5.2 B. Practice . . . 410

6.5.5.3 C. Going deeper . . . 411
6.5.5.4 R. Review . . . 412

7 Chapter 7 **416**

Constructing the Real Numbers by Dedekind Cuts **416**
7.0.1 Roadmap . . . 417
7.0.2 Named moves . . . 418

7.1 Cuts **418**
7.1.1 The set below the gap . . . 418
7.1.2 Dedekind cuts . . . 419
7.1.3 Inside and outside . . . 422
7.1.4 Walking up through a cut . . . 423
7.1.5 Exercises . . . 425
7.1.5.1 A. Check your understanding . . . 425
7.1.5.2 B. Practice . . . 425
7.1.5.3 C. Going deeper . . . 426

7.2 Order by Inclusion, and Least Upper Bounds by Unions **427**
7.2.1 Comparing cuts . . . 427
7.2.2 Least upper bounds are unions . . . 428
7.2.3 Exercises . . . 430
7.2.3.1 A. Check your understanding . . . 430
7.2.3.2 B. Practice . . . 431
7.2.3.3 C. Going deeper . . . 432

7.3 Addition and Negatives **434**
7.3.1 Adding two cuts . . . 434
7.3.2 The zero cut . . . 436
7.3.3 Negatives . . . 436
7.3.4 A cut plus its negative . . . 439
7.3.5 An ordered group . . . 440
7.3.6 Exercises . . . 441
7.3.6.1 A. Check your understanding . . . 441
7.3.6.2 B. Practice . . . 442
7.3.6.3 C. Going deeper . . . 443

7.4 Multiplication **444**
7.4.1 Why not the set of products . . . 445
7.4.2 Positive cuts . . . 445
7.4.3 Signs . . . 448
7.4.4 Reciprocals . . . 451
7.4.5 The cuts form an ordered field . . . 453
7.4.6 Exercises . . . 454
7.4.6.1 A. Check your understanding . . . 454
7.4.6.2 B. Practice . . . 455
7.4.6.3 C. Going deeper . . . 456

7.5 The Rationals Inside the Cuts, and a Complete Ordered Field **457**
7.5.1 A copy of the rationals . . . 458
7.5.2 Between two cuts there is a rational cut . . . 460

7.5.3 Archimedes in the cuts . . . 461
7.5.4 The square root of two . . . 462
7.5.5 A complete ordered field . . . 463
7.5.6 Exercises . . . 464
7.5.6.1 A. Check your understanding . . . 464
7.5.6.2 B. Practice . . . 464
7.5.6.3 C. Going deeper . . . 465

7.6 The Two Roads Give the Same Real Numbers 467
7.6.1 Two models on one page . . . 467
7.6.2 The isomorphism . . . 468
7.6.3 The square root of two, twice . . . 470
7.6.4 What each road pays for . . . 470
7.6.5 Exercises . . . 472
7.6.5.1 A. Check your understanding . . . 472
7.6.5.2 B. Practice . . . 472
7.6.5.3 C. Going deeper . . . 473
7.6.5.4 R. Review . . . 475

Preface

The two earlier volumes used numbers as familiar objects. Their arithmetic and order were school facts, named where they were used and never proved. This volume proves them. It starts from a few axioms about sets and builds the natural numbers, then the integers, the rationals and the real numbers, each out of the one before. At the end the real numbers form a complete ordered field, and any two complete ordered fields are the same up to exactly one isomorphism.

Who it is for. The reader we have in mind has worked through *Logic and Proof* and *Sets, Relations, and Cardinality*, or has their habits from elsewhere. That means negating a quantified statement, running an induction, and checking that a rule on equivalence classes does not depend on the representative chosen. The last of these is used in almost every chapter. The chapters point back to both volumes where they use a definition or a rule, so a reader who is unsure of a step can look it up.

What it covers.

Chapter	Title	What it does
1	The ZF Fragment and the von Neumann Natural Numbers	Seven axioms about sets, ordered pairs built from sets, the natural numbers as $0 = \emptyset$ and $n + 1 = n \cup \{n\}$, and Peano's axioms, recursion, arithmetic and order proved for them
2	Building the Integers and the Rationals from Equivalence Classes	$\mathbb{Z}$ and $\mathbb{Q}$ as quotient sets of pairs, well-defined operations, $\mathbb{Z}$ an ordered ring and $\mathbb{Q}$ an Archimedean ordered field, and the gap in $\mathbb{Q}$ where $\sqrt{2}$ should be
3	Divisibility, Division with Remainder, and Modular Arithmetic	Division with remainder, Bézout's theorem and the Euclidean algorithm, unique prime factorization, quotient rings by an ideal, $\mathbb{Z}/n\mathbb{Z}$, and the Chinese remainder theorem

Chapter	Title	What it does
4	Ordered Fields and Sequences in an Ordered Field	Limits and Cauchy sequences in an arbitrary ordered field, the Archimedean property, and a Cauchy sequence of rationals with no rational limit
5	Constructing the Real Numbers from Cauchy Sequences	$\mathbb{R}$ as classes of rational Cauchy sequences, proved an ordered field in which every Cauchy sequence converges and every bounded set has a supremum
6	The Complete Ordered Field Is Unique	Exactly one order-preserving isomorphism between any two complete ordered fields, two fields that show what each hypothesis buys, and what *Analysis on the Real Line* takes from this volume
7	Constructing the Real Numbers by Dedekind Cuts (optional)	$\mathbb{R}$ as sets of rationals closed downward, a second complete ordered field, and the isomorphism that links it to Chapter 5's

Chapters 1 and 2 build the numbers that algebra needs, and Chapter 3 does their arithmetic. Chapters 4 and 5 fill the gap that Chapter 2 found in $\mathbb{Q}$: first the rules for sequences, proved without the number they aim at, then the numbers made from the sequences themselves. Chapter 6 shows that the construction does not matter, and Chapter 7 checks this on a second construction.

How to read it. The main line is Chapters 1–6; read them in order, since each uses the ones before it. Chapter 7 is optional, and nothing outside it depends on it. Read it for a second construction of $\mathbb{R}$, and to see the uniqueness theorem of Chapter 6 at work on two concrete models. Every section has Quick checks with folded answers; answer each one before you unfold it. The exercises of a section come in three groups, A to check understanding, B for practice and C to go deeper, and the last section of each chapter adds R items that reach back to earlier chapters. A chapter's opening page lists what it needs, with links, and the time it takes. Second-reading folds, in Chapters 5–7, hold edge cases and longer justifications. They can stay closed on a first reading, but they are not optional, since later proofs may use what they contain. Results marked optional can be skipped, and nothing outside them depends on them.

Time. The chapter openings add up to about 150 hours on the core track for the main line, Chapters 1–6, and about 210 hours for everything in it. Chapter 7 adds about 25 hours on the core track and about 30 for everything.

What comes next. *Analysis on the Real Line* does not need this volume. It starts from the axioms of a complete ordered field, the axioms this volume proves for its $\mathbb{R}$, and Chapter 6 explains why it does not matter which model of them one has in mind. A reader who wants to know that those axioms describe something real, and only one thing, will find both answers

here. *Algebraic Preliminaries* builds on Chapters 3, 5 and 6: quotient rings and $\mathbb{Z}/n\mathbb{Z}$, the real numbers as a complete ordered field, and square roots in it, together with the ring and field rules of Chapters 2 and 4 that come with them.

Chapter 1

The ZF Fragment and the von Neumann Natural Numbers

Logic and Proof, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 took the natural numbers as a set $\mathbb{N}$ inside the familiar real line, with induction as an axiom (Logic and Proof, Axiom 4.2.1) and with their arithmetic and order borrowed from the real numbers. This volume builds the real numbers, so it cannot borrow from them: where, then, do the natural numbers come from? This chapter starts again from sets alone. Six axioms say how sets are made and when two are equal; with them we build ordered pairs, and so functions, out of sets. Then we build the natural numbers out of nothing but the empty set. Zero is $\emptyset$, and "one further" is the step $x \mapsto x \cup \{x\}$, so that $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$: each number **is** the set of the numbers before it. A seventh axiom, Infinity, says that the steps can be completed into a set, and the least such set is ω. From there, everything that *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 assumed about $\mathbb{N}$ becomes a theorem about ω: Peano's axioms, induction, the recursion theorem, the laws of $+$ and $\cdot$, and the order with its well-ordering. A final theorem says that any other system satisfying Peano's axioms is a relabeling of this one, by exactly one matching, and the chapter closes with the rule by which the later chapters of this volume write $\mathbb{N}$ for ω and cite *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapters 2 and 3.

Two pictures. *Sets, Relations, and Cardinality*, Chapter 1 adopted two working conventions: numbers are not sets (*Sets, Relations, and Cardinality*, Section 1.1), and a pair is not a number, nor a triple a pair (the remark after Sets, Relations, and Cardinality, Definition 1.3.1). They belong to the picture of *Sets, Relations, and Cardinality*, Chapters 1 to 3, in which numbers are given objects. In this chapter every object is a set, so $0 \in 5$ and $2 \subseteq 3$ are true statements, and the pair $(0, 1)$ is the set $\{1, 2\}$. The two pictures do not conflict. *Sets, Relations, and Cardinality*, Section 1.1 said that nothing in the main line depends on how numbers are made, and *Sets, Relations, and Cardinality*, Chapter 1 used nothing about pairs beyond their characteristic property and that convention; Section 1.2 proves the characteristic property for pairs built from sets. What survives of the conventions becomes a theorem: no pair is a natural number (Section 1.4), and no n-tuple with $n \geq 3$ is a pair or a natural number (Section 1.6). Statements such as $0 \in 5$ are facts about this one construction; the matching of Section 1.7 carries 0, successor, $+$, $\cdot$ and $<$, and nothing else.

What you need. *Logic and Proof*, Chapters 1–5, and *Sets, Relations, and Cardinality*, Chapters 1–3, chiefly its Chapters 1 and 2. From *Logic and Proof*, Chapters 2 and 4: the rules $\neg$I and RAA (Logic and Proof, Definition 2.2.1, Logic and Proof, Definition 2.5.1), the rules for equality (Logic and Proof, Definition 5.7.1), and first-order formulas (Logic and Proof, Definition 5.2.2),

which are what a "property" in the axiom of Separation means; Section 1.3's exercises also use De Morgan's laws and negate quantifiers, bounded ones included (Logic and Proof, Proposition 2.5.5, Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3). From *Logic and Proof*, Chapter 3, the principle of induction (Logic and Proof, Axiom 4.2.1), which Section 1.3 proves as a theorem about ω and which is never used on ω; and, for review exercise R3 of Section 1.7, structural recursion and unique readability (Logic and Proof, Theorem 4.3.4, Logic and Proof, Theorem 4.3.3). From *Logic and Proof*, Chapter 5, the catalogue of what a proof may not do (Logic and Proof, Example (What is not allowed)). From *Sets, Relations, and Cardinality*, Chapter 1: equality of sets, subsets and sets cut out by a property (Sets, Relations, and Cardinality, Definition 1.1.1, Sets, Relations, and Cardinality, Definition 1.1.2, Sets, Relations, and Cardinality, Definition 1.1.3), Russell's theorem (Sets, Relations, and Cardinality, Theorem 1.1.4), unions and intersections, of two sets and of families (Sets, Relations, and Cardinality, Definition 1.2.1, Sets, Relations, and Cardinality, Definition 1.2.3, Sets, Relations, and Cardinality, Definition 1.2.4), the smallest closed set (Sets, Relations, and Cardinality, Proposition 1.2.2), ordered pairs, products and relations (Sets, Relations, and Cardinality, Definition 1.3.1, Sets, Relations, and Cardinality, Definition 1.3.2, Sets, Relations, and Cardinality, Definition 1.3.3), functions and when two are equal (Sets, Relations, and Cardinality, Definition 1.4.1, Sets, Relations, and Cardinality, Proposition 1.4.1), injective, surjective and bijective maps, composites and inverses (Sets, Relations, and Cardinality, Definition 1.4.3, Sets, Relations, and Cardinality, Definition 1.4.4, Sets, Relations, and Cardinality, Definition 1.4.5, Sets, Relations, and Cardinality, Definition 1.4.2, Sets, Relations, and Cardinality, Theorem 1.4.4, Sets, Relations, and Cardinality, Definition 1.5.2, Sets, Relations, and Cardinality, Theorem 1.5.4), images (Sets, Relations, and Cardinality, Definition 1.6.1), and, for review exercise R2 of Section 1.7, left and right inverses (Sets, Relations, and Cardinality, Definition 1.5.1, Sets, Relations, and Cardinality, Proposition 1.5.2, Sets, Relations, and Cardinality, Theorem 1.5.1). From *Sets, Relations, and Cardinality*, Chapter 2: the recursion theorem and its proof (Sets, Relations, and Cardinality, Theorem 2.3.1), which Section 1.5 repeats on ω; and well-ordering (Sets, Relations, and Cardinality, Theorem 2.1.3), induction from a base (Sets, Relations, and Cardinality, Theorem 2.2.1) and the arithmetic of $\mathbb{N}$ (Sets, Relations, and Cardinality, Lemma 2.1.1), which Sections 1.6 and 1.7 prove again on ω and compare with. From *Sets, Relations, and Cardinality*, Chapter 3, the axiom of choice (Sets, Relations, and Cardinality, Axiom 3.3.1), only to say that this chapter never uses it; and, for review exercise R1 of Section 1.7, the definition of an infinite set and the theorem that a finite set is not equinumerous with a proper subset (Sets, Relations, and Cardinality, Definition 3.1.3, Sets, Relations, and Cardinality, Theorem 3.1.9, read on ω through Section 1.7's dictionary), with Hilbert's hotel (Sets, Relations, and Cardinality, Example (Hilbert's Hotel)). Section 1.7 also names results of *Sets, Relations, and Cardinality*, Chapters 2 and 3 about finite and countable sets, to say how the later chapters of this volume read them on ω; apart from R1, it does not use them.

Time. About 30 hours on the core track (all of A, one B and one C per section, the exit checks); about 40 hours for everything.

1.0.1 Roadmap

- **Axioms for Sets.** Six axioms, (Z1)–(Z6), for making sets and telling them apart, why Separation needs a set to cut from, and the sets they name: the empty set, unordered pairs, unions, intersections and differences, each unique by Extensionality.
- **Ordered Pairs Built from Sets.** Kuratowski's pair $\{\{a\},\{a,b\}\}$, three shorter encodings that fail, the proof of the characteristic property with the case $a = b$ on its own, the product $A \times B$ as a set, and the working rules (W1)–(W5) of *Sets, Relations, and Cardinality*, Section 1.1 recovered from the axioms.
- **Successors, Inductive Sets and the Natural Numbers.** The successor $\mathrm{S}(x) = x \cup \{x\}$

and the numerals $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, inductive sets and the axiom of infinity (Z7), ω as the least inductive set, and induction on ω proved as a theorem.

- **The Natural Numbers Satisfy Peano's Axioms.** Transitive sets, the facts that no natural number is an element of itself and that every non-zero one is a successor, Peano's axioms (P1)–(P5) for ω, Peano systems with a structure for each of (P3)–(P5) that fails it alone, and the theorem that no pair is a natural number.
- **The Recursion Theorem; Addition and Multiplication.** The recursion theorem proved on ω by gluing partial solutions on the sets $\mathrm{S}(n)$, with no order needed; addition and multiplication defined by recursion, $n + 1 = \mathrm{S}(n)$, and the laws of arithmetic, each by an induction on the variable the recursion runs on.
- **Order and Well-Ordering of the Natural Numbers.** "Less than" defined as "element of" and proved a strict total order, each number the set of the numbers below it; well-ordering, by an induction on the numbers that contain no element of the set; order tied to arithmetic, with subtraction and induction from any base; powers and parity; and n-tuples and lists of variable length as sets.
- **All Peano Systems Are Alike.** Isomorphisms of Peano systems, the categoricity theorem (every Peano system is a copy of ω by exactly one isomorphism), addition, multiplication and order carried along it, the identification of ω with *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N}$, and the dictionary by which the later chapters cite *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapters 2 and 3 on ω; the review exercises reach back to *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapters 1 and 3.

1.0.2 Named moves

- **Show the set is inductive.** To prove $P(n)$ for every $n \in \omega$, form $\{n \in \omega : P(n)\}$ and show that it contains 0 and is closed under S. When the claim fails at 0, carry a disjunction "$n = 0$ or ..." instead.
- **Fix the others, induct on the recursion variable.** In $m + n$ and $m \cdot n$ the recursion runs on the second argument n; fix m (and any third variable) and induct on n. A fact about the first argument, such as $0 + n = n$, is a lemma to prove first, not a defining equation.
- **Build by recursion, compare by induction.** A map out of ω comes from the recursion theorem; facts about it (one-to-one, onto, preserves $+$) are proved by induction, or by (P5) of the system it maps into.

1.1 Axioms for Sets

Sets, Relations, and Cardinality, Chapters 1 to 3 worked with sets under six working rules, (W1)–(W6) of *Sets, Relations, and Cardinality*, Section 1.1, and took the natural numbers as a given set inside the familiar real line, with induction as an axiom (Logic and Proof, Axiom 4.2.1). This volume builds the real numbers, so it may not lean on them: we start again from sets alone, and in this chapter **every object is a set**. This section asks which rules about sets are basic. It states six axioms, shows that they settle when two sets are equal, and produces the first sets from them: the empty set, unordered pairs, unions, intersections and differences.

1.1.1 Starting again from sets

Sets, Relations, and Cardinality, Section 1.1 listed its working rules and used them without asking where they came from. One of them, specification (Sets, Relations, and Cardinality,

Definition 1.1.3), was the repair of a rule that fails: "every property defines a set" is contradictory, because the property $x \notin x$ would define a set that both is and is not an element of itself, and Sets, Relations, and Cardinality, Theorem 1.1.4 turns that contradiction into the fact that **no set has every set as an element**. So rules about sets cannot be chosen by what sounds harmless. They must be chosen with care and stated once, and everything else must then be derived from them.

That raises two questions, which this chapter answers in order. Which of the working rules are basic, and which follow from the others? And can the basic rules **produce** the objects *Sets, Relations, and Cardinality*, Chapters 1 to 3 took as given (ordered pairs, functions, the natural numbers), rather than only describe them? Section 1.2 builds ordered pairs out of sets; Section 1.3 builds the natural numbers.

To build numbers out of sets we must let sets be the only objects. From here to the end of the chapter, **every object is a set**, and every element of a set is again a set. *Sets, Relations, and Cardinality*, Section 1.1 used a different picture, in which numbers are given objects that are not sets, and it said that nothing in the main line would depend on how numbers are made. That picture stays valid for *Sets, Relations, and Cardinality*, Chapters 1 to 3; this chapter works in the other one, where 0 will be the empty set and $0 \in 1$ will be a true statement. The chapter's introduction says how the two pictures fit together.

The language. Statements about sets are written in the first-order language of *Logic and Proof*, Chapter 4 (Logic and Proof, Definition 5.2.2) with **one** predicate symbol, the two-place $\in$, together with equality. There are no constant symbols and no function symbols. The variables range over sets. So "$x \in y$" and "$x = y$" are the atomic formulas, and everything else is built from them with the connectives and the quantifiers. A **property** $P(x)$ of sets is a formula of this language in which the variable x may occur free; other free variables, called **parameters**, are allowed too, and stand for sets fixed in advance.

1.1.2 The axioms

We want sets that can be built, and we know one way not to build them: by an arbitrary property, with nothing to cut the set out of. So instead of one rule for every property, we list a few specific ways of making a new set from sets we already have, together with one set to start from.

A few ways to make sets, and one way to tell them apart.

Axiom 1.1.1. The Axioms (Z1)–(Z6)

(Z1) **Extensionality.** Sets with the same elements are equal: for all sets A and B, if $x \in A \leftrightarrow x \in B$ for every x, then $A = B$.

(Z2) **Empty set. There is** a set with no elements.

(Z3) **Pairing.** For all sets a and b **there is** a set whose elements are exactly a and b.

(Z4) **Union.** For every set $\mathcal{S}$ **there is** a set whose elements are exactly the elements of members of $\mathcal{S}$.

(Z5) **Power set.** For every set X **there is** a set whose elements are exactly the subsets of X.

(Z6) **Separation.** For every set A and every property $P(x)$ **there is** a set whose elements are exactly the $x \in A$ with $P(x)$.

In symbols, with $S \subseteq X$ as an abbreviation of $\forall y\,(y \in S \to y \in X)$ (Sets, Relations, and Cardinality, Definition 1.1.2):

$$\begin{aligned}
&\text{(Z1)} \quad \forall A\,\forall B\,\big(\forall x\,(x \in A \leftrightarrow x \in B) \to A = B\big),\\
&\text{(Z2)} \quad \exists E\,\forall x\,\neg(x \in E),\\
&\text{(Z3)} \quad \forall a\,\forall b\,\exists C\,\forall x\,\big(x \in C \leftrightarrow (x = a \vee x = b)\big),\\
&\text{(Z4)} \quad \forall \mathcal{S}\,\exists U\,\forall x\,\big(x \in U \leftrightarrow \exists S\,(S \in \mathcal{S} \wedge x \in S)\big),\\
&\text{(Z5)} \quad \forall X\,\exists Q\,\forall S\,(S \in Q \leftrightarrow S \subseteq X).
\end{aligned}$$

(Z6) is not one axiom but one for each property, an **axiom schema**. For every formula φ whose free variables are among $x, p_1, \ldots, p_k$ and in which the letter B does not appear,

$$\forall p_1 \cdots \forall p_k\,\forall A\,\exists B\,\forall x\,\big(x \in B \leftrightarrow (x \in A \wedge \varphi)\big). \tag{Z6}$$

The letter B here is bound, and may be renamed (*Logic and Proof*, Section 4.2) to any letter that appears nowhere else in (Z6) (not in φ, and neither x nor A). So a property may mention a given set called B, as $A \setminus B$ below does, once the new set is given another letter, say D: $\forall B\,\forall A\,\exists D\,\forall x\,\big(x \in D \leftrightarrow (x \in A \wedge x \notin B)\big)$.

In words, clause by clause, with what each one licenses.

- *(Z1)* is the only axiom that proves two sets **equal** by comparing their elements: it turns "same elements" into $A = B$. *Sets, Relations, and Cardinality*, Section 1.1 stated it as the meaning of equality of sets (Sets, Relations, and Cardinality, Definition 1.1.1); here it is an axiom, and only its direction "same elements, hence equal" needs to be assumed. The other direction, from $A = B$ to $x \in A \leftrightarrow x \in B$, is the equality rule =E of Logic and Proof, Definition 5.7.1, a rule of logic.
- *(Z2)* gives the one set we start from. Each of (Z3)–(Z6) makes a set out of sets already given.
- *(Z3)* makes a set out of two given sets a, b. The two need not differ: with $b = a$ its set has the single element a.
- *(Z4)* flattens one level: from a set $\mathcal{S}$ of sets it makes the set of what lies **inside** the members of $\mathcal{S}$. With (Z3) it gives the union of two sets (below).
- *(Z5)* collects **all** subsets of X into one set; it is the working rule (W4).
- *(Z6)* cuts a set out of a set A already given, by a property. It is specification (Sets, Relations, and Cardinality, Definition 1.1.3), now an axiom. The letter B must not appear in φ, or the property could refer to the set being made: with φ the formula $x \notin B$, and B the very set that (Z6) produces, it would say $x \in B \leftrightarrow (x \in A \wedge x \notin B)$, which no set B satisfies when A has an element.

Properties written with defined symbols. The formulas of (Z6) contain only $\in$ and $=$. We shall write properties with defined symbols as well ($\subseteq$, $\emptyset$, $\{a, b\}$, and later $\cup$, pairs and functions), and that is harmless: each defined symbol is shorthand that unwinds into $\in$ and $=$. For instance "$x = \emptyset$" unwinds to $\forall y\,\neg(y \in x)$, and "$x = \{a, b\}$" to $\forall y\,\big(y \in x \leftrightarrow (y = a \vee y = b)\big)$. A property is admissible in (Z6) when its unwinding is a formula of the language, which is the case for every property in this chapter.

One way to tell sets apart, and one way to make them unique. Each of (Z2)–(Z6) says that **there is** a set with a certain membership condition. By (Z1) there is then **exactly one**: two sets with the same membership condition have the same elements, so they are equal. We use this move so often that we state it once:

To show that a set is unique, show that any two sets with its membership condition have the same elements, and apply (Z1).

The first sets. Here are the axioms at work, each set checked against the membership condition that made it.

- *The empty set.* (Z2) gives a set with no elements; Proposition 1.1.2 below shows it is unique, and we write it $\emptyset$.
- *A singleton.* (Z3) with $a = b = \emptyset$ gives a set whose elements are exactly the x with $x = \emptyset \vee x = \emptyset$, that is, $x = \emptyset$. This is $\{\emptyset\}$. It is not $\emptyset$: it has the element $\emptyset$, and $\emptyset$ has none.
- *A two-element set.* (Z3) with $a = \emptyset$ and $b = \{\emptyset\}$ gives $\{\emptyset, \{\emptyset\}\}$, whose elements are exactly $\emptyset$ and $\{\emptyset\}$. These two are different, by the previous item, so this set has two elements.
- *A union.* (Z4) with $\mathcal{S} = \{\emptyset, \{\emptyset\}\}$ gives the set of everything that lies in $\emptyset$ or in $\{\emptyset\}$. Nothing lies in $\emptyset$, and only $\emptyset$ lies in $\{\emptyset\}$. So the union is $\{\emptyset\}$: one element, although $\mathcal{S}$ has two.
- *A power set.* (Z5) with $X = \{\emptyset\}$ gives the set of subsets of $\{\emptyset\}$. There are two, $\emptyset$ and $\{\emptyset\}$ itself (the argument of Sets, Relations, and Cardinality, Example (Small power sets), which uses only extensionality and excluded middle, applies word for word). So $\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$, the set of the third item.
- *A separation.* (Z6) with $A = \{\emptyset, \{\emptyset\}\}$ and the property $x \neq \emptyset$ gives $\{\{\emptyset\}\}$: of the two elements of A, only $\{\emptyset\}$ differs from $\emptyset$.

The degenerate case is the empty set itself, and it matters: each natural number will be built from $\emptyset$ by these axioms, one at a time. The set of all of them needs one more starting point, the Axiom of Infinity (Section 1.3).

A non-example by minimal change. Delete the ambient set from (Z6): "for every property $P(x)$ there is a set whose elements are exactly the x with $P(x)$". This is the naive rule of *Sets, Relations, and Cardinality*, Section 1.1, and it fails at once. With $P(x)$ the property $x \notin x$ it gives a set R with $R \in R \leftrightarrow R \notin R$, a contradiction; and even with the harmless-looking property $x = x$ it gives a set of which every set is an element, which Sets, Relations, and Cardinality, Theorem 1.1.4 rules out. What fails is exactly the clause "$x \in A$" that was deleted. Every other part of (Z6) survives the change; the ambient set is what makes it safe.

Why this list. Each axiom earns its place by something later in the chapter. Pairing and Union give finite sets and the union of two sets; Power set and Separation give the product $A \times B$ (Section 1.2) and, with it, relations and functions; Extensionality makes every construction unique. Four further axioms complete the usual system of set theory with choice, and only one of them is needed here:

- **Infinity** says that some set contains $\emptyset$ and is closed under one particular step. It is (Z7), stated in Section 1.3, once there is a step to iterate. Without it, the axioms above produce each finite set we write down, one at a time. That they cannot produce a set holding infinitely many of them is a theorem of set theory, which this book states and does not prove.
- **Replacement** and **Foundation** are not needed to build the natural numbers, and are left to the volume on set theory, which revisits this chapter.
- **Choice** is Sets, Relations, and Cardinality, Axiom 3.3.1 of *Sets, Relations, and Cardinality*, Chapter 3. Nothing in this chapter uses it.

We do not ask whether the axioms are consistent, or (apart from exercise C1) whether one of them follows from the others. Those are questions about the axioms, and this chapter only uses them.

Warning. **Two traps in the list.** First, (Z3) gives an **unordered** pair: $\{a, b\} = \{b, a\}$, since $x = a \vee x = b$ and $x = b \vee x = a$ say the same thing. So $\{a, b\}$ cannot record which of

a, b came first, and Section 1.2 needs a different construction for the ordered pair. Second, (Z6) needs the ambient set A: the rule without it is exactly the one Sets, Relations, and Cardinality, Theorem 1.1.4 refutes.

1.1.3 The first consequences

The existence axioms say "there is a set"; the first proposition says "there is exactly one", and so lets us give each such set a name. It also rebuilds from the axioms the operations of Sets, Relations, and Cardinality, Definition 1.2.1 and, for a set of sets $\mathcal{S}$, the $\bigcup\mathcal{S}$ and $\bigcap\mathcal{S}$ of Sets, Relations, and Cardinality, Definition 1.2.4; indexed families wait for Section 1.2.

Proposition 1.1.2. The Sets Named by the Axioms

(a) There is **exactly one** set with no elements. It is written $\emptyset$.

(b) For all sets a and b the set of (Z3) is unique. It is written $\{a, b\}$, and $\{a\} \coloneqq \{a, a\}$, the set whose only element is a.

(c) For every set $\mathcal{S}$ the set of (Z4) is unique. It is written $\bigcup\mathcal{S}$, and $A \cup B \coloneqq \bigcup\{A, B\}$ for sets A, B, so that $x \in A \cup B$ if and only if $x \in A$ or $x \in B$. The sets $A \cap B \coloneqq \{x \in A : x \in B\}$ and $A \setminus B \coloneqq \{x \in A : x \notin B\}$ exist by (Z6). For a **non-empty** set $\mathcal{S}$, the set

$$\bigcap\mathcal{S} \coloneqq \{x \in A_0 : x \in S \text{ for every } S \in \mathcal{S}\}$$

exists by (Z6) for any $A_0 \in \mathcal{S}$, it **does not depend** on A_0, and $x \in \bigcap\mathcal{S}$ if and only if $x \in S$ for every $S \in \mathcal{S}$.

(d) For all sets a, b, c, d,

$$\{a, b\} = \{c, d\} \iff (a = c \wedge b = d) \vee (a = d \wedge b = c).$$

In particular $\{a\} = \{c, d\}$ if and only if $a = c$ and $c = d$.

Idea. Parts (a)–(c) are the uniqueness move three times: the axiom named gives a set, and any two sets with the same membership condition are equal by (Z1). For $\bigcap\mathcal{S}$, the condition "$x \in A_0$" adds nothing, since A_0 is itself one of the $S \in \mathcal{S}$; so the condition does not mention A_0 at all. Part (d) is different: it compares elements. From $a \in \{a, b\} = \{c, d\}$ we get $a = c$ or $a = d$, and each case is finished by locating b and then the remaining element of $\{c, d\}$.

Proof.

(a) A set with no elements exists by (Z2). If E and E' both have no elements, then for every x, both $x \in E$ and $x \in E'$ are false, so $x \in E \leftrightarrow x \in E'$; hence $E = E'$ by (Z1). This is the argument of Sets, Relations, and Cardinality, Proposition 1.1.3, with (Z1) in place of Sets, Relations, and Cardinality, Definition 1.1.1.

(b) Let C and C' be sets as in (Z3) for the same a, b. For every x, $x \in C$ and $x \in C'$ are each equivalent to $x = a \vee x = b$, so $C = C'$ by (Z1). With $b = a$ the condition is $x = a \vee x = a$, which is equivalent to $x = a$; so the only element of $\{a\}$ is a.

(c) The set of (Z4) is unique by the same argument, its condition being $\exists S\,(S \in \mathcal{S} \wedge x \in S)$. For $\mathcal{S} = \{A, B\}$, an $S \in \{A, B\}$ is A or B, by (b); so $x \in A \cup B$ if and only if $x \in A$

or $x \in B$. The sets $A \cap B$ and $A \setminus B$ are instances of (Z6), with ambient set A and parameter B (after renaming the bound letter of (Z6)), and are unique by (Z1).
Now let $\mathcal{S}$ be non-empty, and let $A_0 \in \mathcal{S}$. By (Z6) with ambient set A_0 and parameter $\mathcal{S}$, the set $\bigcap \mathcal{S}$ above exists. Its membership condition simplifies:

$$x \in A_0 \wedge \forall S\,(S \in \mathcal{S} \to x \in S) \iff \forall S\,(S \in \mathcal{S} \to x \in S).$$

From left to right, drop the first conjunct. From right to left, apply the condition at $S = A_0$, which is a member of $\mathcal{S}$, to get $x \in A_0$. The right-hand side does not mention A_0. So if A_1 is another member of $\mathcal{S}$, the sets built from A_0 and from A_1 have the same elements, and they are equal by (Z1).

(d) ($\Leftarrow$) If $a = c$ and $b = d$, then $\{a, b\} = \{c, d\}$ by =E. If $a = d$ and $b = c$, then $\{a, b\} = \{d, c\}$, and $\{d, c\} = \{c, d\}$ because $x = d \vee x = c$ is equivalent to $x = c \vee x = d$, so the two sets have the same elements (Z1).

($\Rightarrow$) Suppose $\{a, b\} = \{c, d\}$. Since $a \in \{a, b\}$, we have $a \in \{c, d\}$, so $a = c$ or $a = d$. We take the two cases in turn ($\vee$E).

Case 1: $a = c$. Since $b \in \{c, d\}$, $b = c$ or $b = d$. If $b = d$, then $a = c$ and $b = d$, the first disjunct. If $b = c$, then $b = c = a$, so $\{a, b\} = \{a\}$. Now $d \in \{c, d\} = \{a, b\} = \{a\}$, so $d = a$, and hence $b = a = d$. Again $a = c$ and $b = d$.

Case 2: $a = d$. Since $b \in \{c, d\}$, $b = c$ or $b = d$. If $b = c$, then $a = d$ and $b = c$, the second disjunct. If $b = d$, then $b = d = a$, so $\{a, b\} = \{a\}$. Now $c \in \{a\}$, so $c = a$, and $b = a = c$. So $a = d$ and $b = c$, the second disjunct again.

In each case one of the two disjuncts holds.

For the last claim, apply what was just proved with $b = a$: $\{a\} = \{a, a\} = \{c, d\}$ if and only if $(a = c \wedge a = d) \vee (a = d \wedge a = c)$. Both disjuncts say that $a = c$ and $a = d$, that is, $a = c$ and $c = d$.

■

Two consequences of (d) are used constantly. With $c = d$ it says that $\{a\} = \{c\}$ only if $a = c$: a singleton determines its element. And with $c = a$ it says that $\{a, b\} = \{a, d\}$ only if $b = d$: in the first disjunct $b = d$ directly, and in the second $a = d$ and $b = a$, so again $b = d$. Section 1.2 uses both.

The sets of *Sets, Relations, and Cardinality*, Chapter 1, rebuilt. The union $A \cup B$ of part (c) has the membership condition of Sets, Relations, and Cardinality, Definition 1.2.1 (a), and the intersection and difference are defined exactly as in parts (b) and (c) of that definition; so by (Z1) these are the sets *Sets, Relations, and Cardinality*, Chapter 1 used. Likewise, if $\mathcal{S}$ is a non-empty set of subsets of a set X, the $\bigcap \mathcal{S}$ of part (c) and the $\bigcap \mathcal{S}$ of Sets, Relations, and Cardinality, Definition 1.2.4, cut out of X, have the same membership condition, "x lies in every member of $\mathcal{S}$", and so they are the same set. The new version needs no ambient X: any member A_0 of $\mathcal{S}$ does the job. For the empty $\mathcal{S}$ it would have to contain every set, and no set does (Sets, Relations, and Cardinality, Theorem 1.1.4); exercise B2 returns to this.

Quick check. Compute $\bigcup \emptyset$, $\bigcup \{\emptyset\}$ and $\bigcup \{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$.

Solution. $\bigcup \emptyset = \emptyset$: an element would have to lie in some member of $\emptyset$, and $\emptyset$ has no members. $\bigcup\{\emptyset\} = \emptyset$: the only member is $\emptyset$, which has no elements. $\bigcup\{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}\}$: the first member contributes $\emptyset$, the second contributes $\emptyset$ and $\{\emptyset\}$, and repeats do not count.

With Proposition 1.1.2 every set we have named so far is a set of the axioms, and so are the operations of *Sets, Relations, and Cardinality*, Section 1.2 on sets and on sets of sets: $\cup$, $\cap$, $\setminus$, the power set (Z5), $\bigcup \mathcal{S}$, and $\bigcap \mathcal{S}$ for non-empty $\mathcal{S}$. Unions and intersections of an indexed family $(A_i)_{i \in I}$ need the family as a function, which needs ordered pairs; Section 1.2 returns to them. What is not yet built is an **ordered** pair; the warning above says why $\{a, b\}$ will not do, and Section 1.2 builds one.

1.1.4 Exercises

1.1.4.1 A. Check your understanding

Exercise. A1

(a) For each of the sets $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, $\mathcal{P}(\emptyset)$, $\bigcup\{\emptyset\}$, and $\{x \in A : x \notin x\}$ for a given set A, name the axioms that produce it.

(b) True or false: for all sets a and b, $\{a\} = \{b\}$ implies $a = b$. Justify your answer.

(c) Prove that $\bigcup\{a\} = a$ for every set a.

(d) Is there a set T such that $\{x\} \in T$ for **every** set x? Justify your answer.

Solution.

(a) $\{\emptyset\}$: (Z2) for $\emptyset$, then (Z3) with both inputs $\emptyset$. $\{\emptyset, \{\emptyset\}\}$: (Z2), then (Z3) twice, once for $\{\emptyset\}$ and once with the inputs $\emptyset$ and $\{\emptyset\}$. $\mathcal{P}(\emptyset)$: (Z2), then (Z5); it is $\{\emptyset\}$, since the only subset of $\emptyset$ is $\emptyset$. $\bigcup\{\emptyset\}$: (Z2), (Z3), then (Z4); it is $\emptyset$. $\{x \in A : x \notin x\}$: (Z6) with ambient set A and the property $\neg(x \in x)$. Uniqueness, in every case, is (Z1).

(b) True. Apply Proposition 1.1.2 (d) with $c = d = b$: $\{a\} = \{b, b\}$ gives $a = b$ (and $b = b$).

(c) For every x: $x \in \bigcup\{a\}$ if and only if $x \in S$ for some $S \in \{a\}$. The only $S \in \{a\}$ is a (Proposition 1.1.2 (b)), so this holds if and only if $x \in a$. The two sets have the same elements, so $\bigcup\{a\} = a$ by (Z1).

(d) No. Suppose T were such a set. By (Z4), $\bigcup T$ is a set. Let x be any set. Then $x \in \{x\}$ and $\{x\} \in T$, so $x \in \bigcup T$. So every set is an element of $\bigcup T$, which contradicts the second claim of Sets, Relations, and Cardinality, Theorem 1.1.4. Hence no such T exists.

1.1.4.2 B. Practice

Exercise. B1: Power sets and unions

(a) Compute $\mathcal{P}(\mathcal{P}(\emptyset))$.

(b) Prove that $\bigcup \mathcal{P}(X) = X$ for every set X.

Solution.

(a) $\mathcal{P}(\emptyset) = \{\emptyset\}$, since the only subset of $\emptyset$ is $\emptyset$. So $\mathcal{P}(\mathcal{P}(\emptyset)) = \mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$, as computed in the list of first sets after Axiom 1.1.1.

(b) We show both inclusions and apply Sets, Relations, and Cardinality, Proposition 1.1.1 (whose proof is (Z1) and logic).

($\subseteq$) Let $x \in \bigcup \mathcal{P}(X)$. Then $x \in S$ for some $S \in \mathcal{P}(X)$, that is, for some $S \subseteq X$. Hence $x \in X$.

($\supseteq$) Let $x \in X$. The set $\{x\}$ is a subset of X, since its only element is x (Proposition 1.1.2 (b)). So $\{x\} \in \mathcal{P}(X)$, and $x \in \{x\}$; hence $x \in \bigcup \mathcal{P}(X)$.

Exercise. B2: Intersections of sets of sets

(a) Let $\mathcal{S} = \{X, Y\}$ with $X = \{\emptyset, \{\emptyset\}\}$ and $Y = \{\{\emptyset\}, \{\{\emptyset\}\}\}$. Compute $\bigcap \mathcal{S}$ twice: once with $A_0 = X$ in Proposition 1.1.2 (c), once with $A_0 = Y$.

(b) Explain in one or two sentences why the two computations had to agree.

(c) Explain why the definition of $\bigcap \mathcal{S}$ makes no sense for $\mathcal{S} = \emptyset$.

Solution.

(a) With $A_0 = X$: of the two elements of X, $\emptyset$ is not in Y (the elements of Y are $\{\emptyset\}$ and $\{\{\emptyset\}\}$, both non-empty), and $\{\emptyset\}$ is in Y and in X. So $\bigcap \mathcal{S} = \{\{\emptyset\}\}$. With $A_0 = Y$: $\{\emptyset\}$ lies in X and in Y; $\{\{\emptyset\}\}$ does not lie in X, since it differs from $\emptyset$ (it has an element) and from $\{\emptyset\}$ (its element $\{\emptyset\}$ is not $\emptyset$). So again $\bigcap \mathcal{S} = \{\{\emptyset\}\}$.

(b) An element of either set must lie in **every** member of $\mathcal{S}$, and that condition already forces it into A_0, because A_0 is a member. So the condition "$x \in A_0$" adds nothing, and both computations describe the same elements.

(c) There is no $A_0 \in \emptyset$ to cut the set out of. Without an ambient set, the condition "$x \in S$ for every $S \in \emptyset$" holds for every set x, vacuously; so $\bigcap \emptyset$ would be a set of which every set is an element, and there is no such set (Sets, Relations, and Cardinality, Theorem 1.1.4).

1.1.4.3 C. Going deeper

Exercise. C1: What the empty-set axiom really says; intersections as greatest lower bounds

(a) Show that (Z2) follows from (Z6), once **some** set A is known to exist.

(b) Let $\mathcal{S}$ be a non-empty set. Show that $\bigcap \mathcal{S} \subseteq S$ for every $S \in \mathcal{S}$, and that every set C with $C \subseteq S$ for all $S \in \mathcal{S}$ satisfies $C \subseteq \bigcap \mathcal{S}$. Deduce that $\bigcap \{A, B\} = A \cap B$ for all sets A, B.

Solution (Hint 1). For (a), cut a set out of A by a property that nothing has. For (b), unwind the membership condition of $\bigcap \mathcal{S}$ from Proposition 1.1.2 (c).

Solution (Hint 2). For (a): consider $\{x \in A : x \neq x\}$. For (b): let $x \in \bigcap \mathcal{S}$ and $S \in \mathcal{S}$; what does the condition say at this S?

Solution.

(a) Let A be a set. By (Z6) with the property $\neg(x = x)$, the set $E \coloneqq \{x \in A : x \neq x\}$ exists. If $x \in E$, then $x \neq x$, which contradicts $x = x$ (=I, Logic and Proof, Definition 5.7.1). So E has no elements, and (Z2) holds. Thus in the presence of (Z6), (Z2) says only "there is a set". The Axiom of Infinity of Section 1.3 also says that a set exists, so it could replace (Z2) altogether; that is how the full system is often stated, with the clause "$\emptyset \in I$" of Section 1.3 read as "I has an element with no elements", so that the symbol $\emptyset$ is not needed before (Z2) is. (In the logic of *Logic and Proof*, Chapter 4 even "there is a set" can be derived: =I gives $x = x$, and $\exists$I gives $\exists x\,(x = x)$, because every domain is non-empty, *Logic and Proof*, Section 4.11. We keep (Z2) anyway, so that the axioms say plainly where the first set comes from.)

(b) Let $x \in \bigcap \mathcal{S}$ and $S \in \mathcal{S}$. By the membership condition of Proposition 1.1.2 (c), x lies in every member of $\mathcal{S}$, in particular in S. So $\bigcap \mathcal{S} \subseteq S$.

Now let C be a set with $C \subseteq S$ for every $S \in \mathcal{S}$, and let $x \in C$. Then $x \in S$ for every $S \in \mathcal{S}$. Choose $A_0 \in \mathcal{S}$ (possible, as $\mathcal{S}$ is non-empty); then $x \in A_0$ as well, and so x satisfies the defining condition of $\bigcap \mathcal{S}$. Hence $C \subseteq \bigcap \mathcal{S}$.

For $\mathcal{S} = \{A, B\}$, which is non-empty, the members are exactly A and B (Proposition 1.1.2 (b)). By the first part with $S = A$ and with $S = B$, $\bigcap\{A, B\} \subseteq A$ and $\bigcap\{A, B\} \subseteq B$, so $\bigcap\{A, B\} \subseteq A \cap B$. Conversely, $A \cap B \subseteq A$ and $A \cap B \subseteq B$, so $A \cap B \subseteq S$ for every $S \in \{A, B\}$, and the second part gives $A \cap B \subseteq \bigcap\{A, B\}$. Hence $\bigcap\{A, B\} = A \cap B$ (Sets, Relations, and Cardinality, Proposition 1.1.1).

In words, (b) says that $\bigcap \mathcal{S}$ is the **largest** set contained in every member of $\mathcal{S}$: it is contained in each of them, and it contains every other set that is.

Marking scheme. (10 points) - 3: (a) produces a set with no elements by (Z6) from a given set, with a property that no set has, and proves it has no elements. - 2: (b) shows $\bigcap \mathcal{S} \subseteq S$ from the membership condition, at the given S. - 3: (b) shows $C \subseteq \bigcap \mathcal{S}$, placing x in the ambient member A_0 as well as in every S. - 2: deduces $\bigcap\{A, B\} = A \cap B$ using that the members of $\{A, B\}$ are exactly A and B: by double inclusion, $\bigcap\{A, B\} \subseteq A \cap B$ from the first half of (b) at $S = A$ and $S = B$, and $A \cap B \subseteq \bigcap\{A, B\}$ from the second half with $C = A \cap B$, then Sets, Relations, and Cardinality, Proposition 1.1.1 (or (Z1)). A solution that compares membership conditions directly earns the same 2.

Fatal errors (0 for the step): using a set of all sets, or $\bigcap \emptyset$; in (a), assuming $\emptyset$ exists in order to prove (Z2).

1.2 Ordered Pairs Built from Sets

Sets, Relations, and Cardinality, Section 1.3 introduced the ordered pair (Sets, Relations, and Cardinality, Definition 1.3.1) as an object with one property, the characteristic property, and said that such an object could be built from sets. Section 1.1 has only sets to build with, and its pair $\{a, b\}$ forgets which element came first. This section repairs that with one idea: mark the

first coordinate by putting it in a set of its own. It proves that the result has the characteristic property, so that every pair of *Sets, Relations, and Cardinality*, Chapters 1 to 3 can be read as a set, and then recovers the product $A \times B$ and the working rules (W1)–(W5) of *Sets, Relations, and Cardinality*, Section 1.1 from the axioms, in the form proved here: (W2) for lists written out in the text, (W3) for families of subsets of a given set, and (W5) for pairs. The rest of (W2) and (W5) comes in Section 1.6; (W3) for families with no ambient set needs the axiom of Replacement.

1.2.1 Order from sets

The problem. The set $\{a, b\}$ of Proposition 1.1.2 (b) is symmetric: $\{a, b\} = \{b, a\}$. A pair must not be. We want a set, built from a and b by the axioms, from which we can read off **which** of the two came first, and the second coordinate after that. Here is one way to see what to do. In $\{a, b\}$ the two elements are on an equal footing; we must give a a role that b does not have. Put a into **every** member of the pair, and b into one member only.

The first coordinate is the one in every member.

Definition 1.2.1. Ordered Pair as a Set

For sets a and b, the **ordered pair** of a and b is the set

$$(a, b) := \{\{a\}, \{a, b\}\}.$$

Its **first coordinate** is a and its **second coordinate** is b.

The pair is a set of the axioms: (Z3) gives $\{a\} = \{a, a\}$ and $\{a, b\}$, and (Z3) once more gives the set whose elements are these two (Proposition 1.1.2 (b)). In words: the members of (a, b) are the sets $\{a\}$ and $\{a, b\}$, which coincide when $a = b$. Sets of this form are called **Kuratowski pairs**. The first coordinate lies in both members; the second lies in the member $\{a, b\}$, and in $\{a\}$ only if it equals a.

Examples, simplest first.

- $(a, a) = \{\{a\}, \{a, a\}\} = \{\{a\}, \{a\}\} = \{\{a\}\}$. This is the degenerate case: the two members coincide, and the pair has **one** member. It matters, because a proof that speaks of "the two members of a pair" silently excludes it.
- $(\{\emptyset\}, \emptyset) = \{\{\{\emptyset\}\}, \{\{\emptyset\}, \emptyset\}\}$. Its two members are different: $\emptyset$ lies in the second and not in the first, whose only element is $\{\emptyset\}$.
- Swap the coordinates: $(\emptyset, \{\emptyset\}) = \{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$. The second member is the same as before, $\{\emptyset, \{\emptyset\}\}$; the singleton member has changed from $\{\{\emptyset\}\}$ to $\{\emptyset\}$. So $(\{\emptyset\}, \emptyset) \neq (\emptyset, \{\emptyset\})$, and the singleton member is what records the order.

Why this definition: shorter encodings fail. It is natural to try something simpler first. The next example shows three simpler sets, built from a and b with a written first, none of which is a pair.

Example. Three Encodings That Lose Information

Consider the claim: *any set built from a and b in which a is mentioned first is an ordered*

pair, that is, has the characteristic property. Test it on the three encodings

$$\text{(i) } \{a, b\}, \qquad \text{(ii) } \{\{a\}, \{b\}\}, \qquad \text{(iii) } \{a, \{b\}\}.$$

Solution. The claim is **false**, and each encoding has its own counterexample.

(i) With $a = \emptyset$, $b = \{\emptyset\}$ and with the coordinates swapped, the encoding gives $\{\emptyset, \{\emptyset\}\}$ both times, although $\emptyset \neq \{\emptyset\}$. The order is lost.

(ii) The same two inputs give $\{\{\emptyset\}, \{\{\emptyset\}\}\}$ both times: the encoding is $\{\{a\}, \{b\}\} = \{\{b\}, \{a\}\}$, again symmetric in a and b.

(iii) This one is not symmetric, and it fails in a less obvious way. Take $a = \{\{\emptyset\}\}$, $b = \emptyset$, and $c = \{\emptyset\}$, $d = \{\emptyset\}$. Then

$$\{a, \{b\}\} = \{\{\{\emptyset\}\}, \{\emptyset\}\}, \qquad \{c, \{d\}\} = \{\{\emptyset\}, \{\{\emptyset\}\}\},$$

the same set, by (Z1). But $a \neq c$: the only element of $a = \{\{\emptyset\}\}$ is $\{\emptyset\}$, which is not $\emptyset$, so $\emptyset \in c$ and $\emptyset \notin a$. So equal encodings do not force equal first coordinates. The trouble is that $\{b\}$ can coincide with a first coordinate, and the encoding cannot tell which of its elements was meant as "a".

The definition avoids both failures. Both of its members are built from a, so neither can be mistaken for a lone coordinate, and the member $\{a\}$ singles out a. Another encoding, often called the Wiener pair, uses $\emptyset$ as a marker; exercise C3 asks you to show that it has the characteristic property.

The lesson of (iii) is a habit: a proposed encoding is tested against the characteristic property with **all** inputs, including inputs that are themselves made of the sets being encoded.

Warning. **"The smaller member" names nothing when $a = b$.** It is tempting to read off the first coordinate as "the element of the smaller member" and the second as "the new element of the larger member". When $a = b$, the pair $(a, a) = \{\{a\}\}$ has a **single** member, and there is no smaller and larger. An argument that compares the two members of (a, b) must treat $a = b$ as a case of its own.

1.2.2 The characteristic property

Here is what makes the definition the right one: the set (a, b) determines a and b, each in its place.

Theorem 1.2.1. Characteristic Property of Pairs

For all sets a, b, c, d,

$$(a, b) = (c, d) \iff a = c \text{ and } b = d.$$

Try it first (ten minutes): given $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$, get $a = c$ first; then ask what changes when $a = b$.

Idea. The attempt goes well at first. The member $\{a\}$ of the left side is a member of the right side, so $\{a\} = \{c\}$ or $\{a\} = \{c, d\}$, and either way Proposition 1.1.2 (d) gives $a = c$. Then $\{a, b\}$ must be one of $\{c\}$, $\{c, d\}$, and the natural move is to say "it is $\{c, d\}$, the larger one", giving $\{a, b\} = \{a, d\}$ and so $b = d$. That move is where the attempt runs into the warning above. If $a = b$, then $\{a, b\} = \{a\}$ and it may well equal $\{c\}$; if $c = d$, the right side has a single member. So the proof splits: when $a = b$, the left side is $\{\{a\}\}$, and both members of the right side must equal $\{a\}$; when $a \neq b$, both sides have two different members, and the matching above goes through.

Before reading on: where is $a \neq b$ used?

Proof. ($\Leftarrow$) If $a = c$ and $b = d$, then $(a, b) = (c, d)$ by =E.

($\Rightarrow$) Suppose $(a, b) = (c, d)$, that is, $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. Either $a = b$ or $a \neq b$, and we take the cases in turn.

Case 1: $a = b$. Then $\{a, b\} = \{a\}$, so $(a, b) = \{\{a\}\}$. The members $\{c\}$ and $\{c, d\}$ of (c, d) are members of $\{\{a\}\}$, so $\{c\} = \{a\}$ and $\{c, d\} = \{a\}$. By Proposition 1.1.2 (d), the second equation gives $a = c$ and $c = d$. Hence $a = c$, and $b = a = c = d$.

Case 2: $a \neq b$. First, $c \neq d$. Indeed, if $c = d$, then $(c, d) = \{\{c\}\}$ as in Case 1, so the member $\{a, b\}$ of (a, b) equals $\{c\}$, and Proposition 1.1.2 (d) gives $c = a$ and $a = b$, which contradicts $a \neq b$.

Next, $\{a\} \in (a, b) = (c, d)$, so $\{a\} = \{c\}$ or $\{a\} = \{c, d\}$. The second would give $c = d$ by Proposition 1.1.2 (d), which we have excluded. So $\{a\} = \{c\}$, and $a = c$ by the same proposition.

Finally, $\{a, b\} \in (c, d)$, so $\{a, b\} = \{c\}$ or $\{a, b\} = \{c, d\}$. The first would give $a = c$ and $c = b$ by Proposition 1.1.2 (d), hence $a = b$, which is excluded. So $\{a, b\} = \{c, d\} = \{a, d\}$, using $c = a$. By Proposition 1.1.2 (d), $(a = a \wedge b = d) \vee (a = d \wedge b = a)$; in the first disjunct $b = d$, and in the second $b = a = d$. So $b = d$.

In both cases $a = c$ and $b = d$. This proves the theorem. ∎

So $a \neq b$ is used twice in Case 2: to exclude $c = d$, and to exclude $\{a, b\} = \{c\}$. Once both are excluded, each side has exactly two different members, and they can be matched.

Alternatively, the coordinates can be read off a pair $p = (a, b)$ by two set operations. The members of p are $\{a\}$ and $\{a, b\}$ (Proposition 1.1.2 (b)). An object lies in both if and only if it equals a, since it must lie in $\{a\}$, and $a \in \{a, b\}$; so $\bigcap p = \{a\}$ (Proposition 1.1.2 (c)). And $\bigcup p = \{a\} \cup \{a, b\} = \{a, b\}$. So a is the unique element of $\bigcap p$. And $\bigcup p \setminus \bigcap p = \{a, b\} \setminus \{a\}$ is $\{b\}$ if $b \neq a$, and $\emptyset$ if $b = a$; so b is the unique element of $\bigcup p \setminus \bigcap p$ when that set is non-empty, and $b = a$ when it is empty. Both coordinates are thus determined by the set p alone, which is the characteristic property again. The case split has not gone away here: it is the empty and non-empty cases of $\bigcup p \setminus \bigcap p$. Comparing the unions of two pairs directly avoids it: from $\{a, b\} = \{c, d\}$ with $c = a$, Proposition 1.1.2 (d) gives $b = d$ in both of its disjuncts.

What this pays. *Sets, Relations, and Cardinality*, Section 1.3 took the ordered pair as given, with the characteristic property, and said after Sets, Relations, and Cardinality, Definition 1.3.1

that nothing else about pairs is used in *Sets, Relations, and Cardinality*, Chapter 1, apart from one working convention. From here on, the pair (a, b) of Sets, Relations, and Cardinality, Definition 1.3.1 **is** the set of Definition 1.2.1, and the property it was assumed to have is Theorem 1.2.1. Everything built on pairs becomes a set built by the axioms: a relation from A to B is a subset of $A \times B$ (Sets, Relations, and Cardinality, Definition 1.3.3), and a function $X \to Y$ is the pair (G, Y) of its graph and codomain (Sets, Relations, and Cardinality, Definition 1.4.1), now a Kuratowski pair of two sets. The working convention said that **a pair is not a number**, and that a triple is not a pair. In the picture of this chapter numbers and tuples will be sets too, so both clauses become statements to be checked: Section 1.4 checks the first for the natural numbers (Proposition 1.4.3), and Section 1.6 checks the second (Proposition 1.6.5 (c)).

1.2.3 Products, and the working rules recovered

The working rule (W5) said that the pairs with coordinates in given sets form a set. With pairs built from sets, this is now something to prove.

Proposition 1.2.2. The Product Exists

For all sets A and B there is **exactly one** set $A \times B$ whose elements are exactly the pairs (a, b) with $a \in A$ and $b \in B$:

$$p \in A \times B \iff \exists a \in A\ \exists b \in B\ \ p = (a, b).$$

Idea. Separation can only cut a set out of a set we already have, so we first need **some** set containing every pair (a, b) with $a \in A$ and $b \in B$. The members $\{a\}$ and $\{a, b\}$ of such a pair are subsets of $A \cup B$, so the pair is a set of subsets of $A \cup B$: an element of $\mathcal{P}(\mathcal{P}(A \cup B))$. Cut $A \times B$ out of that.

Proof. Let $U := A \cup B$, a set by Proposition 1.1.2 (c), and let $W := \mathcal{P}(\mathcal{P}(U))$, a set by (Z5) twice. Let $a \in A$ and $b \in B$. Then $a, b \in U$, so $\{a\} \subseteq U$ and $\{a, b\} \subseteq U$; that is, $\{a\} \in \mathcal{P}(U)$ and $\{a, b\} \in \mathcal{P}(U)$. Hence $(a, b) = \{\{a\}, \{a, b\}\} \subseteq \mathcal{P}(U)$, that is, $(a, b) \in W$.

By (Z6) with ambient set W and parameters A, B, let

$$A \times B := \{p \in W : \exists a \in A\ \exists b \in B\ \ p = (a, b)\}.$$

(The property is admissible: "$p = (a, b)$" unwinds into $\in$ and $=$, as Section 1.1 explained for defined symbols.) If $p \in A \times B$, then $p = (a, b)$ for some $a \in A$, $b \in B$. Conversely, if $p = (a, b)$ with $a \in A$ and $b \in B$, then $p \in W$ by the first paragraph, so $p \in A \times B$. This is the displayed membership condition. Any two sets with this condition are equal by (Z1).

■

The product of Sets, Relations, and Cardinality, Definition 1.3.2 has the same membership condition, so it is this set; and $X^2 := X \times X$ is a set too. The construction never needed a "set of all pairs": every pair it collects lies in the set W, which depends on A and B.

The working rules of *Sets, Relations, and Cardinality*, Section 1.1 can now be checked against the axioms one by one.

Proposition 1.2.3. The Working Rules from the Axioms

The axioms (Z1)–(Z6) give the working rules (W1)–(W5) of *Sets, Relations, and Cardinality*, Section 1.1 in the following form.

(W1) For every set A and every admissible property $P(x)$ (one whose unwinding is a formula of the language, Section 1.1), $\{x \in A : P(x)\}$ is a set.

(W2) For each list $a_1, \ldots, a_n$ written out in the text, entry by entry, there is a set $\{a_1, \ldots, a_n\}$ whose elements are exactly $a_1, \ldots, a_n$.

(W3) For sets A and B, $A \cup B$ is a set. For a set X and a family $(A_i)_{i \in I}$ of subsets of X, given as a function $I \to \mathcal{P}(X)$ (Sets, Relations, and Cardinality, Definition 1.2.3), the union $\bigcup_{i \in I} A_i$ and the intersection $\bigcap_{i \in I} A_i$ of Sets, Relations, and Cardinality, Definition 1.2.4 are sets.

(W4) For every set X, $\mathcal{P}(X)$ is a set.

(W5) For sets A and B, $A \times B$ is a set.

Idea. Three of the rules are already in hand. For a written list, add one entry at a time with $\cup$. For a union of a family of subsets of X, the ambient set X is what lets (Z6) apply.

Proof. (W1) is (Z6), applied to the unwinding of $P(x)$, with the sets P mentions as parameters; (W4) is (Z5), and (W5) is Proposition 1.2.2.

(W2) We build the set one entry at a time. The first set is $\{a_1\}$, by (Z3). If $\{a_1, \ldots, a_k\}$ has been built and the list has a further entry a_{k+1}, let

$$\{a_1, \ldots, a_{k+1}\} \coloneqq \{a_1, \ldots, a_k\} \cup \{a_{k+1}\},$$

a set by Proposition 1.1.2 (c). Its elements are those of $\{a_1, \ldots, a_k\}$ together with a_{k+1}, that is, the x with $x = a_1 \vee \cdots \vee x = a_{k+1}$. A list written out in the text has a definite number of entries, and the step is applied once for each entry after the first.

(W3) $A \cup B$ is a set by Proposition 1.1.2 (c). Now let $f\colon I \to \mathcal{P}(X)$ be the function with $f(i) = A_i$, and let G be its graph. By (Z6) with ambient set X and parameters I, G,

$$V \coloneqq \{x \in X : \exists i \in I \;\; x \in f(i)\}$$

is a set; "$x \in f(i)$" unwinds to $\exists S\,((i, S) \in G \wedge x \in S)$. Every element of every A_i lies in X, since $A_i \subseteq X$; so $x \in V$ if and only if $x \in A_i$ for some $i \in I$, the membership condition of Sets, Relations, and Cardinality, Definition 1.2.4 (a). Hence V is the union. The intersection of Sets, Relations, and Cardinality, Definition 1.2.4 (b) is defined there by a set-builder inside X, so it exists by (Z6).

■

The proof of (W2) deserves a second look, because of what it does **not** prove. It is not a proof "for every $n \in \mathbb{N}$": we have no $\mathbb{N}$ yet, so the number of entries cannot be a variable. It is a recipe which, applied to any list actually written down, produces its set in as many steps as the list has entries. That is all (W2) is used for in *Sets, Relations, and Cardinality*, Section 1.1, but Chapters 2 and 3 of that volume also write $\{x_1, \ldots, x_n\}$ for a **variable** n, for a function

x defined on $\{1, \ldots, n\}$ (Sets, Relations, and Cardinality, Proposition 2.3.8, Sets, Relations, and Cardinality, Definition 3.1.2). That second half of (W2) needs the natural numbers and $\{1, \ldots, n\}$ as sets; it is part (d) of a proposition in Section 1.6, which also builds n-tuples and so gives (W5) for them.

The rest of the list is paid elsewhere. A family of sets with **no** ambient set X needs the axiom of Replacement, which belongs to the volume on set theory. And (W6), "$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ are sets", becomes a series of constructions: $\mathbb{N}$ in Section 1.3, $\mathbb{Z}$ and $\mathbb{Q}$ in Chapter 2, and $\mathbb{R}$ later in this volume, from Cauchy sequences of rationals.

Quick check. Write $(\emptyset, \emptyset)$ and $(\emptyset, \{\emptyset\})$ as sets, using only braces and $\emptyset$.

Solution. $(\emptyset, \emptyset) = \{\{\emptyset\}, \{\emptyset, \emptyset\}\} = \{\{\emptyset\}\}$, a set with one member, as for every pair (a, a). And $(\emptyset, \{\emptyset\}) = \{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, with two different members.

1.2.4 Exercises

1.2.4.1 A. Check your understanding

Exercise. A1

(a) In the proof of Theorem 1.2.1, where is the hypothesis $a \neq b$ of Case 2 used, and what replaces the argument when $a = b$?

(b) Run the proof of Theorem 1.2.1 on the equation $(\{\emptyset\}, \emptyset) = (c, d)$: which case applies, and what does each step give?

(c) Which axioms make (a, b) a set?

(d) State the idea of the proof of Theorem 1.2.1 in two sentences.

Solution.

(a) Twice: to show $c \neq d$ (if $c = d$, the member $\{a, b\}$ would equal $\{c\}$, forcing $a = b$), and to rule out $\{a, b\} = \{c\}$ (which would force $a = b$). When $a = b$, the pair $(a, b) = \{\{a\}\}$ has one member, so both members of (c, d) must equal $\{a\}$, and $\{c, d\} = \{a\}$ gives $c = d = a$ by Proposition 1.1.2 (d).

(b) Here $a = \{\emptyset\}$ and $b = \emptyset$, and $a \neq b$ because $\emptyset \in a$ and $\emptyset \notin b$; so Case 2 applies. Step 1: $c \neq d$. Step 2: $\{a\} = \{\{\emptyset\}\}$ is a member of (c, d), and it cannot be $\{c, d\}$, so $\{\{\emptyset\}\} = \{c\}$ and $c = \{\emptyset\}$. Step 3: $\{a, b\} = \{\{\emptyset\}, \emptyset\}$ is a member of (c, d) and is not $\{c\}$, so $\{\{\emptyset\}, \emptyset\} = \{c, d\} = \{\{\emptyset\}, d\}$, which gives $d = \emptyset$. So $(c, d) = (\{\emptyset\}, \emptyset)$ forces $c = \{\emptyset\}$ and $d = \emptyset$.

(c) (Z3), three times: once for $\{a\} = \{a, a\}$, once for $\{a, b\}$, and once for the set with these two elements. Uniqueness of each is (Z1).

(d) The singleton member $\{a\}$ of the pair must be one of the members of the other pair, which pins down the first coordinate. When $a \neq b$ the remaining members must match and give the second coordinate; when $a = b$ the pair has a single member, and that case is handled on its own.

1.2.4.2 B. Practice

Exercise. B1: Coordinates from unions and intersections

Let $p = (a, b)$. Compute $\bigcap p$ and $\bigcup p$, and show that, when $a \neq b$, the set $\bigcup p \setminus \bigcap p$ has exactly one element, namely b.

Solution. The members of p are $\{a\}$ and $\{a, b\}$. An object lies in both if and only if it equals a (it is in $\{a\}$), and a is in $\{a, b\}$; so $\bigcap p = \{a\}$. An object lies in at least one member if and only if it equals a or b; so $\bigcup p = \{a, b\}$.

Now let $a \neq b$. Then $x \in \bigcup p \setminus \bigcap p$ if and only if ($x = a$ or $x = b$) and $x \neq a$, that is, if and only if $x = b$ and $x \neq a$. Since $b \neq a$, the element b satisfies this, and any x satisfying it equals b. Hence $\bigcup p \setminus \bigcap p = \{b\}$.

Exercise. B2: Nested pairs

Write $((a, b), c)$ for the pair whose first coordinate is the pair (a, b).

(a) Prove that $((a, b), c) = ((a', b'), c')$ if and only if $a = a'$, $b = b'$ and $c = c'$.

(b) Determine whether $((\emptyset, \emptyset), \emptyset) = (\emptyset, (\emptyset, \emptyset))$. Justify your answer.

Solution.

(a) ($\Leftarrow$) by =E. ($\Rightarrow$) By Theorem 1.2.1 applied to the outer pairs, $(a, b) = (a', b')$ and $c = c'$. By Theorem 1.2.1 again, $(a, b) = (a', b')$ gives $a = a'$ and $b = b'$.

(b) No. Both sides are pairs, so by Theorem 1.2.1 they are equal only if their first coordinates are equal. The first coordinates are $(\emptyset, \emptyset) = \{\{\emptyset\}\}$ and $\emptyset$, and $\{\{\emptyset\}\} \neq \emptyset$, since $\{\emptyset\} \in \{\{\emptyset\}\}$. Hence the two pairs are different.

1.2.4.3 C. Going deeper

Exercise. C1: Exit check: Kuratowski pairs

Prove that for all sets a, b, c, d, $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ if and only if $a = c$ and $b = d$. You may use Proposition 1.1.2.

Solution. This is Theorem 1.2.1; its proof above is a complete answer. In outline: ($\Leftarrow$) is substitution. For ($\Rightarrow$), if $a = b$, the left side is $\{\{a\}\}$, so $\{c, d\} = \{a\}$, giving $c = d = a = b$. If $a \neq b$, then $c \neq d$ (else the left side would have a single member); $\{a\}$ equals $\{c\}$ or $\{c, d\}$, the second being excluded by $c \neq d$, so $a = c$; and $\{a, b\}$ equals $\{c\}$ or $\{c, d\}$, the first being excluded by $a \neq b$, so $\{a, b\} = \{a, d\}$ and $b = d$.

Marking scheme. (10 points) - 1: proves ($\Leftarrow$) by substitution. - 2: in the case $a = b$, identifies $(a, b) = \{\{a\}\}$ and concludes $c = d = a$, hence $a = c$ and $b = d$. - 3: in the case $a \neq b$, shows $\{a\} = \{c\}$ or $\{a\} = \{c, d\}$, excludes the second (via $c \neq d$, itself proved), and concludes $a = c$. - 3: obtains $b = d$ from the remaining members, excluding $\{a, b\} = \{c\}$ and

justifying $\{a,b\} = \{a,d\} \Rightarrow b = d$. - 1: joins the two cases into the conclusion.

Fatal errors (0 for the step): treating $\{a,b\}$ as ordered, or matching "smaller with smaller and larger with larger" without justification; omitting the case $a = b$ in an argument that matches the members of the two pairs (or reads b off $\bigcup p \setminus \bigcap p$).

Second route (through $\bigcap p$ and $\bigcup p$, as in the "Alternatively" paragraph): reading a off $\bigcap p = \{a\}$ earns the 3 points for $a = c$; reading b off $\bigcup p \setminus \bigcap p$, with its empty case $a = b$ treated separately, earns the 3 points for $b = d$ and the 2 points of the case $a = b$.

Third route (comparing $\bigcap$ and $\bigcup$ of the two pairs): showing $\bigcap(x,y) = \{x\}$ and $\bigcup(x,y) = \{x,y\}$ for every pair (no case split is needed), then getting $a = c$ from $\{a\} = \{c\}$ and $b = d$ from $\{a,b\} = \{c,d\} = \{a,d\}$ by Proposition 1.1.2 (d), earns the 3 + 3 points and also the 2 + 1 points of the case structure; with ($\Leftarrow$) this is 10.

Exercise. C2: Find the error

Here is an argument for Theorem 1.2.1, direction ($\Rightarrow$).

> (1) Let $(a,b) = (c,d)$. (2) Matching the members of $\{\{a\},\{a,b\}\}$ and $\{\{c\},\{c,d\}\}$ in the order written, $\{a\} = \{c\}$ and $\{a,b\} = \{c,d\}$. (3) From $\{a\} = \{c\}$, $a = c$. (4) Then $\{a,b\} = \{a,d\}$, and so $b = d$.

Is the conclusion true? Does the argument prove it? If not, name the first line that fails, the rule it breaks, and give a counterexample to that step.

Solution. The conclusion is **true**: it is Theorem 1.2.1.

The argument does not prove it. The first failing line is (2). It uses the rule "if $\{P,Q\} = \{P',Q'\}$, then $P = P'$ and $Q = Q'$", which treats a set with two members as if it were ordered. No result of Section 1.1 gives this rule: Proposition 1.1.2 (d) allows the crossed matching $P = Q'$ and $Q = P'$ as well. A counterexample to the step: take $P = Q' = \emptyset$ and $Q = P' = \{\emptyset\}$. Then $\{P,Q\} = \{\emptyset,\{\emptyset\}\} = \{P',Q'\}$ by (Z1), but $P \neq P'$, since $\emptyset \in \{\emptyset\}$ and $\emptyset \notin \emptyset$.

For pairs the two equations of line (2) happen to be true, since they follow from the theorem. But showing that the members of (a,b) and (c,d) match in this way is the work of the proof: Case 1 handles $a = b$, where $(a,b) = \{\{a\}\}$ has a single member, and Case 2 rules out $c = d$ and $\{a,b\} = \{c\}$ before matching. So line (2) assumes what is to be proved. This is entry (N9) of Logic and Proof, Example (What is not allowed): line (2) cites a general rule that is false, and for pairs the matching it asserts is the goal itself.

A second, lesser point concerns line (4). "$\{a,b\} = \{a,d\}$, so $b = d$" is true for all a, b, d, but it needs a reason: Proposition 1.1.2 (d) with $c = a$. The disjunct $a = a \wedge b = d$ gives $b = d$ directly, and the disjunct $a = d \wedge b = a$ gives $b = a = d$. No hypothesis $b \neq a$ is needed, and an answer saying so is correct.

Repair: split into the cases $a = b$ and $a \neq b$, as in the proof of the theorem.

Exercise. C3: Wiener's encoding

For sets a and b let $W(a,b) := \{\{\{a\},\emptyset\},\{\{b\}\}\}$. Prove that $W(a,b) = W(c,d)$ if and only if $a = c$ and $b = d$.

Solution (Hint 1). The empty set is a marker. Which member of $W(a,b)$ has $\emptyset$ as an element?

Solution (Hint 2). Let $P := \{\{a\},\emptyset\}$ and $Q := \{\{b\}\}$. Show first that $\emptyset \in P$ and $\emptyset \notin Q$.

Solution. ($\Leftarrow$) by =E.

($\Rightarrow$) Let $P := \{\{a\},\emptyset\}$, $Q := \{\{b\}\}$, $P' := \{\{c\},\emptyset\}$, $Q' := \{\{d\}\}$, and suppose $\{P,Q\} = \{P',Q'\}$. We have $\emptyset \in P$ and $\emptyset \in P'$. On the other hand $\emptyset \notin Q$: the only element of Q is $\{b\}$, and $\{b\} \neq \emptyset$ because $b \in \{b\}$. Likewise $\emptyset \notin Q'$. So $P \neq Q$ and $P' \neq Q'$, and in each set of members exactly one member has $\emptyset$ as an element.

Since $P \in \{P',Q'\}$, $P = P'$ or $P = Q'$; the second is impossible, since $\emptyset \in P$ and $\emptyset \notin Q'$. So $P = P'$. Since $Q \in \{P',Q'\}$, $Q = P'$ or $Q = Q'$; the first is impossible, since $\emptyset \in P'$ and $\emptyset \notin Q$. So $Q = Q'$.

From $\{\{a\},\emptyset\} = \{\{c\},\emptyset\}$, Proposition 1.1.2 (d) gives $\{a\} = \{c\}$ (and $\emptyset = \emptyset$), or $\{a\} = \emptyset$ (and $\emptyset = \{c\}$). The second is impossible, as $a \in \{a\}$. So $\{a\} = \{c\}$, and $a = c$ by Proposition 1.1.2 (d) again. From $\{\{b\}\} = \{\{d\}\}$, the same proposition gives $\{b\} = \{d\}$ and then $b = d$.

Unlike Theorem 1.2.1, this proof needs no case $a = b$: the marker $\emptyset$ tells the two members apart whatever a and b are.

Marking scheme. (10 points) - 3: shows $\emptyset \in P$ and $\emptyset \notin Q$ (with the reason $\{b\} \neq \emptyset$), and likewise for P', Q'. - 3: matches $P = P'$ and $Q = Q'$, excluding the crossed matches by the marker. - 2: derives $a = c$ from $P = P'$, excluding $\{a\} = \emptyset$. - 2: derives $b = d$ from $Q = Q'$.

Fatal errors (0 for the step): assuming the members match in the order written; using Theorem 1.2.1, which is about a different set.

1.3 Successors, Inductive Sets and the Natural Numbers

Logic and Proof, Chapter 3 described the natural numbers as what we reach from 0 by adding 1, again and again, and nothing else; induction (Logic and Proof, Axiom 4.2.1) was the precise form of "nothing else". In this chapter every object is a set, so we need a set to play 0 and an operation on sets to play "add 1". This section chooses both, adds the one axiom that lets the repetition be completed into a set, and proves induction as a theorem about the set it produces.

1.3.1 Going one step further

The empty set is the only set we can name without first having another one, so it will be 0. For "add one" we want an operation that, started from $\emptyset$, keeps producing new sets, never returning to one already met. Putting a set inside itself as an element does that, as we shall see.

The successor of a set adds the set itself as an element.

Definition 1.3.1. Successor of a set

For **every** set x, the **successor** of x is the set

$$\mathrm{S}(x) := x \cup \{x\}.$$

In words: the elements of $\mathrm{S}(x)$ are the elements of x, together with x itself. So, for every set y,

$$y \in \mathrm{S}(x) \iff y \in x \quad \text{or} \quad y = x.$$

The set $\mathrm{S}(x)$ exists and is unique: $\{x\}$ is the set $\{x, x\}$ of Pairing (Z3), and $x \cup \{x\} = \bigcup\{x, \{x\}\}$ comes from Union (Z4), both by Proposition 1.1.2 (b), (c). Two facts are visible at once. First, $x \in \mathrm{S}(x)$ for every x. Second, $x \subseteq \mathrm{S}(x)$: the successor keeps every old element and adds x itself, which is a new element whenever $x \notin x$. The next section proves $n \notin n$ for every natural number n.

The letter is an upright S. *Sets, Relations, and Cardinality*, Chapter 1 used an italic S for a set in Sets, Relations, and Cardinality, Proposition 1.2.2, and we are about to use that proposition with this operation as its step; the two letters never stand for the same thing.

Start from $\emptyset$ and apply S repeatedly.

Example. The first natural numbers as sets

Define

$$0 := \emptyset, \qquad 1 := \mathrm{S}(0), \quad 2 := \mathrm{S}(1),$$
$$3 := \mathrm{S}(2), \quad 4 := \mathrm{S}(3).$$

Write each of $1, 2, 3, 4$ as a set of the earlier numerals, and then using only $\emptyset$ and braces. Check that $0, 1, 2, 3$ are four different sets.

Solution. Each step adds the previous set as a new element:

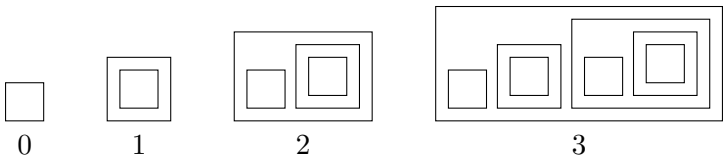
$$1 = \emptyset \cup \{0\} = \{0\} = \{\emptyset\},$$
$$2 = 1 \cup \{1\} = \{0, 1\} = \{\emptyset, \{\emptyset\}\},$$
$$3 = 2 \cup \{2\} = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},$$
$$4 = 3 \cup \{3\} = \{0, 1, 2, 3\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}.$$

Written with $\emptyset$ and braces only, 4 already needs eight pairs of braces and eight copies of $\emptyset$, which is why the numerals are worth having.

The four sets differ, by Extensionality (Z1): two sets are equal only if they have the same elements, so it is enough to find an element of one that is not an element of the other. The set 0 has no elements, while 1, 2 and 3 all have the element 0; so 0 differs from each of them. Next, $1 \in 2$ and $1 \in 3$, but $1 \notin 1$: the only element of $1 = \{0\}$ is 0, and $1 \neq 0$. So 1 differs

from 2 and from 3. Finally $2 \in 3$, but $2 \notin 2$: the elements of 2 are 0 and 1, and 2 differs from both, as just shown. So $2 \neq 3$.

Each number is the set of the numbers before it: $3 = \{0, 1, 2\}$, and $0 = \emptyset$ because nothing comes before 0. The picture draws a set as a box and its elements as the boxes directly inside it.



The box for 3 holds three boxes, which are copies of the boxes for 0, 1 and 2. In general a natural number n will turn out to be the set of the natural numbers less than n, so that $n = \{0, 1, \ldots, n-1\}$ for $n \neq 0$. That is a theorem, proved later in this chapter once the order on the natural numbers is defined. For now the pattern is only something we have seen in the first five cases.

From here to the end of the chapter, the numerals $0, 1, 2, 3, \ldots$ name these sets. They are not real numbers; where the chapter compares the two pictures, it says so.

Other choices of "add one" are possible. The rule $x \mapsto \{x\}$ also starts a list of different sets, $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$. With an axiom of infinity stated for that rule, it too would give a set satisfying Peano's axioms, but not by the proofs of the next section. What it loses is the pattern above: under it, the set playing 3 has one element, not three, and it does not contain the sets playing 0 and 1. The next section's proofs rest on that pattern, and one of its results fails outright: under $x \mapsto \{x\}$ the set playing 2 is $\{\{\emptyset\}\}$, which is the pair $(\emptyset, \emptyset)$ of Definition 1.2.1, while the next section shows that no pair is a natural number in our sense. The choice $x \cup \{x\}$ makes the order of the numbers into membership, which the chapter uses repeatedly.

Warning. The successor is defined for **every** set, not only for numerals, and its result need not look like a number. For $x = \{1\} = \{\{\emptyset\}\}$, the successor is $\mathrm{S}(x) = \{1, \{1\}\}$. It does not contain 0, while every numeral after 0 computed above does. A statement about $\mathrm{S}(x)$ that holds for the numerals may fail for other sets; this is why the results below are stated for elements of one particular set.

1.3.2 Inductive sets and the axiom of infinity

We want a set whose elements are exactly $0, 1, 2, \ldots$. The dots are the problem: Pairing and Union make each numeral, one at a time, and no axiom so far collects them all. The way out is to describe the property such a set must have, ask for a set with that property, and then cut it down to size.

An inductive set contains 0 *and never runs out of successors.*

Definition 1.3.2. Inductive set

A set I is **inductive** if $\emptyset \in I$ and, **for every** $x \in I$, also $\mathrm{S}(x) \in I$.

In words: the first clause gives the set a starting point, and the second says it is **closed** under

the successor, so that applying S never leads out of it. In symbols, with bounded quantifiers,

$$\emptyset \in I \ \wedge \ \forall x \in I \ \ \mathrm{S}(x) \in I.$$

The word "inductive" has another use in the analysis volume of this series, for subsets of the real numbers that contain the number 0 and are closed under $x \mapsto x + 1$. The shape is the same, but here the starting point is the set $\emptyset$, the step is S, and there is no ambient set of real numbers.

An inductive set contains all the numerals: $0 = \emptyset$ by the first clause, then $1 = \mathrm{S}(0)$ by the second, then $2 = \mathrm{S}(1)$, and so on, one application of the second clause per numeral. So an inductive set cannot be finite in any naive sense, and nothing we have so far produces one.

Non-examples by minimal change come cheaply. The set $4 = \{0, 1, 2, 3\}$ contains $\emptyset$ and the successors of 0, 1 and 2; only the last successor is missing, since $\mathrm{S}(3) = 4 \notin 4$: the set 4 has the element 3, and none of $0, 1, 2, 3$ does (0 has no elements, and the elements of 1, 2 and 3 are among $0, 1, 2$, each different from 3 by Example (The first natural numbers as sets)). And $\{1, 2, 3, \dots\}$, if we had it, would be closed under S but would miss the starting point. Both clauses do work.

Quick check. Is $\{\emptyset\}$ inductive? Is $\{0, 1, 2\}$?

Solution. Neither. $\{\emptyset\}$ contains $\emptyset$, but $\mathrm{S}(\emptyset) = \{\emptyset\} = 1$ is not an element of $\{\emptyset\}$: its only element is $\emptyset = 0$, and $1 \neq 0$ because $0 \in 1$. So the closure clause fails at $x = \emptyset$.

$\{0, 1, 2\}$ contains 0 and the successors of 0 and 1, but $\mathrm{S}(2) = 3$ is not among $0, 1, 2$: the set 3 has the element 2, and none of $0, 1, 2$ does (0 has no elements, the only element of 1 is 0, and the elements of 2 are 0 and 1, both different from 2 by Example (The first natural numbers as sets)). So the closure clause fails at $x = 2$.

Axioms (Z1)–(Z6) of Axiom 1.1.1 produce each numeral separately, never all of them at once. That they cannot do better, that is, that no inductive set can be shown to exist from (Z1)–(Z6), is a theorem of set theory proper; we state it without proof, and nothing in this book depends on it. So the existence of an inductive set has to be assumed.

Axiom 1.3.1. Axiom of Infinity

(Z7) There is an inductive set.

Everything else the chapter builds, the natural numbers and their arithmetic included, comes from the inductive set of (Z7) by the six axioms of Axiom 1.1.1.

The axiom provides **some** inductive set and says nothing else about it. It may contain sets other than $0, 1, 2, \dots$. We cannot simply call it "the set of natural numbers"; we have to carve that set out of it.

1.3.3 The least inductive set

An inductive set contains the numerals, and perhaps more. To remove anything extra, keep only what lies in **every** inductive set: whatever a set must contain merely by being inductive.

The natural numbers are what every inductive set is forced to contain.

There is one obstacle. Separation (Z6) can only cut a set out of a set already given, so we need an ambient set for the natural numbers; and the smallest-closed-set construction of *Sets, Relations, and Cardinality*, Chapter 1 (Sets, Relations, and Cardinality, Proposition 1.2.2) intersects a set of subsets, but the inductive sets are not handed to us as the elements of any set. So we fix one inductive set I and look only at the inductive subsets of I, which do form a set, by Power set and Separation; the theorem below shows that the choice of I does not matter.

Definition 1.3.3. The set of natural numbers

Fix an inductive set I, which exists by (Z7). Define

$$\omega := \{x \in I : x \in J \text{ for every inductive set } J \subseteq I\}.$$

The elements of ω are the **natural numbers**.

In words: a natural number is an element of the fixed inductive set I that survives every attempt to shrink I to a smaller inductive set.

The set ω exists by Separation (Z6), with ambient set I, once the condition is a property in the sense of (Z6): a formula of *Logic and Proof*, Chapter 4's first-order language with $\in$ and $=$ (Logic and Proof, Definition 5.2.2), with I as a parameter. The condition is "for every J, if $J \subseteq I$ and J is inductive, then $x \in J$", and, since $J \subseteq I$ abbreviates $\forall y\,(y \in J \to y \in I)$ as in (Z5) (Sets, Relations, and Cardinality, Definition 1.1.2), the only part not yet written in that language is "J is inductive". Written out with no abbreviations it is

$$\begin{aligned}&\exists e\,(e \in J \wedge \forall y\,\neg(y \in e))\\ &\wedge \forall u\,\Big(u \in J \to \exists z\,(z \in J \wedge \forall w\,(w \in z \leftrightarrow (w \in u \vee w = u)))\Big).\end{aligned}$$

The first line says that J has an element with no elements, which is $\emptyset \in J$, because $\emptyset$ is the only set with no elements (Proposition 1.1.2 (a)). The second says that for each $u \in J$ some element z of J has exactly the elements of $u \cup \{u\}$, which is $\mathrm{S}(u) \in J$ by (Z1).

Here is what "least" means, and why the choice of I is harmless.

Theorem 1.3.2. The Least Inductive Set

(a) ω is inductive.

(b) $\omega \subseteq J$ for **every** inductive set J, whether or not $J \subseteq I$.

(c) ω is the only set with properties (a) and (b). In particular, ω does not depend on the inductive set I fixed in Definition 1.3.3.

Idea. Part (a) is *Sets, Relations, and Cardinality*, Chapter 1's smallest closed set, with I as the ambient set, $\emptyset$ as the starting element and S as the step: the closed subsets of I are exactly its inductive subsets, and ω is their intersection. For (b), an inductive J need not lie inside I, so we cut it down to $\omega \cap J$, which does; then the definition of ω applies. Part (c) is two inclusions, each an instance of (b).

Proof. *(a)* Apply Sets, Relations, and Cardinality, Proposition 1.2.2 with $X := I$, the single starting element $\emptyset$, which lies in I because I is inductive, and the single step $x \mapsto \mathrm{S}(x)$, which turns elements of I into elements of I for the same reason. A subset $J \subseteq I$ is closed in the sense of that proposition exactly when $\emptyset \in J$ and $\mathrm{S}(x) \in J$ for every $x \in J$, that is, when J is inductive. So the set of closed subsets is

$$\mathcal{S} := \{J \in \mathcal{P}(I) : J \text{ is inductive}\},$$

a set by Power set (Z5) and Separation (Z6); it is non-empty, since $I \in \mathcal{S}$. Its intersection $\bigcap \mathcal{S}$, formed inside I as in Proposition 1.1.2 (c), is $\{x \in I : x \in J \text{ for every } J \in \mathcal{S}\}$, which is ω. By Sets, Relations, and Cardinality, Proposition 1.2.2, $\bigcap \mathcal{S}$ is closed; that is, ω is inductive.

(b) Let J be an inductive set. The set $\omega \cap J$ exists by (Z6) (Proposition 1.1.2 (c)), and $\omega \cap J \subseteq \omega \subseteq I$. It is inductive: $\emptyset \in \omega$ by (a) and $\emptyset \in J$ because J is inductive; and if $x \in \omega \cap J$, then $\mathrm{S}(x) \in \omega$ by (a) and $\mathrm{S}(x) \in J$ because J is inductive. So $\omega \cap J$ is an inductive subset of I, and by Definition 1.3.3 every element of ω lies in it. Hence $\omega \subseteq \omega \cap J \subseteq J$.

(c) Let W be a set that is inductive and is contained in every inductive set. Then $W \subseteq \omega$, because ω is inductive by (a); and $\omega \subseteq W$ by (b), because W is inductive. So $W = \omega$ by Sets, Relations, and Cardinality, Proposition 1.1.1. Now let I' be another inductive set, and let ω' be the set that Definition 1.3.3 produces from I'. The proofs of (a) and (b) used nothing about I except that it is inductive, so they apply to ω' as well: ω' is inductive and is contained in every inductive set. By what we just proved, $\omega' = \omega$. This proves the theorem. ∎

So ω is **the** least inductive set: it is inductive, and it lies inside every inductive set. From now on we forget I. The numerals $0, 1, 2, 3, 4$ of Example (The first natural numbers as sets) are natural numbers, since ω is inductive and contains each of them by the argument given after Definition 1.3.2.

Warning. "An inductive set is the set of natural numbers" is **false** in general. Theorem 1.3.2 gives $\omega \subseteq J$ for an inductive J, and nothing more. The set-theory volume of this series, using the axiom of Replacement (not part of this chapter), builds the inductive set $\omega \cup \{\omega, \mathrm{S}(\omega), \mathrm{S}(\mathrm{S}(\omega)), \dots\}$; it contains ω itself as an element, and the next section shows that $\omega \notin \omega$. Axiom (Z7) does not say that its I is not a set of this kind. What distinguishes ω among inductive sets is "least", and every argument about natural numbers must go through ω, not through an arbitrary inductive set.

1.3.4 Induction on the natural numbers

Leastness has an immediate consequence. To show that every natural number belongs to some set A, it is enough to show that A is inductive, since then $\omega \subseteq A$.

Theorem 1.3.3. Induction on ω

(a) (Set form.) Let $A \subseteq \omega$. If $0 \in A$, and $\mathrm{S}(n) \in A$ for every $n \in A$, then $A = \omega$.

(b) (Property form.) Let $P(n)$ be a property, that is, a formula of the first-order language

with $\in$ and $=$, possibly with other variables as parameters. If $P(0)$ holds and, **for every** $n \in \omega$, $P(n)$ implies $P(\mathrm{S}(n))$, then $P(n)$ holds **for every** $n \in \omega$.

Proof. *(a)* The hypotheses say that A is inductive. By Theorem 1.3.2 (b), $\omega \subseteq A$. With $A \subseteq \omega$, Sets, Relations, and Cardinality, Proposition 1.1.1 gives $A = \omega$.

(b) Let $A \coloneqq \{n \in \omega : P(n)\}$, a set by (Z6). Then $0 \in A$, because $0 \in \omega$ by Theorem 1.3.2 (a) and $P(0)$ holds. Let $n \in A$. Then $n \in \omega$ and $P(n)$, so $P(\mathrm{S}(n))$ by hypothesis, and $\mathrm{S}(n) \in \omega$ because ω is inductive; hence $\mathrm{S}(n) \in A$. By (a), $A = \omega$, which says that $P(n)$ holds for every $n \in \omega$.

■

Compare part (b) with the principle of induction, Logic and Proof, Axiom 4.2.1. Read with ω in place of $\mathbb{N}$, and with $\mathrm{S}(n)$ in place of $n+1$, the two statements are the same; and once addition is defined, later in this chapter, $n+1$ will be $\mathrm{S}(n)$. The difference lies in their status. In *Logic and Proof*, Chapter 3, induction was an axiom, a basic property of $\mathbb{N}$ taken on trust. Here it is a **theorem**, proved from the seven axioms (Z1)–(Z7) about sets, and the proof is three lines long because all the work went into the definition of ω: the natural numbers were defined as what every inductive set contains, and induction says exactly that.

The proof of (b) is the model for the inductions in the rest of this chapter.

Show the set is inductive. To prove $P(n)$ for every $n \in \omega$, form $\{n \in \omega : P(n)\}$ and show that it has 0 as an element and is closed under S.

Here is the move once, on a statement about the shape of natural numbers: **every** $n \in \omega$ **is** 0**, or is** 1**, or has** 1 **as an element**. Let $A \coloneqq \{n \in \omega : n = 0 \text{ or } n = 1 \text{ or } 1 \in n\}$, a set by (Z6). Then $0 \in A$ by the first alternative. Let $n \in A$; there are three cases. If $n = 0$, then $\mathrm{S}(n) = 1$, which satisfies the second alternative. If $n = 1$, then $\mathrm{S}(n) = 2 = \{0, 1\}$, which has 1 as an element. If $1 \in n$, then $1 \in n \subseteq \mathrm{S}(n)$. In every case $\mathrm{S}(n) \in A$, since also $\mathrm{S}(n) \in \omega$ because ω is inductive (Theorem 1.3.2 (a)). By Theorem 1.3.3 (a), $A = \omega$.

Two features of this small proof recur. The property carried along is a disjunction, because "$1 \in n$" alone is false at $n = 0$ and $n = 1$, and an induction can only carry a property that holds at the start. And the step is a case split along the disjunction, with one case for each alternative; dropping a case would be the error (N7) of Logic and Proof, Example (What is not allowed).

Quick check. In the proof just given, which case of the step produces the first natural number that satisfies the third alternative "$1 \in n$", and which number is it?

Solution. The case $n = 1$. It produces $\mathrm{S}(1) = 2 = \{0, 1\}$, which has 1 as an element. Neither 0 nor 1 has 1 as an element (0 has no elements, and the only element of 1 is $0 \neq 1$), so 2 is the first. From then on the third case passes "$1 \in n$" to $\mathrm{S}(n)$, because $n \subseteq \mathrm{S}(n)$.

The next section uses the move five times in a row, to show that ω satisfies Peano's axioms.

1.3.5 Exercises

1.3.5.1 A. Check your understanding

Exercise. A1

(a) List the elements of 3 and of 4 as sets of numerals.

(b) Is $2 \subseteq 3$? Is $2 \in 3$?

(c) True or false, with a reason: every inductive set has $5 := \mathrm{S}(4)$ as an element.

(d) Write "J is not inductive" in symbols, with the negation moved inward until it stands only in front of a membership statement.

Solution.

(a) $3 = \{0, 1, 2\}$ and $4 = \{0, 1, 2, 3\}$, by Example (The first natural numbers as sets).

(b) Both are true. The elements of 2 are 0 and 1, and both are elements of $3 = \{0, 1, 2\}$; so $2 \subseteq 3$ (Sets, Relations, and Cardinality, Definition 1.1.2). And 2 is itself one of the listed elements of 3, so $2 \in 3$. A set can be both an element and a subset of another set; for numerals this will turn out to be the rule, not an accident.

(c) True. Let J be inductive. Then $0 = \emptyset \in J$ by the first clause of Definition 1.3.2, and the second clause applied five times gives $1, 2, 3, 4, 5 \in J$, each being the successor of the one before.

(d) "J is inductive" is $\emptyset \in J \wedge \forall x \in J \;\; \mathrm{S}(x) \in J$. Its negation is

$$\neg(\emptyset \in J \wedge \forall x \in J \;\; \mathrm{S}(x) \in J),$$

which by De Morgan's law Logic and Proof, Proposition 2.5.5 (d) and (c) is equivalent to $\neg(\emptyset \in J) \vee \neg\forall x \in J \;\; \mathrm{S}(x) \in J$, and by Logic and Proof, Proposition 6.2.3 (a), the bounded form of Logic and Proof, Proposition 6.2.1, to

$$\emptyset \notin J \;\vee\; \exists x \in J \;\; \mathrm{S}(x) \notin J.$$

In words: J lacks the starting point, or some element of J has its successor outside J.

1.3.5.2 B. Practice

Exercise. B1: Intersections and differences

Let J and K be inductive sets. Prove that $J \cap K$ is inductive. Is $J \setminus \{\emptyset\}$ inductive? Justify your answer, and say which clause of the definition holds and which fails.

Solution. The set $J \cap K$ exists by (Z6) (Proposition 1.1.2 (c)). It contains $\emptyset$, since $\emptyset \in J$ and $\emptyset \in K$. If $x \in J \cap K$, then $\mathrm{S}(x) \in J$ because J is inductive and $x \in J$, and $\mathrm{S}(x) \in K$ likewise; so $\mathrm{S}(x) \in J \cap K$. Hence $J \cap K$ is inductive.

The set $J \setminus \{\emptyset\}$ is **not** inductive: it does not contain $\emptyset$, so the first clause fails. The second clause holds: if $x \in J \setminus \{\emptyset\}$, then $\mathrm{S}(x) \in J$, and $\mathrm{S}(x) \neq \emptyset$ because $x \in \mathrm{S}(x)$; so $\mathrm{S}(x) \in J \setminus \{\emptyset\}$. This is a non-example by minimal change: removing one element breaks exactly one clause.

Exercise. B2: Zero or a successor

Prove that every $n \in \omega$ is $\emptyset$ or is $\mathrm{S}(k)$ for some $k \in \omega$, by showing that the set

$$A \coloneqq \{n \in \omega : n = \emptyset \text{ or } n = \mathrm{S}(k) \text{ for some } k \in \omega\}$$

is inductive.

Solution. The set A exists by (Z6), and $A \subseteq \omega$. It contains $0 = \emptyset$ by the first alternative. Let $n \in A$. Then $n \in \omega$, so $\mathrm{S}(n) \in \omega$ because ω is inductive (Theorem 1.3.2 (a)), and $\mathrm{S}(n)$ satisfies the second alternative with $k \coloneqq n \in \omega$. So $\mathrm{S}(n) \in A$. By Theorem 1.3.3 (a), $A = \omega$, which is the claim.

The step never used the hypothesis that n satisfies one of the alternatives, only that $n \in \omega$. That is legitimate: the step's conclusion was reached outright, and induction still does the work of saying that every natural number is reached from 0 by successors.

1.3.5.3 C. Going deeper

Exercise. C1: Two steps back

(a) Prove that every $n \in \omega$ is 0, or is 1, or is $\mathrm{S}(\mathrm{S}(k))$ for some $k \in \omega$.

(b) Let I be an arbitrary inductive set. Running the argument of (a) on $\{x \in I : x = 0 \text{ or } x = 1 \text{ or } x = \mathrm{S}(\mathrm{S}(k)) \text{ for some } k \in I\}$, what does it prove about the elements of I, and what does it **not** prove?

Solution.

(a) Let

$$A \coloneqq \{n \in \omega : n = 0 \text{ or } n = 1 \text{ or } n = \mathrm{S}(\mathrm{S}(k)) \text{ for some } k \in \omega\},$$

a set by (Z6). Then $0 \in A$ by the first alternative. Let $n \in A$; we show $\mathrm{S}(n) \in A$, in three cases along the disjunction. If $n = 0$, then $\mathrm{S}(n) = 1$, the second alternative. If $n = 1$, then $\mathrm{S}(n) = \mathrm{S}(\mathrm{S}(0))$, the third alternative with $k \coloneqq 0 \in \omega$. If $n = \mathrm{S}(\mathrm{S}(k))$ with $k \in \omega$, then $\mathrm{S}(n) = \mathrm{S}(\mathrm{S}(\mathrm{S}(k)))$, the third alternative with $\mathrm{S}(k)$ in place of k, and $\mathrm{S}(k) \in \omega$ because ω is inductive. In each case $\mathrm{S}(n) \in \omega$, so $\mathrm{S}(n) \in A$. By Theorem 1.3.3 (a), $A = \omega$.

(b) Call the set A_I. The same three cases show that A_I has 0 as an element and is closed under S, so A_I is inductive; by Theorem 1.3.2 (b), $\omega \subseteq A_I$. That says only that every **natural number** is 0, 1 or $\mathrm{S}(\mathrm{S}(k))$ for some $k \in I$, which is weaker than (a), whose k lies in ω. It does not say that every element of I is 0, 1 or $\mathrm{S}(\mathrm{S}(k))$ for some $k \in I$. The set form of induction concludes $A = \omega$ only because $A \subseteq \omega$ was given; for $A_I \subseteq I$ the most we get is $\omega \subseteq A_I \subseteq I$, and the elements of I outside ω, if any, are not reached. Induction proves statements about ω, not about arbitrary inductive sets.

1.4 The Natural Numbers Satisfy Peano's Axioms

The previous section built ω and proved induction on it (Theorem 1.3.3). *Sets, Relations, and Cardinality*, Chapter 2 used two further facts about the natural numbers without a second thought: 0 is not of the form $n+1$, and $m+1 = n+1$ forces $m = n$. For the sets of this chapter neither is automatic, and the obvious argument for the second fails, as we shall see. This section proves both for natural numbers, using one property of their membership structure; it then names the list of properties we have reached, Peano's axioms, and shows that each of its last three items, (P3), (P4) and (P5), does a job that the other two cannot do.

1.4.1 Transitive sets

We want to prove: if $m, n \in \omega$ and $\mathrm{S}(m) = \mathrm{S}(n)$, then $m = n$. The tempting move is to cancel, reading $m \cup \{m\} = n \cup \{n\}$ as "the same thing was added to both". But union does not cancel: $\{1\} \cup \{0\} = \{0\} \cup \{1\}$, and $0 \neq 1$. What $\mathrm{S}(m) = \mathrm{S}(n)$ does give is membership. Since $m \in \mathrm{S}(m) = \mathrm{S}(n)$, either $m \in n$ or $m = n$; and likewise $n \in m$ or $n = m$. So the one bad case is $m \in n$ and $n \in m$ at once, two sets each inside the other.

For the numerals, membership only ever points to earlier numerals, and the elements of an element are earlier still. In $3 = \{0, 1, 2\}$, the element 2 has the elements 0 and 1, and both are again elements of 3. This is the property we name.

A transitive set contains the elements of its elements.

Definition 1.4.1. Transitive set

A set x is **transitive** if **every** element of x is a subset of x: for every $y \in x$, $y \subseteq x$.

In words: unpacking $y \subseteq x$ (Sets, Relations, and Cardinality, Definition 1.1.2), x is transitive when, for all sets y and z,

$$z \in y \quad \text{and} \quad y \in x \implies z \in x.$$

Membership passes through x in two steps: whatever is an element of an element of x is already an element of x. The name comes from this display, which has the shape of the transitivity of a relation (Sets, Relations, and Cardinality, Definition 2.4.1) with $\in$ as the relation and x fixed at the end.

Examples, simplest first.

- $\emptyset$ is transitive, vacuously: it has no element y at which the condition could fail. This degenerate case is the base of every induction below.
- $1 = \{0\}$ is transitive: its only element is $0 = \emptyset$, and $\emptyset \subseteq 1$.
- $3 = \{0, 1, 2\}$ is transitive: $0 = \emptyset \subseteq 3$, $1 = \{0\} \subseteq 3$, and $2 = \{0, 1\} \subseteq 3$.

A non-example by minimal change: remove 1 from 3. The set $\{0, 2\}$ is not transitive, because its element $2 = \{0, 1\}$ has the element 1, and $1 \notin \{0, 2\}$ (1 differs from 0 and from 2, Example (The first natural numbers as sets)). Deleting one element broke the property at exactly one place.

Remark. Transitivity is a property of the set, and it fails for sets that look harmless. The set $\{1\} = \{\{\emptyset\}\}$ is not transitive: $1 \in \{1\}$, but $1 = \{0\}$ is not a subset of $\{1\}$, since $0 \notin \{1\}$ (as $0 \neq 1$).

If m is transitive, then $n \in m$ gives $n \subseteq m$, so $m \in n$ and $n \in m$ together give $m \in m$. So the bad case above reduces to a set that is an element of itself. The lemma below rules that out for natural numbers, and with it the bad case.

Lemma 1.4.1. Natural Numbers Are Transitive Sets

(a) Every $n \in \omega$ is a transitive set.

(b) ω is a transitive set: every element of a natural number is a natural number.

(c) $n \notin n$ for every $n \in \omega$.

(d) Every $n \in \omega$ with $n \neq 0$ is $\mathrm{S}(k)$ for some $k \in \omega$.

(e) $0 \in n$ for every $n \in \omega$ with $n \neq 0$.

Idea. Each part is one use of the move of the previous section, **show the set is inductive**, in its property form Theorem 1.3.3 (b). The base is always $0 = \emptyset$, where the claim is empty or immediate. The steps of (a)–(c) start from "$y \in \mathrm{S}(n)$ means $y \in n$ or $y = n$" and treat the two cases. Part (c) is the only one that uses an earlier part: it argues by contradiction inside the step, and transitivity (a) is what turns a membership one level down into the forbidden $n \in n$. In (d) and (e) the statement is false at 0, so the property carried is a disjunction "$n = 0$ or ...". *Before reading on: in the proof of (c), why does* $\mathrm{S}(n) \in n$ *give* $n \in n$*?*

Proof. Each part is proved by Theorem 1.3.3 (b), with the property named at its start.

(a) Let $P(n)$ be "n is transitive". $P(0)$ holds because $\emptyset$ has no elements. Let $n \in \omega$ be transitive, and let $y \in \mathrm{S}(n)$. If $y \in n$, then $y \subseteq n$ because n is transitive, and $n \subseteq \mathrm{S}(n)$; so $y \subseteq \mathrm{S}(n)$. If $y = n$, then $y = n \subseteq \mathrm{S}(n)$. In both cases $y \subseteq \mathrm{S}(n)$, so $\mathrm{S}(n)$ is transitive.

(b) Let $P(n)$ be "$n \subseteq \omega$". $P(0)$ holds because $\emptyset \subseteq \omega$. Let $n \in \omega$ with $n \subseteq \omega$. The elements of $\mathrm{S}(n) = n \cup \{n\}$ are the elements of n, which lie in ω by hypothesis, and n itself, which lies in ω by the choice of n. So $\mathrm{S}(n) \subseteq \omega$. Hence $n \subseteq \omega$ for every $n \in \omega$, which says that ω is transitive.

(c) Let $P(n)$ be "$n \notin n$". $P(0)$ holds because $\emptyset$ has no elements. Let $n \in \omega$ with $n \notin n$, and suppose $\mathrm{S}(n) \in \mathrm{S}(n)$. Then $\mathrm{S}(n) \in n$ or $\mathrm{S}(n) = n$.

- If $\mathrm{S}(n) \in n$, then $\mathrm{S}(n) \subseteq n$, because n is transitive by (a); and $n \in \mathrm{S}(n)$, so $n \in n$.
- If $\mathrm{S}(n) = n$, then $n \in \mathrm{S}(n) = n$.

Both cases contradict $n \notin n$. So $\mathrm{S}(n) \notin \mathrm{S}(n)$, by ¬I (Logic and Proof, Definition 2.2.1).

(d) Let $P(n)$ be "$n = 0$, or $n = \mathrm{S}(k)$ for some $k \in \omega$". $P(0)$ holds by the first alternative. For every $n \in \omega$, $P(\mathrm{S}(n))$ holds by the second alternative, with $k := n$. So $P(n)$ holds for every $n \in \omega$, and for $n \neq 0$ the second alternative is the claim.

(e) Let $P(n)$ be "$n = 0$ or $0 \in n$". $P(0)$ holds by the first alternative. Let $n \in \omega$ with $P(n)$.

If $n = 0$, then $\mathrm{S}(n) = \{0\}$, which has 0 as an element. If $0 \in n$, then $0 \in n \subseteq \mathrm{S}(n)$. Either way $0 \in \mathrm{S}(n)$, so $P(\mathrm{S}(n))$ holds. So $P(n)$ holds for every $n \in \omega$, and for $n \neq 0$ the second alternative is the claim. This proves the lemma.

■

The step of (d) never used $P(n)$. As with the predecessor in *Sets, Relations, and Cardinality*, Chapter 2 (Sets, Relations, and Cardinality, Example (An induction that ignores its hypothesis)), the step's conclusion was reached outright, and induction is still what guarantees that 0 and the successors are all there is. Part (b) says that the natural numbers are built only from natural numbers: unwrapping a natural number never leads outside ω. Part (c) is the fact the bad case needed.

1.4.2 Peano's axioms for the natural numbers

Here is the list. Its five items are traditionally called **Peano's axioms**; for ω they are theorems.

Theorem 1.4.2. Peano's Axioms Hold in ω

(P1) $0 \in \omega$.

(P2) For every $n \in \omega$, $\mathrm{S}(n) \in \omega$.

(P3) For every $n \in \omega$, $\mathrm{S}(n) \neq 0$.

(P4) For all $m, n \in \omega$, if $\mathrm{S}(m) = \mathrm{S}(n)$, then $m = n$.

(P5) Every $A \subseteq \omega$ with $0 \in A$ and $\mathrm{S}(n) \in A$ for every $n \in A$ is equal to ω.

Moreover, $\{(n, \mathrm{S}(n)) : n \in \omega\}$ is a set, and with codomain ω it is a function $\mathrm{S}\colon \omega \to \omega$, which is injective and does not take the value 0.

Idea. (P1), (P2) and (P5) are the previous section restated: ω is inductive, and induction holds. (P3) is one line, since $n \in \mathrm{S}(n)$ and nothing is an element of $\emptyset$. (P4) is the argument of the opening of this section: $\mathrm{S}(m) = \mathrm{S}(n)$ leaves only the case $m \in n$ and $n \in m$ to exclude, and transitivity with $m \notin m$ excludes it.

Proof. (P1) and (P2) say that ω is inductive, which is Theorem 1.3.2 (a). (P5) is Theorem 1.3.3 (a).

(P3) Let $n \in \omega$. Then $n \in \mathrm{S}(n)$. If $\mathrm{S}(n) = 0$, then $n \in 0$ by =E (Logic and Proof, Definition 5.7.1), which is impossible, since $0 = \emptyset$ has no elements. So $\mathrm{S}(n) \neq 0$, by ¬I (Logic and Proof, Definition 2.2.1).

(P4) Let $m, n \in \omega$ with $\mathrm{S}(m) = \mathrm{S}(n)$. Since $m \in \mathrm{S}(m) = \mathrm{S}(n)$, we have $m \in n$ or $m = n$; likewise, since $n \in \mathrm{S}(n) = \mathrm{S}(m)$, we have $n \in m$ or $n = m$. Suppose $m \neq n$. Then $m \in n$ and $n \in m$. Since m is transitive (Lemma 1.4.1 (a)) and $n \in m$, we get $n \subseteq m$; so $m \in n \subseteq m$, that is, $m \in m$. This contradicts Lemma 1.4.1 (c). Hence $m = n$, by RAA (Logic and Proof, Definition 2.5.1).

For the function: let $G \coloneqq \{p \in \omega \times \omega : p = (n, \mathrm{S}(n)) \text{ for some } n \in \omega\}$, a set by Separation (Z6) inside the product $\omega \times \omega$ of Proposition 1.2.2; the pairs $(n, \mathrm{S}(n))$ lie in that product by

(P2). We check the two clauses of Sets, Relations, and Cardinality, Definition 1.4.1. (Fn1): for $n \in \omega$, $(n, \mathrm{S}(n)) \in G$. (Fn2): if $(n, y) \in G$ and $(n, y') \in G$, then $(n, y) = (k, \mathrm{S}(k))$ and $(n, y') = (k', \mathrm{S}(k'))$ for some $k, k' \in \omega$; by Theorem 1.2.1, $k = n = k'$, and then $y = \mathrm{S}(n) = y'$. So (G, ω) is a function $\omega \to \omega$, whose value at n is $\mathrm{S}(n)$. It is injective by (P4) (Sets, Relations, and Cardinality, Definition 1.4.3), and 0 is not a value by (P3). This proves the theorem.

■

The function $\mathrm{S}\colon \omega \to \omega$ and the operation $x \mapsto \mathrm{S}(x)$ on all sets share a letter, and they agree wherever both are defined. Only the first is a set. The operation has no graph: a set containing the pair $(x, \mathrm{S}(x)) = \{\{x\}, \{x, \mathrm{S}(x)\}\}$ for **every** set x would give, after two unions, a set containing every set, which Sets, Relations, and Cardinality, Theorem 1.1.4 forbids.

One more consequence is used constantly. By Lemma 1.4.1 (d), every $n \in \omega$ with $n \neq 0$ is $\mathrm{S}(k)$ for some $k \in \omega$, and by (P4) this k is **unique**: if $\mathrm{S}(k) = n = \mathrm{S}(k')$, then $k = k'$. So every non-zero natural number has exactly one predecessor, and 0 has none, by (P3).

Quick check. Is the function $\mathrm{S}\colon \omega \to \omega$ surjective? Is $\omega \in \omega$?

Solution. No to both. S is not surjective: $0 \in \omega$, and by (P3) no $n \in \omega$ has $\mathrm{S}(n) = 0$. (Every other natural number is a value, by Lemma 1.4.1 (d).)

If $\omega \in \omega$, then ω would be a natural number n with $n \in n$, which Lemma 1.4.1 (c) rules out. So $\omega \notin \omega$. In particular ω is a set that is not a natural number, although all its elements are.

1.4.3 Peano systems

Sets, Relations, and Cardinality, Chapter 2 worked with a set $\mathbb{N}$ of real numbers, and this chapter with a set ω of sets. Both come with a starting element and a step. The last section of the chapter asks whether they are "the same", and to ask it we need to say what any such structure must satisfy. The answer is the list just proved, with ω, 0 and S replaced by letters.

A Peano system is a set with a start and a step, where the step is one-to-one, never returns to the start, and reaches everything.

Definition 1.4.2. Peano system

A **Peano system** (M, e, s) consists of a set M, an element $e \in M$, and a function $s\colon M \to M$, such that:

- (P3) $s(x) \neq e$ for **every** $x \in M$;
- (P4) s is injective;
- (P5) **every** subset $A \subseteq M$ with $e \in A$ and $s(x) \in A$ for every $x \in A$ is equal to M.

In words: (P1) and (P2) are built into the data, since $e \in M$ and s maps M into M. (P3) says the step never leads back to the start. (P4) says two different elements never step to the same element. (P5) is induction for M: a subset that contains the start and is closed under the step is everything, so every element of M is reached from e by steps. We write (M, e, s) and not (N, e, s), so that the letter never suggests $\mathbb{N}$. The notation (M, e, s) lists the three pieces

of data; it is not a set built from them (if one object is wanted, take the pair $((M, e), s)$ of Definition 1.2.1).

The first example is the one we have, and the second shows that the elements need not be numbers.

- $(\omega, 0, \mathrm{S})$ is a Peano system, by Theorem 1.4.2: $\mathrm{S}\colon \omega \to \omega$ is a function, and (P3)–(P5) are those of the theorem.
- **Tagged numbers.** Let $M := \omega \times \{0\}$, a set by Proposition 1.2.2, let $e := (0, 0)$, and let $s(n, 0) := (\mathrm{S}(n), 0)$. Every element of M is $(n, 0)$ for exactly one $n \in \omega$ (Theorem 1.2.1), so s is well defined, and its graph is a set by (Z6) inside $M \times M$. (P3): $(\mathrm{S}(n), 0) = (0, 0)$ would give $\mathrm{S}(n) = 0$ by Theorem 1.2.1, against (P3) for ω. (P4): $(\mathrm{S}(m), 0) = (\mathrm{S}(n), 0)$ gives $\mathrm{S}(m) = \mathrm{S}(n)$, so $m = n$ by (P4) for ω. (P5): let $A \subseteq M$ contain e and be closed under s. The set $B := \{n \in \omega : (n, 0) \in A\}$ has 0 as an element and is closed under S, so $B = \omega$ by Theorem 1.3.3 (a), and $A = M$. Here $e = (0, 0) = \{\{0\}\}$ is not $\emptyset$, and no element of M is a natural number, as we shall see at the end of the section.

Why these three clauses and no others? Each one does work that the other two cannot do: for each clause there is a structure satisfying the other two and failing it.

Example. Each axiom is needed

Show that each of the following structures (M, e, s), with $s\colon M \to M$, satisfies two of (P3), (P4), (P5) and fails the third.

(i) $M := \{0\}$, $e := 0$, $s(0) := 0$.

(ii) $M := \{0, 1\}$, $e := 0$, $s(0) := 1$, $s(1) := 1$.

(iii) (**The natural numbers beside a two-sided chain.**) $M := \omega \times 3$, whose elements are the pairs (n, i) with $n \in \omega$ and $i \in 3 = \{0, 1, 2\}$; $e := (0, 0)$; and

$$s(n, 0) := (\mathrm{S}(n), 0), \qquad s(n, 1) := (\mathrm{S}(n), 1),$$
$$s(0, 2) := (0, 1), \qquad s(\mathrm{S}(k), 2) := (k, 2),$$

for $n, k \in \omega$.

Solution.

(i) **Fails (P3)**: $s(0) = 0 = e$. (P4) holds: M has one element, so two inputs with equal outputs are equal. (P5) holds: a subset of $\{0\}$ containing $e = 0$ is $\{0\}$.

(ii) **Fails (P4)**: $s(0) = s(1)$ with $0 \neq 1$. (P3) holds: both values are $1 \neq 0 = e$. (P5) holds: if $A \subseteq \{0, 1\}$ has 0 as an element and is closed under s, then $s(0) = 1 \in A$, so $A = \{0, 1\}$.

(iii) First, s is a function $M \to M$. Each element of M is (n, i) for exactly one $n \in \omega$ and one $i \in 3$, by Theorem 1.2.1. For $i = 2$, the first coordinate is 0 or $\mathrm{S}(k)$ with $k \in \omega$ (Lemma 1.4.1 (d)); never both, by (P3) for ω; and the k is unique, by (P4) for ω. So exactly one line of the definition applies to each element, and it names one value. The values lie in M, since $\mathrm{S}(n), 0, k \in \omega$. The graph is a set by (Z6) inside $M \times M$.

To check (P3) and (P4), sort the values by their second coordinates, which are 0, 1 and 2,

three different sets (Example (The first natural numbers as sets)):

$$\begin{aligned} &\text{second coordinate } 0: && s(n,0) = (\mathrm{S}(n),0), \\ &\text{second coordinate } 1: && s(n,1) = (\mathrm{S}(n),1) \quad \text{and} \quad s(0,2) = (0,1), \\ &\text{second coordinate } 2: && s(\mathrm{S}(k),2) = (k,2). \end{aligned}$$

(P3) holds. A value equal to $e = (0,0)$ would have second coordinate 0, so it would be $(\mathrm{S}(n),0)$ with $\mathrm{S}(n) = 0$, against (P3) for ω.

(P4) holds. Suppose $s(x) = s(x')$. The two values have the same second coordinate, so they come from the same row of the display. In the first row, $(\mathrm{S}(n),0) = (\mathrm{S}(n'),0)$ gives $n = n'$ by Theorem 1.2.1 and (P4) for ω. In the third row, $(k,2) = (k',2)$ gives $k = k'$, so $\mathrm{S}(k) = \mathrm{S}(k')$. In the second row, $(\mathrm{S}(n),1) \neq (0,1)$ because $\mathrm{S}(n) \neq 0$; so either both values are of the form $(\mathrm{S}(n),1)$, and $n = n'$ as before, or both are $(0,1)$, and both inputs are $(0,2)$. In every case $x = x'$.

(P5) fails. Let $A \coloneqq \omega \times \{0\}$, a subset of M (Proposition 1.2.2). It contains $e = (0,0)$, and it is closed under s, since $s(n,0) = (\mathrm{S}(n),0) \in A$. But $(0,1) \in M$ is not in A, because its second coordinate 1 is not 0. So $A \neq M$.

The picture shows (iii), with an arrow from each element to its image under s.

$$\begin{array}{l} \ e \\ (0,0) \to (1,0) \to (2,0) \to (3,0) \to \cdots \end{array}$$

$$\cdots \to (2,2) \to (1,2) \to (0,2) \to (0,1) \to (1,1) \to (2,1) \to \cdots$$

The top row is a copy of ω, starting at e. The bottom row is a chain running forever in both directions: read $(n,1)$ as "n" and $(k,2)$ as "$-(k+1)$", and s adds one everywhere, with $(0,2)$, playing -1, stepping to $(0,1)$, playing 0. Every element has exactly one arrow out and at most one arrow in, which is (P4); only e has no arrow in, which is (P3). What fails is that the bottom row cannot be reached from e. Induction, (P5), is the clause that says "nothing else".

So (P3) and (P4) alone cannot prevent extra elements: in (iii) the whole bottom row is extra. That is why the definition of ω needed "least", and why induction is part of the list.

1.4.4 Pairs are not natural numbers

Sets, Relations, and Cardinality, Chapter 1 adopted a working convention, stated after Sets, Relations, and Cardinality, Definition 1.3.1: a pair is not a number. In that chapter numbers were given and pairs were postulated, and the convention said they were different kinds of object. Now both are sets: a pair is the set $\{\{a\},\{a,b\}\}$ of Definition 1.2.1, and a natural number is an element of ω. For natural numbers the convention becomes a theorem.

Proposition 1.4.3. No Pair Is a Natural Number

For all sets a and b, the pair (a,b) is not an element of ω.

Idea. Every natural number other than 0 has $\emptyset$ as an element, by Lemma 1.4.1 (e). A pair is not $\emptyset$, and it does not have $\emptyset$ as an element, because both of its elements contain a.

Proof. Let a, b be sets, and suppose $(a, b) = \{\{a\}, \{a, b\}\} \in \omega$ (Definition 1.2.1). The pair has the element $\{a\}$, so $(a, b) \neq \emptyset = 0$. By Lemma 1.4.1 (e), $0 = \emptyset \in (a, b)$, so $\emptyset = \{a\}$ or $\emptyset = \{a, b\}$. Both are impossible, since $a \in \{a\}$ and $a \in \{a, b\}$, while $\emptyset$ has no elements. This contradiction shows $(a, b) \notin \omega$.

■

So, for instance, the tagged numbers $(n, 0)$ in the second example of a Peano system above are never natural numbers, and *Sets, Relations, and Cardinality*, Chapter 1's answer that $((1, 2), 3) \neq (1, (2, 3))$ because "a pair is not a number" is borne out by Proposition 1.4.3 once the numbers 1, 2, 3 are read as elements of ω, as the later chapters of this volume read $\mathbb{N}$ (Section 1.7). The same statement for n-tuples with $n \geq 3$ needs a definition of n-tuples, which comes later in this chapter, together with the order on ω.

Warning. The encoding produces true statements that sound absurd. Since $\{0\} = 1$ and $\{0, 1\} = 2$,

$$(0, 1) = \{\{0\}, \{0, 1\}\} = \{1, 2\},$$

so "$1 \in (0, 1)$" is **true** of the pair $(0, 1)$, although 1 is not in the open interval $(0, 1)$ of real numbers, and "$0 \in (0, 1)$" is **false** of the pair. Such a statement is a fact about how pairs were encoded as sets, not about pairs: with Wiener's encoding $\{\{\{a\}, \emptyset\}, \{\{b\}\}\}$, met in Section 1.2 (Exercise (C3: Wiener's encoding)), the pair of 0 and 1 is $\{2, \{\{1\}\}\}$, which does not have 1 as an element. A proof about pairs should use only their characteristic property, Theorem 1.2.1. The places in this section that look inside a pair (the remark after Theorem 1.4.2 that the operation $x \mapsto \mathrm{S}(x)$ has no graph, the closing remark of the tagged-numbers example that $(0, 0) = \{\{0\}\}$ is not $\emptyset$, Proposition 1.4.3, and this warning) are about the encoding itself.

With Peano's axioms in hand, the next section proves the recursion theorem on ω and uses it to define addition and multiplication.

1.4.5 Exercises

1.4.5.1 A. Check your understanding

Exercise. A1

(a) For each of the three structures in Example (Each axiom is needed), which of (P3), (P4), (P5) fails?

(b) In the proof of (P4) in Theorem 1.4.2, where is Lemma 1.4.1 (a) used, and where is (c)?

(c) Is $\omega \cup \{\omega\}$ inductive? Justify your answer.

(d) Is $\{0, 1, \{1\}\}$ a transitive set? Is it an element of ω?

Solution.

(a) (i) fails (P3); (ii) fails (P4); (iii) fails (P5). Each satisfies the other two (Example (Each axiom is needed)).

(b) Part (a) is used to pass from $n \in m$ to $n \subseteq m$: m is a natural number, hence a transitive set. Part (c) is used at the end: the argument reaches $m \in m$, and (c) says $m \notin m$, which gives the contradiction.

(c) No. The set $\omega \cup \{\omega\} = \mathrm{S}(\omega)$ has $\emptyset$ as an element and is closed under S on its elements from ω, but $\mathrm{S}(\omega)$ itself is not an element of $\omega \cup \{\omega\}$. First, $\mathrm{S}(\omega) \neq \omega$: $\omega \in \mathrm{S}(\omega)$, while $\omega \notin \omega$ (the Quick check after Theorem 1.4.2). Second, $\mathrm{S}(\omega) \notin \omega$: otherwise $\mathrm{S}(\omega) \subseteq \omega$ because ω is transitive (Lemma 1.4.1 (b)), and $\omega \in \mathrm{S}(\omega)$ would give $\omega \in \omega$. So the closure clause fails at the element ω.

(d) It is transitive: $0 = \emptyset$ is a subset of every set; $1 = \{0\}$ is a subset, since 0 is listed; $\{1\}$ is a subset, since 1 is listed. It is not an element of ω: by Lemma 1.4.1 (b), every element of a natural number is a natural number, and $\{1\}$ is not one. Indeed $\{1\} \neq 0$, and if $\{1\} = \mathrm{S}(k) = k \cup \{k\}$ with $k \in \omega$, then $k \in \{1\}$, so $k = 1$, and $\mathrm{S}(1) = \{0, 1\} \neq \{1\}$ since $0 \notin \{1\}$; this contradicts Lemma 1.4.1 (d). So being transitive does not make a set a natural number.

1.4.5.2 B. Practice

Exercise. B1: Put the proof in order

Let $m, n \in \omega$. The following lines contain a proof that **if** $\mathrm{S}(m) = \mathrm{S}(n)$**, then** $m = n$, shuffled, together with two lines that do not belong. Write the seven correct lines in an order that makes a proof, say which other orders are also admissible, and say why each of the other two lines fails.

(i) So $m \in n \subseteq m$, that is, $m \in m$.

(ii) Suppose $\mathrm{S}(m) = \mathrm{S}(n)$.

(iii) We have $m \cup \{m\} = n \cup \{n\}$, so $m = n$ by canceling $\cup\{\cdot\}$ on both sides.

(iv) Since $n \in \mathrm{S}(n) = \mathrm{S}(m)$, we have $n \in m$ or $n = m$.

(v) Suppose $m \neq n$; then $m \in n$ and $n \in m$.

(vi) This contradicts $m \notin m$ (Lemma 1.4.1 (c)); so $m = n$.

(vii) From $m \in n$ and $n \in m$ we get $m = n$, since $\in$ is antisymmetric.

(viii) Since m is transitive (Lemma 1.4.1 (a)) and $n \in m$, we have $n \subseteq m$.

(ix) Since $m \in \mathrm{S}(m) = \mathrm{S}(n)$, we have $m \in n$ or $m = n$.

Solution. An admissible order is (ii), (ix), (iv), (v), (viii), (i), (vi):

> Suppose $\mathrm{S}(m) = \mathrm{S}(n)$. Since $m \in \mathrm{S}(m) = \mathrm{S}(n)$, we have $m \in n$ or $m = n$. Since $n \in \mathrm{S}(n) = \mathrm{S}(m)$, we have $n \in m$ or $n = m$. Suppose $m \neq n$; then $m \in n$ and $n \in m$. Since m is transitive and $n \in m$, we have $n \subseteq m$. So $m \in n \subseteq m$, that is, $m \in m$. This contradicts $m \notin m$; so $m = n$.

The only other admissible order exchanges (ix) and (iv), which use only (ii) and not each other. Everything else is forced: (v) needs both (ix) and (iv) to reduce each disjunction to its first alternative; (viii) needs the $n \in m$ of (v); (i) needs the $m \in n$ of (v) and the inclusion

of (viii); (vi) closes the RAA box opened by (v), whose assumption is $m \neq n$, the negation of the goal (Logic and Proof, Definition 2.5.1).

The distractor (iii) cancels where no cancellation law exists. For sets in general, $x \cup \{y\} = x' \cup \{y'\}$ does not give $y = y'$: $\{1\} \cup \{0\} = \{0\} \cup \{1\}$ with $0 \neq 1$. For natural numbers the conclusion $m = n$ happens to be true, but that is the statement being proved, not a reason for it.

The distractor (vii) cites a "fact" that is not available. No result so far says that $m \in n$ and $n \in m$ cannot both hold, and for natural numbers that is exactly what lines (viii), (i) and (vi) establish. Citing it is the circular step (N9) of Logic and Proof, Example (What is not allowed): the line uses, as a reason, what the proof is there to show.

Exercise. B2: Two steps at a time

Let $d \coloneqq \mathrm{S} \circ \mathrm{S} \colon \omega \to \omega$ (Sets, Relations, and Cardinality, Definition 1.4.2), so that $d(n) = \mathrm{S}(\mathrm{S}(n))$. Show that $(\omega, 0, d)$ satisfies (P3) and (P4) of Definition 1.4.2, and that it does not satisfy (P5).

Hint: look at the set $\omega \setminus \{1\}$.

Solution. (P3): for $n \in \omega$, $d(n) = \mathrm{S}(\mathrm{S}(n)) \neq 0$ by (P3) of Theorem 1.4.2, applied to $\mathrm{S}(n) \in \omega$.

(P4): if $d(m) = d(n)$, then $\mathrm{S}(\mathrm{S}(m)) = \mathrm{S}(\mathrm{S}(n))$, so $\mathrm{S}(m) = \mathrm{S}(n)$ by (P4) of Theorem 1.4.2, and then $m = n$ by (P4) again. (Equivalently: d is a composite of two injections, so it is injective by Sets, Relations, and Cardinality, Theorem 1.4.4 (a).)

(P5) fails. Let $A \coloneqq \omega \setminus \{1\}$. It contains 0, since $0 \neq 1$. It is closed under d: let $n \in A$; then $d(n) \in \omega$, and $d(n) \neq 1$, because $\mathrm{S}(\mathrm{S}(n)) = 1 = \mathrm{S}(0)$ would give $\mathrm{S}(n) = 0$ by (P4), against (P3). But $A \neq \omega$, since $1 \in \omega$ and $1 \notin A$. So a subset of ω with 0 as an element and closed under d need not be all of ω. The step d skips 1, and starting from 0 it reaches only the "even" numbers.

1.4.5.3 C. Going deeper

Exercise. C1: A Peano system with a different start

Give an example of a Peano system (M, e, s) with $M \subseteq \omega$ and $e \neq 0$. Check every clause of Definition 1.4.2.

Solution. Any correct answer has: a set $M \subseteq \omega$; an element $e \in M$ with $e \neq 0$; a function $s \colon M \to M$, so that every value lies in M; and the three clauses (P3), (P4), (P5) for these data.

Take $M \coloneqq \omega \setminus \{0\}$, $e \coloneqq 1$, and $s(n) \coloneqq \mathrm{S}(n)$ for $n \in M$.

- **The data.** $1 \in M$, since $1 \in \omega$ and $1 \neq 0$. For $n \in M$, $\mathrm{S}(n) \in \omega$ by (P2) and $\mathrm{S}(n) \neq 0$ by (P3) of Theorem 1.4.2, so $\mathrm{S}(n) \in M$. The graph $\{(n, \mathrm{S}(n)) : n \in M\}$ is a set by (Z6) inside $M \times M$. (Fn1) holds, since $(n, \mathrm{S}(n))$ is in the graph for each $n \in M$; (Fn2) holds as in the proof of Theorem 1.4.2, by Theorem 1.2.1. So $s \colon M \to M$ is a function

(Sets, Relations, and Cardinality, Definition 1.4.1).

- **(P3).** If $s(n) = 1$ for some $n \in M$, then $\mathrm{S}(n) = \mathrm{S}(0)$, so $n = 0$ by (P4) for ω; but $0 \notin M$.
- **(P4).** If $\mathrm{S}(m) = \mathrm{S}(n)$ with $m, n \in M$, then $m = n$ by (P4) for ω.
- **(P5).** Let $A \subseteq M$ with $1 \in A$ and $\mathrm{S}(n) \in A$ for every $n \in A$. The set $A' \coloneqq A \cup \{0\} \subseteq \omega$ is inductive: $0 \in A'$; $\mathrm{S}(0) = 1 \in A \subseteq A'$; and for $n \in A$, $\mathrm{S}(n) \in A \subseteq A'$. By Theorem 1.3.3 (a), $A' = \omega$. Every $n \in M$ lies in A' and is not 0, so $n \in A$. Hence $A = M$.

Other answers can be checked against the four properties at the start. For instance, $M \coloneqq \omega \setminus \{0, 1\}$, $e \coloneqq 2$, $s(n) \coloneqq \mathrm{S}(n)$ also works, by the same argument with $\{0, 1\}$ added in (P5).

Exercise. C2: Closed subsets of the chain

Let (M, e, s) be the structure (iii) of Example (Each axiom is needed). Call $A \subseteq M$ **good** if $e \in A$ and $s(x) \in A$ for every $x \in A$.

(a) Show that every good set contains $\omega \times \{0\}$.

(b) Show that each of the following sets is good, where $m, k \in \omega$:

$$G_0 \coloneqq \omega \times \{0\}, \qquad G_\infty \coloneqq M,$$
$$R_m \coloneqq G_0 \cup \{(n, 1) : n \in \omega,\ m \in \mathrm{S}(n)\},$$
$$L_k \coloneqq G_0 \cup (\omega \times \{1\}) \cup \{(j, 2) : j \in \mathrm{S}(k)\}.$$

(c) Show that a good set that contains $(k, 2)$ contains L_k.

Solution (Hint 1). For (a), apply induction to the set of $n \in \omega$ with $(n, 0) \in A$. For (c), the arrows of the bottom row point toward $(0, 1)$: from $(\mathrm{S}(k), 2)$ the step goes to $(k, 2)$.

Solution (Hint 2). For (c), prove by induction on $k \in \omega$ the property: "for every good A, if $(k, 2) \in A$, then $L_k \subseteq A$".

Solution.

(a) Let A be good, and let $B \coloneqq \{n \in \omega : (n, 0) \in A\}$. Then $0 \in B$, because $(0, 0) = e \in A$; and if $n \in B$, then $(\mathrm{S}(n), 0) = s(n, 0) \in A$, so $\mathrm{S}(n) \in B$. By Theorem 1.3.3 (a), $B = \omega$, so $\omega \times \{0\} \subseteq A$.

(b) All four sets contain $e = (0, 0) \in G_0$. Closure is checked element by element, using the definition of s.

- G_0: $s(n, 0) = (\mathrm{S}(n), 0) \in G_0$.
- $G_\infty = M$: s maps M into M.
- R_m: elements of G_0 are handled as before. If $(n, 1) \in R_m$ with $m \in \mathrm{S}(n)$, then $s(n, 1) = (\mathrm{S}(n), 1)$, and $m \in \mathrm{S}(n) \subseteq \mathrm{S}(\mathrm{S}(n))$; so $s(n, 1) \in R_m$.
- L_k: elements of G_0 as before; $s(n, 1) = (\mathrm{S}(n), 1) \in \omega \times \{1\}$. Every $j \in \mathrm{S}(k)$ lies in ω, since $\mathrm{S}(k) \in \omega$ by (P2) of Theorem 1.4.2 and ω is transitive (Lemma 1.4.1 (b)); so $L_k \subseteq M$. For $(j, 2)$ with $j \in \mathrm{S}(k)$: if $j = 0$, then $s(0, 2) = (0, 1) \in L_k$; if $j = \mathrm{S}(i)$

with $i \in \omega$ (Lemma 1.4.1 (d)), then $s(j,2) = (i,2)$, and $i \in \mathrm{S}(i) = j \in \mathrm{S}(k)$, so $i \in \mathrm{S}(k)$ because $\mathrm{S}(k)$ is transitive (Lemma 1.4.1 (a)); hence $(i,2) \in L_k$.

(c) We prove by Theorem 1.3.3 (b) the property $P(k)$: "for every good A, if $(k,2) \in A$, then $L_k \subseteq A$".

Base. Let A be good with $(0,2) \in A$. Then $(0,1) = s(0,2) \in A$. The set $\{n \in \omega : (n,1) \in A\}$ has 0 as an element and is closed under S, since $s(n,1) = (\mathrm{S}(n),1)$; so it is ω by Theorem 1.3.3 (a), and $\omega \times \{1\} \subseteq A$. With (a) and $\mathrm{S}(0) = \{0\}$, this gives $L_0 \subseteq A$.

Step. Let $k \in \omega$ with $P(k)$, and let A be good with $(\mathrm{S}(k),2) \in A$. Then $(k,2) = s(\mathrm{S}(k),2) \in A$, so $L_k \subseteq A$ by $P(k)$. Now $L_{\mathrm{S}(k)} = L_k \cup \{(\mathrm{S}(k),2)\}$, because $j \in \mathrm{S}(\mathrm{S}(k))$ means $j \in \mathrm{S}(k)$ or $j = \mathrm{S}(k)$. Both parts lie in A, so $L_{\mathrm{S}(k)} \subseteq A$, which is $P(\mathrm{S}(k))$.

So the good sets include G_0, the sets R_m (whose part in the bottom row starts at $(m,1)$ and runs to the right), the sets L_k (whose part in the bottom row starts at $(k,2)$), and M. Read on the two-sided chain, these are the empty part, the rays starting at some point, and the whole chain. That they are **all** the good sets is also true, but the proof needs a least element of a non-empty set of natural numbers, which comes later in this chapter with the order on ω. Part (c) already shows that a good set containing every $(k,2)$ contains every L_k, hence all of M. Every good set other than M is a witness to the failure of (P5).

1.5 The Recursion Theorem; Addition and Multiplication

The set ω now has a zero, a successor map and an induction theorem, and Theorem 1.4.2 shows that together they satisfy Peano's axioms. It has no addition yet. In *Sets, Relations, and Cardinality*, Chapter 2 the sum $m + n$ of two natural numbers came for free, from the real numbers; in ω a natural number is a set, and nothing we know about sets adds two of them. This section defines addition and multiplication by recursion, as school does ("$m + (n + 1)$ is one more than $m + n$"), after first proving, from the axioms, that recursion defines something. Then it proves the laws of arithmetic one by one, most of them by an induction.

1.5.1 Recursion on ω

Sets, Relations, and Cardinality, Section 2.3 proved Sets, Relations, and Cardinality, Theorem 2.3.1: a starting value and a rule for the next value determine exactly one function on $\mathbb{N}$. That proof ran inside the familiar real numbers, where $\mathbb{N}$ is given. Its ingredients, though, were few: induction, and some facts about the sets $\{0,\dots,n\}$ on which the partial solutions live. On ω there is a natural replacement for $\{0,\dots,n\}$, namely the set $\mathrm{S}(n) = n \cup \{n\}$ itself. (The next section proves that $\mathrm{S}(n)$ is exactly the set of $k \in \omega$ with $k \leq n$; here we need no order, only $\mathrm{S}(n)$ as a set.) So the proof can be repeated on ω, with one change, which we point out when we reach it.

A starting value and a rule for the next value determine exactly one function on ω.

Theorem 1.5.1. Recursion Theorem on ω

Let X be a set, let $a \in X$, and let $g\colon \omega \times X \to X$ be a function. Then there is **exactly one**

function $f\colon \omega \to X$ such that

$$f(0) = a \quad \text{and} \quad f(\mathrm{S}(n)) = g(n, f(n)) \quad \text{for every } n \in \omega.$$

This is the statement of Sets, Relations, and Cardinality, Theorem 2.3.1 word for word, with ω in place of $\mathbb{N}$ and $\mathrm{S}(n)$ in place of $n+1$. The set $\omega \times X$ exists by Proposition 1.2.2. As before, a is the starting value and g turns the stage n and the current value into the next value; the conclusion has an existence half and a uniqueness half.

Idea. Uniqueness is an induction. For existence, induction cannot build f, so we prove by induction that the finite pieces of f exist and are unique, and then glue them. The roadmap is that of *Sets, Relations, and Cardinality*, Section 2.3:

① **Uniqueness**, by Theorem 1.3.3.

② **Partial solutions.** A partial solution up to n is a function on $\mathrm{S}(n)$ obeying the two equations wherever they make sense. A restriction of a partial solution is again one.

③ **Claim:** for every n there is exactly one partial solution up to n; the next one adds a single pair at the new input $\mathrm{S}(n)$.

④ **Glue:** the union of all their graphs is the graph of f.

The glue step of *Sets, Relations, and Cardinality*, Section 2.3 compared two indices: if $m \leq n$, the piece up to m lies inside the piece up to n. We have no order on ω yet, so at each point k where both pieces are defined we cut both down to $\mathrm{S}(k)$: by ② and ③, each restriction is the unique partial solution up to k.

Before reading on: where is $\mathrm{S}(n) \notin \mathrm{S}(n)$ *used?*

Proof. **Step 0 (facts about $\mathrm{S}(n)$).** Let $n, k, m \in \omega$. Each of $\mathrm{S}(n)$, $\mathrm{S}(m)$ is an element of ω by (P2) of Theorem 1.4.2. We use five facts.

(i) $0 \in \mathrm{S}(n)$, by Lemma 1.4.1 (e), since $\mathrm{S}(n) \neq 0$ by (P3) of Theorem 1.4.2.

(ii) If $\mathrm{S}(k) \in \mathrm{S}(n)$, then $k \in \mathrm{S}(n)$: we have $k \in k \cup \{k\} = \mathrm{S}(k)$, and $\mathrm{S}(k) \subseteq \mathrm{S}(n)$ because $\mathrm{S}(n)$ is a transitive set (Lemma 1.4.1 (a), Definition 1.4.1).

(iii) If $k \in \mathrm{S}(m)$, then $\mathrm{S}(k) \subseteq \mathrm{S}(m)$: $k \subseteq \mathrm{S}(m)$ because $\mathrm{S}(m)$ is transitive, and $\{k\} \subseteq \mathrm{S}(m)$, so $k \cup \{k\} \subseteq \mathrm{S}(m)$.

(iv) $\mathrm{S}(n) \notin \mathrm{S}(n)$, by Lemma 1.4.1 (c) applied to $\mathrm{S}(n) \in \omega$.

(v) The elements of $\mathrm{S}(\mathrm{S}(n)) = \mathrm{S}(n) \cup \{\mathrm{S}(n)\}$ are the elements of $\mathrm{S}(n)$ and $\mathrm{S}(n)$ itself (Definition 1.3.1). Every element of $\mathrm{S}(n)$ lies in ω, by Lemma 1.4.1 (b).

Step 1 (uniqueness). Let $f, f'\colon \omega \to X$ both satisfy the two equations, and let $P(n)$ be "$f(n) = f'(n)$". $P(0)$ holds, since $f(0) = a = f'(0)$. If $P(n)$ holds, then

$$f(\mathrm{S}(n)) = g(n, f(n)) = g(n, f'(n)) = f'(\mathrm{S}(n)),$$

so $P(\mathrm{S}(n))$ holds. By Theorem 1.3.3 (b), $P(n)$ holds for every $n \in \omega$. The two functions have the same domain and codomain, so $f = f'$ by Sets, Relations, and Cardinality, Proposition 1.4.1.

Step 2 (partial solutions and their restrictions). For $n \in \omega$, a **partial solution up**

to n is a function $h\colon \mathrm{S}(n) \to X$ with $h(0) = a$ and $h(\mathrm{S}(k)) = g(k, h(k))$ for every $k \in \omega$ with $\mathrm{S}(k) \in \mathrm{S}(n)$. The equations make sense: $0 \in \mathrm{S}(n)$ by (i), and for such k, also $k \in \mathrm{S}(n)$ by (ii).

Claim. If h is a partial solution up to m and $j \in \mathrm{S}(m)$, then the restriction $h|_{\mathrm{S}(j)}$ is a partial solution up to j.

Proof of Claim. Here $j \in \omega$ by (v), and $\mathrm{S}(j) \subseteq \mathrm{S}(m)$ by (iii), so the restriction (*Sets, Relations, and Cardinality*, Section 1.4) is a function $\mathrm{S}(j) \to X$ with the same values as h. Its value at 0 is $h(0) = a$, since $0 \in \mathrm{S}(j)$ by (i). Let $i \in \omega$ with $\mathrm{S}(i) \in \mathrm{S}(j)$. Then $\mathrm{S}(i) \in \mathrm{S}(m)$, because $\mathrm{S}(j) \subseteq \mathrm{S}(m)$, and $i \in \mathrm{S}(j)$ by (ii). So the equation at $\mathrm{S}(i)$ is one of the equations of h, and both of its sides are values at points of $\mathrm{S}(j)$:

$$h|_{\mathrm{S}(j)}(\mathrm{S}(i)) = g(i, h(i)) = g(i, h|_{\mathrm{S}(j)}(i)).$$

■

Step 3.

Claim. For every $n \in \omega$ there is exactly one partial solution up to n.

Proof of Claim. We use Theorem 1.3.3 (b) on n.

Base case. $\mathrm{S}(0) = \emptyset \cup \{0\} = \{0\}$. The graph $\{(0, a)\}$ with codomain X is a function on $\{0\}$, since $a \in X$. It is a partial solution up to 0: its value at 0 is a, and the second condition is vacuous, because $\mathrm{S}(k) \in \{0\}$ would mean $\mathrm{S}(k) = 0$, which (P3) of Theorem 1.4.2 excludes. Any partial solution up to 0 has the value a at 0, so it equals this one by Sets, Relations, and Cardinality, Proposition 1.4.1.

Inductive step. Let $n \in \omega$, and suppose there is exactly one partial solution up to n; call it f_n, with graph $G_n \subseteq \mathrm{S}(n) \times X$. Since $n \in \mathrm{S}(n)$, the value $f_n(n)$ is defined. Let

$$G' \coloneqq G_n \cup \Big\{\big(\mathrm{S}(n),\ g(n, f_n(n))\big)\Big\}.$$

We check Sets, Relations, and Cardinality, Definition 1.4.1 for G' on $\mathrm{S}(\mathrm{S}(n))$ with codomain X. By (v) and because g takes values in X, every pair of G' lies in $\mathrm{S}(\mathrm{S}(n)) \times X$. (Fn1): by (v), an input is an element of $\mathrm{S}(n)$, which has an output in G_n, or is $\mathrm{S}(n)$, which has the added output. (Fn2): every first coordinate in G_n lies in $\mathrm{S}(n)$, while $\mathrm{S}(n) \notin \mathrm{S}(n)$ by (iv). So the added pair is the only pair of G' with first coordinate $\mathrm{S}(n)$, and for an input in $\mathrm{S}(n)$ all pairs of G' lie in G_n, which satisfies (Fn2). Hence $f' \coloneqq (G', X)$ is a function $\mathrm{S}(\mathrm{S}(n)) \to X$ that agrees with f_n on $\mathrm{S}(n)$.

It is a partial solution up to $\mathrm{S}(n)$. Its value at 0 is $f_n(0) = a$, since $0 \in \mathrm{S}(n)$. Let $k \in \omega$ with $\mathrm{S}(k) \in \mathrm{S}(\mathrm{S}(n))$. By (v), there are two cases.

Case 1: $\mathrm{S}(k) \in \mathrm{S}(n)$. Then $k \in \mathrm{S}(n)$ by (ii), and since f_n is a partial solution,

$$f'(\mathrm{S}(k)) = f_n(\mathrm{S}(k)) = g(k, f_n(k)) = g(k, f'(k)).$$

Case 2: $\mathrm{S}(k) = \mathrm{S}(n)$. Then $k = n$ by (P4) of Theorem 1.4.2, so the equation at the new input asks for $g(n, f'(n))$ and for nothing else; and $f'(\mathrm{S}(n)) = g(n, f_n(n)) = g(n, f'(n))$ by the choice of the added pair.

It is the only one. Let h be a partial solution up to $\mathrm{S}(n)$. Since $n \in \mathrm{S}(\mathrm{S}(n))$ by (v), the claim of Step 2 (with $m = \mathrm{S}(n)$ and $j = n$) makes $h|_{\mathrm{S}(n)}$ a partial solution up to n, so it is f_n by the induction hypothesis. Thus $h(k) = f_n(k) = f'(k)$ for every $k \in \mathrm{S}(n)$, and, by the equation of h at $\mathrm{S}(n) \in \mathrm{S}(\mathrm{S}(n))$,

$$h(\mathrm{S}(n)) = g(n, h(n)) = g(n, f_n(n)) = f'(\mathrm{S}(n)).$$

By (v) these are all the inputs, so $h = f'$ by Sets, Relations, and Cardinality, Proposition 1.4.1. This proves the claim.

■

For each $n \in \omega$, let f_n be the partial solution up to n, and G_n its graph. One consequence of the two claims is used below. If $k \in \mathrm{S}(m)$, then $f_m|_{\mathrm{S}(k)}$ is a partial solution up to k (Step 2), hence equals f_k (Step 3), and since $k \in \mathrm{S}(k)$,

$$f_m(k) = f_k(k) \qquad \text{whenever } k \in \mathrm{S}(m). \tag{$*$}$$

Step 4 (glue). Each G_n is a subset of $\mathrm{S}(n) \times X \subseteq \omega \times X$ (by (v)). Let

$$\mathcal{S} \coloneqq \{G \in \mathcal{P}(\omega \times X) : (G, X) \text{ is a partial solution up to some } n \in \omega\},$$

a set by (Z5) and (Z6) of Axiom 1.1.1. By Step 3, the elements of $\mathcal{S}$ are exactly the graphs G_n. Let $F \coloneqq \bigcup \mathcal{S}$, a set by (Z4); a pair lies in F if and only if it lies in some G_n, and $F \subseteq \omega \times X$. We check Sets, Relations, and Cardinality, Definition 1.4.1 for F.

(Fn1): for $k \in \omega$, the pair $(k, f_k(k))$ lies in $G_k \in \mathcal{S}$, hence in F.

(Fn2): let $(k, y) \in F$ and $(k, y') \in F$, say $(k, y) \in G_m$ and $(k, y') \in G_n$. Then k lies in the domains of both, $k \in \mathrm{S}(m)$ and $k \in \mathrm{S}(n)$, and by $(*)$,

$$y = f_m(k) = f_k(k) = f_n(k) = y'.$$

No comparison of m with n was needed: both pieces were cut down to $\mathrm{S}(k)$, where each is the unique partial solution up to k.

So $f \coloneqq (F, X)$ is a function $\omega \to X$. If $k \in \mathrm{S}(m)$, then $(k, f_m(k)) \in G_m \subseteq F$, so $f(k) = f_m(k)$. Hence $f(0) = f_0(0) = a$. For $n \in \omega$, both n and $\mathrm{S}(n)$ lie in $\mathrm{S}(\mathrm{S}(n))$ by (v), and $f_{\mathrm{S}(n)}$ is a partial solution up to $\mathrm{S}(n)$, so

$$f(\mathrm{S}(n)) = f_{\mathrm{S}(n)}(\mathrm{S}(n)) = g(n, f_{\mathrm{S}(n)}(n)) = g(n, f(n)).$$

This proves existence, and with Step 1, the theorem.

■

The proof is that of *Sets, Relations, and Cardinality*, Section 2.3 with $\{0, \dots, n\}$ replaced by $\mathrm{S}(n)$, and with one departure, in (Fn2) of Step 4. There, *Sets, Relations, and Cardinality*, Section 2.3 used that two indices m, n satisfy $m \leq n$ or $n < m$, and that $\{0, \dots, m\} \subseteq \{0, \dots, n\}$ when $m \leq n$. On ω those are facts about order, which the next section proves. Cutting both pieces down to $\mathrm{S}(k)$ avoids the comparison altogether, so the recursion theorem does not have to wait for the order. In Step 3, (P4) did a quiet job: the new input $\mathrm{S}(n)$ is the successor of n and of nothing else, so the added pair had to satisfy exactly one equation.

As in *Sets, Relations, and Cardinality*, Section 2.3, no choice is made: each partial solution is the only one, and $\mathcal{S}$ is cut out of a power set by a condition, so no family $n \mapsto f_n$ had to be formed first. From now on a definition "$f(0) \coloneqq a$, $f(\mathrm{S}(n)) \coloneqq g(n, f(n))$" on ω is legitimate whenever $a \in X$ and $g\colon \omega \times X \to X$ is a function, and f is **defined by recursion**.

Remark. The functions g that we feed to the theorem are usually given by a formula, such as $g(n, x) = \mathrm{S}(x)$. Such a g is a function as soon as each pair (n, x) gets exactly one value in X: its graph, the set of pairs $((n, x), y) \in (\omega \times X) \times X$ such that y is the value the formula assigns to (n, x), is a set by (Z6) of Axiom 1.1.1 inside a product (Proposition 1.2.2), and (Fn1), (Fn2) say exactly that the formula gives one value in X for each pair.

Warning. The hypothesis that g takes its values in X cannot be dropped. Let $X = \{0\}$, $a = 0$, and $g(n, x) \coloneqq \mathrm{S}(0)$ for $n \in \omega$, $x \in X$, so that g maps $\omega \times X$ into ω but not into X. Any $f\colon \omega \to X$ with the two equations would need $f(\mathrm{S}(0)) = g(0, f(0)) = \mathrm{S}(0)$, and $\mathrm{S}(0) \notin X$ since $\mathrm{S}(0) \neq 0$ by (P3) of Theorem 1.4.2; so no such f exists. The proof breaks in Step 3, where the added pair $(\mathrm{S}(n), g(n, f_n(n)))$ must lie in $\mathrm{S}(\mathrm{S}(n)) \times X$.

1.5.2 Addition and multiplication

With the theorem, addition is the old school definition made official. To add n to m, start at m and take n successor steps: $m + 0$ is m, and $m + \mathrm{S}(n)$ is one step further than $m + n$. Multiplication is repeated addition in the same way.

Adding n to m is taking n successor steps from m.

Definition 1.5.1. Addition and Multiplication on ω

(a) For each $m \in \omega$, let $A_m\colon \omega \to \omega$ be the function given by Theorem 1.5.1 with $X = \omega$, $a = m$ and $g(n, x) = \mathrm{S}(x)$. The **sum** of m and n is $m + n \coloneqq A_m(n)$. Thus, for all $m, n \in \omega$,

$$m + 0 = m, \qquad m + \mathrm{S}(n) = \mathrm{S}(m + n).$$

(b) For each $m \in \omega$, let $M_m\colon \omega \to \omega$ be the function given by Theorem 1.5.1 with $X = \omega$, $a = 0$ and $g(n, x) = x + m$. The **product** of m and n is $m \cdot n \coloneqq M_m(n)$, also written mn. Thus, for all $m, n \in \omega$,

$$m \cdot 0 = 0, \qquad m \cdot \mathrm{S}(n) = m \cdot n + m.$$

(c) The **numerals** are $0 \coloneqq \emptyset$ and $1 \coloneqq \mathrm{S}(0)$, $2 \coloneqq \mathrm{S}(1)$, $3 \coloneqq \mathrm{S}(2)$, and so on, as in Example (The first natural numbers as sets). For every $n \in \omega$, $n + 1 = \mathrm{S}(n)$; in particular $1 + 1 = 2$ and $k + 2 = (k + 1) + 1$ for every $k \in \omega$.

In words: (a) and (b) are the two clauses we would write for a school definition, and the theorem guarantees that they define exactly one function A_m, respectively M_m, for each m. Multiplication binds more tightly than addition, so $m \cdot n + m$ is $(m \cdot n) + m$.

Well-definedness. Three things need checking. First, each g is a function $\omega \times \omega \to \omega$ (the remark above): $\mathrm{S}(x) \in \omega$ for $x \in \omega$ by (P2) of Theorem 1.4.2, and $x + m \in \omega$ because A_x maps into ω. So (b) uses (a), and (a) uses nothing but the successor. Second, $+$ is a function

$\omega \times \omega \to \omega$: its graph is the set of $((m,n),k) \in (\omega \times \omega) \times \omega$ such that $k = A(n)$ for some function $A\colon \omega \to \omega$ with $A(0) = m$ and $A(\mathrm{S}(j)) = \mathrm{S}(A(j))$ for every $j \in \omega$; this is a set by (Z6) of Axiom 1.1.1, and each (m,n) gets exactly one k because A_m exists and is unique. The same holds for $\cdot$. Third, nothing was chosen: each A_m and M_m is the only function with its property.

The successor is "plus one". Proof of the last sentence of (c): for every $n \in \omega$,

$$n + 1 = n + \mathrm{S}(0) = \mathrm{S}(n + 0) = \mathrm{S}(n),$$

by (c) and the two equations of (a). From now on we write $n + 1$ for $\mathrm{S}(n)$ whenever $n \in \omega$; the defining equations then read as in school:

$$m + 0 = m, \quad m + (n + 1) = (m + n) + 1,$$
$$m \cdot 0 = 0, \quad m \cdot (n + 1) = m \cdot n + m.$$

Two small computations are used repeatedly later: $1 + 1 = \mathrm{S}(1) = 2$, and, for every $k \in \omega$,

$$k + 2 = k + \mathrm{S}(1) = \mathrm{S}(k + 1) = (k + 1) + 1.$$

A non-example by minimal change: keep $a = m$ in (a), but replace $g(n,x) = \mathrm{S}(x)$ by $g(n,x) = \mathrm{S}(n)$, which ignores the current value. This g is still a function $\omega \times \omega \to \omega$, so Theorem 1.5.1 gives exactly one $f\colon \omega \to \omega$ with $f(0) = m$ and $f(\mathrm{S}(n)) = \mathrm{S}(n)$. The first equation $m + 0 = m$ still holds for this "sum", but the second, $m + \mathrm{S}(n) = \mathrm{S}(m + n)$, fails: with $m = 2$, $f(1) = \mathrm{S}(0) = 1$, while $\mathrm{S}(f(0)) = \mathrm{S}(2) = 3$. So it would give $2 + 1 = 1$, and it is not addition.

The defining equations immediately give the two Peano facts in the "plus one" notation.

Corollary 1.5.2. Successor Cancellation

For all $k, l \in \omega$: if $k + 1 = l + 1$, then $k = l$; and $k + 1 \neq 0$.

Proof. By (c) of Definition 1.5.1, $k + 1 = \mathrm{S}(k)$ and $l + 1 = \mathrm{S}(l)$. So the first claim is (P4) of Theorem 1.4.2, and the second is (P3). ∎

Quick check. Compute $2 + 2$ and $2 \cdot 2$ from the defining equations alone.

Solution. By (a) of Definition 1.5.1, three times,

$$2 + 2 = \mathrm{S}(2 + 1) = \mathrm{S}(\mathrm{S}(2 + 0)) = \mathrm{S}(\mathrm{S}(2)) = \mathrm{S}(3) = 4.$$

By (b), twice and then at 0,

$$2 \cdot 2 = 2 \cdot 1 + 2 = (2 \cdot 0 + 2) + 2 = (0 + 2) + 2,$$

and $0 + 2 = \mathrm{S}(0 + 1) = \mathrm{S}(\mathrm{S}(0 + 0)) = \mathrm{S}(\mathrm{S}(0)) = 2$. So $2 \cdot 2 = 2 + 2 = 4$.

1.5.3 The laws of arithmetic

Every law of arithmetic that *Sets, Relations, and Cardinality*, Chapter 2 took from the real numbers must now be proved, and the only tools are the four defining equations and induction. The equations are lopsided: they say what happens when the **second** argument goes up by one, and nothing about the first. So in every induction below we fix all variables but one and induct on the variable that the recursion runs on, the second argument of $+$ or $\cdot$.

Theorem 1.5.3. Laws of Arithmetic on ω

For all $k, m, n \in \omega$:

(a) $0 + n = n$, and $(m + 1) + n = (m + n) + 1$;
(b) $m + n = n + m$;
(c) $(k + m) + n = k + (m + n)$;
(d) $m(n + k) = mn + mk$ and $(n + k)m = nm + km$;
(e) $0 \cdot n = 0$, $(m + 1) \cdot n = mn + n$, and $n \cdot 1 = 1 \cdot n = n$;
(f) $mn = nm$;
(g) $(km)n = k(mn)$.

Idea. The move is **fix the others, induct on the recursion variable**: in $m + n$ and $m \cdot n$ the recursion runs on n, so we fix m (and k) and induct on n. Then the base case is a defining equation at 0, and the step can open $m + (n + 1)$ or $m \cdot (n + 1)$ by the other defining equation, apply the induction hypothesis inside, and close up again. The order of the clauses matters: commutativity (b) needs the two facts (a) about the **first** argument, which no defining equation provides.

Proof. Each clause except the second half of (d) and the last part of (e), which the table derives from earlier clauses without induction, is proved by Theorem 1.3.3 (b), applied to the property of n stated in the clause, with the other variables fixed; for the first half of (d) the variable is k, the recursion variable of $m(n + k)$. "Def" refers to the equations of Definition 1.5.1, "IH" to the induction hypothesis.

We prove (c), associativity of addition, in full. Fix $k, m \in \omega$, and let $P(n)$ be "$(k+m)+n = k + (m + n)$". *Base case:* by the first equation of (a) of Definition 1.5.1, applied twice,

$$(k + m) + 0 = k + m = k + (m + 0).$$

Inductive step: suppose $P(n)$. Then

$$\begin{aligned}(k + m) + (n + 1) &= ((k + m) + n) + 1 = (k + (m + n)) + 1 \\ &= k + ((m + n) + 1) = k + (m + (n + 1)),\end{aligned}$$

where the first equality is Def with first argument $k+m$, the second is $P(n)$, the third is Def with first argument k and second argument $m+n$, and the fourth is Def with first argument m. So $P(n + 1)$ holds, and by Theorem 1.3.3 (b), $P(n)$ holds for every $n \in \omega$.

The other clauses follow the same pattern; the table gives, for each, the property inducted

on, the base case, the step, and the earlier clauses used. They are proved in the order of the table, so each row uses only rows above it.

Clause, variable	Base case	Inductive step (IH at the marked equality)	Uses
(a) $0+n=n$; on n	$0+0=0$ (Def)	$0+(n+1)=$ $(0+n)+1 \overset{\text{IH}}{=} n+1$	Def
(a) $(m+1)+n=$ $(m+n)+1$; m fixed, on n	both sides are $m+1$ (Def)	$(m+1)+(n+1)=$ $((m+1)+n)+1 \overset{\text{IH}}{=}$ $((m+n)+1)+1=$ $(m+(n+1))+1$	Def
(b) $m+n=n+m$; m fixed, on n	$m+0=m=0+m$	$m+(n+1)=$ $(m+n)+1 \overset{\text{IH}}{=} (n+$ $m)+1=(n+1)+m$	Def; (a), both halves
(c)	in full above		Def
(d) $m(n+k)=mn+mk$; m,n fixed, on k	$m(n+0)=mn=$ $mn+0=mn+m\cdot 0$	$m(n+(k+1))=$ $m((n+k)+1)=$ $m(n+k)+m \overset{\text{IH}}{=}$ $(mn+mk)+m=$ $mn+(mk+m)=$ $mn+m(k+1)$	Def; (c)
(e) $0\cdot n=0$; on n	$0\cdot 0=0$ (Def)	$0\cdot(n+1)=$ $0\cdot n+0 \overset{\text{IH}}{=} 0+0=0$	Def
(e) $(m+1)n=mn+n$; m fixed, on n	$(m+1)\cdot 0=0=$ $0+0=m\cdot 0+0$	$(m+1)(n+1)=$ $(m+1)n+(m+1) \overset{\text{IH}}{=}$ $(mn+n)+(m+1)=$ $mn+(n+(m+1))=$ $mn+((n+m)+1)=$ $mn+((m+n)+1)=$ $mn+(m+(n+1))=$ $(mn+m)+(n+1)=$ $m(n+1)+(n+1)$	Def; (b), (c)
(e) $n\cdot 1=n$, $1\cdot n=n$; no induction		$n\cdot 1=n\cdot 0+n=$ $0+n=n$; $1\cdot n=(0+1)\cdot n=$ $0\cdot n+n=0+n=n$	Def; (a); first two parts of (e)
(f) $mn=nm$; m fixed, on n	$m\cdot 0=0=0\cdot m$	$m(n+1)=mn+m \overset{\text{IH}}{=}$ $nm+m=(n+1)m$	Def; (e)
(d), second half; no induction		$(n+k)m=$ $m(n+k)=$ $mn+mk=nm+km$	(d), first half; (f)
(g) $(km)n=k(mn)$; k,m fixed, on n	$(km)\cdot 0=0=$ $k\cdot 0=k(m\cdot 0)$	$(km)(n+1)=$ $(km)n+km \overset{\text{IH}}{=}$ $k(mn)+km=$ $k(mn+m)=$ $k(m(n+1))$	Def; (d)

In the step of (a), second half, the third equality is $(m+n)+1 = m+(n+1)$ (Def), read inside the outer $+1$; in the step of (b), the last equality is (a), second half, read from right to left. In the long step of (e), the equalities after IH use, in order, (c), Def, (b), Def, (c), Def. In the second half of (e), $1 = 0+1$ is the first half of (a), and $(0+1)\cdot n = 0\cdot n + n$ is the second part of (e) with $m = 0$. In (g), $k(mn+m) = k(mn) + km$ is the first half of (d). This proves the theorem.

■

The laws (a)–(g) are the rules of school arithmetic for $+$ and $\cdot$ on the natural numbers, now proved from the seven axioms for sets. The next section defines the order of ω and ties it to them: $m \le n$ exactly when $n = m + j$ for some $j \in \omega$.

Warning. **Do not assume $0+n = n$.** It is clause (a) of Theorem 1.5.3, proved by induction; the defining equation is $n + 0 = n$. For the same reason, inducting on m in $m+n = n+m$ gets stuck: the step would need $(m+1)+n$, and the definition says nothing about a sum whose **first** argument is a successor until (a) is proved.

The order of the table is forced by the lopsided definition. Commutativity, which lets us move a successor from the second place to the first, comes only after the two facts in (a) that move it by hand; once it is available, (e) and (f) follow the same way for multiplication.

1.5.4 Exercises

1.5.4.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 1.5.1.

(a) Where in the proof is $\mathrm{S}(n) \notin \mathrm{S}(n)$ used, and what would fail without it?

(b) Run the construction for $X = \omega$, $a = 0$ and $g(n,x) = \mathrm{S}(\mathrm{S}(x))$: write the graphs of the partial solutions up to 0, 1 and 2.

(c) Why does the uniqueness half need only an induction, while existence needs partial solutions?

(d) State the idea of the proof in two sentences.

Solution.

(a) In the inductive step of Step 3, in checking (Fn2) for the extended graph G'. The added pair has first coordinate $\mathrm{S}(n)$, and every first coordinate in G_n lies in $\mathrm{S}(n)$; because $\mathrm{S}(n) \notin \mathrm{S}(n)$, the input $\mathrm{S}(n)$ is new, so adding the pair cannot give an old input a second output. If $\mathrm{S}(n)$ were an element of $\mathrm{S}(n)$, then G_n would already contain a pair $(\mathrm{S}(n), y)$, and G' would contain both it and $(\mathrm{S}(n), g(n, f_n(n)))$, so G' might fail to be a function.

(b) Here $g(n,x) = \mathrm{S}(\mathrm{S}(x))$ adds two steps. The partial solution up to 0 has graph $\{(0,0)\}$. The next adds the pair at $\mathrm{S}(0) = 1$ with value $g(0,0) = \mathrm{S}(\mathrm{S}(0)) = 2$: the graph up to 1 is $\{(0,0),(1,2)\}$. The next adds the pair at 2 with value $g(1,2) = \mathrm{S}(\mathrm{S}(2)) = 4$: the

graph up to 2 is $\{(0,0),(1,2),(2,4)\}$. Each graph contains the previous one, as Step 3 says.

(c) Uniqueness compares two functions that are **given**, value by value, and comparing values at $n+1$ through the values at n is exactly an induction on n. Existence asks for a function that is not given; induction proves statements and builds nothing. So the proof proves by induction a statement (the claim of Step 3) about finite pieces, and builds f from them as a union, using (Z4).

(d) For each n, induction shows that there is exactly one function on $\mathrm{S}(n)$ obeying the equations, and any two of these agree at every point k where both are defined, because on $\mathrm{S}(k)$ each restricts to the unique one up to k. Their union is the required function, and an induction shows there is no other.

1.5.4.2 B. Practice

Exercise. B1: Small products

Compute $3 \cdot 2$ and $2 \cdot 3$ from the defining equations of Definition 1.5.1 alone, without using Theorem 1.5.3. Here $5 := \mathrm{S}(4)$ and $6 := \mathrm{S}(5)$.

Solution. For $3 \cdot 2$, by (b) twice and then at 0:

$$3 \cdot 2 = 3 \cdot 1 + 3 = (3 \cdot 0 + 3) + 3 = (0 + 3) + 3.$$

By (a), $0 + 3 = \mathrm{S}(0 + 2) = \mathrm{S}(\mathrm{S}(0 + 1)) = \mathrm{S}(\mathrm{S}(\mathrm{S}(0 + 0))) = \mathrm{S}(\mathrm{S}(\mathrm{S}(0))) = 3$. Then

$$3 + 3 = \mathrm{S}(3 + 2) = \mathrm{S}(\mathrm{S}(3 + 1)) = \mathrm{S}\Big(\mathrm{S}(\mathrm{S}(3 + 0))\Big) = \mathrm{S}(\mathrm{S}(\mathrm{S}(3))) = 6.$$

So $3 \cdot 2 = 6$.

For $2 \cdot 3$, by (b), $2 \cdot 3 = 2 \cdot 2 + 2$, and $2 \cdot 2 = 4$ by the Quick check above. Then

$$4 + 2 = \mathrm{S}(4 + 1) = \mathrm{S}(\mathrm{S}(4 + 0)) = \mathrm{S}(\mathrm{S}(4)) = 6.$$

So $2 \cdot 3 = 6 = 3 \cdot 2$, as (f) of Theorem 1.5.3 predicts; but the two computations took different routes, and neither used a law.

Exercise. B2: Distributivity, last step blank

Here is the proof of $m(n + k) = mn + mk$ for all $m, n, k \in \omega$, with the last step left blank. *Name the move, then complete it.*

> Fix $m, n \in \omega$ and induct on k. Base case: $m(n+0) = mn = mn+0 = mn+m \cdot 0$. Inductive step: suppose $m(n + k) = mn + mk$. Then
>
> $$m(n+(k+1)) = m((n+k)+1) = m(n+k)+m = (mn+mk)+m = \_\_\_\_\_.$$

Justify each equality you add.

Solution. The move is **fix the others, induct on the recursion variable**: in $m(n+k)$ the recursion of $+$ runs on k, so m and n are fixed and the induction is on k.

The completion is

$$(mn+mk)+m = mn+(mk+m) = mn+m(k+1),$$

where the first equality is associativity, (c) of Theorem 1.5.3, and the second is the defining equation $m(k+1) = mk+m$ of Definition 1.5.1 (b), read from right to left. This is the statement for $k+1$, so by Theorem 1.3.3 (b) it holds for every $k \in \omega$. (This is clause (d) of Theorem 1.5.3, already proved; the exercise practices the move.)

1.5.4.3 C. Going deeper

Exercise. C1: Exit check: addition is commutative

From the two defining equations of addition, $m+0 = m$ and $m+(n+1) = (m+n)+1$, and Theorem 1.3.3 alone, prove that $m+n = n+m$ for all $m, n \in \omega$. Prove every auxiliary fact you need.

Solution (Hint 1). The definition says what happens when the second argument is 0 or a successor. To use it on $n+m$, you need to know the same for the first argument: prove two lemmas first.

Solution (Hint 2). The lemmas are $0+n = n$ and $(m+1)+n = (m+n)+1$, each by induction on n.

Solution. All inductions are by Theorem 1.3.3 (b).

Lemma 1: $0+n = n$ for every n. Induct on n. Base: $0+0 = 0$, by the first equation. Step: if $0+n = n$, then $0+(n+1) = (0+n)+1 = n+1$, by the second equation and the hypothesis.

Lemma 2: $(m+1)+n = (m+n)+1$ for all m, n. Fix m and induct on n. Base: $(m+1)+0 = m+1 = (m+0)+1$, by the first equation twice. Step: if $(m+1)+n = (m+n)+1$, then

$$(m+1)+(n+1) = \big((m+1)+n\big)+1 = \big((m+n)+1\big)+1 = \big(m+(n+1)\big)+1,$$

by the second equation, the hypothesis, and the second equation read from right to left inside the outer $+1$.

Main induction. Fix m and induct on n for "$m+n = n+m$". Base: $m+0 = m = 0+m$, by the first equation and Lemma 1. Step: if $m+n = n+m$, then

$$m+(n+1) = (m+n)+1 = (n+m)+1 = (n+1)+m,$$

by the second equation, the hypothesis, and Lemma 2 (with n in the role of m and m in the role of n), read from right to left. Hence $m+n = n+m$ for all $m, n \in \omega$.

Marking scheme. (10 points) - 2: proves $0+n=n$ by induction on n, from the first equation (base) and the second (step). - 3: proves, by an induction with the other variable fixed, a lemma that moves a successor out of the first argument (the book's $(m+1)+n=(m+n)+1$, or $1+n=n+1$ together with associativity of addition), base and step both justified. - 2: in the main induction, the base case $m+0=0+m$, using the first lemma. - 2: in the main induction, the step, using the second equation, the hypothesis and that lemma. - 1: names the variable inducted on in each induction and fixes the other. Fatal errors (0 for the step): citing commutativity, or any law of Theorem 1.5.3, without proving it from the two equations; treating $0+n=n$ as if it were a defining equation. A second route inducts on m in the main induction with n fixed; it needs the same two lemmas, and its base and step map to the same 2 + 2 points. A third route proves associativity and $1+n=n+1$ (each by induction on n, from the equations and $0+n=n$) and then closes the step as $m+(n+1)=(m+n)+1=(n+m)+1=n+(m+1)=n+(1+m)=(n+1)+m$; those two lemmas together earn the 3 points.

Exercise. C2: Associativity of multiplication, last half blank

Prove that $(km)n=k(mn)$ for all $k,m,n\in\omega$. The first half is given; *name the move, then complete it.*

> Fix $k,m\in\omega$ and induct on n. Base case: $(km)\cdot 0=0=k\cdot 0=k(m\cdot 0)$. Inductive step: _____

You may use (a)–(f) of Theorem 1.5.3.

Solution. The move is **fix the others, induct on the recursion variable**: in $(km)n$ and $k(mn)$ the variable n is the second argument of the outer and the inner product, where the recursion runs, so k and m are fixed.

Inductive step. Suppose $(km)n=k(mn)$. Then

$$\begin{aligned}(km)(n+1)&=(km)n+km=k(mn)+km\\&=k(mn+m)=k(m(n+1)),\end{aligned}$$

where the first equality is the defining equation of Definition 1.5.1 (b) with first factor km, the second is the induction hypothesis, the third is (d) of Theorem 1.5.3 read from right to left ($k(x+y)=kx+ky$ with $x=mn$, $y=m$), and the fourth is the defining equation with first factor m. By Theorem 1.3.3 (b), $(km)n=k(mn)$ for every n, and k,m were arbitrary.

The base case used the defining equation $x\cdot 0=0$ three times, with $x=km$, $x=k$ and $x=m$. Commutativity (f) is not needed for this route; a proof that first rewrites km as mk would need it.

Exercise. C3: Doubling

Show that there is exactly one function $D\colon\omega\to\omega$ with $D(0)=0$ and $D(n+1)=(D(n)+1)+1$ for every $n\in\omega$, and prove that $D(n)=n+n$ for every $n\in\omega$.

Solution. Existence and uniqueness. The function $g\colon \omega \times \omega \to \omega$, $g(n,x) := \mathrm{S}(\mathrm{S}(x)) = (x+1)+1$, maps into ω by (P2) of Theorem 1.4.2 applied twice. By Theorem 1.5.1 with $X = \omega$ and $a = 0$, there is exactly one $D\colon \omega \to \omega$ with $D(0) = 0$ and $D(n+1) = g(n, D(n)) = (D(n)+1)+1$. (Its partial solutions up to $0, 1, 2$ are those of Exercise (A1) (b), so $D(1) = 2$ and $D(2) = 4$.)

The formula. Induct on n by Theorem 1.3.3 (b) for "$D(n) = n+n$". Base: $D(0) = 0 = 0+0$, by the defining equation $m + 0 = m$. Step: suppose $D(n) = n + n$. Then

$$\begin{aligned}(n+1)+(n+1) &= \big((n+1)+n\big)+1 = \big((n+n)+1\big)+1 \\ &= \big(D(n)+1\big)+1 = D(n+1),\end{aligned}$$

where the first equality is the defining equation of Definition 1.5.1 (a), the second is (a) of Theorem 1.5.3, $(m+1)+n = (m+n)+1$ with $m = n$, the third is the hypothesis, and the last is the equation of D. Hence $D(n) = n + n$ for every $n \in \omega$.

Alternatively, $n \mapsto n+n$ satisfies the two equations that define D (the computation above, read without the hypothesis, shows $(n+1)+(n+1) = ((n+n)+1)+1$), so it is D by the uniqueness half of Theorem 1.5.1. Either way the double $n+n$ is $2n$ too: $2n = n \cdot 2 = n \cdot 1 + n = n + n$ by (f), the defining equation of $\cdot$, and (e) of Theorem 1.5.3.

1.6 Order and Well-Ordering of the Natural Numbers

Sets, Relations, and Cardinality, Chapter 2 compared natural numbers with the order of the real line, and proved that every non-empty set of natural numbers has a least element (Sets, Relations, and Cardinality, Theorem 2.1.3) from that order and induction. In ω there is no real line to borrow from. But the von Neumann numerals carry their order inside them: $3 = \{0, 1, 2\}$ is the set of the numbers before it. This section takes "is an element of" as the order, proves that it is a strict total order and a well-ordering, ties it to the arithmetic of the previous section, and then collects three groups of facts about ω: powers and parity, n-tuples, and lists of variable length.

1.6.1 Less than is "element of"

The numerals of Example (The first natural numbers as sets) were built so that each one collects all the earlier ones: $1 = \{0\}$, $2 = \{0, 1\}$, $4 = \{0, 1, 2, 3\}$. So "m comes before n" can be read off the sets themselves, without any new structure.

m is less than n when m is one of the elements of n.

Definition 1.6.1. Order on ω

For $m, n \in \omega$, we write $m < n$ if $m \in n$, and $m \leq n$ if $m < n$ or $m = n$. As usual, $n > m$ means $m < n$, and $n \geq m$ means $m \leq n$.

In words: $m < n$ is membership, read as an order. For example $2 < 4$, because $2 \in \{0, 1, 2, 3\}$; $0 < n$ for every $n \neq 0$, by Lemma 1.4.1 (e); and no m satisfies $m < 0$, because $0 = \emptyset$ has no elements. A minimal change gives a non-example: $4 < 2$ fails, since 4 is not one of $0, 1$. The

relation $\leq$ is "$\in$ or $=$"; exercise B1 shows that on ω it is also $\subseteq$.

Nothing in the definition says that $<$ behaves like an order. For arbitrary sets, membership does not: it can fail to be transitive, as the warning below shows. That it is a strict total order on ω is the first theorem of this section, and the induction theorem does the work.

Theorem 1.6.1. Order on ω

(a) $<$ is a strict total order on ω: for all $k, m, n \in \omega$, never $n < n$; if $k < m$ and $m < n$, then $k < n$; and **exactly one** of $m < n$, $m = n$, $n < m$ holds.

(b) For $m, n \in \omega$, $m < \mathrm{S}(n)$ if and only if $m \leq n$. Hence no $k \in \omega$ satisfies $n < k < \mathrm{S}(n)$.

(c) Every $n \in \omega$ is the set of the natural numbers less than it: $n = \{m \in \omega : m < n\}$. Hence $\{0, \ldots, n\} \coloneqq \{k \in \omega : k \leq n\}$ equals $\mathrm{S}(n)$.

(d) $0 \leq n$ for every $n \in \omega$; if $n \neq 0$, then $1 \leq n$; and $0 < 1$.

(e) For $n \in \omega$, $\{1, \ldots, n\} \coloneqq \{k \in \omega : 1 \leq k \leq n\}$ equals $\mathrm{S}(n) \setminus \{0\}$.

Idea. Irreflexivity and transitivity are facts already proved about the sets in ω: no n is an element of itself, and every n is a transitive set (Lemma 1.4.1). The real content is trichotomy. We prove "for every m: $m \in n$, $m = n$ or $n \in m$" by induction on n, keeping m universally quantified inside the property. In the step, the awkward case is $n \in m$: we must place $\mathrm{S}(n)$ relative to m. A separate induction shows that the successor preserves $\in$, and that settles it. The rest unwinds the definition of $\mathrm{S}(n)$.

Proof.

(a) *Irreflexive:* $n \notin n$ by Lemma 1.4.1 (c). *Transitive:* if $k \in m$ and $m \in n$, then $m \subseteq n$ because n is a transitive set (Lemma 1.4.1 (a)), so $k \in n$.
Claim. For all $m, n \in \omega$: if $m \in n$, then $\mathrm{S}(m) \in \mathrm{S}(n)$.
Proof of Claim. We use Theorem 1.3.3 (b) on n, for the property "for every $m \in \omega$, if $m \in n$, then $\mathrm{S}(m) \in \mathrm{S}(n)$". For $n = 0$ it holds vacuously, since no m lies in $0 = \emptyset$. Suppose it holds for n, and let $m \in \mathrm{S}(n) = n \cup \{n\}$. If $m \in n$, then $\mathrm{S}(m) \in \mathrm{S}(n)$ by the hypothesis, and $\mathrm{S}(n) \subseteq \mathrm{S}(\mathrm{S}(n))$, so $\mathrm{S}(m) \in \mathrm{S}(\mathrm{S}(n))$. If $m = n$, then $\mathrm{S}(m) = \mathrm{S}(n) \in \mathrm{S}(\mathrm{S}(n))$. So the property holds for $\mathrm{S}(n)$. ■
Trichotomy, at least one. We use Theorem 1.3.3 (b) on n, for the property $T(n)$: "for every $m \in \omega$, $m \in n$ or $m = n$ or $n \in m$".
Base case. Let $m \in \omega$. Then $m = 0$ or $0 \in m$, by Lemma 1.4.1 (e). So $T(0)$ holds.
Inductive step. Suppose $T(n)$, and let $m \in \omega$. By $T(n)$ there are three cases. If $m \in n$ or $m = n$, then $m \in n \cup \{n\} = \mathrm{S}(n)$. If $n \in m$, then $\mathrm{S}(n) \in \mathrm{S}(m) = m \cup \{m\}$ by the claim, so $\mathrm{S}(n) \in m$ or $\mathrm{S}(n) = m$. In every case $m \in \mathrm{S}(n)$, $m = \mathrm{S}(n)$ or $\mathrm{S}(n) \in m$, which is $T(\mathrm{S}(n))$.
Trichotomy, at most one. If $m \in n$ and $m = n$, then $n \in n$; if $m = n$ and $n \in m$, then $n \in n$; if $m \in n$ and $n \in m$, then $m \in m$ by transitivity. Each contradicts irreflexivity.

(b) By Definition 1.3.1, $m \in \mathrm{S}(n) = n \cup \{n\}$ if and only if $m \in n$ or $m = n$, that is, $m \leq n$. Now suppose $n < k$ and $k < \mathrm{S}(n)$. Then $k \leq n$, so $k = n$, giving $n < n$, or $k < n$, giving $n < n$ by transitivity. Both contradict irreflexivity.

(c) ($\supseteq$): if $m \in \omega$ and $m < n$, then $m \in n$ by definition. ($\subseteq$): if $m \in n$, then $m \in \omega$, because ω is a transitive set (Lemma 1.4.1 (b)), and $m < n$. Applying this to $\mathrm{S}(n)$,

which lies in ω by (P2) of Theorem 1.4.2, and using (b), $\mathrm{S}(n) = \{k \in \omega : k < \mathrm{S}(n)\} = \{k \in \omega : k \leq n\}$.

(d) By Lemma 1.4.1 (e), $n = 0$ or $0 \in n$; that is, $0 \leq n$. Let $n \neq 0$. By Lemma 1.4.1 (d), $n = \mathrm{S}(k)$ for some $k \in \omega$, and $0 \leq k$. If $k = 0$, then $n = \mathrm{S}(0) = 1$. If $0 < k$, then $1 = \mathrm{S}(0) < \mathrm{S}(k) = n$ by the claim. So $1 \leq n$. Finally $0 \in \{0\} = \mathrm{S}(0) = 1$, so $0 < 1$.

(e) Let $k \in \mathrm{S}(n) \setminus \{0\}$. Then $k \in \omega$ (as in (c)), $k \leq n$ by (b), and $1 \leq k$ by (d), since $k \neq 0$. Conversely, let $k \in \omega$ with $1 \leq k \leq n$. Then $k \in \mathrm{S}(n)$ by (b). And $k \neq 0$: otherwise $1 \leq 0$, but $1 \neq 0$ by (P3) of Theorem 1.4.2 and $1 \notin 0 = \emptyset$. This proves the theorem.

■

So the notation of *Sets, Relations, and Cardinality*, Chapter 2, $\{0, \ldots, n\}$ and $\{1, \ldots, n\}$, keeps its meaning on ω, and the two sets are $\mathrm{S}(n)$ and $\mathrm{S}(n) \setminus \{0\}$; in particular $\{1, \ldots, 0\} = \emptyset$. In the "plus one" notation of Definition 1.5.1, (b) says that $n + 1$ is the next number after n, and (d) with Corollary 1.5.2 gives $1 \leq k + 1$ for every $k \in \omega$, since $k + 1 \neq 0$.

Warning. **On arbitrary sets, $\in$ is not transitive.** $\emptyset \in \{\emptyset\}$ and $\{\emptyset\} \in \{\{\emptyset\}\}$, but $\emptyset \notin \{\{\emptyset\}\}$, whose only element is $\{\emptyset\} \neq \emptyset$. The conclusion "$k \in m$ and $m \in n$ give $k \in n$" is false in general; on ω it holds by a theorem, Lemma 1.4.1 (a): the elements of ω are transitive sets. Membership by itself does not give it.

1.6.2 Well-ordering

In *Sets, Relations, and Cardinality*, Chapter 2, Sets, Relations, and Cardinality, Theorem 2.1.3 was derived from Logic and Proof, Axiom 4.2.1, which there was an axiom. On ω, induction is a theorem (Theorem 1.3.3), so well-ordering becomes one too. The proof here takes a different route from *Sets, Relations, and Cardinality*, Chapter 2's strong induction, and the order of ω as membership makes it short.

Theorem 1.6.2. Well-Ordering of ω

Every **non-empty** subset $A \subseteq \omega$ has a least element: an $a \in A$ with $a \leq x$ for every $x \in A$. The least element is unique.

Idea. Suppose A has no least element. Then we can show, by induction, that no natural number n contains an element of A: $n \cap A = \emptyset$. Passing from n to $\mathrm{S}(n) = n \cup \{n\}$ adds only the element n, and if n were in A while nothing below it is, n would be least. But $\mathrm{S}(a)$ contains a, so no a can lie in A, and A is empty.

Proof. Let $A \subseteq \omega$ be non-empty, and suppose, for a contradiction, that A has no least element. Let

$$B \coloneqq \{n \in \omega : n \cap A = \emptyset\},$$

a set by (Z6) of Axiom 1.1.1. We show that B is inductive in the sense of Theorem 1.3.3 (a). First, $0 \in B$, since $0 = \emptyset$. Next, let $n \in B$. Then

$$\mathrm{S}(n) \cap A = (n \cap A) \cup (\{n\} \cap A) = \{n\} \cap A.$$

Suppose $n \in A$. Let $x \in A$. Then $x \notin n$, because $n \cap A = \emptyset$; that is, $x < n$ fails, so $x = n$ or $n < x$ by trichotomy (Theorem 1.6.1 (a)), and $n \leq x$. Hence n is a least element of A, contrary to our supposition. So $n \notin A$, $\mathrm{S}(n) \cap A = \emptyset$, and $\mathrm{S}(n) \in B$. By Theorem 1.3.3 (a), $B = \omega$.

Since A is non-empty, there is some $a \in A$. Then $\mathrm{S}(a) \in \omega = B$, so $\mathrm{S}(a) \cap A = \emptyset$; but $a \in \mathrm{S}(a) \cap A$. This contradiction shows, by RAA (Logic and Proof, Definition 2.5.1), that A has a least element.

Uniqueness. If a and a' are least elements, then $a \leq a'$ and $a' \leq a$. If $a \neq a'$, these give $a < a'$ and $a' < a$, so $a < a$ by transitivity, contradicting irreflexivity (Theorem 1.6.1 (a)). So $a = a'$.

■

The proof concludes an existence statement from a contradiction, so its last step is RAA, not ¬I: what is refuted is "A has no least element", and the double negation is removed classically. The statement is that of Sets, Relations, and Cardinality, Theorem 2.1.3, read on ω; there the proof went through strong induction on $\mathbb{N} \subseteq \mathbb{R}$, here through the inductive set B.

Both hypotheses are needed. The empty set is a subset of ω with no least element, since a least element must be an element of the set. The hypothesis $A \subseteq \omega$ is needed too. Read with the order of Definition 1.6.1, "$\in$ or $=$", the non-empty set $\{0, \{1\}\} = \{\emptyset, \{\{\emptyset\}\}\}$ has no least element: neither $\emptyset \in \{\{\emptyset\}\}$ (the only element of $\{\{\emptyset\}\}$ is $\{\emptyset\} \neq \emptyset$) nor $\{\{\emptyset\}\} \in \emptyset$ holds, and the two elements are different, so neither is $\leq$ the other. Outside ω, membership need not even compare two sets (compare the warning above). For sets of numbers, the warning after Sets, Relations, and Cardinality, Theorem 2.1.3 gives $\mathbb{Z}$ and the positive rationals.

Quick check. Is $3 < 5$? What is the $j \in \omega$ with $5 = 2 + j$?

Solution. Yes: $3 \in \mathrm{S}(3) = 4$ and $4 \subseteq 4 \cup \{4\} = 5$, so $3 \in 5$; by Theorem 1.6.1 (c), $5 = \{0, 1, 2, 3, 4\}$. For the second question, $j = 3$: by Definition 1.5.1 (a), $2 + 3 = \mathrm{S}(2 + 2) = \mathrm{S}(4) = 5$, using $2 + 2 = 4$ from the Quick check of the previous section. By (a) of the next proposition, 3 is the only such j.

1.6.3 Order and arithmetic

The order was defined by membership and the arithmetic by recursion. They are tied together by clause (a) of the next proposition: $m \leq n$ exactly when n is reached from m by adding something. Clauses (c), (e), (f) and (g) follow from it, with cancellation (b), the laws of Theorem 1.5.3 and trichotomy, together with (d) for (e) and (f) and an induction for (g); (d) itself needs only that a nonzero natural number is a successor and that a successor is never 0.

Proposition 1.6.3. Order and Arithmetic on ω

For all $k, m, n, b \in \omega$:

(a) $m \leq n$ if and only if $n = m + j$ for some $j \in \omega$; this j is then unique, and is written $n - m$. Moreover $m < n$ if and only if $n = m + j$ for some $j \neq 0$.

(b) If $m + k = n + k$, then $m = n$.

(c) If $m < n$, then $m + k < n + k$.

(d) If $mn = 0$, then $m = 0$ or $n = 0$.

(e) If $m < n$ and $k \neq 0$, then $mk < nk$.

(f) If $mk = nk$ and $k \neq 0$, then $m = n$.

(g) *(Induction from b.)* Let $P(n)$ be a property of natural numbers. If $P(b)$ holds and, for every $n \in \omega$ with $n \geq b$, $P(n)$ implies $P(n + 1)$, then $P(n)$ holds for every $n \in \omega$ with $n \geq b$.

Idea. Cancellation (b) needs no order: induct on k and peel off one successor at a time with Corollary 1.5.2. For (a), "$m \leq m + j$" is an induction on j, and the converse is an induction on n whose step uses that nothing lies between n and $n + 1$. Then (c), (e) and (f) translate an inequality $m < n$ into an equation $n = m + j$ with $j \neq 0$, compute, and translate back; (d) writes m and n as successors and uses that a successor is never 0. Finally (g) shifts the start of an induction from 0 to b.

Proof. All inductions are by Theorem 1.3.3 (b), and "Law (x)" refers to clause (x) of Theorem 1.5.3. We prove (b) first, since (a) uses it.

(b) Induct on k for "for all m, n: $m + k = n + k \Rightarrow m = n$". For $k = 0$, $m + 0 = m$ and $n + 0 = n$. If the statement holds for k and $m + (k + 1) = n + (k + 1)$, then $(m + k) + 1 = (n + k) + 1$ by Definition 1.5.1 (a), so $m + k = n + k$ by Corollary 1.5.2, and $m = n$ by the hypothesis.

(c) ($\Leftarrow$) Fix m and induct on j for "$m \leq m + j$". For $j = 0$, $m + 0 = m$. If $m \leq m + j$, then, since $m + j < \mathrm{S}(m + j)$ by Theorem 1.6.1 (b), transitivity (or substitution, if $m = m + j$) gives

$$m < \mathrm{S}(m + j) = m + (j + 1),$$

the equality by Definition 1.5.1 (a). So $m \leq m + (j + 1)$.

($\Rightarrow$) Induct on n for "for every $m \leq n$ there is $j \in \omega$ with $n = m + j$". For $n = 0$: $m < 0$ is impossible, since $0 = \emptyset$, so $m = 0$, and $0 = 0 + 0$. Suppose the statement holds for n, and let $m \leq n+1$. If $m = n+1$, take $j = 0$. Otherwise $m < n+1 = \mathrm{S}(n)$, so $m \leq n$ by Theorem 1.6.1 (b); by the hypothesis $n = m + j$ for some j, and then $n + 1 = (m + j) + 1 = m + (j + 1)$ by Definition 1.5.1 (a).

Uniqueness. If $m + j = m + j'$, then $j + m = j' + m$ by Law (b), so $j = j'$ by (b) of this proposition.

The strict case. If $n = m + j$, then $j = 0$ if and only if $n = m$: if $j = 0$, then $n = m + 0 = m$; if $n = m$, then $m + j = m = m + 0$, so $j = 0$ by uniqueness. Since $m < n$ holds exactly when $m \leq n$ and $m \neq n$ (by Definition 1.6.1 and irreflexivity, Theorem 1.6.1 (a)), $m < n$ holds if and only if $n = m + j$ with $j \neq 0$.

(c) Let $m < n$, so $n = m + j$ with $j \neq 0$ by (a). Then

$$n + k = (m + j) + k = m + (k + j) = (m + k) + j$$

by Laws (c), (b) and (c). Since $j \neq 0$, (a) gives $m + k < n + k$.

(d) Let $mn = 0$ and $n \neq 0$; we show $m = 0$. By Lemma 1.4.1 (d), $n = j + 1$ for some $j \in \omega$, so $0 = mn = mj + m$ by Definition 1.5.1 (b). If $m \neq 0$, then $m = i + 1$ for some $i \in \omega$, and

$$0 = mj + (i + 1) = (mj + i) + 1,$$

by Definition 1.5.1 (a), which contradicts Corollary 1.5.2. So $m = 0$.

(e) Let $m < n$ and $k \neq 0$. By (a), $n = m + j$ with $j \neq 0$, so $nk = (m + j)k = mk + jk$ by Law (d). By (d), $jk \neq 0$, since $j \neq 0$ and $k \neq 0$. By (a), $mk < nk$.

(f) Let $mk = nk$ and $k \neq 0$. If $m < n$, then $mk < nk$ by (e), contradicting $mk = nk$ and irreflexivity; likewise if $n < m$. By trichotomy, $m = n$.

(g) Let $Q(i)$ be "$P(b + i)$", for $i \in \omega$. $Q(0)$ is $P(b + 0) = P(b)$. If $Q(i)$ holds, then $b + i \geq b$ by (a), so the hypothesis gives $P((b + i) + 1)$; and $(b + i) + 1 = b + (i + 1)$ by Definition 1.5.1 (a), so $Q(i + 1)$ holds. By Theorem 1.3.3 (b), $Q(i)$ holds for every $i \in \omega$. Finally, if $n \geq b$, then $n = b + i$ for some i by (a), and $P(n)$ is $Q(i)$.

■

Clause (a) is how subtraction appears on ω: $n - m$ is defined only when $m \leq n$, as the unique j with $m + j = n$. For example $5 - 2 = 3$, by the Quick check above, and $n - 0 = n$. Clause (g) is the "induction from a base" that *Sets, Relations, and Cardinality*, Chapter 2 used as Sets, Relations, and Cardinality, Theorem 2.2.1, now proved on ω.

1.6.4 Powers and parity

Sets, Relations, and Cardinality, Chapter 3 counted with powers $2^m 3^n$ and with even and odd numbers, and took both from the real numbers and the integers. Here are the facts it used, proved on ω. They are recorded for the citing rule at the end of this chapter, which says how *Sets, Relations, and Cardinality*, Chapter 3's counting arguments read on ω.

For $b \in \omega$, the powers b^n ($n \in \omega$) are defined by Theorem 1.5.1 with $X = \omega$, $a = 1$ and $g(n, x) = x \cdot b$:

$$b^0 := 1, \qquad b^{n+1} := b^n \cdot b.$$

These are the two equations of Sets, Relations, and Cardinality, Definition 2.3.2, now on ω; so $b^1 = 1 \cdot b = b$ by Theorem 1.5.3 (e).

Proposition 1.6.4. Powers and Parity on ω

For all $b, m, n \in \omega$:

(a) $b^{m+n} = b^m \cdot b^n$;

(b) if $b \neq 0$, then $b^n \neq 0$;

(c) if $b > 1$ and $m < n$, then $b^m < b^n$;

(d) there is exactly one pair (q, r) with $q \in \omega$, $r \in \{0, 1\}$ and $n = 2q + r$. In particular, no $n \in \omega$ is both of the form $2q$ and of the form $2q' + 1$.

Idea.

(a) and (b) are inductions on n following the definition of b^n. For (c), first $b^j > 1$ for $j \geq 1$, by induction from 1; then write $n = m + j$ and use (a). For (d), existence is an induction (a number is even or odd, and adding one swaps the two), and uniqueness

reduces, after canceling, to "$r = 2j + r'$ with $j \geq 1$", which makes $r \geq 2$.

Proof. "Law (x)" refers to Theorem 1.5.3, and "Order (x)" to Proposition 1.6.3.

(a) Fix m and induct on n (Theorem 1.3.3 (b)). For $n = 0$: $b^{m+0} = b^m = b^m \cdot 1 = b^m b^0$, by Law (e). If $b^{m+n} = b^m b^n$, then

$$b^{m+(n+1)} = b^{(m+n)+1} = b^{m+n} b = (b^m b^n) b = b^m (b^n b) = b^m b^{n+1},$$

by Definition 1.5.1 (a), the definition of powers, the hypothesis, Law (g), and the definition again.

(b) Let $b \neq 0$; induct on n. $b^0 = 1 \neq 0$ by Corollary 1.5.2 ($1 = 0 + 1$). If $b^n \neq 0$, then $b^{n+1} = b^n b \neq 0$ by Order (d).

(c) Let $b > 1$. Then $b > 0$, by $0 < 1$ (Theorem 1.6.1 (d)) and transitivity, so $b \neq 0$. First, $b^j > 1$ for every $j \geq 1$, by Order (g) with base 1: $b^1 = b > 1$; and if $b^j > 1$, then

$$b^{j+1} = b^j b > 1 \cdot b = b > 1$$

by Order (e) (with $b \neq 0$), Law (e) and transitivity. Now let $m < n$. By Order (a), $n = m + j$ with $j \neq 0$, so $j \geq 1$ by Theorem 1.6.1 (d), and $b^j > 1$. By (b), $b^m \neq 0$. Then, by Order (e), Law (e), Law (f) and (a),

$$b^m = 1 \cdot b^m < b^j b^m = b^m b^j = b^{m+j} = b^n.$$

(d) *Existence,* by Theorem 1.3.3 (b). For $n = 0$: $0 = 2 \cdot 0 + 0$. Suppose $n = 2q + r$ with $r \in \{0, 1\}$. If $r = 0$, then $n = 2q$ and $n + 1 = 2q + 1$. If $r = 1$, then

$$n + 1 = (2q + 1) + 1 = 2q + 2 = 2q + 2 \cdot 1 = 2(q + 1) = 2(q + 1) + 0,$$

by Law (c) and $1 + 1 = 2$ (after Definition 1.5.1), Law (e), and Law (d).

Uniqueness. Let $2q + r = 2q' + r'$ with $r, r' \in \{0, 1\}$. Suppose first that $q < q'$. By Order (a), $q' = q + j$ with $j \neq 0$, so

$$2q + r = 2(q + j) + r' = 2q + (2j + r')$$

by Laws (d) and (c), and canceling $2q$ (Order (b), after Law (b)) gives $r = 2j + r'$. Since $j \geq 1$ (Theorem 1.6.1 (d)), Order (a) gives $j = 1 + i$ for some $i \in \omega$, and $2j + r' = (2 + 2i) + r' = 2 + (2i + r')$ by Laws (d), (e) and (c). So $r = 2 + (2i + r') \geq 2$ by Order (a). But $r = 0$ or $r = 1$, and both are $< 2 = \mathrm{S}(1)$ by Theorem 1.6.1 (b); this contradicts trichotomy. So $q < q'$ is impossible, and $q' < q$ is impossible by the same argument with the roles exchanged. By trichotomy $q = q'$, and then $r = r'$ by Order (b) (after Law (b)). The pairs (q, r) and (q', r') are therefore equal (Theorem 1.2.1).

If $n = 2q = 2q + 0$ and $n = 2q' + 1$, the pairs $(q, 0)$ and $(q', 1)$ would both represent n, and they differ since $0 \neq 1$; so no n is of both forms.

■

Any other definition of powers on ω by the same two equations, $c(0) = 1$ and $c(n + 1) = c(n) \cdot b$, gives these powers, by the uniqueness half of Theorem 1.5.1. No proof in this chapter uses the

proposition; it is here for the citing rule.

1.6.5 Tuples and lists as sets

Sets, Relations, and Cardinality, Chapter 1 used n-tuples $(x_1, \dots, x_n)$ with n a variable, as working rule (W5), and stated working rule (W2) for lists (*Sets, Relations, and Cardinality*, Section 1.1); *Sets, Relations, and Cardinality*, Chapters 2 and 3 write $\{x_1, \dots, x_n\}$ with n a variable. With pairs built from sets (Theorem 1.2.1) and with $\{1, \dots, n\} = \mathrm{S}(n) \setminus \{0\}$ (Theorem 1.6.1 (e)), we can now say what they are. An n-tuple, for $n \geq 3$, is a list of entries labeled by $1, \dots, n$, and a function on $\{1, \dots, n\}$ is exactly that.

Proposition 1.6.5. Tuples and Lists as Sets

Let X be a set. For $n = 2$, an **n-tuple in X** is a pair (x_1, x_2) with $x_1, x_2 \in X$ (Definition 1.2.1), and $X^2 \coloneqq X \times X$. For $n \in \omega$ with $n \geq 3$, an **n-tuple in X** is the graph G of a function $x\colon \{1, \dots, n\} \to X$; with $x_k \coloneqq x(k)$ it is written

$$(x_1, \dots, x_n) \coloneqq \{(k, x_k) : k \in \{1, \dots, n\}\} = G.$$

(a) For $n \geq 2$, there is exactly one set X^n whose elements are exactly the n-tuples in X; for $n = 2$ it is $X \times X$.

(b) For $n \geq 2$, two n-tuples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$ in X are equal if and only if $x_k = y_k$ for every $k \in \{1, \dots, n\}$.

(c) For $n \geq 3$, no n-tuple is a pair; and for $n, n' \geq 3$ with $n \neq n'$, no n-tuple is an n'-tuple.

(d) For every $n \in \omega$ ($n = 0, 1$ allowed) and every function $x\colon \{1, \dots, n\} \to X$, the list

$$\{x_1, \dots, x_n\} \coloneqq \{y \in X : y = x_k \text{ for some } k \in \{1, \dots, n\}\}$$

is a set: the image $x(\{1, \dots, n\})$ of Sets, Relations, and Cardinality, Definition 1.6.1.

(e) For $n \geq 3$, no n-tuple is an element of ω.

The tuple is the graph alone, without the codomain X, so the same list of entries is the same tuple in every set that contains them.

Since a natural number is also a set, an exponent $n \in \omega$ could be read in more than one way. In X^n the exponent is always read as a number: X^n is the set of n-tuples of (a), not the set Y^X of *Sets, Relations, and Cardinality*, Section 1.4 taken with $Y = X$ and with the set n as exponent, whose elements are pairs (graph, codomain) with graphs on $n = \{0, \dots, n-1\}$. And X^n is written only for a declared set of entries X, never for $X \in \omega$; for $b \in \omega$, b^n is always the power defined above. For example, the power 2^2 is $4 = \{0, 1, 2, 3\}$, while 2×2 is a set of four pairs.

Idea. Each clause unwinds a definition. Sets come from (Z5) and (Z6); equality of tuples is equality of graphs, which is agreement of values. For (c) and (e), look at the elements: an n-tuple with $n \geq 3$ has at least three elements, all of them pairs, while a pair has at most two elements and every non-zero natural number has the element $\emptyset$, which is not a pair.

Proof. In (a)–(c) and (e), $n \geq 3$ unless $n = 2$ is named. First, $\{1, \dots, n\}$ is a set, by (Z6) of Axiom 1.1.1 inside ω, and $1, 2, 3$ are distinct elements of it: $1 < 2 < 3$ by Theorem 1.6.1 (b)

$(2 = \mathrm{S}(1),\ 3 = \mathrm{S}(2))$, so the three are distinct by irreflexivity and transitivity (Theorem 1.6.1 (a)); $1 \leq 1, 2, 3$, and each of them is $\leq 3 \leq n$, so they lie in $\{1, \ldots, n\}$.

(a) In each case the set is unique by (Z1), since two sets with the same elements are equal. For $n = 2$, $X \times X$ is a set by Proposition 1.2.2, and its elements are exactly the pairs (x_1, x_2) with $x_1, x_2 \in X$. For $n \geq 3$, every n-tuple is a subset of $\{1, \ldots, n\} \times X$, so the set

$$\{G \in \mathcal{P}(\{1, \ldots, n\} \times X) : (G, X) \text{ is a function } \{1, \ldots, n\} \to X\}$$

exists by Proposition 1.2.2 and (Z5), (Z6) of Axiom 1.1.1, and its elements are exactly the n-tuples in X. This is X^n.

(b) For $n = 2$ this is Theorem 1.2.1. For $n \geq 3$, let G and G' be the graphs of x and y. If $x_k = y_k$ for every k, then each element (k, x_k) of G is $(k, y_k) \in G'$, and vice versa, so $G = G'$ by (Z1). Conversely, if $G = G'$, then $(k, x_k) \in G'$ for each k, so $x_k = y_k$ by (Fn2) of Sets, Relations, and Cardinality, Definition 1.4.1 for y.

(c) An n-tuple contains the three pairs $(1, x_1)$, $(2, x_2)$, $(3, x_3)$, which are distinct by Theorem 1.2.1, since their first coordinates are. Every element of a pair $(a, b) = \{\{a\}, \{a, b\}\}$ is $\{a\}$ or $\{a, b\}$. If a pair contained three distinct elements u, v, w, then u and v, being different, would be $\{a\}$ and $\{a, b\}$ in some order, and w, being one of these two, would equal u or v. So a pair has no three distinct elements, and an n-tuple is not a pair.
Let $3 \leq n < n'$. The n'-tuple contains a pair with first coordinate n', since $1 \leq n' \leq n'$ (Theorem 1.6.1 (d), with $n' \neq 0$). Every element of the n-tuple is a pair (k, x_k) with $k \leq n$, and $k = n'$ would give $n' \leq n$, which trichotomy rules out since $n < n'$. By Theorem 1.2.1, no element of the n-tuple has first coordinate n', so the two sets differ. The case $n' < n$ is the same with the roles exchanged.

(d) Let $n \in \omega$ and $x \colon \{1, \ldots, n\} \to X$. The set $\{x_1, \ldots, x_n\}$ is cut out of X by a condition, so it is a set by (Z6) of Axiom 1.1.1; by Sets, Relations, and Cardinality, Definition 1.6.1 it is $x(\{1, \ldots, n\})$. For $n = 0$ it is $\emptyset$, since $\{1, \ldots, 0\} = \emptyset$.

(e) An n-tuple t contains $(1, x_1)$, so $t \neq \emptyset = 0$. Each element of t is a pair $(k, x_k) = \{\{k\}, \{k, x_k\}\}$, which has the element $\{k\}$, so no element of t is $\emptyset$. Every $m \in \omega$ with $m \neq 0$ has $0 = \emptyset$ as an element (Lemma 1.4.1 (e)). So t is neither 0 nor any other element of ω.

■

Clause (d) is the second half of working rule (W2): Sets, Relations, and Cardinality, Proposition 2.3.8 and Sets, Relations, and Cardinality, Definition 3.1.2 in *Sets, Relations, and Cardinality*, Chapters 2 and 3, speak of $\{x_1, \ldots, x_n\}$ for a variable n, and this is the set they mean. Clauses (a) and (b) are working rule (W5) for n-tuples. Clause (c) bears out the convention of *Sets, Relations, and Cardinality*, Chapter 1, that a triple is not a pair, and (e), with Proposition 1.4.3 for $n = 2$, that no pair or n-tuple is a natural number.

1.6.6 Exercises

1.6.6.1 A. Check your understanding

Exercise. A1

(a) List the elements of $\{k \in \omega : k \leq 3\}$. Which natural number is this set?

(b) True or false: for $m, n \in \omega$, $m < n$ implies $m \subseteq n$. Justify your answer.

(c) Find the least element of $\{n \in \omega : n \cdot n > 10\}$, and say why it is least.

(d) Which clause of Theorem 1.6.1 says that the notation $\{0, \ldots, n\}$ of *Sets, Relations, and Cardinality*, Chapter 2 still means what it meant there, and what set is it on ω?

Solution.

(a) $0, 1, 2, 3$. By Theorem 1.6.1 (c), the set is $\mathrm{S}(3) = 4$.

(b) True. If $m < n$, then $m \in n$, and n is a transitive set (Lemma 1.4.1 (a)), so $m \subseteq n$.

(c) The least element is 4. From the defining equations, $0 \cdot 0 = 0$, $1 \cdot 1 = 1$, $2 \cdot 2 = 4$, $3 \cdot 3 = 9$ and $4 \cdot 4 = 16$; and $16 > 10$, since $10 \in 16 = \{0, 1, \ldots, 15\}$, while $0, 1, 4, 9$ are all < 10, since each is an element of $10 = \{0, 1, \ldots, 9\}$ (both by Theorem 1.6.1 (c)). So 4 lies in the set. If x lies in the set and $x < 4$, then $x \in 4 = \{0, 1, 2, 3\}$, and $x \cdot x \in \{0, 1, 4, 9\}$, so $x \cdot x < 10$, and $x \cdot x > 10$ fails by trichotomy. Hence $x < 4$ is impossible, and $4 \leq x$ by trichotomy. The set is non-empty, so Theorem 1.6.2 promised a least element; this is it.

(d) Clause (c): $\{0, \ldots, n\} \coloneqq \{k \in \omega : k \leq n\}$ is $\mathrm{S}(n)$. (Clause (e) does the same for $\{1, \ldots, n\} = \mathrm{S}(n) \setminus \{0\}$.)

1.6.6.2 B. Practice

Exercise. B1: Order as inclusion

Prove that for all $m, n \in \omega$, $m \leq n$ if and only if $m \subseteq n$.

Solution. ($\Rightarrow$) If $m = n$, then $m \subseteq n$. If $m < n$, then $m \in n$, and $m \subseteq n$ because n is a transitive set (Lemma 1.4.1 (a)).

($\Leftarrow$) Let $m \subseteq n$, and suppose $m \leq n$ fails. By trichotomy (Theorem 1.6.1 (a)), $n < m$, that is, $n \in m$. Then $n \in m \subseteq n$, so $n \in n$, contradicting Lemma 1.4.1 (c). Hence $m \leq n$.

So on ω, "$\leq$" can be read as "$\subseteq$", just as "$<$" is "$\in$". For arbitrary sets the two readings come apart: $\{1\} \subseteq \{0, 1\} = 2$, but $\{1\}$ is not a natural number at all: $0 \in 1 \in \{1\}$ while $0 \notin \{1\}$ (its only element is $1 \neq 0$), so $\{1\}$ is not a transitive set, as every natural number is (Lemma 1.4.1 (a)).

Exercise. B2: A product is at least its factor

Prove that for all $m, n \in \omega$ with $n \geq 1$, $m \leq mn$. Show by an example that the hypothesis $n \geq 1$ is needed.

Solution. Since $1 \leq n$, Proposition 1.6.3 (a) gives $n = 1 + j$ for some $j \in \omega$. Then

$$mn = m(1 + j) = m \cdot 1 + mj = m + mj,$$

by Theorem 1.5.3 (d) and (e). By Proposition 1.6.3 (a) again, $m \leq m + mj = mn$.

The hypothesis is needed: for $m = 1$ and $n = 0$, $mn = 0$, and $1 \leq 0$ is false, since $0 < 1$

(Theorem 1.6.1 (d)) and trichotomy.

1.6.6.3 C. Going deeper

Exercise. C1: Strong induction on ω

Let $P(n)$ be a property of natural numbers. Suppose that, for every $n \in \omega$, if $P(k)$ holds for every $k \in \omega$ with $k < n$, then $P(n)$ holds. Prove that $P(n)$ holds for every $n \in \omega$.

Solution (Hint 1). Consider the set of natural numbers at which P fails.

Solution (Hint 2). If that set is non-empty, Theorem 1.6.2 gives it a least element a. What do you know about $P(k)$ for $k < a$?

Solution. Let $A := \{n \in \omega : P(n) \text{ fails}\}$, a set by (Z6) of Axiom 1.1.1, and suppose, for a contradiction, that $A \neq \emptyset$. By Theorem 1.6.2, A has a least element a. Let $k \in \omega$ with $k < a$. If $k \in A$, then $a \leq k$, and with $k < a$ this gives $k < k$ (by transitivity, or directly if $a = k$), contradicting irreflexivity (Theorem 1.6.1 (a)). So $k \notin A$, that is, $P(k)$ holds. This holds for every $k < a$, so the hypothesis gives $P(a)$, contradicting $a \in A$. Hence $A = \emptyset$, and $P(n)$ holds for every $n \in \omega$.

This is the statement of Logic and Proof, Theorem 4.2.2, read on ω. There it was derived from Logic and Proof, Axiom 4.2.1 directly; here the route goes through well-ordering, which was itself proved from Theorem 1.3.3. As there, the hypothesis at $n = 0$ is vacuous (no $k < 0$), so it hands us $P(0)$ outright.

Marking scheme. (10 points) - 2: considers the set of failures (or argues by contradiction from a failure) and makes it a set. - 3: applies Theorem 1.6.2, with the set non-empty, to obtain a least failure a. - 3: shows $P(k)$ for every $k < a$, using that a is least and the order facts. - 2: applies the hypothesis at a and reaches the contradiction. Fatal errors (0 for the step): citing Logic and Proof, Theorem 4.2.2 (a statement about *Logic and Proof*, Chapter 3's $\mathbb{N}$, not proved on ω) instead of proving the result; taking a least element of a set not shown to be non-empty. A route by Theorem 1.3.3 (b) on "$P(k)$ for every $k < n$" earns the same points: 2 for the property, 3 for the base (vacuous) and the use of Theorem 1.6.1 (b) in the step, 3 for the step, 2 for deducing $P(n)$ from the property at $n + 1$.

Exercise. C2: Find the error

Here is an attempted proof that any two natural numbers are comparable. Is the conclusion true? Is the proof correct? If not, give the first line that fails, the rule it breaks, and a counterexample to the step.

> **Claim.** For all $m, n \in \omega$: $m \in n$, $m = n$ or $n \in m$.
>
> (1) We use Theorem 1.3.3 (b) on n, for the property "for every $m \in \omega$: $m \in n$, $m = n$ or $n \in m$". (2) For $n = 0$: every m is 0 or contains 0, by Lemma 1.4.1

(e). (3) Suppose the property holds for n, and let $m \in \omega$. (4) If $m \in n$ or $m = n$, then $m \in n \cup \{n\} = \mathrm{S}(n)$. (5) If $n \in m$, then $\mathrm{S}(n) \leq m$, because nothing lies strictly between n and $\mathrm{S}(n)$: no $k \in \omega$ has $n \in k \in \mathrm{S}(n)$. (6) In every case $m \in \mathrm{S}(n)$, $m = \mathrm{S}(n)$ or $\mathrm{S}(n) \in m$, which is the property for $\mathrm{S}(n)$. (7) By Theorem 1.3.3 (b), the claim holds.

Solution. The conclusion is **true**: it is trichotomy, Theorem 1.6.1 (a). The proof is **not** correct.

The failing line is (5). Its reason, "nothing lies strictly between n and $\mathrm{S}(n)$", is true (it follows from irreflexivity and transitivity, as in Theorem 1.6.1 (b)), but it does not give $\mathrm{S}(n) \leq m$. To pass from "m is not strictly between n and $\mathrm{S}(n)$" to "$\mathrm{S}(n) \in m$ or $\mathrm{S}(n) = m$", one must know that if neither holds, then $m \in \mathrm{S}(n)$. That is trichotomy for $\mathrm{S}(n)$ and m: an instance of the claim being proved, at a number the induction has not reached. So the stated reason gives (5) only by assuming the goal (circular). Without that, (5) asserts the general fact "$n \in m \Rightarrow \mathrm{S}(n) \leq m$", which is true but not yet proved. Either way, line (5) rests on an unavailable fact: entry (N9) of Logic and Proof, Example (What is not allowed).

Counterexample to the step. The inference "$x \in y$ and nothing lies strictly between x and $x \cup \{x\}$, so $x \cup \{x\} \in y$ or $x \cup \{x\} = y$" fails for sets in general. Take $x = \emptyset$ and $y = \{\emptyset, \{\{\emptyset\}\}\}$. Then $x \in y$, and nothing lies strictly between x and $x \cup \{x\} = \{\emptyset\}$ (the only element of $\{\emptyset\}$ is $\emptyset$, and $\emptyset \notin \emptyset$). But $\{\emptyset\} \notin y$, since $\{\emptyset\} \neq \emptyset$ and $\{\emptyset\} \neq \{\{\emptyset\}\}$; and $\{\emptyset\} \neq y$, since $\{\{\emptyset\}\} \in y$ is not an element of $\{\emptyset\}$. So the step needs something about ω beyond "nothing in between".

Repair. Replace (5) by the claim inside the proof of Theorem 1.6.1: for all $m, n \in \omega$, $m \in n$ implies $\mathrm{S}(m) \in \mathrm{S}(n)$, proved by its own induction. Then $n \in m$ gives $\mathrm{S}(n) \in \mathrm{S}(m) = m \cup \{m\}$, so $\mathrm{S}(n) \in m$ or $\mathrm{S}(n) = m$. Lines (1)–(4), (6) and (7) are correct.

Marking scheme. (10 points) - 2: says that the conclusion is true and that the proof is not correct. - 3: names line (5) as the failing line and explains why its reason does not give $\mathrm{S}(n) \leq m$ (it needs trichotomy for $\mathrm{S}(n)$ and m, or the unproved fact $n \in m \Rightarrow \mathrm{S}(n) \leq m$). - 2: classifies the error as (N9): a line resting on an unavailable fact, either the goal itself (circular) or an unproved general fact. - 2: gives a counterexample to the step on arbitrary sets (any $x \in y$ with $x \cup \{x\}$ neither in nor equal to y). - 1: gives the repair through "$m \in n \Rightarrow \mathrm{S}(m) \in \mathrm{S}(n)$". Fatal errors (0 for the step): declaring the conclusion false; blaming line (2) or (4), which are correct.

1.7 All Peano Systems Are Alike

Section 1.4 proved that $(\omega, 0, \mathrm{S})$ is a Peano system (Theorem 1.4.2, Definition 1.4.2), and Sections 1.5 and 1.6 built $+$, $\cdot$ and $<$ on it. One question is left. *Sets, Relations, and Cardinality*, Chapter 2 worked with a set $\mathbb{N}$ inside the familiar real numbers, and this chapter has built a quite different set ω out of $\emptyset$. In what sense are they the same? This section answers: any two Peano systems are copies of each other, by exactly one matching, which also carries $+$, $\cdot$ and $<$ across. Then it says what the later chapters may take from *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 when they write $\mathbb{N}$ for ω.

1.7.1 Two Peano systems, and what "alike" should mean

Here is a second Peano system, inside ω itself: $(\omega \setminus \{0\}, 1, \mathrm{S})$, the natural numbers from 1 on, with the same successor. Exercise C1 of Section 1.4 asked you to check it; the check is three lines.

- S maps $\omega \setminus \{0\}$ into itself, since $\mathrm{S}(n) \neq 0$ by Theorem 1.4.2 (P3).
- (P3) holds with $e = 1$: if $\mathrm{S}(n) = 1 = \mathrm{S}(0)$, then $n = 0$ by (P4) of Theorem 1.4.2, and $0 \notin \omega \setminus \{0\}$. (P4) holds because S is injective on all of ω.
- (P5) holds: if $A \subseteq \omega \setminus \{0\}$ contains 1 and is closed under S, then $\{0\} \cup A$ contains 0 and is closed under S (as $\mathrm{S}(0) = 1 \in A$), so it is ω by Theorem 1.3.3 (a), and $A = \omega \setminus \{0\}$.

Its starting point is 1, not $\emptyset$, yet it counts the same way: $n \mapsto \mathrm{S}(n)$ pairs $0, 1, 2, \ldots$ with $1, 2, 3, \ldots$, sends start to start, and turns one step into one step.

That is the right meaning of "alike". For Peano systems (M, e, s) and (M', e', s'), call a bijection $\psi\colon M \to M'$ an **isomorphism** if

$$\psi(e) = e' \quad \text{and} \quad \psi(s(x)) = s'(\psi(x)) \quad \text{for every } x \in M.$$

In words: ψ renames the elements, sends the starting point to the starting point, and it does not matter whether we first step and then rename, or first rename and then step.

Warning. Being equinumerous is not enough. The bijection $\tau\colon \omega \to \omega$ that exchanges 1 and 2 and fixes every other number satisfies $\tau(0) = 0$, but $\tau(\mathrm{S}(0)) = \tau(1) = 2$, while $\mathrm{S}(\tau(0)) = \mathrm{S}(0) = 1$. So τ is a bijection from a Peano system to itself that is **not** an isomorphism: it does not respect the successor. Categoricity, below, is a statement about isomorphisms, and it says much more than "same size".

1.7.2 The categoricity theorem

The theorem says that every Peano system is an isomorphic copy of ω, and so any two are copies of each other, in only one way.

Theorem 1.7.1. Categoricity of the Peano Axioms

Let (M, e, s) be a Peano system.

(a) There is **exactly one** function $\varphi\colon \omega \to M$ with $\varphi(0) = e$ and $\varphi(\mathrm{S}(n)) = s(\varphi(n))$ for every $n \in \omega$.

(b) This φ is a bijection.

(c) For any two Peano systems (M, e, s) and (M', e', s') there is **exactly one** bijection $\psi\colon M \to M'$ with $\psi(e) = e'$ and $\psi(s(x)) = s'(\psi(x))$ for every $x \in M$.

Try it first (ten minutes): given a Peano system (M, e, s), which map $\omega \to M$ would you build, with which theorem, and how would you show that it is onto?

Idea. The attempt meets three obstacles, one for each part.

① **Which map.** It is forced: 0 must go to e, then 1 to $s(e)$, then 2 to $s(s(e))$, and so on. But "and so on" is not a definition. Theorem 1.5.1 turns it into one, with $X = M$, starting value e and step $x \mapsto s(x)$; it gives existence and uniqueness at once.

② **Onto.** We know nothing about the elements of M one by one, so we cannot show "each $x \in M$ is $s(\dots s(e))$" directly. What we have is (P5) for M, which speaks about subsets. So look at the image of φ: it contains e and is closed under s, and (P5) does the rest.

③ **One-to-one.** Prove "$\varphi(m) = \varphi(n) \Rightarrow m = n$" by induction on n. If m is fixed before the induction, the hypothesis at n speaks about that one m, while the step needs it at the predecessor of m. So "for every m" goes inside the property. The base and the step then use (P3) and (P4) for M: a successor is never e, and equal successors have equal predecessors.

For (c), compose: go back from M to ω along φ^{-1}, then forward to M' along φ'.

Before reading on: which axiom of M gives surjectivity, and which give injectivity?

Proof. (a) Let $g\colon \omega \times M \to M$ be $g(n, x) \coloneqq s(x)$; its graph $\{((n, x), y) \in (\omega \times M) \times M : y = s(x)\}$ is a set by (Z6) inside the product of Proposition 1.2.2, and each (n, x) has exactly one partner $s(x)$ because s is a function. By Theorem 1.5.1 with $X = M$ and $a = e$, there is exactly one $\varphi\colon \omega \to M$ with $\varphi(0) = e$ and $\varphi(\mathrm{S}(n)) = g(n, \varphi(n)) = s(\varphi(n))$ for every $n \in \omega$.

(b), onto. Let $B \coloneqq \varphi(\omega) \subseteq M$, the image of ω under φ (Sets, Relations, and Cardinality, Definition 1.6.1). Then $e = \varphi(0) \in B$. If $x \in B$, say $x = \varphi(n)$ with $n \in \omega$, then $\mathrm{S}(n) \in \omega$ by Theorem 1.4.2 (P2), and $s(x) = s(\varphi(n)) = \varphi(\mathrm{S}(n)) \in B$. So B contains e and is closed under s, and (P5) for M gives $B = M$. Hence φ is surjective (Sets, Relations, and Cardinality, Definition 1.4.4).

(b), one-to-one. Let $Q(n)$ be "for every $m \in \omega$, if $\varphi(m) = \varphi(n)$, then $m = n$". We prove $Q(n)$ for every $n \in \omega$ by Theorem 1.3.3 (b).

Base case. Let $m \in \omega$ with $\varphi(m) = \varphi(0) = e$. Suppose $m \neq 0$. Then $m = \mathrm{S}(k)$ for some $k \in \omega$ (Lemma 1.4.1 (d)), and $\varphi(m) = s(\varphi(k)) \neq e$ by (P3) for M, a contradiction. So $m = 0$.

Inductive step. Let $n \in \omega$ with $Q(n)$, and let $m \in \omega$ with $\varphi(m) = \varphi(\mathrm{S}(n)) = s(\varphi(n))$. If $m = 0$, then $s(\varphi(n)) = \varphi(0) = e$, against (P3) for M. So $m \neq 0$, and $m = \mathrm{S}(k)$ for some $k \in \omega$ (Lemma 1.4.1 (d)). Then $s(\varphi(k)) = \varphi(m) = s(\varphi(n))$, so $\varphi(k) = \varphi(n)$ by (P4) for M, and $k = n$ by $Q(n)$, applied to k in the place of m. Hence $m = \mathrm{S}(k) = \mathrm{S}(n)$, which is $Q(\mathrm{S}(n))$.

So φ is injective (Sets, Relations, and Cardinality, Definition 1.4.3), and with the first half it is a bijection (Sets, Relations, and Cardinality, Definition 1.4.5).

(c) Let $\varphi\colon \omega \to M$ and $\varphi'\colon \omega \to M'$ be the maps of (a) for the two systems; both are bijections by (b). By Sets, Relations, and Cardinality, Theorem 1.5.4, φ has an inverse $\varphi^{-1}\colon M \to \omega$, and we set $\psi \coloneqq \varphi' \circ \varphi^{-1}$ (Sets, Relations, and Cardinality, Definition 1.4.2). The two equations $\varphi \circ \varphi^{-1} = \mathrm{id}_M$ and $\varphi^{-1} \circ \varphi = \mathrm{id}_\omega$ of Sets, Relations, and Cardinality, Definition 1.5.2 also say that φ is an inverse of φ^{-1}, so φ^{-1} is a bijection by Sets, Relations, and Cardinality, Theorem 1.5.4. Hence ψ is a bijection as a composite of bijections (Sets, Relations, and Cardinality, Theorem 1.4.4 (c)).

ψ has the two properties. Since $\varphi(0) = e$, we have $\varphi^{-1}(e) = 0$, so $\psi(e) = \varphi'(0) = e'$. Let $x \in M$ and $n \coloneqq \varphi^{-1}(x)$, so $x = \varphi(n)$. Then $s(x) = \varphi(\mathrm{S}(n))$, so $\varphi^{-1}(s(x)) = \mathrm{S}(n)$, and

$$\psi(s(x)) = \varphi'(\mathrm{S}(n)) = s'(\varphi'(n)) = s'(\psi(x)).$$

Uniqueness. Let $\chi\colon M \to M'$ be **any** function with $\chi(e) = e'$ and $\chi(s(x)) = s'(\chi(x))$ for every $x \in M$. Then $\chi \circ \varphi\colon \omega \to M'$ satisfies $(\chi \circ \varphi)(0) = \chi(e) = e'$ and

$$(\chi \circ \varphi)(\mathrm{S}(n)) = \chi(s(\varphi(n))) = s'((\chi \circ \varphi)(n))$$

for every $n \in \omega$. By the uniqueness in (a), for the system M', $\chi \circ \varphi = \varphi'$. So for every $x \in M$, $\chi(x) = \chi(\varphi(\varphi^{-1}(x))) = \varphi'(\varphi^{-1}(x)) = \psi(x)$, and $\chi = \psi$ by Sets, Relations, and Cardinality, Proposition 1.4.1. This proves the theorem.

■

The proof is the third named move of the chapter, **build by recursion, compare by induction**: the map out of ω came from the recursion theorem, and every fact about it (onto, one-to-one) came from an induction or from (P5). The uniqueness in (c) is stronger than stated: no function at all, bijective or not, other than ψ respects e and s.

Each axiom of M has a job in (b), and the structures of Example (Each axiom is needed) show each job failing. These are not Peano systems, so the theorem does not apply to them. But for each of them Theorem 1.5.1, with $X = M$, $a = e$ and $g(n, x) = s(x)$ exactly as in the proof of (a), still gives exactly one $\varphi\colon \omega \to M$ with $\varphi(0) = e$ and $\varphi(\mathrm{S}(n)) = s(\varphi(n))$.

- In structure (i), $M = \{0\}$ with $s(0) = 0$, which fails (P3) only, φ is the constant map 0, not injective: $\varphi(1) = s(\varphi(0)) = 0 = \varphi(0)$.
- In structure (ii), $M = \{0, 1\}$ with $s(0) = s(1) = 1$, which fails (P4) only, $\varphi(1) = \varphi(2) = 1$, again not injective.
- In the natural numbers beside a two-sided chain, structure (iii), which fails (P5) only, φ is $n \mapsto (n, 0)$ (remark below), which is injective by Theorem 1.2.1, but not onto, as the remark below shows.

Remark. **What induction buys.** In structure (iii) of Example (Each axiom is needed), the natural numbers beside a two-sided chain (a copy of ω next to a copy of $\mathbb{Z}$), $M = \omega \times 3$ with $e = (0, 0)$, the map φ still exists, by Theorem 1.5.1 as above. It is $\varphi(n) = (n, 0)$, by the uniqueness half of Theorem 1.5.1: $n \mapsto (n, 0)$ sends 0 to e and satisfies the recursion, as $s(n, 0) = (\mathrm{S}(n), 0)$. Its image $\omega \times \{0\}$ misses $(0, 1)$, so φ is not onto. Exercise C2 asks the reverse question: is there a map $M \to \omega$ that respects the two structures?

Quick check. For the Peano system $(\omega \setminus \{0\}, 1, \mathrm{S})$ of the first subsection, what is the map φ of Theorem 1.7.1 (a)? What is $\varphi(3)$?

Solution. $\varphi(n) = \mathrm{S}(n) = n + 1$. Indeed $n \mapsto \mathrm{S}(n)$ maps ω into $\omega \setminus \{0\}$ by (P3), sends 0 to $\mathrm{S}(0) = 1$, and satisfies $\mathrm{S}(\mathrm{S}(n)) = \mathrm{S}(\varphi(n))$; by the uniqueness in (a), it is φ. So $\varphi(3) = \mathrm{S}(3) = 4$. Step by step: $\varphi(0) = 1$, $\varphi(1) = \mathrm{S}(1) = 2$, $\varphi(2) = 3$, $\varphi(3) = 4$.

1.7.3 Arithmetic and order travel along the isomorphism

An isomorphism matches e and s. Addition and multiplication on ω were built from 0 and S alone, by recursion (Definition 1.5.1), and $<$ can be described by $+$ (Proposition 1.6.3 (a)). So we expect the matching to carry them across, as long as the other system has operations defined by the same equations.

Proposition 1.7.2. Isomorphisms Preserve Arithmetic and Order

Let (M, e, s) be a Peano system, and let $\oplus, \otimes \colon M \times M \to M$ be functions such that, for all $x, y \in M$,

$$x \oplus e = x, \quad x \oplus s(y) = s(x \oplus y),$$
$$x \otimes e = e, \quad x \otimes s(y) = (x \otimes y) \oplus x.$$

Let $\varphi \colon \omega \to M$ be the bijection of Theorem 1.7.1. Then for all $m, n \in \omega$:

(a) $\varphi(m+n) = \varphi(m) \oplus \varphi(n)$;
(b) $\varphi(mn) = \varphi(m) \otimes \varphi(n)$;
(c) $m < n$ if and only if $\varphi(n) = \varphi(m) \oplus z$ for some $z \in M$ with $z \neq e$.

Idea. For (a) and (b), fix m and induct on n, the variable the recursions run on: each side obeys the same two equations, the base from 0 and the step from S, and φ turns one set of equations into the other. For (c), translate Proposition 1.6.3 (a), "$m < n$ iff $n = m + j$ with $j \neq 0$", through (a), using that φ is a bijection with $\varphi(0) = e$.

Proof. *(a)* Fix $m \in \omega$; we induct on n (Theorem 1.3.3 (b)). *Base case:* $\varphi(m+0) = \varphi(m) = \varphi(m) \oplus e = \varphi(m) \oplus \varphi(0)$, by Definition 1.5.1 (a) and the first equation for $\oplus$. *Step:* if $\varphi(m+n) = \varphi(m) \oplus \varphi(n)$, then

$$\begin{aligned}\varphi(m + \mathrm{S}(n)) &= \varphi(\mathrm{S}(m+n)) = s(\varphi(m+n)) \\ &= s(\varphi(m) \oplus \varphi(n)) = \varphi(m) \oplus s(\varphi(n)) = \varphi(m) \oplus \varphi(\mathrm{S}(n)),\end{aligned}$$

by Definition 1.5.1 (a), the equation of φ, the hypothesis, the second equation for $\oplus$, and the equation of φ again.

(b) Fix $m \in \omega$ and induct on n. *Base case:* $\varphi(m \cdot 0) = \varphi(0) = e = \varphi(m) \otimes e = \varphi(m) \otimes \varphi(0)$, by Definition 1.5.1 (b) and the first equation for $\otimes$. *Step:* if $\varphi(mn) = \varphi(m) \otimes \varphi(n)$, then

$$\begin{aligned}\varphi(m \cdot \mathrm{S}(n)) &= \varphi(mn + m) = \varphi(mn) \oplus \varphi(m) \\ &= (\varphi(m) \otimes \varphi(n)) \oplus \varphi(m) = \varphi(m) \otimes s(\varphi(n)) = \varphi(m) \otimes \varphi(\mathrm{S}(n)),\end{aligned}$$

by Definition 1.5.1 (b), part (a) just proved, the hypothesis, the second equation for $\otimes$, and the equation of φ.

(c) By Proposition 1.6.3 (a), $m < n$ if and only if $n = m + j$ for some $j \in \omega$ with $j \neq 0$.

($\Rightarrow$) Let $n = m + j$ with $j \neq 0$. By (a), $\varphi(n) = \varphi(m) \oplus \varphi(j)$. Put $z := \varphi(j)$. Since φ is injective (Theorem 1.7.1 (b)), $j \neq 0$ gives $z \neq \varphi(0) = e$.

($\Leftarrow$) Let $z \in M$ with $z \neq e$ and $\varphi(n) = \varphi(m) \oplus z$. Since φ is onto (Theorem 1.7.1 (b)), $z = \varphi(j)$ for some $j \in \omega$, and $j \neq 0$ because $\varphi(0) = e \neq z$. By (a), $\varphi(n) = \varphi(m) \oplus \varphi(j) = \varphi(m + j)$, so $n = m + j$ by injectivity. Hence $m < n$. This proves the proposition.

■

So once a Peano system carries an addition and a multiplication obeying the recursive equations, φ is an isomorphism for them too, and the order defined from $\oplus$ as in (c) is carried along as well. Nothing about $\oplus$ and $\otimes$ was assumed beyond the four equations.

1.7.4 Identifying the two sets of natural numbers

Now apply all this to the set $\mathbb{N}$ of *Sets, Relations, and Cardinality*, Chapter 2. Two different things are true, and it matters which one the later chapters lean on. One is a statement about *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N}$, a subset of the familiar real numbers: it is a Peano system, so it is a copy of ω. The other is a statement about ω alone: the principles *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 assumed or proved for $\mathbb{N}$ are theorems about ω, proved from the axioms (Z1)–(Z7).

Corollary 1.7.3. ω Is the Natural Numbers

(a) **About *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N}$.** Let $\mathbb{N} \subseteq \mathbb{R}$ be the set of natural numbers as *Sets, Relations, and Cardinality*, Section 2.1 describes it, with (($\mathbb{N}$)0), ($\mathbb{N}$S) and Logic and Proof, Axiom 4.2.1. With $e \coloneqq 0$ and $s(n) \coloneqq n + 1$, it is a Peano system. There is exactly one bijection $\varphi\colon \omega \to \mathbb{N}$ with $\varphi(0) = 0$ and $\varphi(\mathrm{S}(n)) = \varphi(n) + 1$ for every $n \in \omega$. For all $m, n \in \omega$ it satisfies

$$\varphi(m + n) = \varphi(m) + \varphi(n), \qquad \varphi(mn) = \varphi(m)\varphi(n),$$

and $m < n$ if and only if $\varphi(m) < \varphi(n)$; and it sends each numeral of ω to the natural number of the same name.

(b) **About ω, proved from (Z1)–(Z7).** The statements of Logic and Proof, Axiom 4.2.1 and Sets, Relations, and Cardinality, Theorem 2.3.1, read with $\mathbb{N} \coloneqq \omega$ and $n{+}1 \coloneqq \mathrm{S}(n)$, are Theorem 1.3.3 (b) and Theorem 1.5.1. Induction started at a base $b \in \omega$ (the form of Sets, Relations, and Cardinality, Theorem 2.2.1 with base $n_0 = b$) is Proposition 1.6.3 (g).

(c) Each clause of Sets, Relations, and Cardinality, Lemma 2.1.1 holds on ω, with $n - m$, for $m \leq n$, read as the unique $j \in \omega$ with $n = m + j$ of Proposition 1.6.3 (a).

Idea. For (a), check (P3)–(P5) for *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N}$ with the tools of *Sets, Relations, and Cardinality*, Section 2.1, and the equations of Proposition 1.7.2 with school arithmetic; then Theorem 1.7.1 and that proposition do the rest. Parts (b) and (c) are a matter of reading statements side by side and pointing to the label that proves each one on ω.

Before reading on: which clause do the later chapters actually use, when they build $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$?

Proof. *(a)* The map s sends $\mathbb{N}$ into $\mathbb{N}$ by ($\mathbb{N}$S). (P3): for $n \in \mathbb{N}$, $n \geq 0$ by Sets, Relations, and Cardinality, Lemma 2.1.1 (a), so $n+1 \geq 1 > 0$ and $n+1 \neq 0$. (P4): $m+1 = n+1$ gives $m = n$, by subtracting 1 in $\mathbb{R}$. (P5): a subset of $\mathbb{N}$ that contains 0 and is closed under $n \mapsto n+1$ is $\mathbb{N}$; this is the set form of Logic and Proof, Axiom 4.2.1, the remark at the start of *Sets, Relations, and Cardinality*, Section 2.1. So $(\mathbb{N}, 0, s)$ is a Peano system (Definition 1.4.2), and Theorem 1.7.1 (a), (b) give exactly one φ as stated, and it is a bijection.

For the operations, take $\oplus$ and $\otimes$ to be the addition and multiplication of $\mathbb{N}$; they are functions $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by Sets, Relations, and Cardinality, Lemma 2.1.1 (b). The four equations of Proposition 1.7.2 read $x + 0 = x$, $x + (y+1) = (x+y)+1$, $x \cdot 0 = 0$ and $x(y+1) = xy + x$, which hold in $\mathbb{R}$. That proposition gives $\varphi(m+n) = \varphi(m) + \varphi(n)$ and $\varphi(mn) = \varphi(m)\varphi(n)$.

For the order we first show: for $x, y \in \mathbb{N}$, $x < y$ if and only if $y = x + z$ for some $z \in \mathbb{N}$ with $z \neq 0$. If $x < y$, then $z := y - x \in \mathbb{N}$ by Sets, Relations, and Cardinality, Lemma 2.1.1 (d), and $z \neq 0$ since $y \neq x$. Conversely, $z \neq 0$ gives $z \geq 1 > 0$ by Sets, Relations, and Cardinality, Lemma 2.1.1 (a), so $y = x + z > x$. Now Proposition 1.7.2 (c), with $e = 0$, says that $m < n$ if and only if $\varphi(n) = \varphi(m) + z$ for some $z \in \mathbb{N}$ with $z \neq 0$, that is, if and only if $\varphi(m) < \varphi(n)$.

Finally, $\varphi(0) = 0$, and each numeral of ω is the successor of the one before (Definition 1.5.1 (c)), so $\varphi(1) = \varphi(\mathrm{S}(0)) = 0 + 1 = 1$, $\varphi(2) = \varphi(\mathrm{S}(1)) = 1 + 1 = 2$, and so on, one step for each numeral written.

(b) By the two equations of Definition 1.5.1 (a), $n + 1 = n + \mathrm{S}(0) = \mathrm{S}(n + 0) = \mathrm{S}(n)$ for every $n \in \omega$ (Section 1.5, "the successor is plus one"). With $\mathbb{N} := \omega$, the statement of Logic and Proof, Axiom 4.2.1 becomes Theorem 1.3.3 (b), and the statement of Sets, Relations, and Cardinality, Theorem 2.3.1 is word for word Theorem 1.5.1. Induction from a base $b \in \omega$ is Proposition 1.6.3 (g).

(c) We take the clauses of Sets, Relations, and Cardinality, Lemma 2.1.1 in turn, for $m, n \in \omega$.

- (a) "$n \geq 0$; if $n \neq 0$, then $n \geq 1$" is Theorem 1.6.1 (d).
- (b) "$m + n \in \omega$ and $mn \in \omega$" holds because $+$ and $\cdot$ are functions $\omega \times \omega \to \omega$ (Definition 1.5.1).
- (c) If $n \neq 0$, then $1 \leq n$ by Theorem 1.6.1 (d), so $n = 1 + j$ for exactly one $j \in \omega$ by Proposition 1.6.3 (a); this j is $n - 1$, and it lies in ω.
- (d) If $m \leq n$, then $n - m \in \omega$ by Proposition 1.6.3 (a).
- (e) No $k \in \omega$ has $n < k < \mathrm{S}(n)$, by Theorem 1.6.1 (b). For the second form, let $m < n$. By trichotomy (Theorem 1.6.1 (a)), $\mathrm{S}(m) < n$, $\mathrm{S}(m) = n$ or $n < \mathrm{S}(m)$. The last gives $n \leq m$ by Theorem 1.6.1 (b), and then $n \leq m < n$ gives $n < n$ by transitivity, which (a) of that theorem rules out. So $m + 1 = \mathrm{S}(m) \leq n$.

This proves the corollary.

■

Clause (a) is the honest sense in which "ω **is** $\mathbb{N}$": the two are copies of each other by a unique isomorphism, which also matches $+$, $\cdot$ and $<$. But (a) is a theorem about *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N}$, and its proof used the school arithmetic of $\mathbb{R}$ (in (P3), (P4), the four equations, the step describing $<$ by $+$, and the numerals). The answer to the prompt is therefore (b), with (c): when a later chapter applies induction or recursion to ω, it uses a theorem proved here from the axioms for sets, and nothing about the school real numbers enters.

Warning. The isomorphism carries 0, successor, $+$, $\cdot$ and $<$, and nothing else. In ω, $0 \in 5$ and $2 \subseteq 3$ are true, because natural numbers here are sets (Example (The first natural numbers as sets)). In *Sets, Relations, and Cardinality*, Chapter 2's picture numbers are not sets, and "$0 \in 5$" has no meaning there. Such statements are facts about one particular construction, and no theorem of the later chapters asks them.

1.7.5 How later chapters cite *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2

In the rest of this volume, $\mathbb{N}$ **means** ω, with the 0, $+$, $\cdot$ and $<$ of this chapter. The later chapters build $\mathbb{Z}$, $\mathbb{Q}$ and finally $\mathbb{R}$ from it, and they cite results of *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapters 2 and 3 about $\mathbb{N}$. This subsection says how such a citation is to be read, so that the construction of $\mathbb{R}$ never rests on the school $\mathbb{R}$ it is meant to replace.

The proofs in *Sets, Relations, and Cardinality*, Chapters 2 and 3 use the arithmetic of the real numbers freely: *Sets, Relations, and Cardinality*, Chapter 2's "What you need" paragraph takes the arithmetic and order of $\mathbb{R}$ as school facts, *Sets, Relations, and Cardinality*, Section 2.1 says so for itself in "the rule for this section", and *Sets, Relations, and Cardinality*, Chapter 3's proofs do the same (Sets, Relations, and Cardinality, Theorem 3.4.1, for example, uses $1 - b$). So they do **not** rest only on (($\mathbb{N}$)0), ($\mathbb{N}$S), Logic and Proof, Axiom 4.2.1, Sets, Relations, and Cardinality, Theorem 2.3.1 and Sets, Relations, and Cardinality, Lemma 2.1.1, and "they hold unchanged on ω" would be false. What is true is a **dictionary**: each fact about $\mathbb{N}$ those proofs use has a counterpart proved on ω here.

used for $\mathbb{N}$ in *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapters 2, 3	read on ω as
Logic and Proof, Axiom 4.2.1; Sets, Relations, and Cardinality, Theorem 2.3.1	Corollary 1.7.3 (b): Theorem 1.3.3 (b); Theorem 1.5.1
Sets, Relations, and Cardinality, Theorem 2.2.1 with a base $n_0 = b \in \mathbb{N}$, induction from a base	Proposition 1.6.3 (g)
Sets, Relations, and Cardinality, Theorem 2.1.3	Theorem 1.6.2
Sets, Relations, and Cardinality, Lemma 2.1.1	Corollary 1.7.3 (c)
other school arithmetic and order of natural numbers: closure under $+$, $\cdot$; cancellation; trichotomy; $n - m$ for $m \leq n$; "subtract 1 from $m > n + 1$"	Theorem 1.5.3, Theorem 1.6.1, Proposition 1.6.3
powers b^n of a natural number, Sets, Relations, and Cardinality, Definition 2.3.2, Sets, Relations, and Cardinality, Proposition 2.3.4 (a), (d), (e)	Proposition 1.6.4 (a)–(c); closure of ω under powers is the definition there
"no natural number is both even and odd"	Proposition 1.6.4 (d)

With the dictionary, the results of *Sets, Relations, and Cardinality*, Chapters 2 and 3 that later chapters cite fall into three groups.

(i) **Results that run on ω through the dictionary.** These are Sets, Relations, and Cardinality, Proposition 2.3.8 (for its $\mathbb{N}$ side, which is Sets, Relations, and Cardinality, Theorem 2.2.1 with base 1 and Sets, Relations, and Cardinality, Lemma 2.1.1 (e)), Sets, Relations, and Cardinality, Definition 3.1.2, Sets, Relations, and Cardinality, Definition 3.1.3, Sets, Relations, and Cardinality, Corollary 3.1.3, Sets, Relations, and Cardinality, Theorem 3.1.9, Sets, Relations, and Cardinality, Lemma 3.2.1, Sets, Relations, and Cardinality, Theorem 3.2.2, Sets, Relations, and Cardinality, Proposition 3.2.3, Sets, Relations, and Cardinality, Proposition 3.2.4 and Sets, Relations, and Cardinality, Theorem 3.4.1 (where $\{0,1\}^{\mathbb{N}}$ is indexed by ω, and $1-b$ on $\{0,1\}$ is read by Proposition 1.6.3 (a)), and Sets, Relations, and Cardinality, Exercise (C2: Counting products and functions) (a), cited later only from an exercise solution (its $\mathbb{N}$ side is an induction on m, read by Corollary 1.7.3 (b), $mn \in \mathbb{N}$, read by Corollary 1.7.3 (c), and $0 = 0 \cdot n$ and $mn + n = (m+1)n$, read by Theorem 1.5.3 (e)). Running a result on ω includes the results its proof cites: Sets, Relations, and Cardinality, Theorem 3.1.2, Sets, Relations, and Cardinality, Proposition 3.1.5, Sets, Relations, and Cardinality, Proposition 3.1.6, Sets, Relations, and Cardinality, Corollary 3.1.4, Sets, Relations, and Cardinality, Corollary 3.1.7 and the count "$\{0,\dots,x\}$ has $x+1$ elements" in the examples after Sets, Relations, and Cardinality, Definition 3.1.2 run through the same dictionary, and they belong to group (i) themselves. A later chapter may cite the results of group (i) anywhere.

(ii) **Results that use school facts, and need Proposition 1.6.4 on ω.** These are Sets, Relations, and Cardinality, Theorem 3.2.6 (powers by Sets, Relations, and Cardinality, Definition 2.3.2 and Sets, Relations, and Cardinality, Proposition 2.3.4 (a), (d), (e) over the school $\mathbb{R}$, parity through Sets, Relations, and Cardinality, Theorem 2.2.3 on the school $\mathbb{Z}$, and cancellation in $\mathbb{R}$), Sets, Relations, and Cardinality, Proposition 3.2.5 (parity through Sets, Relations, and Cardinality, Theorem 2.2.3), and through them Sets, Relations, and Cardinality, Corollary 3.2.7 (a). On ω, the powers are read as Proposition 1.6.4 (a)–(c), the parity claim as its (d), and cancellation as Proposition 1.6.3 (b), (f); with that, each of these proofs runs on ω. A later chapter may cite them only in a review (R) exercise; anywhere else, a construction that needs one of them proves it on ω first. Part (b) of Sets, Relations, and Cardinality, Corollary 3.2.7 is not in this group and is not checked on ω here: *Sets, Relations, and Cardinality*, Chapter 3 treats its list $A_1,\dots,A_k$ as a family in the everyday sense of Sets, Relations, and Cardinality, Definition 1.2.3, with no set holding it, and Section 1.2 left such families to the axiom of Replacement. A list given as a function on $\{1,\dots,k\}$, with graph G, does have an ambient set: each A_i lies in $\bigcup\bigcup G$ by (Z4), so the union of the A_i exists by (Z4) and (Z6), and the proof of (b) can then be run on ω, with its base case (a) read as above and its Sets, Relations, and Cardinality, Theorem 2.2.1 from base 2 read by Proposition 1.6.3 (g). A later chapter that needs (b) proves it on ω first, in that form. Sets, Relations, and Cardinality, Theorem 3.2.8 is a statement about the school $\mathbb{Z}$ and $\mathbb{Q}$, and is never applied to ω.

(iii) **Templates.** Sets, Relations, and Cardinality, Definition 2.3.1 and Sets, Relations, and Cardinality, Proposition 2.3.4 are used later only as patterns, with the school $\mathbb{R}$ replaced by an ordered field F. Their $\mathbb{N}$ side is read through the dictionary, and their field side is proved again from the axioms of F. The parts that need $\mathbb{Z}$ (Sets, Relations, and Cardinality, Proposition 2.3.4 (d) for an integer base, and the general index of Sets, Relations, and Cardinality, Definition 2.3.1, which uses Sets, Relations, and Cardinality, Proposition 2.1.2) are not part of the pattern.

Finally, the axiom of choice. Apart from *Sets, Relations, and Cardinality*, Chapter 2's $\mathbb{N} \subseteq \mathbb{R}$ and its school operations, which Corollary 1.7.3 (a) takes as given, every set in Sections 1.1–1.7 is produced by one of (Z1)–(Z7). Every function is produced by Theorem 1.5.1, by (Z6) applied to a graph described by a property (as g was in the proof of Theorem 1.7.1 (a)), or by *Sets,*

Relations, and Cardinality, Chapter 1's inverses and composites of such functions. Nothing is chosen from infinitely many sets, and Sets, Relations, and Cardinality, Axiom 3.3.1 is used nowhere in this chapter; it is mentioned only to say where it lives.

So everything that *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 assumed about $\mathbb{N}$ (that 0 is a natural number, that $n+1$ is one whenever n is, induction, recursion, and the arithmetic of Sets, Relations, and Cardinality, Lemma 2.1.1) is now proved, from seven axioms about sets. The set-theory volume returns to those axioms and adds the rest of the system. Chapter 2 builds $\mathbb{Z}$ and $\mathbb{Q}$ from ω by the same kind of construction as here: pairs, and classes of pairs.

1.7.6 Exercises

1.7.6.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 1.7.1.

(a) Where in the proof is (P5) for M used, and where are (P3) and (P4) for M used?

(b) Where is (P4) for ω used? (It is not visible in the proof itself.)

(c) Run φ of part (a) for the system $(\omega \setminus \{0\}, 1, \mathrm{S})$ on 0, 1 and 2, from the two equations alone.

(d) State the idea of the proof of (a) and (b) in two sentences.

Solution.

(a) (P5) for M is used once, for surjectivity: the image $\varphi(\omega)$ contains e and is closed under s, so it is M. (P3) for M is used for injectivity, in the base case ($\varphi(\mathrm{S}(k)) = s(\varphi(k)) \neq e$, so only $m = 0$ has $\varphi(m) = e$) and at the start of the inductive step (to exclude $m = 0$). (P4) for M is used in the inductive step, to pass from $s(\varphi(k)) = s(\varphi(n))$ to $\varphi(k) = \varphi(n)$.

(b) Inside part (a), through Theorem 1.5.1: its proof extends a partial solution by one new pair at $\mathrm{S}(n)$, and (P4) for ω makes $\mathrm{S}(n)$ the successor of n only, so the equation at $\mathrm{S}(n)$ asks for $g(n, h(n))$ and nothing else.

(c) $\varphi(0) = 1$; $\varphi(1) = \varphi(\mathrm{S}(0)) = \mathrm{S}(\varphi(0)) = \mathrm{S}(1) = 2$; $\varphi(2) = \varphi(\mathrm{S}(1)) = \mathrm{S}(2) = 3$.

(d) The map is forced at 0 and at each successor, so the recursion theorem builds it and makes it unique. It is onto because its image contains e and is closed under s, so (P5) for M applies; it is one-to-one by an induction whose base and step are (P3) and (P4) for M.

1.7.6.2 B. Practice

Exercise. B1: The even numbers

Let $E \coloneqq \{2n : n \in \omega\}$, with $e \coloneqq 0$ and $s(x) \coloneqq x + 2$ for $x \in E$. Show that $(E, 0, s)$ is a Peano system, and that the map φ of Theorem 1.7.1 is $\varphi(n) = 2n$.

Solution. E is a set by (Z6) inside ω: it is $\{x \in \omega : x = 2n \text{ for some } n \in \omega\}$. By Definition 1.5.1 (b), $2 \cdot \mathrm{S}(n) = 2n + 2$ for every $n \in \omega$. So s maps E into E: $s(2n) = 2n + 2 = 2 \cdot \mathrm{S}(n) \in E$.

(P3): $2n + 2 = 2n + \mathrm{S}(1) = \mathrm{S}(2n + 1)$ by Definition 1.5.1 (a), and a successor is never 0 (Theorem 1.4.2 (P3)). (P4): if $x + 2 = y + 2$, then $x = y$ by Proposition 1.6.3 (b). (P5): let $A \subseteq E$ with $0 \in A$ and $x + 2 \in A$ for every $x \in A$. Let $B \coloneqq \{n \in \omega : 2n \in A\}$. Then $0 \in B$, since $2 \cdot 0 = 0 \in A$; and if $n \in B$, then $2 \cdot \mathrm{S}(n) = 2n + 2 \in A$, so $\mathrm{S}(n) \in B$. By Theorem 1.3.3 (a), $B = \omega$, so every $2n$ lies in A, and $A = E$.

The map $n \mapsto 2n$ from ω to E sends 0 to $2 \cdot 0 = 0 = e$, and $2 \cdot \mathrm{S}(n) = 2n + 2 = s(2n)$. By the uniqueness in Theorem 1.7.1 (a), it is φ. In particular, by (b) of that theorem, $n \mapsto 2n$ is a bijection $\omega \to E$.

Exercise. B2: The identification at work

Let $\varphi\colon \omega \to \mathbb{N}$ be the bijection of Corollary 1.7.3 (a).

(a) Compute $\varphi(3)$ step by step from the two equations of φ.

(b) Compute $\varphi(2 \cdot 3)$ in two ways: from (a) of the corollary, and by unwinding $2 \cdot 3$ in ω to a numeral first.

Solution.

(a) Using $\varphi(0) = 0$ and $\varphi(\mathrm{S}(n)) = \varphi(n) + 1$, with $3 = \mathrm{S}(2)$, $2 = \mathrm{S}(1)$, $1 = \mathrm{S}(0)$ (Definition 1.5.1 (c)):

$$\varphi(3) = \varphi(2) + 1 = (\varphi(1) + 1) + 1 = ((0 + 1) + 1) + 1 = 3.$$

(b) By the corollary, $\varphi(2 \cdot 3) = \varphi(2)\varphi(3) = 2 \cdot 3 = 6$ in $\mathbb{N}$, where $\varphi(2) = 2$ as in (a). In ω, $2 \cdot 3 = 2 \cdot 2 + 2 = (2 \cdot 1 + 2) + 2 = ((2 \cdot 0 + 2) + 2) + 2 = 6$, by Definition 1.5.1 (b) three times and then $2 \cdot 0 = 0$, $0 + 2 = 2$, $2 + 2 = 4$, $4 + 2 = 6$ from the equations for $+$. Then $\varphi(6) = 6$ by six steps as in (a). The two answers agree, as the corollary says they must.

1.7.6.3 C. Going deeper

Exercise. C1: Writing out categoricity

Without looking at the text, prove part (b) of Theorem 1.7.1: for a Peano system (M, e, s), the map $\varphi\colon \omega \to M$ with $\varphi(0) = e$ and $\varphi(\mathrm{S}(n)) = s(\varphi(n))$ is a bijection.

Solution (Hint 1). For surjectivity, use (P5) for M on a set built from φ. For injectivity, induct on n; decide first where "for every m" goes.

Solution (Hint 2). The induction for injectivity starts: "Let $Q(n)$ be: for every $m \in \omega$, $\varphi(m) = \varphi(n)$ implies $m = n$."

Solution. Onto. The image $B := \varphi(\omega)$ (Sets, Relations, and Cardinality, Definition 1.6.1) contains $e = \varphi(0)$, and $x = \varphi(n) \in B$ gives $s(x) = \varphi(\mathrm{S}(n)) \in B$. By (P5) for M, $B = M$.

One-to-one. By Theorem 1.3.3 (b) we prove, for every n, $Q(n)$: "for every $m \in \omega$, $\varphi(m) = \varphi(n)$ implies $m = n$". For $Q(0)$, let $\varphi(m) = e$; an $m \neq 0$ is $\mathrm{S}(k)$ (Lemma 1.4.1 (d)), with $\varphi(m) = s(\varphi(k)) \neq e$ by (P3) for M; so $m = 0$. For the step, assume $Q(n)$ and $\varphi(m) = s(\varphi(n))$. Then $m \neq 0$, as e is not a value of s (P3 for M); so $m = \mathrm{S}(k)$, and $s(\varphi(k)) = s(\varphi(n))$ gives $\varphi(k) = \varphi(n)$ by (P4) for M, hence $k = n$ by $Q(n)$ used at k, and $m = \mathrm{S}(n)$. So φ is injective, and a bijection.

Marking scheme. (10 points) - 3: surjectivity: shows the image of φ contains e and is closed under s, and cites (P5) for M. - 2: sets up an induction (on either variable, or by least counterexample) whose hypothesis is strong enough to use at the predecessor, e.g. $Q(n)$ with "for every m" inside, or the difference $j \neq 0$ fixed and induction on m. - 2: base case, via $m = \mathrm{S}(k)$ for $m \neq 0$ and (P3) for M. - 3: inductive step, using (P3) for M to get $m \neq 0$, (P4) for M to cancel s, and the induction hypothesis at the predecessor k. Fatal errors (0 for the part concerned): an induction whose hypothesis is then applied at a value it does not cover (typically: m fixed outside Q, then used at k); using injectivity of φ to prove injectivity; using (P3), (P4) or (P5) **for** ω where the axiom for M is needed. A proof of injectivity by Theorem 1.6.2 (a least n with some $m \neq n$, $\varphi(m) = \varphi(n)$) earns the same 7 points, mapped to base, step and hypothesis in the same way. So does the fixed-difference route (reducing to $\varphi(m + j) \neq \varphi(m)$ for $j \neq 0$ by Theorem 1.6.1 (a) and Proposition 1.6.3 (a), and inducting on m): its base uses (P3) for M, its step (P4) for M and the hypothesis at m, and the points are mapped in the same way.

Exercise. C2: Without induction, no recursion

Let (M, e, s) be structure (iii) of Example (Each axiom is needed), the natural numbers beside a two-sided chain: $M = \omega \times 3$, $e = (0, 0)$, $s(n, 0) = (\mathrm{S}(n), 0)$, $s(n, 1) = (\mathrm{S}(n), 1)$, $s(0, 2) = (0, 1)$ and $s(\mathrm{S}(k), 2) = (k, 2)$. Show that there is **no** function $f\colon M \to \omega$ with $f(e) = 0$ and $f(s(x)) = \mathrm{S}(f(x))$ for every $x \in M$. Which Peano axiom, missing in M, is responsible?

Solution. Suppose such an f exists. For every $k \in \omega$, $s(\mathrm{S}(k), 2) = (k, 2)$, so

$$f((k, 2)) = \mathrm{S}\Big(f((\mathrm{S}(k), 2))\Big).$$

Let $T := \{f((k, 2)) : k \in \omega\} \subseteq \omega$, a set by (Z6) inside ω; it is non-empty, since it contains $f((0, 2))$. By Theorem 1.6.2, T has a least element $t = f((k_0, 2))$. But $u := f((\mathrm{S}(k_0), 2)) \in T$, and $t = \mathrm{S}(u)$, so $u \in \mathrm{S}(u) = t$, that is, $u < t$ (Definition 1.6.1). This contradicts the leastness of t, since $t \leq u$ and $u < t$ would give $u < u$ (Theorem 1.6.1 (a)). So no such f exists. (The condition $f(e) = 0$ was not even used: no map at all turns s into S.)

M satisfies (P3) and (P4) and fails only (P5) (Example (Each axiom is needed)). So (P5) is what makes the recursion theorem possible: the downward chain $(0, 2), (1, 2), (2, 2), \ldots$, each

element the successor of the next, would need values in ω that decrease forever.

1.7.6.4 R. Review

Exercise. R1

Prove that ω is infinite.

Solution. We read Sets, Relations, and Cardinality, Definition 3.1.3 and Sets, Relations, and Cardinality, Theorem 3.1.9 on ω, through the dictionary of this section (group (i)). Consider S with codomain $\omega \setminus \{0\}$; its values lie there by Theorem 1.4.2 (P3). It is injective by (P4) of that theorem, and onto $\omega \setminus \{0\}$ by Lemma 1.4.1 (d): every $n \neq 0$ is $\mathrm{S}(k)$ for some $k \in \omega$. So it is a bijection $\omega \to \omega \setminus \{0\}$. The set $\omega \setminus \{0\}$ is a proper subset of ω, since $0 \in \omega$ and $0 \notin \omega \setminus \{0\}$. By Sets, Relations, and Cardinality, Theorem 3.1.9, a finite set is not equinumerous with any proper subset; so ω is not finite, that is, it is infinite (Sets, Relations, and Cardinality, Definition 3.1.3). This is the Quick check after Sets, Relations, and Cardinality, Example (Hilbert's Hotel) (every guest moves up one room), built from (P3) and (P4).

Exercise. R2

Show that the successor function $\mathrm{S}\colon \omega \to \omega$ has a left inverse and no right inverse. Write a left inverse down explicitly. Is it the only one?

Solution. No right inverse. A function with a right inverse is surjective (Sets, Relations, and Cardinality, Proposition 1.5.2), and S is not: 0 is not a value, by Theorem 1.4.2 (P3).

A left inverse. Let $L\colon \omega \to \omega$ have the graph

$$G := \{(n, k) \in \omega \times \omega : (n = 0 \text{ and } k = 0) \text{ or } n = \mathrm{S}(k)\},$$

a set by (Z6). Each $n \in \omega$ has exactly one partner. If $n = 0$, the partner $k = 0$ works, and no k has $0 = \mathrm{S}(k)$, by (P3). If $n \neq 0$, then $n = \mathrm{S}(k)$ for some k (Lemma 1.4.1 (d)), and only one, by (P4). So L is a function (Sets, Relations, and Cardinality, Definition 1.4.1) with $L(0) = 0$ and $L(\mathrm{S}(k)) = k$. Hence $L(\mathrm{S}(n)) = n$ for every $n \in \omega$, that is, $L \circ \mathrm{S} = \mathrm{id}_\omega$: L is a left inverse of S (Sets, Relations, and Cardinality, Definition 1.5.1 (a)). This matches Sets, Relations, and Cardinality, Theorem 1.5.1: S is injective and $\omega \neq \emptyset$.

Not unique. The value of a left inverse at 0 is never used, since 0 is not of the form $\mathrm{S}(n)$; setting $L'(0) := 5$ and $L'(\mathrm{S}(k)) := k$ gives another left inverse.

Exercise. R3

Let $M = \{0, 1\}$, $e = 0$ and $s(0) = s(1) = 1$. Show that there is no function $f\colon M \to \omega$ with $f(0) = 0$ and $f(s(x)) = \mathrm{S}(f(x))$ for every $x \in M$. Which of the Peano axioms play the part, for definitions on ω, that a theorem about reading formulas plays for definitions on formulas?

Solution. Suppose f exists. Taking $x = 1$, $f(1) = f(s(1)) = \mathrm{S}(f(1))$. But no $n \in \omega$ has $n = \mathrm{S}(n)$: $n \in \mathrm{S}(n)$ (Definition 1.3.1), so $n = \mathrm{S}(n)$ would give $n \in n$, against Lemma 1.4.1 (c). So no such f exists. (Taking $x = 0$ instead gives $f(1) = \mathrm{S}(0) = 1$, so the two equations at the input 1 also contradict each other.)

The trouble is that 1 is built in two ways, as $s(0)$ and as $s(1)$, and the two ways ask for different values. In *Logic and Proof*, Chapter 3, Logic and Proof, Theorem 4.3.4 defines functions on formulas, and what makes it work is Logic and Proof, Theorem 4.3.3: each compound formula is built from its parts in exactly one way. On ω the same job is done by (P3) and (P4): 0 is not a successor, so its value is not asked for twice, and a successor is the successor of exactly one number, so the equation at $\mathrm{S}(n)$ asks for one thing. The structure M fails (P4), since $s(0) = s(1)$, though it satisfies (P5).

Exercise. R4

For $n \in \omega$, compute $\bigcup \mathrm{S}(n)$ and $\bigcup n$. Then find a set x with $\bigcup \mathrm{S}(x) \neq x$.

Solution. A formula for every set. For any set x, $\bigcup \mathrm{S}(x) = \bigcup(x \cup \{x\}) = \bigcup x \cup x$: an element of a member of $x \cup \{x\}$ is an element of a member of x, or an element of x itself.

For $n \in \omega$. The tool is transitivity. By Lemma 1.4.1 (a), n is transitive, so every element of a member of n is a member of n, that is, $\bigcup n \subseteq n$. Hence $\bigcup \mathrm{S}(n) = \bigcup n \cup n = n$. For $\bigcup n$ itself: $\bigcup 0 = \bigcup \emptyset = \emptyset = 0$; and if $n \neq 0$, then $n = \mathrm{S}(k)$ (Lemma 1.4.1 (d)), and $\bigcup n = \bigcup \mathrm{S}(k) = k$ by what we just proved. So $\bigcup$ undoes the successor: it is the "predecessor" on ω, with 0 sent to 0.

A set where it fails. Let $x \coloneqq \{1\} = \{\{\emptyset\}\}$. Then $\bigcup x = 1 = \{\emptyset\}$, and $\bigcup \mathrm{S}(x) = \{\emptyset\} \cup \{\{\emptyset\}\} = \{0, 1\} = 2$. This is not x: $0 = \emptyset \in 2$, but $\emptyset \notin x$, since the only element of x is $\{\emptyset\} \neq \emptyset$.

Why the identity alone is not enough. The look-alike tool is the identity $\bigcup \mathrm{S}(x) = \bigcup x \cup x$, which holds for every set. On its own it does not give x: it gives $\bigcup \mathrm{S}(x) = x$ **exactly** when $\bigcup x \subseteq x$, that is, when x is transitive (Definition 1.4.1), and $\{1\}$, which is not transitive, is a witness that this can fail. Nor is an induction on n a route that avoids transitivity: its step $\bigcup \mathrm{S}(\mathrm{S}(n)) = \bigcup \mathrm{S}(n) \cup \mathrm{S}(n) = n \cup \mathrm{S}(n) = \mathrm{S}(n)$ proves $\bigcup \mathrm{S}(n) = n$, which is equivalent to transitivity of n, so it proves Lemma 1.4.1 (a) again along the way.

Chapter 2

Building the Integers and the Rationals from Equivalence Classes

"Subtract 5 from 3" has no answer in $\mathbb{N}$, and "halve 1" has none among the integers. School answers by inventing -2 and $\frac{1}{2}$. This chapter **builds** them, from pairs and nothing else. The pairs $(3,5)$ and $(0,2)$ both record "two less than nothing", and the integer -2 is the set of all pairs that record it; the pairs $(1,2)$ and $(2,4)$ both record "one divided by two", and the rational $\frac{1}{2}$ is the set of all pairs that record that. The tool is the quotient of a set by an equivalence relation, from *Sets, Relations, and Cardinality*, Section 2.5, used twice: once on $\mathbb{N} \times \mathbb{N}$ to make $\mathbb{Z}$, once on $\mathbb{Z} \times \mathbb{Z}_{>0}$ to make $\mathbb{Q}$. The one thing that can go wrong is an operation that depends on which pair was used to name a class, and the chapter's central skill is checking that it does not. At the end, $\mathbb{Q}$ is an ordered field, every rule for inequalities used in this chapter is proved from the order axioms, and every rational lies below a natural number. Yet $\mathbb{Q}$ still has a hole where $\sqrt{2}$ should be: no rational squares to 2, and the rationals whose square is below 2 have no least upper bound in $\mathbb{Q}$. Later chapters of this volume fill the hole.

The school picture. *Logic and Proof*, Chapters 1–5, and *Sets, Relations, and Cardinality*, Chapters 1–3 treated $\mathbb{Z}$ and $\mathbb{Q}$ as known subsets of the familiar real line (*Sets, Relations, and Cardinality*, Section 2.1), and used their arithmetic and order as school facts. From this chapter on they are the sets built here, and the familiar facts are theorems. Section 2.2 proves $\mathbb{Z} = \mathbb{N} \cup \{-n : n \in \mathbb{N}\}$, the description that *Sets, Relations, and Cardinality*, Section 2.1 took as a definition, and Section 2.4 proves $\mathbb{Q} = \{m/n : m \in \mathbb{Z},\ n \in \mathbb{Z}_{>0}\}$ in the same way. Nothing earlier was circular: *Logic and Proof*, Chapters 1–5, and *Sets, Relations, and Cardinality*, Chapters 1–3 used these facts as stated facts, never as proved ones. No proof in this chapter uses a school fact about the integers, the rationals or the real numbers; school decimals appear once, in a table in Section 2.6 marked as being for orientation only. The results of *Sets, Relations, and Cardinality*, Chapters 2 and 3 about their integers, such as the division algorithm and congruences, are not cited for the constructed $\mathbb{Z}$; a later chapter of this volume rebuilds them.

What you need. Chapter 1 and *Sets, Relations, and Cardinality*, Chapters 1 and 2, chiefly Chapter 1 and *Sets, Relations, and Cardinality*, Section 2.5. From Chapter 1, the natural numbers ω with the citing rule of Section 1.7 (Corollary 1.7.3), by which $\mathbb{N}$ means ω and induction (Logic and Proof, Axiom 4.2.1) and the recursion theorem (Sets, Relations, and Cardinality, Theorem 2.3.1) of *Logic and Proof*, Chapter 3, and *Sets, Relations, and Cardinality*, Chapter 2 apply to it; the arithmetic of ω (Definition 1.5.1, Theorem 1.5.3, Corollary 1.5.2), its order, cancellation and subtraction (Theorem 1.6.1, Proposition 1.6.3), well-ordering (Theorem 1.6.2), induction and recursion on ω (Theorem 1.3.3, Theorem 1.5.1), and the theorem that no pair is a natural number (Proposition 1.4.3). From *Sets, Relations, and Cardinality*, Section 2.5: equiva-

lence relations, classes and quotient sets (Sets, Relations, and Cardinality, Definition 2.5.1, Sets, Relations, and Cardinality, Definition 2.5.2, Sets, Relations, and Cardinality, Definition 2.5.4, Sets, Relations, and Cardinality, Theorem 2.5.1), and the two examples that this chapter generalizes, arithmetic modulo n (Sets, Relations, and Cardinality, Proposition 2.5.5) and a rule on classes that is not a function (Sets, Relations, and Cardinality, Example (A rule on classes that is not a function)). From *Sets, Relations, and Cardinality*, Section 2.6: partial orders, upper bounds and suprema (Sets, Relations, and Cardinality, Definition 2.6.1, Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5). From *Sets, Relations, and Cardinality*, Chapter 1: sets cut out by a property (Sets, Relations, and Cardinality, Definition 1.1.3), ordered pairs, products and relations (Sets, Relations, and Cardinality, Definition 1.3.1, Sets, Relations, and Cardinality, Definition 1.3.2, Sets, Relations, and Cardinality, Definition 1.3.3), and functions, their equality and injectivity (Sets, Relations, and Cardinality, Definition 1.4.1, Sets, Relations, and Cardinality, Proposition 1.4.1, Sets, Relations, and Cardinality, Definition 1.4.3). From *Logic and Proof*, Chapter 5, the catalogue of what a proof may not do (Logic and Proof, Example (What is not allowed)), for the find-the-error exercises. Section 2.6 compares its proof with *Logic and Proof*, Chapter 2's parity proof that $\sqrt{2}$ is irrational (Logic and Proof, Theorem 2.2.1), which it does not use, and names the rule ¬I (Logic and Proof, Definition 2.2.1, Logic and Proof, Definition 2.5.1). A few exercises reach further back: the negation laws of *Logic and Proof*, Chapters 2 and 4 (Logic and Proof, Proposition 2.5.5, Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2) in Section 2.5's A1 and in review exercise R3; finite and infinite sets (Sets, Relations, and Cardinality, Definition 3.1.2, Sets, Relations, and Cardinality, Definition 3.1.3, Sets, Relations, and Cardinality, Corollary 3.1.3, Sets, Relations, and Cardinality, Proposition 3.1.6) in Section 2.3's C1; a quotient by the fibers of a function (Sets, Relations, and Cardinality, Example (The fibers of a function)) in Section 2.2's C1; and countable sets (Sets, Relations, and Cardinality, Theorem 3.2.6, Sets, Relations, and Cardinality, Corollary 3.2.7, Sets, Relations, and Cardinality, Proposition 3.2.3, Sets, Relations, and Cardinality, Proposition 3.2.4) in review exercise R2, read on ω through Section 1.7 (Proposition 1.6.4).

Time. About 25 hours on the core track (all of A, one B and one C per section, the exit checks); about 35 hours for everything.

2.0.1 Roadmap

- **The Integers as Classes of Pairs.** Pairs (a, b) of natural numbers as notes saying "a minus b", the cross-sum relation $a+d = b+c$ proved an equivalence relation by cancellation in $\mathbb{N}$, $\mathbb{Z}$ as its quotient set, the classes drawn as diagonals of the quarter plane, and the normal form $[(n, 0)]$ or $[(0, n)]$ of each integer.
- **Well-Definedness, and the Arithmetic of the Integers.** The lemma that a rule on classes is a function as soon as equivalent inputs give equivalent outputs (for maps, for binary operations one argument at a time, and for properties), the sum, product and negative of integers checked against it, rules that fail it, and the copy of $\mathbb{N}$ inside $\mathbb{Z}$ that makes $\mathbb{Z} = \mathbb{N} \cup \{-n : n \in \mathbb{N}\}$ a theorem.
- **The Integers as an Ordered Ring.** The axioms (F1)–(F9) of a commutative ring and of a field, $\mathbb{Z}$ a commutative ring by reduction of each law to $\mathbb{N}$, ordered rings and the orders that come from a positive set, the order of $\mathbb{Z}$ and its discreteness, the rules for inequalities in every ordered ring, least elements of bounded sets of integers, and a field with two elements that cannot be ordered.
- **The Rationals as Classes of Fractions.** Pairs (p, q) with $q > 0$ identified by cross-multiplication, $\mathbb{Q}$ as their quotient set with a well-defined sum and product, the proof that $\mathbb{Q}$ is a field, the copy of $\mathbb{Z}$ inside it that turns the class $[(p, q)]$ into the quotient p/q, and why non-zero denominators would give the same set but spoil the order.

- **The Rationals as an Ordered Field; the Archimedean Property.** The positive rationals as the classes with a positive numerator, $\mathbb{Q}$ an ordered field with the comparison formula $p/q < r/s \iff ps < rq$, reciprocals reversing order, a midpoint between any two rationals, the Archimedean property proved from the construction with no supremum in sight, its form $1/n < \varepsilon$ for rational $\varepsilon > 0$, and the integer part $\lfloor x \rfloor$.
- **The Gaps in the Rationals.** No rational squares to 2, proved by descent on a least denominator rather than by parity, the set $\{x \in \mathbb{Q} : x^2 < 2\}$ bounded above with no supremum in $\mathbb{Q}$, and the rational sequence $2, \frac{3}{2}, \frac{17}{12}, \frac{577}{408}, \ldots$ that closes in on the gap from above; the review exercises reach back to Chapter 1, *Logic and Proof*, Chapters 2, 4, and *Sets, Relations, and Cardinality*, Chapters 2, 3.

2.0.2 Named moves

- **Check on representatives, one argument at a time.** To show that $[x]*[y] \coloneqq [g(x,y)]$ is well defined, change x to an equivalent x' with y fixed, then y to an equivalent y' with x' fixed, and join the two changes by transitivity, as the lemma of Section 2.2 does. Each change is one computation at the level below. When the formula is symmetric in its two inputs, the second change comes free. A formula that merely looks right is not yet a definition.
- **Reduce to the level below.** A law about classes, written on representatives, becomes an identity in $\mathbb{N}$ (for $\mathbb{Z}$) or in $\mathbb{Z}$ (for $\mathbb{Q}$), which is already proved. The two sides are either the same pair or two related pairs, and in both cases the classes agree.
- **Order from a positive set.** To order a commutative ring, say which elements are positive, check that they are closed under $+$ and $\cdot$ and that every element is exactly one of positive, zero or minus a positive, and let $a < b$ mean that $b - a$ is positive. Every rule for inequalities then follows by turning the comparison into one positive element.

2.1 The Integers as Classes of Pairs

Chapter 1 built the natural numbers as the set ω, with its 0, $+$, $\cdot$ and $<$, and proved the laws of arithmetic, the order and well-ordering from the axioms for sets. From this chapter on, $\mathbb{N}$ **means** that set (Corollary 1.7.3 and the citing rule of Section 1.7). In $\mathbb{N}$ the equation $5 + x = 3$ has no solution: $5 \leq 5 + x$ for every $x \in \mathbb{N}$ by Proposition 1.6.3 (a), while $3 < 5$ by the same result, as $5 = 3 + 2$ with $2 \neq 0$. So $5 + x = 3$ would give $5 \leq 3 < 5$, hence $5 < 5$, which Theorem 1.6.1 (a) rules out. This section builds a larger set $\mathbb{Z}$, out of $\mathbb{N}$ alone, with the one tool of *Sets, Relations, and Cardinality*, Section 2.5: the quotient of a set by an equivalence relation.

2.1.1 What we use about the natural numbers

Every fact about $\mathbb{N}$ in this section and the next is one of the following results of Chapter 1, and each is cited where a proof uses it. The defining equation $n + 0 = n$ and the evaluation of numeral sums and comparisons such as $2 + 3 = 5$ and $0 \neq 2$ rest on Definition 1.5.1 and Theorem 1.6.1 and are not cited each time.

- **Defining equations** (Definition 1.5.1): $n + 0 = n$, $m + \mathrm{S}(n) = \mathrm{S}(m + n)$ and $n \cdot 0 = 0$, and the numerals $0 \coloneqq \emptyset$, 1, 2, $\ldots$, with $n + 1 = \mathrm{S}(n)$.
- **Laws of arithmetic** (Theorem 1.5.3): $+$ and $\cdot$ are commutative and associative, $\cdot$ distributes over $+$, and $0 + n = n$, $1 \cdot n = n$, $0 \cdot n = 0$.
- **Cancellation** (Proposition 1.6.3 (b), (f)): $m + k = n + k$ gives $m = n$; and $mk = nk$ with $k \neq 0$ gives $m = n$.

- **Order** (Theorem 1.6.1 (a)): $<$ is a strict total order, so for all m, n **exactly one** of $m < n$, $m = n$, $n < m$ holds, and $<$ is transitive and irreflexive.
- **Order and addition** (Proposition 1.6.3 (a)): $m \leq n$ if and only if $n = m + j$ for some $j \in \mathbb{N}$; and $m < n$ if and only if $n = m + j$ for some $j \neq 0$. Adding preserves $<$ (Proposition 1.6.3 (c)).
- **Small facts and well-ordering**: $0 < 1$ (Theorem 1.6.1 (d)); no k satisfies $n < k < \mathrm{S}(n)$ (Theorem 1.6.1 (b)); every non-empty subset of $\mathbb{N}$ has a least element (Theorem 1.6.2).

That is the whole list for now. Nothing about the integers or the rationals of school enters a proof: from here on they are the sets this chapter builds, and their familiar properties become theorems.

2.1.2 A naive attempt

The school picture of an integer is "a natural number with a sign": $+3$, -2, 0. It is tempting to take that literally and let an integer be a pair (s, n) with a sign $s \in \{0, 1\}$ (read 0 as "plus" and 1 as "minus") and a size $n \in \mathbb{N}$. Two things go wrong at once.

First, zero gets two names, $(0, 0)$ and $(1, 0)$: a "plus zero" and a "minus zero". Either both stay, and every theorem must treat them as the same, or one is banned, and every definition must check that it never produces the banned one.

Second, addition needs cases. Adding $+m$ and $-n$ should give $+(m - n)$ when $n \leq m$ and $-(n - m)$ when $m < n$, where $m - n$ is the subtraction of Proposition 1.6.3 (a), defined on $\mathbb{N}$ only when $n \leq m$. There are four sign patterns for a sum, and two of them split again by the order of the sizes. A proof that addition is associative would split by the signs of all three terms and then by the order of their sizes, and every case would subtract in $\mathbb{N}$ in its own direction.

The trouble in both is the same: the definition **performs** a subtraction that $\mathbb{N}$ can only do one way round. The construction below never performs it. It records the subtraction as a pair, and lets an equivalence relation decide when two recorded subtractions should count as equal.

2.1.3 Pairs that record a difference

Think of a pair $(a, b) \in \mathbb{N} \times \mathbb{N}$ as a note saying "a minus b", to be computed later or never. The note $(3, 5)$ is legitimate even though $3 - 5$ is not a natural number. Many notes say the same thing: $(3, 5)$, $(0, 2)$ and $(10, 12)$ all describe "two less than nothing", which school calls -2.

When do two notes (a, b) and (c, d) describe the same thing? School would write $a - b = c - d$, and then add $b + d$ to both sides to get $a + d = b + c$. That last equation is the one we keep, because it uses only addition, which $\mathbb{N}$ has for all inputs. The computation with minus signs is a **heuristic**, borrowed from school to find the definition; the definition does not depend on it.

An integer is a difference you have not computed yet.

Definition 2.1.1. The Integers

For $(a, b), (c, d) \in \mathbb{N} \times \mathbb{N}$, let

$$(a, b) \sim (c, d) \iff a + d = b + c.$$

The set of **integers** is the quotient set

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\!\sim$$

of Sets, Relations, and Cardinality, Definition 2.5.4, and an **integer** is an element of $\mathbb{Z}$. The class of the pair (a, b) is written $[(a, b)]$:

$$[(a, b)] = \{(c, d) \in \mathbb{N} \times \mathbb{N} : (c, d) \sim (a, b)\}.$$

In words: two pairs are related when their **cross sums** agree, the first coordinate of each added to the second coordinate of the other. The relation is a subset of $(\mathbb{N} \times \mathbb{N}) \times (\mathbb{N} \times \mathbb{N})$, described by a property (Sets, Relations, and Cardinality, Definition 1.3.3, Sets, Relations, and Cardinality, Definition 1.3.2). An integer is one whole class: the set of all pairs that record the same difference. The notation $[(a, b)]$ is the class notation $[x]$ of Sets, Relations, and Cardinality, Definition 2.5.2 with $x = (a, b)$; no new bracket is introduced.

Well-definedness. Sets, Relations, and Cardinality, Definition 2.5.4 is stated for an **equivalence relation**. Until $\sim$ is shown to be one, the definition of $\mathbb{Z}$ is a plan, not a definition. The next proposition supplies the missing check, and it is placed directly after the definition for that reason.

An integer is a set of pairs. It is not a natural number with a sign, and it is not a pair. *Sets, Relations, and Cardinality*, Chapter 1 adopted the convention that a pair is not a number (after Sets, Relations, and Cardinality, Definition 1.3.1), and on ω that became a theorem, Proposition 1.4.3. The same care is needed here: $[(0, 2)]$ is the set $\{(0, 2), (1, 3), (2, 4), \dots\}$, and the school -2 is what those pairs have in common. The construction takes "what they have in common" to be **the set of them**.

Proposition 2.1.1. The Relation Behind the Integers

The relation $\sim$ of Definition 2.1.1 is an equivalence relation on $\mathbb{N} \times \mathbb{N}$ (Sets, Relations, and Cardinality, Definition 2.5.1).

Idea. Reflexivity and symmetry are rearrangements of one equation, by commutativity. Transitivity has two hypotheses, $a + d = b + c$ and $c + f = d + e$, and a goal $a + f = b + e$ in which c and d do not appear. In school one would subtract them. In $\mathbb{N}$ we add the two equations, so that $c + d$ appears on both sides, and then remove it by cancellation.

Before reading on: which line uses cancellation in $\mathbb{N}$, and why is it more direct than subtracting $c + d$?

Proof. We check (Eq1)–(Eq3) of Sets, Relations, and Cardinality, Definition 2.5.1. Let $(a, b), (c, d), (e, f) \in \mathbb{N} \times \mathbb{N}$.

(Eq1) $a + b = b + a$ by commutativity (Theorem 1.5.3 (b)), so $(a, b) \sim (a, b)$.

(Eq2) Suppose $(a, b) \sim (c, d)$, that is, $a + d = b + c$. By commutativity, $c + b = b + c = a + d = d + a$, and $c + b = d + a$ says $(c, d) \sim (a, b)$.

(Eq3) Suppose $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, that is,

$$a + d = b + c \quad \text{and} \quad c + f = d + e.$$

Adding the two equations gives $(a + d) + (c + f) = (b + c) + (d + e)$. By associativity and commutativity (Theorem 1.5.3 (b), (c)), the left side is $(a + f) + (c + d)$ and the right side

is $(b+e)+(c+d)$, so

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

By cancellation (Proposition 1.6.3 (b)) with $k=c+d$, $a+f=b+e$. That is $(a,b)\sim(e,f)$.

This proves that $\sim$ is an equivalence relation.

■

The answer to the prompt is the last line of (Eq3). It is tempting to "subtract $c+d$ from both sides" at the end. In $\mathbb{N}$ the difference $n-m$ of Proposition 1.6.3 (a) exists only when $m\le n$. Here it does exist, since $c+d\le(a+f)+(c+d)$ and $c+d\le(b+e)+(c+d)$ by the same result and commutativity, but only after that check; and reading off $((a+f)+(c+d))-(c+d)=a+f$ uses the uniqueness in (a), which is itself proved by cancellation. Cancellation removes $c+d$ directly, and it holds for all k, with no size check.

So $\mathbb{Z}=(\mathbb{N}\times\mathbb{N})/\sim$ is now defined, and by Sets, Relations, and Cardinality, Theorem 2.5.1 its elements partition $\mathbb{N}\times\mathbb{N}$: every pair lies in exactly one integer, and $[(a,b)]=[(c,d)]$ if and only if $a+d=b+c$.

2.1.4 Pictures of classes

The quarter plane $\mathbb{N}\times\mathbb{N}$ is a grid of lattice points, and the classes are easy to see in it.

Example. Three classes, listed and drawn

List the elements (x,y) with $x\le 4$ of each of the classes $[(3,5)]$, $[(0,0)]$ and $[(4,1)]$, and describe each class as a set of lattice points.

Solution. The class $[(3,5)]$. A pair (x,y) lies in it when $x+5=y+3$. This holds if and only if $y=x+2$. Indeed, if $y=x+2$, then $y+3=x+(2+3)=x+5$ by associativity (Theorem 1.5.3 (c)). Conversely, $x+5=(x+2)+3$, so $x+5=y+3$ gives $(x+2)+3=y+3$, and cancellation (Proposition 1.6.3 (b)) gives $y=x+2$. So

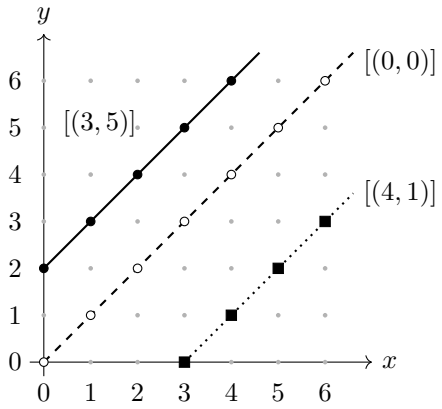
$$[(3,5)]=\{(x,x+2):x\in\mathbb{N}\},$$

and the elements with $x\le 4$ are $(0,2)$, $(1,3)$, $(2,4)$, $(3,5)$, $(4,6)$.

The class $[(0,0)]$. Here the condition is $x+0=y+0$, that is, $x=y$, since $n+0=n$ (Definition 1.5.1 (a)). So $[(0,0)]=\{(x,x):x\in\mathbb{N}\}$, the diagonal, and the elements with $x\le 4$ are $(0,0)$, $(1,1)$, $(2,2)$, $(3,3)$, $(4,4)$.

The class $[(4,1)]$. The condition $x+1=y+4$ holds if and only if $x=y+3$, by the same two steps as for $[(3,5)]$. So $[(4,1)]=\{(y+3,y):y\in\mathbb{N}\}$. Its elements with $x\le 4$ are $(3,0)$ and $(4,1)$ only: $x=y+3=3+y$ (Theorem 1.5.3 (b)) gives $3\le x$ (Proposition 1.6.3 (a)); with $x\le 4=\mathrm{S}(3)$, x is 3 or 4, since no natural number lies strictly between 3 and $\mathrm{S}(3)$ (Theorem 1.6.1 (b)); and then $y+3=3=0+3$ or $y+3=4=1+3$, so y is 0 or 1 by cancellation (Proposition 1.6.3 (b)).

Each of these three classes is the set of lattice points on one diagonal line $y=x+k$ or $x=y+k$ of the quarter plane, starting on an axis and running up and to the right.



The picture shows the three classes as three diagonals: $[(3,5)]$ solid with filled dots, $[(0,0)]$ dashed with open dots, $[(4,1)]$ dotted with squares. Every pair lies in exactly one class (Sets, Relations, and Cardinality, Theorem 2.5.1), so no two of these diagonals share a lattice point; after the normal form below we see that every class is a diagonal of this kind, so every lattice point lies on exactly one of them. The class $[(0,0)]$ is the degenerate case: it is the integer zero, and it is the only class whose line passes through the origin.

A non-example by minimal change. Replace the sums in the definition by products: on $\mathbb{N} \times \mathbb{N}$, let $(a,b) \approx (c,d)$ mean $ad = bc$. This is still reflexive ($ab = ba$) and symmetric ($ad = bc$ gives $cb = da$), both by commutativity of $\cdot$ (Theorem 1.5.3 (f)). It is **not** transitive:

$$(1,1) \approx (0,0), \qquad (0,0) \approx (2,1), \qquad \text{but} \qquad (1,1) \not\approx (2,1),$$

since $1 \cdot 0 = 1 \cdot 0$, $0 \cdot 1 = 0 \cdot 2$ and $1 \cdot 1 \neq 1 \cdot 2$. The clause that fails is (Eq3), and the line of the proof above that has no counterpart is cancellation. Additive cancellation holds for every k; multiplicative cancellation (Proposition 1.6.3 (f)) needs $k \neq 0$, and the pair $(0,0)$ supplies a middle factor that is 0. When Section 2.4 builds fractions with this product relation, it excludes zero second coordinates for exactly this reason.

2.1.5 Why this definition

Why cross sums. "$a - b = c - d$" cannot be the definition, because $a - b$ is defined in $\mathbb{N}$ only when $b \leq a$. The cross-sum equation $a + d = b + c$ says the same thing whenever both differences exist, and it makes sense for every two pairs.

Why whole classes. Each class has a simplest pair, one with a coordinate 0, as the next proposition shows. One could define $\mathbb{Z}$ as the set of these simplest pairs instead. But then the sum of $(3,0)$ and $(0,5)$ would have to be brought back to a simplest pair, $(0,2)$, by a subtraction in $\mathbb{N}$ in the right direction: the case analysis of the naive attempt returns. With classes, any representatives may be combined, and the equivalence relation absorbs the bookkeeping. The price is the check of the next section: a definition through representatives must not depend on which ones were used.

The degenerate class and the name. The class of $(0,0)$ is the diagonal; it will be the zero of $\mathbb{Z}$. The word "integer" is Latin for "whole" or "untouched", a number not broken into fractions; Section 2.4 builds those.

The first result about the new set says that each class has exactly one simplest name.

Proposition 2.1.2. Normal Form of an Integer

(a) Every class $[(a,b)] \in \mathbb{Z}$ contains **exactly one** pair of the form $(n,0)$ or $(0,n)$ with $n \in \mathbb{N}$ (the pair $(0,0)$ has both forms and counts once). Precisely, for each $(a,b) \in \mathbb{N} \times \mathbb{N}$ **exactly one** of the following holds:
- $a = b + k$ for some $k \neq 0$, and then $[(a,b)] = [(k,0)]$;
- $a = b$, and then $[(a,b)] = [(0,0)]$;
- $b = a + k$ for some $k \neq 0$, and then $[(a,b)] = [(0,k)]$.

(b) For all $m, n \in \mathbb{N}$: if $[(m,0)] = [(n,0)]$, then $m = n$; and if $[(0,m)] = [(0,n)]$, then $m = n$.

Idea. The three cases are the three cases of trichotomy for a and b, and Proposition 1.6.3 (a) turns each inequality into an equation with a non-zero k. In each case one line of arithmetic shows that (a,b) is related to the simplest pair. For "exactly one pair", two simplest pairs of the same form are handled by (b), so we prove (b) first. Two of opposite forms, $(k,0)$ and $(0,l)$ with $k, l \neq 0$, would need $k + l = 0$, which the order forbids.

Proof. (b) Suppose $[(m,0)] = [(n,0)]$. By Sets, Relations, and Cardinality, Theorem 2.5.1 (b), $(m,0) \sim (n,0)$, that is, $m + 0 = 0 + n$. Since $m + 0 = m$ (Definition 1.5.1 (a)) and $0 + n = n$ (Theorem 1.5.3 (a)), $m = n$. Likewise $[(0,m)] = [(0,n)]$ gives $0 + n = m + 0$, so $n = m$.

(a), the three cases. Let $(a,b) \in \mathbb{N} \times \mathbb{N}$. By Theorem 1.6.1 (a), exactly one of $b < a$, $a = b$, $a < b$ holds. By Proposition 1.6.3 (a), $b < a$ holds if and only if $a = b + k$ for some $k \neq 0$, and $a < b$ if and only if $b = a + k$ for some $k \neq 0$. So exactly one of the three listed conditions holds. In each case the stated equality of classes follows from Sets, Relations, and Cardinality, Theorem 2.5.1 (b), once the two pairs are shown to be related:

- if $a = b + k$, then $a + 0 = a = b + k$ (Definition 1.5.1 (a)), so $(a,b) \sim (k,0)$;
- if $a = b$, then $a + 0 = b + 0$, so $(a,b) \sim (0,0)$;
- if $b = a + k$, then $a + k = b = b + 0$ (Definition 1.5.1 (a)), so $(a,b) \sim (0,k)$.

In particular every class contains a pair of the form $(n,0)$ or $(0,n)$.

(a), only one such pair. Let C be a class containing two such pairs p and p'. Since both lie in C, $[p] = C = [p']$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). If $p = (m,0)$ and $p' = (n,0)$, then $m = n$ by (b), so $p = p'$; likewise if both have the form $(0,\cdot)$. This covers every case in which one of the pairs is $(0,0)$, since $(0,0)$ has both forms. There remains $p = (k,0)$ and $p' = (0,l)$ with $k \neq 0$ and $l \neq 0$. Then $p \sim p'$ says $k + l = 0 + 0 = 0$. Since $l \neq 0$, Proposition 1.6.3 (a) gives $k < 0$; and since $k = 0 + k$ (Theorem 1.5.3 (a)) with $k \neq 0$, it also gives $0 < k$. By transitivity $k < k$, which Theorem 1.6.1 (a) rules out. So this case cannot occur, and the pair is unique.

■

So every integer has a **normal form**: it is $[(n,0)]$ for a unique $n \in \mathbb{N}$, or $[(0,n)]$ for a unique $n \in \mathbb{N}$ with $n \neq 0$. In the picture, every class is a diagonal: $(x,y) \sim (n,0)$ means $x + 0 = y + n$, that is, $x = y + n$, and $(x,y) \sim (0,n)$ means $x + n = y + 0$, that is, $y = x + n$ (Definition 1.5.1

(a)), so

$$[(n,0)] = \{(y+n, y) : y \in \mathbb{N}\} \quad \text{and} \quad [(0,n)] = \{(x, x+n) : x \in \mathbb{N}\}.$$

The normal form is where the class's diagonal meets an axis. The next section reads $[(n,0)]$ as the natural number n and $[(0,n)]$ as its negative; this is how the description of *Sets, Relations, and Cardinality*, Section 2.1, "an integer is a natural number or the negative of one", becomes a theorem about the constructed $\mathbb{Z}$.

Warning. A property of a class must not be read off **one** representative. "The integer $[(a,b)]$ is small when $a \leq 3$" sounds like a property of integers and is not one: $[(3,3)] = [(5,5)]$, the first pair has $a = 3 \leq 3$, and the second has $a = 5$. The class is "small" or not according to the name we happened to use. The property "$[(a,b)]$ is positive when $b < a$" looks the same and happens to be safe: if $(a,b) \sim (a',b')$ and $b < a$, then $b' < a'$. That needs a proof, and Section 2.3 gives it before it calls any integer positive.

Quick check. Is $(2,7) \sim (10,15)$? Is $(2,7) \sim (15,10)$?

Solution. Yes: the cross sums are $2+15 = 17$ and $7+10 = 17$. No: the cross sums are $2+10 = 12$ and $7+15 = 22$, and $12 \neq 22$. In school terms, $(2,7)$ and $(10,15)$ both record "five less", while $(15,10)$ records "five more". The normal forms agree: $7 = 2+5$ and $15 = 10+5$, so $[(2,7)] = [(0,5)] = [(10,15)]$, while $[(15,10)] = [(5,0)]$, a different class by Proposition 2.1.2 (a).

The set $\mathbb{Z}$ exists, and each element has a simplest name. It has no arithmetic yet. The next section defines the sum and product of two integers by formulas on representatives, and the warning above is the question it must answer for each formula.

2.1.6 Exercises

2.1.6.1 A. Check your understanding

Exercise. A1

(a) State the relation $\sim$ and the set $\mathbb{Z}$ of Definition 2.1.1 in symbols.

(b) Which pairs with a coordinate equal to 0 lie in $[(6,9)]$?

(c) Why does the proof of Proposition 2.1.1 cancel $c+d$ rather than subtract it?

(d) True or false, with a reason: $[(1,1)] = [(5,5)]$.

Solution.

(a) For $(a,b),(c,d) \in \mathbb{N} \times \mathbb{N}$: $(a,b) \sim (c,d) \iff a+d = b+c$; and $\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\!\sim = \{[(a,b)] : (a,b) \in \mathbb{N} \times \mathbb{N}\}$.

(b) Only $(0,3)$. A pair $(x,0)$ in the class would need $x+9 = 0+6 = 6$ (Theorem 1.5.3 (a)); but $9 \leq x+9$ by Proposition 1.6.3 (a) (with Theorem 1.5.3 (b) to write $x+9 = 9+x$), and $6 < 9$, so $x+9 = 6$ would give $9 \leq 6 < 9$, and then $9 < 9$, which Theorem 1.6.1 (a) excludes. A pair $(0,y)$ needs $0+9 = y+6$, that is, $3+6 = y+6$, so $y = 3$ by

cancellation (Proposition 1.6.3 (b)). This agrees with Proposition 2.1.2 (a): $9 = 6 + 3$, so $[(6, 9)] = [(0, 3)]$, and the class has exactly one pair with a coordinate 0.

(c) Subtraction in $\mathbb{N}$ is defined only when the number subtracted is at most the other (Proposition 1.6.3 (a)). Here $c + d$ is at most both sides, so subtracting would be possible, but only after that check, and reading off the differences uses the uniqueness in (a), which is itself proved by cancellation. Cancellation (Proposition 1.6.3 (b)) holds for every $k \in \mathbb{N}$ with no check, so the proof uses it directly.

(d) True. $(1, 1) \sim (5, 5)$ because $1 + 5 = 1 + 5$, so the classes are equal by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). Both are the zero class $[(0, 0)]$.

2.1.6.2 B. Practice

Exercise. B1: Three look-alike relations

On $\mathbb{N} \times \mathbb{N}$, determine which of the following relations are equivalence relations. For each one, decide which of (Eq1)–(Eq3) hold. Justify your answer.

(a) $(a, b) \approx_1 (c, d)$ if and only if $a + c = b + d$.

(b) $(a, b) \approx_2 (c, d)$ if and only if $a + 2d = 2b + c$.

(c) $(a, b) \approx_3 (c, d)$ if and only if $a + d \leq b + c$.

Solution.

(a) **Symmetric, but neither reflexive nor transitive.** Symmetric: $a + c = b + d$ gives $c + a = d + b$ by commutativity (Theorem 1.5.3 (b)). Not reflexive: $(0, 1) \approx_1 (0, 1)$ would need $0 + 0 = 1 + 1$, that is, $0 = 2$, which is false. Not transitive: $(0, 1) \approx_1 (1, 0)$ since $0 + 1 = 1 + 0$, and $(1, 0) \approx_1 (0, 1)$ since $1 + 0 = 0 + 1$; transitivity would give $(0, 1) \approx_1 (0, 1)$, which we just saw is false. Notice that **every** pair is related to something, since $(a, b) \approx_1 (b, a)$ by commutativity. A symmetric relation in which every element is related to something still need not be reflexive: "x related to y, so y related to x, so x related to x" uses transitivity, and that is what fails here.

(b) **An equivalence relation.** (Eq1) $a + 2b = 2b + a$. (Eq2) $a + 2d = 2b + c$ gives $c + 2b = 2d + a$, by commutativity. (Eq3) Suppose $a + 2d = 2b + c$ and $c + 2f = 2d + e$. Adding, and rearranging by Theorem 1.5.3 (b), (c),

$$(a + 2f) + (2d + c) = (2b + e) + (2d + c),$$

and cancellation (Proposition 1.6.3 (b)) gives $a + 2f = 2b + e$. In school terms the pair (a, b) records $a - 2b$, and the relation says $a - 2b = c - 2d$.

(c) **Reflexive and transitive, not symmetric.** (Eq1) $a + b \leq b + a$, since the two sides are equal. (Eq3) First, for $x, y, u, v \in \mathbb{N}$, $x \leq y$ and $u \leq v$ give $x + u \leq y + v$: by Proposition 1.6.3 (a), $y = x + i$ and $v = u + j$ for some $i, j \in \mathbb{N}$, so $y + v = (x + u) + (i + j)$ by Theorem 1.5.3 (b), (c), and (a) again gives $x + u \leq y + v$. Now let $a + d \leq b + c$ and $c + f \leq d + e$. Adding and rearranging,

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

If $b + e < a + f$, then $(b + e) + (c + d) < (a + f) + (c + d)$ by Proposition 1.6.3 (c), contradicting the display by trichotomy (Theorem 1.6.1 (a)). So $a + f \leq b + e$, by trichotomy again. Not symmetric: $(0, 1) \approx_3 (0, 0)$ since $0 + 0 \leq 1 + 0$, that is, $0 \leq 1$;

but $(0,0) \approx_3 (0,1)$ would need $0+1 \leq 0+0$, that is, $1 \leq 0$, which is false since $0 < 1$ (Theorem 1.6.1 (d)).

Exercise. B2: A class in set-builder form

Let $(a,b) \in \mathbb{N} \times \mathbb{N}$. Describe the class $[(a,b)]$ by a set-builder expression, in the two cases $b \leq a$ and $a \leq b$. Hence describe $[(7,2)]$ and $[(2,7)]$.

Solution. Case $b \leq a$. Write $a = b + k$ with $k \in \mathbb{N}$ (Proposition 1.6.3 (a)). We claim

$$[(a,b)] = \{(y+k, y) : y \in \mathbb{N}\}.$$

A pair (x,y) lies in $[(a,b)]$ when $x+b = y+a$, that is, $x+b = (y+k)+b$ by Theorem 1.5.3 (c), (b). By cancellation (Proposition 1.6.3 (b)) this holds if and only if $x = y+k$; the converse direction is substitution. So the elements of the class are exactly the pairs $(y+k, y)$, $y \in \mathbb{N}$.

Case $a \leq b$. Write $b = a + k$ (Proposition 1.6.3 (a)). In the same way, $(x,y) \sim (a,b)$ means $x + (a+k) = y + a$, which by rearranging and canceling a holds if and only if $y = x + k$. So $[(a,b)] = \{(x, x+k) : x \in \mathbb{N}\}$.

The two descriptions agree when $a = b$ (then $k = 0$), as they must. For the examples, $7 = 2 + 5$, so $[(7,2)] = \{(y+5, y) : y \in \mathbb{N}\}$, and $[(2,7)] = \{(x, x+5) : x \in \mathbb{N}\}$.

2.1.6.3 C. Going deeper

Exercise. C1: Exit check: the relation that builds $\mathbb{Z}$

Let $\sim$ be the relation of Definition 2.1.1 on $\mathbb{N} \times \mathbb{N}$.

(a) Prove that $\sim$ is an equivalence relation.

(b) Prove that if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, then $(a+c, b+d) \sim (a'+c', b'+d')$.

(c) Say in one sentence what (b) means for a rule that adds two integers by adding representatives coordinatewise.

Name the property of $\mathbb{N}$ behind each step.

Solution.

(a) This is Proposition 2.1.1; here is the proof once more. Reflexive: $a + b = b + a$ by commutativity. Symmetric: if $a + d = b + c$, then $c + b = d + a$ by commutativity. Transitive: let $a+d = b+c$ and $c+f = d+e$. Adding the two equations and rearranging by associativity and commutativity gives $(a+f) + (c+d) = (b+e) + (c+d)$, and cancellation of $c+d$ (Proposition 1.6.3 (b)) gives $a + f = b + e$.

(b) The hypotheses are $a + b' = b + a'$ and $c + d' = d + c'$. Adding them,

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

By associativity and commutativity (Theorem 1.5.3 (b), (c)), the left side is $(a+c) +$

$(b' + d')$ and the right side is $(b + d) + (a' + c')$. So

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

which is the cross-sum equation for $(a + c, b + d) \sim (a' + c', b' + d')$.

(c) The class of $(a + c, b + d)$ does not depend on which representatives of $[(a, b)]$ and $[(c, d)]$ were used, so "add representatives coordinatewise, then take the class" is **well defined**: it assigns exactly one integer to each pair of integers.

Marking scheme. (10 points)

- 2: reflexive and symmetric, each citing commutativity of $+$ in $\mathbb{N}$.
- 3: transitivity: combines the two hypotheses into one equation with a common summand on both sides (by adding them, or by substituting one into the other) and removes that summand by **cancellation** in $\mathbb{N}$, named as such.
- 2: in (b), combines the two hypotheses (by adding them, or by substituting both into a rearranged sum).
- 2: in (b), rearranges by associativity and commutativity to the cross-sum equation for the two sums.
- 1: (c) concludes that the coordinatewise sum is well defined on classes.

A proof of (b) that changes one argument at a time (first (a, b) to (a', b') with (c, d) fixed, then the other) and joins the two by transitivity earns the four points of (b) if each step is justified. Fatal errors (0 for the step): subtracting in $\mathbb{N}$ (for instance "subtract d from both sides") without showing that the difference exists; proving only reflexivity and symmetry and calling the relation an equivalence relation.

2.2 Well-Definedness, and the Arithmetic of the Integers

Section 2.1 built $\mathbb{Z}$ as the set of classes of $\sim$ on $\mathbb{N} \times \mathbb{N}$, and gave each integer a normal form (Proposition 2.1.2). The integers cannot yet be added. Every operation on classes will be defined by a formula on representatives, and *Sets, Relations, and Cardinality*, Section 2.5 showed what can go wrong: a formula that gives two answers for one class (Sets, Relations, and Cardinality, Example (A rule on classes that is not a function)). This section proves, once and for all, the lemma that says when such a formula defines a function. It then uses the lemma for the sum, product and negative of integers, and finds a copy of $\mathbb{N}$ inside $\mathbb{Z}$.

2.2.1 Definitions through representatives

In *Sets, Relations, and Cardinality*, Section 2.5 the addition of $\mathbb{Z}/n$ was defined by $[a] + [b] \coloneqq [a + b]$, and Sets, Relations, and Cardinality, Proposition 2.5.5 checked that the right side depends only on the classes, not on the integers a and b used to name them. In Sets, Relations, and Cardinality, Example (A rule on classes that is not a function) the same kind of formula, $[a]_6 \mapsto [a]_4$, gave one class two values, and so defined no function.

The same situation will recur throughout this volume: for $\mathbb{Z}$ in this section, for $\mathbb{Q}$ in Section 2.4, and for $\mathbb{R}$ later in this volume. Each time, a set is a quotient $X/\sim$, and something is defined on it by a formula on elements of X. Rather than repeat the argument of Sets, Relations, and Cardinality, Proposition 2.5.5 for every formula, we separate it into two parts: a general lemma,

proved here once, which turns a **check on representatives** into a function on classes; and the check itself, which is different for each formula and is all that remains to do each time.

A rule "$[x] \mapsto$ something computed from x" is called **well defined** when representatives of the same class always give the same result. The lemma below says that being well defined is enough, in three settings: a map from one quotient to another, a binary operation on a quotient, and a property of classes.

2.2.2 The lemma

A rule on classes is a function as soon as equivalent inputs give equivalent outputs.

Theorem 2.2.1. Functions on Classes

Let $\sim$ be an equivalence relation on a set X, and $\approx$ an equivalence relation on a set Y. Classes are taken for $\sim$ in X and for $\approx$ in Y.

(a) Let $f\colon X \to Y$ be a function such that, **for all** $x, x' \in X$, $x \sim x'$ implies $f(x) \approx f(x')$. Then there is **exactly one** function $F\colon X/\!\sim \to Y/\!\approx$ with $F([x]) = [f(x)]$ for every $x \in X$.

(b) Let $g\colon X \times X \to X$ be a function such that, **for all** $x, x', y, y' \in X$,

$$\begin{aligned} x \sim x' &\implies g(x,y) \sim g(x',y), \\ y \sim y' &\implies g(x,y) \sim g(x,y'). \end{aligned}$$

Then there is **exactly one** function $G\colon (X/\!\sim) \times (X/\!\sim) \to X/\!\sim$ with $G([x],[y]) = [g(x,y)]$ for all $x, y \in X$.

(c) Let $S \subseteq X$ be such that, for all $x, x' \in X$, $x \sim x'$ and $x \in S$ imply $x' \in S$. Let $\bar{S} \coloneqq \{[x] : x \in S\} \subseteq X/\!\sim$. Then, **for every** $x \in X$, $[x] \in \bar{S}$ if and only if $x \in S$.

In words: (a) says that a map on elements that respects the two relations passes to a map on classes; (b) says that a binary operation that respects $\sim$ **in each argument separately** passes to an operation on classes; (c) says that a property that respects $\sim$ can be asked of a class, and the answer is the answer for any of its representatives. Clause (a) will give the negative of an integer, (b) its sum and product, and (c) the positive integers in Section 2.3.

Idea. A function is a graph with **exactly one** output for each input (Sets, Relations, and Cardinality, Definition 1.4.1). The obvious graph is the set of all pairs $([x],[f(x)])$. It has at least one output for each class, because every class is some $[x]$. The only danger is two outputs for one class, coming from two representatives x and x' of it; Sets, Relations, and Cardinality, Theorem 2.5.1 (b) turns "same class" into $x \sim x'$, the hypothesis turns that into $f(x) \approx f(x')$, and Sets, Relations, and Cardinality, Theorem 2.5.1 (b) turns it back into "same output". For (b) the same plan works, once the two one-argument hypotheses are chained into one statement about both arguments by transitivity.

Proof. (a) Let

$$G_F \coloneqq \{(C,D) \in (X/\!\sim) \times (Y/\!\approx) : C = [x] \text{ and } D = [f(x)] \text{ for some } x \in X\},$$

a set by specification (Sets, Relations, and Cardinality, Definition 1.1.3). We check the two clauses of Sets, Relations, and Cardinality, Definition 1.4.1. (Fn1): every $C \in X/\sim$ is $[x]$ for some $x \in X$ (Sets, Relations, and Cardinality, Definition 2.5.4), and then $(C, [f(x)]) \in G_F$. (Fn2): suppose $(C, D) \in G_F$ and $(C, D') \in G_F$, say $C = [x]$, $D = [f(x)]$, and $C = [x']$, $D' = [f(x')]$. Then $[x] = [x']$, so $x \sim x'$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). By the hypothesis, $f(x) \approx f(x')$, so $[f(x)] = [f(x')]$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b), that is, $D = D'$. So F, the function with graph G_F and codomain $Y/\approx$, is a function $X/\sim \to Y/\approx$, and $F([x]) = [f(x)]$ because $([x], [f(x)]) \in G_F$.

For uniqueness, let $F' \colon X/\sim \to Y/\approx$ also satisfy $F'([x]) = [f(x)]$ for every x. Every $C \in X/\sim$ is some $[x]$, and then $F'(C) = [f(x)] = F(C)$. The two functions have the same domain and codomain, so $F' = F$ by Sets, Relations, and Cardinality, Proposition 1.4.1.

(b) First, for all $x, x', y, y' \in X$ with $x \sim x'$ and $y \sim y'$,

$$g(x, y) \sim g(x', y) \quad \text{and} \quad g(x', y) \sim g(x', y'),$$

by the first hypothesis (with y fixed) and the second (with x' fixed). By transitivity, (Eq3) of Sets, Relations, and Cardinality, Definition 2.5.1, $g(x, y) \sim g(x', y')$.

Now let G_G be the set of $(P, E) \in ((X/\sim) \times (X/\sim)) \times (X/\sim)$ such that $P = ([x], [y])$ and $E = [g(x, y)]$ for some $x, y \in X$, a set by specification. (Fn1): an element of $(X/\sim) \times (X/\sim)$ is a pair (C, D) of classes (Sets, Relations, and Cardinality, Definition 1.3.2), and $C = [x]$, $D = [y]$ for some $x, y \in X$ (Sets, Relations, and Cardinality, Definition 2.5.4); then $((C, D), [g(x, y)]) \in G_G$. (Fn2): suppose (P, E) and (P, E') lie in G_G, with $P = ([x], [y])$, $E = [g(x, y)]$, and $P = ([x'], [y'])$, $E' = [g(x', y')]$. From $([x], [y]) = ([x'], [y'])$ the characteristic property of pairs (Sets, Relations, and Cardinality, Definition 1.3.1) gives $[x] = [x']$ and $[y] = [y']$, so $x \sim x'$ and $y \sim y'$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). By the first paragraph, $g(x, y) \sim g(x', y')$, so $E = E'$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). So G_G, with codomain $X/\sim$, defines a function G with $G([x], [y]) = [g(x, y)]$. Uniqueness is as in (a), by Sets, Relations, and Cardinality, Proposition 1.4.1.

(c) ($\Leftarrow$) If $x \in S$, then $[x] \in \bar{S}$ by the definition of $\bar{S}$. ($\Rightarrow$) If $[x] \in \bar{S}$, then $[x] = [x']$ for some $x' \in S$. By Sets, Relations, and Cardinality, Theorem 2.5.1 (b), $x \sim x'$, so $x' \sim x$ by (Eq2). The hypothesis, applied to $x' \sim x$ and $x' \in S$, gives $x \in S$.

This proves the theorem.

■

The functions F and G are said to be **induced** by f and g. In practice we write the conclusion of (a) as "the rule $[x] \mapsto [f(x)]$ is well defined", and of (b) as "the operation $[x] * [y] \coloneqq [g(x, y)]$ is well defined", and the work is always the check of the hypotheses.

Remark. **Each hypothesis has a job.** Without the hypothesis of (c) the conclusion fails. On $\mathbb{N} \times \mathbb{N}$ with the relation $\sim$ of Definition 2.1.1, let $S \coloneqq \{(a, b) : 0 < a\}$. Then $(1, 1) \in S$, so $[(1, 1)] \in \bar{S}$; and $[(0, 0)] = [(1, 1)]$, so $[(0, 0)] \in \bar{S}$, while $(0, 0) \notin S$. Membership in $\bar{S}$ does not tell us whether a particular representative lies in S. The hypotheses of (a) and (b) are needed in the same way, and the examples below show them failing.

2.2.3 Sum, product and negative of integers

To guess the formulas, borrow the school picture once more. If (a, b) records "$a - b$" and (c, d) records "$c - d$", then school algebra gives

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

These identities are a **heuristic**: they come from the school integers, which this chapter does not assume, and they are used only to find the definition. Each right side is again a difference of two natural numbers, so it can be recorded as a pair.

Definition 2.2.1. Sum, Product and Negative of Integers

For all $(a, b), (c, d) \in \mathbb{N} \times \mathbb{N}$, let

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

The **sum** $+\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ and **product** $\cdot\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ are the functions that Theorem 2.2.1 (b), with $X = \mathbb{N} \times \mathbb{N}$ and the equivalence relation $\sim$ of Proposition 2.1.1, induces from the maps g_+, $g_\cdot\colon (\mathbb{N} \times \mathbb{N}) \times (\mathbb{N} \times \mathbb{N}) \to \mathbb{N} \times \mathbb{N}$ given by the pairs on the right; the **negative** $-\colon \mathbb{Z} \to \mathbb{Z}$ is the function that Theorem 2.2.1 (a) induces from $(a, b) \mapsto (b, a)$. We also write xy for $x \cdot y$.

In words: to add integers, add representatives coordinatewise; to multiply, use the cross formula; to negate, swap the coordinates. All the arithmetic on the right happens in $\mathbb{N}$, where it is always possible. The definition is legitimate only if the hypotheses of Theorem 2.2.1 hold for the three maps on pairs; the next proposition checks them, and until then the three formulas are rules, not yet functions.

The move used in the proof is the one this chapter is built on.

Check on representatives, one argument at a time. To show that $[x] * [y] \coloneqq [g(x, y)]$ is well defined, change x to an equivalent x' with y fixed, then change y to an equivalent y' with x' fixed. Each change is a computation at the level below, in $\mathbb{N}$ here; Theorem 2.2.1 (b) joins the two by transitivity.

Proposition 2.2.2. The Operations on Integers Are Well Defined

The maps g_+ and $g_\cdot$ of Definition 2.2.1 satisfy the hypotheses of Theorem 2.2.1 (b), and the map $(a, b) \mapsto (b, a)$ satisfies those of Theorem 2.2.1 (a), with $\sim$ on both sides. Hence the sum, product and negative of Definition 2.2.1 are well-defined functions.

Idea. Each hypothesis is an equation in $\mathbb{N}$ to be derived from another. The given equation is the cross-sum equation for $(a, b) \sim (a', b')$; the target is the cross-sum equation for the two outputs. For the sum, adding $c + d$ to both sides of the given equation and rearranging is enough. For the product, the target has c and d in every term, so we multiply the given equation by c, and separately by d, and add the results; the rest is rearranging. Both formulas

are unchanged when the two input pairs are swapped, so the second argument costs nothing.

Before reading on: why may we change one argument at a time, rather than both at once?

Proof. Let $(a,b),(a',b'),(c,d) \in \mathbb{N}\times\mathbb{N}$ with $(a,b)\sim(a',b')$, that is,

$$a+b' = b+a'. \tag{2.2.1}$$

All rearrangements below use Theorem 1.5.3: commutativity and associativity of $+$, commutativity of $\cdot$, and distributivity.

Sum, first argument. We need $(a+c,b+d)\sim(a'+c,b'+d)$, that is, $(a+c)+(b'+d) = (b+d)+(a'+c)$. The left side is $(a+b')+(c+d)$, which by Eq. 2.2.1 equals $(b+a')+(c+d)$, and that is the right side rearranged.

Product, first argument. We need

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

that is, the cross-sum equation

$$(ac+bd)+(a'd+b'c) = (ad+bc)+(a'c+b'd). \tag{2.2.2}$$

Multiply both sides of Eq. 2.2.1 by c and use distributivity: $ac+b'c = bc+a'c$. Multiply both sides of Eq. 2.2.1, read from right to left, by d: $bd+a'd = ad+b'd$. Adding these two equations,

$$(ac+b'c)+(bd+a'd) = (bc+a'c)+(ad+b'd).$$

The left side rearranges to the left side of Eq. 2.2.2, and the right side to its right side. This proves Eq. 2.2.2.

Second arguments. By commutativity in $\mathbb{N}$, $g_+((a,b),(c,d)) = (a+c,b+d) = (c+a,d+b) = g_+((c,d),(a,b))$, and likewise

$$g.((a,b),(c,d)) = (ac+bd,\ ad+bc) = (ca+db,\ cb+da) = g.((c,d),(a,b)).$$

So if $(c,d)\sim(c',d')$, then, by the first-argument case applied with the roles of the two pairs exchanged,

$$g((a,b),(c,d)) = g((c,d),(a,b)) \sim g((c',d'),(a,b)) = g((a,b),(c',d'))$$

for $g = g_+$ and $g = g.$. This is the second hypothesis of Theorem 2.2.1 (b).

Negative. The pair (b,a) is related to (b',a') when $b+a' = a+b'$, and that is Eq. 2.2.1 read from right to left. This is the hypothesis of Theorem 2.2.1 (a).

By Theorem 2.2.1 (b) and (a), the three rules define functions $\mathbb{Z}\times\mathbb{Z}\to\mathbb{Z}$, $\mathbb{Z}\times\mathbb{Z}\to\mathbb{Z}$ and $\mathbb{Z}\to\mathbb{Z}$, as claimed.

■

The answer to the prompt is the first paragraph of the proof of Theorem 2.2.1 (b): two one-argument changes, $g(x,y) \sim g(x',y)$ and $g(x',y) \sim g(x',y')$, are joined by transitivity of $\sim$. Changing both at once is also possible, but the computation then mixes four equations in eight unknowns; one argument at a time keeps each computation to a single hypothesis.

2.2.4 Rules that fail

Every check happens in the hypotheses of the lemma. Here they fail, next to a rule where they hold.

Example. The first coordinate is not a function of the class

(a) Does the rule $[(a,b)] \mapsto a$ define a function $\mathbb{Z} \to \mathbb{N}$?

(b) Does the rule $[(a,b)] \mapsto [(a+1,b)]$ define a function $\mathbb{Z} \to \mathbb{Z}$?

(c) A student argues: "Each integer x has a pair with least first coordinate, by Theorem 1.6.2 applied to the non-empty set of first coordinates of the pairs in x; call that least first coordinate $h(x)$. Then h is a function $\mathbb{Z} \to \mathbb{N}$, since each x determines its own least first coordinate. So $[(a,b)] \mapsto a$ is a function, namely h, with $h([(a,b)]) = a$ for every (a,b)." Find the error.

Solution.

(a) No. $(0,2) \sim (1,3)$, since $0+3 = 2+1$, so $[(0,2)] = [(1,3)]$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). Read with the representative $(0,2)$, the rule gives 0; read with $(1,3)$, it gives 1. One input gets two values, and (Fn2) of Sets, Relations, and Cardinality, Definition 1.4.1 fails. The hypothesis of Theorem 2.2.1 (a), with $f(a,b) \coloneqq a$ and equality on $\mathbb{N}$ as $\approx$ (so that $f(x) \approx f(x')$ means $f(x) = f(x')$), fails at $x = (0,2)$, $x' = (1,3)$.

(b) Yes. Let $(a,b) \sim (a',b')$, so $a + b' = b + a'$. Then

$$(a+1) + b' = (a+b') + 1 = (b+a') + 1 = b + (a'+1)$$

by Theorem 1.5.3 (b), (c), which is $(a+1,b) \sim (a'+1,b')$. By Theorem 2.2.1 (a), the rule is a function. In fact it is $x \mapsto x + [(1,0)]$, since $[(a,b)] + [(1,0)] = [(a+1,b+0)] = [(a+1,b)]$.

(c) The conclusion is false, by (a), so some step fails. The function h is fine: the least first coordinate is computed from the set x alone. The failing step is the last clause, "$h([(a,b)]) = a$ for every (a,b)". It holds only when (a,b) happens to be the pair with least first coordinate. The argument chose a particular representative to define h, and then used an **arbitrary** one. Counterexample to the step: $x = [(1,3)] = [(0,2)]$ contains $(0,2)$, so $h(x) = 0$, while the representative $(1,3)$ has first coordinate 1.

Warning. **A formula that looks right need not be well defined.** "$[(a,b)] \cdot [(c,d)] \coloneqq [(ac,bd)]$, like the sum" is not a definition: $(1,0) \sim (2,1)$, since $1+1 = 0+2$, but with $(c,d) = (1,2)$ the outputs are $(1 \cdot 1, 0 \cdot 2) = (1,0)$ and $(2 \cdot 1, 1 \cdot 2) = (2,2)$, and $(1,0) \not\sim (2,2)$, since $1+2 \neq 0+2$. **And well defined is necessary, not sufficient.** The rule $[(a,b)] * [(c,d)] \coloneqq [(ad+bc,\ ac+bd)]$ is well defined, since it is the negative of the product; but it gives $[(0,1)] * [(0,1)] = [(0+0, 0+1)] = [(0,1)]$, so the "square" of the class of $(0,1)$

is that class itself, not $[(1,0)]$. Well-definedness says that a rule is a function; whether it is the right function is a separate question.

2.2.5 The natural numbers inside the integers

The integers were built to contain the natural numbers and more. They do not literally contain them: the natural number 0 is the empty set (Definition 1.5.1 (c)), while every integer is a non-empty class (Sets, Relations, and Cardinality, Theorem 2.5.1 (a)), so $0 \notin \mathbb{Z}$. What $\mathbb{Z}$ contains is a **copy** of $\mathbb{N}$, the classes in normal form $[(n,0)]$, and the copy has the same arithmetic.

Proposition 2.2.3. The Natural Numbers Embed in the Integers

Let $\iota\colon \mathbb{N} \to \mathbb{Z}$, $\iota(n) \coloneqq [(n,0)]$.

(a) ι is injective, and for all $m, n \in \mathbb{N}$,

$$\iota(m+n) = \iota(m) + \iota(n), \qquad \iota(mn) = \iota(m)\,\iota(n).$$

(b) For every $x \in \mathbb{Z}$, $x + \iota(0) = x$ and $\iota(1) \cdot x = x$. For every $(a,b) \in \mathbb{N} \times \mathbb{N}$, $[(a,b)] = \iota(a) + (-\iota(b))$.

(c) $\mathbb{Z} = \iota(\mathbb{N}) \cup \{-\iota(n) : n \in \mathbb{N}\}$.

Idea. The identities in (a) and (b) are computations with the formulas of Definition 2.2.1 in which at least one operand is some $\iota(n) = [(n,0)]$, and its zero second coordinate makes most terms vanish. Injectivity and (c) are the two halves of Proposition 2.1.2: (b) of it says that different natural numbers give different classes, and (a) says that every class is in normal form.

Proof. *(a)* If $\iota(m) = \iota(n)$, then $[(m,0)] = [(n,0)]$, so $m = n$ by Proposition 2.1.2 (b); hence ι is injective (Sets, Relations, and Cardinality, Definition 1.4.3). By Definition 2.2.1,

$$\iota(m) + \iota(n) = [(m+n,\ 0+0)] = [(m+n,0)] = \iota(m+n),$$

since $0+0=0$ (Definition 1.5.1 (a)), and

$$\iota(m)\,\iota(n) = [(mn + 0\cdot 0,\ m\cdot 0 + 0\cdot n)] = [(mn,0)] = \iota(mn),$$

since $m \cdot 0 = 0$ (Definition 1.5.1 (b)), $0 \cdot n = 0$ (Theorem 1.5.3 (e)) and $k + 0 = k$.

(b) Let $x = [(a,b)]$. Then $x + \iota(0) = [(a+0, b+0)] = [(a,b)] = x$, and

$$\iota(1) \cdot x = [(1\cdot a + 0\cdot b,\ 1\cdot b + 0\cdot a)] = [(a,b)] = x,$$

by Theorem 1.5.3 (e) ($1\cdot n = n$, $0\cdot n = 0$) and $n + 0 = n$. Finally, $-\iota(b) = -[(b,0)] = [(0,b)]$, so $\iota(a) + (-\iota(b)) = [(a+0,\ 0+b)] = [(a,b)]$, using $0 + b = b$ (Theorem 1.5.3 (a)).

(c) ($\supseteq$) Both sets on the right consist of integers. ($\subseteq$) Let $x \in \mathbb{Z}$. By Proposition 2.1.2 (a), $x = [(k,0)]$ or $x = [(0,k)]$ for some $k \in \mathbb{N}$. In the first case $x = \iota(k)$. In the second, $x = [(0,k)] = -[(k,0)] = -\iota(k)$ by Definition 2.2.1. So x lies in the right side.

■

Identification. From now on we write n for $\iota(n)$ and regard $\mathbb{N}$ as a subset of $\mathbb{Z}$. This is a convenient abuse of notation, and (a) makes it honest: because ι is injective, different natural numbers stay different; because ι preserves $+$ and $\cdot$, a sum or product of natural numbers is the same whether computed in $\mathbb{N}$ or in $\mathbb{Z}$. One could rebuild $\mathbb{Z}$ so that it contains $\mathbb{N}$ literally, by putting each n in place of $\iota(n)$, but nothing later needs it. The order of $\mathbb{N}$ is matched to an order on $\mathbb{Z}$ in Section 2.3.

With this convention, (b) reads $[(a,b)] = a + (-b)$: the pair (a,b) records the difference it was meant to record. The notation $a - b$ is defined in Section 2.3, where $x - y := x + (-y)$ is set up in every ring at once. Clause (c) reads

$$\mathbb{Z} = \mathbb{N} \cup \{-n : n \in \mathbb{N}\},$$

which is word for word the definition of the integers in *Sets, Relations, and Cardinality*, Section 2.1. There it was a definition inside the familiar real numbers; here it is a theorem about a set built from ω. The positive integers keep the definition of *Sets, Relations, and Cardinality*, Section 2.1,

$$\mathbb{Z}_{>0} := \{n \in \mathbb{N} : n \neq 0\},$$

which mentions only $\mathbb{N}$, and so is now a subset of $\mathbb{Z}$. In particular $5 + x = 3$ has the solution $x = [(0,2)] = -2$ in $\mathbb{Z}$, since $[(5,0)] + [(0,2)] = [(5,2)] = [(3,0)]$ by Proposition 2.1.2 (a).

Quick check. Compute $[(2,5)] \cdot [(4,1)]$ from Definition 2.2.1, and write the two factors and the product as n or $-n$ with $n \in \mathbb{N}$.

Solution. $[(2,5)] \cdot [(4,1)] = [(2 \cdot 4 + 5 \cdot 1,\ 2 \cdot 1 + 5 \cdot 4)] = [(13,22)]$. By Proposition 2.1.2 (a): $5 = 2 + 3$, so $[(2,5)] = [(0,3)] = -3$; $4 = 1 + 3$, so $[(4,1)] = [(3,0)] = 3$; and $22 = 13 + 9$, so $[(13,22)] = [(0,9)] = -9$. So the product of -3 and 3 is -9, as school predicts; here it is a computation in $\mathbb{N}$, not a rule of signs.

The integers now have a sum, a product and a negative, each a function on classes, and they contain a copy of $\mathbb{N}$ with its arithmetic. Section 2.3 checks that these operations obey the usual laws, by reducing each law to a law of $\mathbb{N}$, and defines the order.

2.2.6 Exercises

2.2.6.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Proposition 2.2.2.

(a) Which hypothesis of Theorem 2.2.1 (b) does the computation leading to Eq. 2.2.2 verify?

(b) Run that computation with $(a,b) = (1,3)$, $(a',b') = (4,6)$ and $(c,d) = (2,0)$: check that $(a,b) \sim (a',b')$, compute both outputs, and check that they are equivalent.

(c) Where in the proof are the laws of arithmetic in $\mathbb{N}$ used?

(d) Give the idea of the proof in two sentences.

Solution.

(a) The first: for all x, x', y, $x \sim x'$ implies $g.(x, y) \sim g.(x', y)$, with $x = (a, b)$, $x' = (a', b')$, $y = (c, d)$.

(b) $1 + 6 = 7 = 3 + 4$, so $(1, 3) \sim (4, 6)$. The outputs are

$$(ac + bd,\ ad + bc) = (2 + 0,\ 0 + 6) = (2, 6),$$
$$(a'c + b'd,\ a'd + b'c) = (8 + 0,\ 0 + 12) = (8, 12),$$

and $2 + 12 = 14 = 6 + 8$, so they are equivalent. The two intermediate equations are $ac + b'c = bc + a'c$, here $2 + 12 = 6 + 8$, and $bd + a'd = ad + b'd$, here $0 = 0$, since $d = 0$. Both outputs are representatives of -4, and indeed $(-2) \cdot 2 = -4$.

(c) In every rearrangement: commutativity and associativity of $+$ (to regroup the four-term sums), distributivity (to multiply Eq. 2.2.1 by c and by d), and commutativity of $\cdot$ and $+$ (in the second-argument step, to see that both formulas are symmetric in the two input pairs).

(d) A change of representative in one argument is an equation in $\mathbb{N}$, and multiplying it by c and by d and adding gives exactly the cross-sum equation for the two products. The second argument follows by symmetry of the formula, and transitivity (inside Theorem 2.2.1 (b)) combines the two changes.

2.2.6.2 B. Practice

Exercise. B1: Two rules that pass the check

Prove that each rule defines a function $\mathbb{Z} \to \mathbb{Z}$, and identify it in terms of the operations of Definition 2.2.1.

(a) $[(a, b)] \mapsto [(b, a)]$.

(b) $[(a, b)] \mapsto [(a + b,\ a + b)]$.

Solution.

(a) Let $(a, b) \sim (a', b')$, that is, $a + b' = b + a'$. The target $(b, a) \sim (b', a')$ is $b + a' = a + b'$, the same equation read from right to left. By Theorem 2.2.1 (a), the rule is a function. It is the negative $x \mapsto -x$ of Definition 2.2.1.

(b) Every output is the zero class: $(a + b, a + b) \sim (0, 0)$, since $(a + b) + 0 = (a + b) + 0$. So for $(a, b) \sim (a', b')$ the outputs $(a+b, a+b)$ and $(a'+b', a'+b')$ are both related to $(0, 0)$, hence to each other by (Eq2) and (Eq3) of Proposition 2.1.1, and Theorem 2.2.1 (a) applies. The function is the constant map $x \mapsto 0$, where $0 = \iota(0) = [(0, 0)]$. In terms of Definition 2.2.1 it is $x \mapsto x + (-x)$, since $[(a, b)] + [(b, a)] = [(a+b,\ b+a)] = [(a+b,\ a+b)]$ by Theorem 1.5.3 (b).

Exercise. B2: Determine which

Determine which of the following rules define a function. Justify your answer.

(a) $\mathbb{Z} \to \mathbb{N}$, $[(a,b)] \mapsto a+b$.

(b) $\mathbb{Z} \to \mathbb{Z}$, $[(a,b)] \mapsto [(2a,2b)]$.

(c) $\mathbb{Z} \to \mathbb{Z}$, $[(a,b)] \mapsto [(a^2,b^2)]$, where $n^2 := n \cdot n$.

Solution.

(a) No. $[(0,0)] = [(1,1)]$, since $0+1 = 0+1$, but the rule gives $0+0 = 0$ for the first representative and $1+1 = 2$ for the second.

(b) Yes. If $a+b' = b+a'$, then by distributivity (Theorem 1.5.3 (d))

$$2a+2b' = 2(a+b') = 2(b+a') = 2b+2a',$$

which is $(2a,2b) \sim (2a',2b')$; Theorem 2.2.1 (a) applies. The function is $x \mapsto x+x$: $[(a,b)]+[(a,b)] = [(a+a,b+b)]$, and $2n = (1+1)n = n+n$ by $1+1=2$ (Definition 1.5.1 (c)) and Theorem 1.5.3 (d), (e).

(c) No. $[(0,1)] = [(1,2)]$, since $0+2 = 1+1$. The rule gives $[(0,1)]$ for the first representative and $[(1,4)]$ for the second, and $(0,1) \not\sim (1,4)$, since $0+4 = 4 \neq 2 = 1+1$.

2.2.6.3 C. Going deeper

Exercise. C1: Properties of classes, and an operation that half works

(a) Prove Theorem 2.2.1 (c), and give an example, different from the one in the remark after the theorem, of a subset $S \subseteq \mathbb{N} \times \mathbb{N}$ for which its conclusion fails for $\sim$ of Definition 2.1.1.

(b) Give an example of an equivalence relation $\sim$ on $\mathbb{N}$ and a function $g\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that the first hypothesis of Theorem 2.2.1 (b) holds and the second fails. Check each property, and show that no function G as in the conclusion of (b) exists for your example.

Solution.

(a) *Proof.* Let $x \in X$. If $x \in S$, then $[x] \in \bar{S}$ by the definition of $\bar{S}$. Conversely, if $[x] \in \bar{S}$, there is $x' \in S$ with $[x] = [x']$; then $x \sim x'$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)), so $x' \sim x$ by symmetry, and the hypothesis on S, applied to $x' \sim x$ and $x' \in S$, gives $x \in S$.
Example. Let $S := \{(a,b) \in \mathbb{N} \times \mathbb{N} : a = 0\}$. Then $(0,0) \in S$, so $[(0,0)] \in \bar{S}$; and $[(1,1)] = [(0,0)]$, so $[(1,1)] \in \bar{S}$, but $(1,1) \notin S$. The hypothesis fails at $(0,0) \sim (1,1)$. Any example must contain a pair of S and an equivalent pair outside S, since that is exactly the negation of the hypothesis.

(b) *Example.* On $\mathbb{N}$, let $x \sim y$ mean "$x = 0$ if and only if $y = 0$". This is an equivalence relation: it says that x and y have the same value under the function sending 0 to 0 and every other number to 1, so it is an equivalence relation by Sets, Relations, and Cardinality, Example (The fibers of a function). Its two classes are $\{0\}$ and $\mathbb{N} \setminus \{0\}$.

Let

$$g(x,y) := \begin{cases} x & \text{if } y = 1, \\ 0 & \text{if } y \neq 1. \end{cases}$$

First hypothesis holds. Let $x \sim x'$ and $y \in \mathbb{N}$. If $y = 1$, then $g(x,y) = x \sim x' = g(x',y)$. If $y \neq 1$, both values are 0, and $0 \sim 0$.

Second hypothesis fails. $1 \sim 2$, since neither is 0; but $g(5,1) = 5$ and $g(5,2) = 0$, and $5 \not\sim 0$.

No induced G. If $G([x],[y]) = [g(x,y)]$ held for all x, y, then, since $[1] = [2]$, we would get $[5] = [g(5,1)] = G([5],[1]) = G([5],[2]) = [g(5,2)] = [0]$, and $[5] \neq [0]$ because $5 \not\sim 0$.

What any example must have. (i) $\sim$ is not equality: some class has two elements $y \sim y'$ with $y \neq y'$, else the second hypothesis holds trivially. (ii) $\sim$ has at least two classes, else every two outputs are related and both hypotheses hold. (iii) For some x and some $y \sim y'$, the values $g(x,y)$ and $g(x,y')$ lie in different classes, while for every y, changing x within its class never moves $g(x,y)$ out of its class. A different answer can be checked against (i)–(iii).

Exercise. C2: One at a time or both at once

Let $\sim$ be an equivalence relation on X and $g\colon X \times X \to X$ a function. Prove that the hypotheses of Theorem 2.2.1 (b) hold if and only if, for all $x, x', y, y' \in X$,

$$x \sim x' \text{ and } y \sim y' \implies g(x,y) \sim g(x',y').$$

Name the property of $\sim$ that each direction uses.

Solution. ($\Rightarrow$) Let $x \sim x'$ and $y \sim y'$. The first hypothesis, with y fixed, gives $g(x,y) \sim g(x',y)$; the second, with x' fixed, gives $g(x',y) \sim g(x',y')$. By **transitivity** (Eq3), $g(x,y) \sim g(x',y')$. This direction is the first paragraph of the proof of Theorem 2.2.1 (b).

($\Leftarrow$) Let $x \sim x'$ and $y \in X$. By **reflexivity** (Eq1), $y \sim y$, so the displayed condition with $y' = y$ gives $g(x,y) \sim g(x',y)$: the first hypothesis. Likewise $x \sim x$ and $y \sim y'$ give $g(x,y) \sim g(x,y')$: the second.

So the two forms are interchangeable. The one-at-a-time form is the one to check, because each check involves a single hypothesis $x \sim x'$; the both-at-once form is the one the proof of the lemma uses.

2.3 The Integers as an Ordered Ring

Sections 2.1 and 2.2 built $\mathbb{Z}$ as a set of classes with well-defined operations $+$, $\cdot$ and $-$, but we do not yet know which laws they obey. This section collects the laws of school algebra into one definition, a **commutative ring**, and checks them for $\mathbb{Z}$ by reducing each to a law of $\mathbb{N}$. Then it orders $\mathbb{Z}$ and derives the rules for inequalities from axioms alone: from here on, an inequality about integers is proved, never read off a picture.

2.3.1 The laws of a ring

Expanding brackets or moving a term across an equation uses a short list of laws: sums and products may be reordered and regrouped, 0 and 1 do nothing, every element has a negative, and products distribute over sums. The list holds in $\mathbb{Z}$, it will hold in $\mathbb{Q}$, and a result proved from the list alone holds wherever the list does. It deserves a name.

A commutative ring is a set where you may add, subtract and multiply by the usual rules; a field is one where you may also divide by anything non-zero.

Definition 2.3.1. Commutative Ring, Field

A **commutative ring** is a set R with two functions, **addition** $+\colon R \times R \to R$ and **multiplication** $\cdot\colon R \times R \to R$ (written $a \cdot b = ab$), such that clauses (F1)–(F7) and (F9) of the following list hold; clause (F8) is used only for fields.

(F1) $a + b = b + a$ **for all** $a, b \in R$.

(F2) $(a + b) + c = a + (b + c)$ **for all** $a, b, c \in R$.

(F3) **There is** an element $0 \in R$ with $a + 0 = a$ **for every** $a \in R$.

(F4) **For every** $a \in R$ there is an element $-a \in R$ with $a + (-a) = 0$.

(F5) $(ab)c = a(bc)$ **for all** $a, b, c \in R$.

(F6) $ab = ba$ **for all** $a, b \in R$.

(F7) **There is** an element $1 \in R$ with $a \cdot 1 = a$ **for every** $a \in R$.

(F8) **For every non-zero** $a \in R$ there is an element $a^{-1} \in R$ with $aa^{-1} = 1$.

(F9) $a(b + c) = ab + ac$ **for all** $a, b, c \in R$.

A **field** is a commutative ring in which, **in addition**, $1 \neq 0$ and (F8) holds. We write $a - b := a + (-b)$, and in a field $a/b := ab^{-1}$ for $b \neq 0$.

In words: (F1)–(F4) say that addition is commutative and associative, with a neutral element 0 and negatives; (F5)–(F7) say the first three things about multiplication; (F9) links the two and lets us expand brackets. (F8) asks for inverses of **non-zero** elements only, and only in a field. The notations $-a$ and a^{-1} anticipate that these elements are unique, which we prove below. These are the axioms (F1)–(F9) of the linear-algebra companion, Chapter 0, and of the ordered fields of *Analysis on the Real Line*, Chapter 1, in the same numbering, except that there (F7) includes $1 \neq 0$.

Examples.

- $\mathbb{Z}$, with the operations of Definition 2.2.1: the next theorem.
- **The zero ring** $R := \{0\}$, with $0 + 0 := 0$, $0 \cdot 0 := 0$ and $1 := 0$. Every axiom other than (F8) is an equation in R, and every such equation reads $0 = 0$. So R is a commutative ring, with $1 = 0$, and not a field. This degenerate ring is why several results below assume $1 \neq 0$.

A non-example by minimal change. The set $\mathbb{N}$, with its addition and multiplication from Chapter 1, satisfies (F1)–(F3), (F5)–(F7) and (F9), by Theorem 1.5.3 and Definition 1.5.1 (a). Only (F4) fails, at $a = 1$: $1 + x = x + 1 \neq 0$ for every $x \in \mathbb{N}$, by Theorem 1.5.3 (b) and Corollary 1.5.2. That one missing clause is why $\mathbb{Z}$ had to be built.

Why these clauses. No axiom says that $ab = 0$ forces $a = 0$ or $b = 0$: that is true in $\mathbb{Z}$ but fails in other commutative rings (the warning below), so it must be proved where it holds. The

condition $1 \neq 0$ is kept out of the ring axioms, so that the zero ring is a ring, and put into the field axioms, so that $\{0\}$, where (F8) holds vacuously, is not a field.

Warning. A commutative ring may have **zero divisors**: non-zero a, b with $ab = 0$. The integers modulo 6 of *Sets, Relations, and Cardinality*, Section 2.5 (built there from the school integers, with well-defined operations by Sets, Relations, and Cardinality, Proposition 2.5.5) form a commutative ring: each of (F1)–(F7) and (F9) passes from the school integers to classes through representatives, a check we do not need and do not carry out here. In this ring $[2] \neq [0]$ and $[3] \neq [0]$, yet $[2][3] = [6] = [0]$. So "$ab = 0$, hence $a = 0$ or $b = 0$" is not a consequence of (F1)–(F9). In an **ordered** ring it is a theorem, proved later in this section from the order axioms.

2.3.2 The integers form a commutative ring

To prove a law for classes, write each side through representatives (Definition 2.2.1) and compare the two pairs of natural numbers: if they are equal, or related by $\sim$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)), the classes are equal. We call this move **reduce to the level below**: a statement about classes becomes an identity in $\mathbb{N}$, which Chapter 1 has proved.

Theorem 2.3.1. The Integers Form a Commutative Ring

$\mathbb{Z}$, with the addition, multiplication and negation of Definition 2.2.1, is a commutative ring, with zero $[(0,0)]$ and one $[(1,0)]$, and in it $1 \neq 0$.

Idea. The operations are functions by Proposition 2.2.2, so only the eight laws remain. On representatives each side of a law is the class of an explicit pair, and we compare the pairs in $\mathbb{N}$. Only associativity of multiplication is laborious; it is written out below the table.

Proof. By Proposition 2.2.2, $+$ and $\cdot$ are functions $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. Let $x, y, z \in \mathbb{Z}$; by Sets, Relations, and Cardinality, Definition 2.5.4, $x = [(a,b)]$, $y = [(c,d)]$, $z = [(e,f)]$ with $a, \ldots, f \in \mathbb{N}$. By Definition 2.2.1, each side of each axiom is the class of the pair shown in the table. "Equal" means the same pair of natural numbers, by Theorem 1.5.3 and Definition 1.5.1 (a), (b). "Related" means related by $\sim$ (Definition 2.1.1: $(p,q) \sim (r,s)$ means $p+s = q+r$), so the classes agree by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

Axiom	left side, as a pair	right side, as a pair	why the classes agree
(F1) $x + y = y + x$	$(a + c, b + d)$	$(c + a, d + b)$	equal: $+$ is commutative on $\mathbb{N}$
(F2) $(x + y) + z = x + (y + z)$	$((a+c)+e, (b+d)+f)$	$(a+(c+e), b+(d+f))$	equal: $+$ is associative on $\mathbb{N}$
(F3) $x + [(0,0)] = x$	$(a + 0, b + 0)$	(a, b)	equal: $m + 0 = m$
(F4) $x + (-x) = [(0,0)]$	$(a + b, b + a)$	$(0, 0)$	related: $(a+b)+0 = (b+a)+0$
(F5) $(xy)z = x(yz)$	see below	see below	equal, after expanding (below)

(F6) $xy = yx$	$(ac+bd, ad+bc)$	$(ca+db, cb+da)$	equal: $\cdot$ and $+$ are commutative
(F7) $x \cdot [(1,0)] = x$	$(a\cdot 1+b\cdot 0, a\cdot 0+b\cdot 1)$	(a,b)	equal: $n\cdot 1 = n$, $n\cdot 0 = 0$, $n+0 = 0+n = n$
(F9) $x(y+z) = xy+xz$	$(a(c+e)+b(d+f), a(d+f)+b(c+e))$	$((ac+bd)+(ae+bf), (ad+bc)+(af+be))$	equal: distribute, then reorder the four terms

For (F5), write $xy = [(g,h)]$ with $g := ac+bd$, $h := ad+bc$. Then $(xy)z = [(ge+hf, gf+he)]$, and in $\mathbb{N}$

$$ge + hf = ace + bde + adf + bcf,$$
$$gf + he = acf + bdf + ade + bce.$$

On the other side, $yz = [(ce+df, cf+de)]$, and $x(yz)$ is the class of the pair whose coordinates are

$$a(ce+df) + b(cf+de) = ace + adf + bcf + bde,$$
$$a(cf+de) + b(ce+df) = acf + ade + bce + bdf.$$

Both first coordinates are sums of ace, adf, bcf, bde, and both second coordinates of acf, ade, bce, bdf, so the pairs are equal (Theorem 1.5.3 (b)–(g)).

Finally, $1 \neq 0$: if $[(1,0)] = [(0,0)]$, then $1 = 0$ by Proposition 2.1.2 (b). But $0 < 1$ in $\mathbb{N}$ (Theorem 1.6.1 (d)), and $<$ is irreflexive (Theorem 1.6.1 (a)), so $1 \neq 0$. This proves the theorem.

■

Seven of the eight rows end in "equal". Only (F4) needs $\sim$: the pair $(a+b, b+a)$ records "$(a+b)-(b+a)$", which is zero, without performing the subtraction, and that is the whole point of passing to classes.

Remark. The zero and one of this ring are $[(0,0)] = \iota(0)$ and $[(1,0)] = \iota(1)$ (Proposition 2.2.3), so after the identification of Section 2.2 they are the natural numbers 0 and 1. With $a-b := a+(-b)$, the formula $[(a,b)] = \iota(a) + (-\iota(b))$ of Proposition 2.2.3 reads $[(a,b)] = a-b$ for $a, b \in \mathbb{N}$: the subtraction the pair recorded is now performed in $\mathbb{Z}$.

What the definition does not list must be derived, and the derivations work in every commutative ring at once.

Proposition 2.3.2. Algebra in a Commutative Ring

Let R be a commutative ring, and let $a, b \in R$.

(a) The elements 0 and 1 of (F3) and (F7) are unique, the element $-a$ of (F4) is unique, and $-(-a) = a$.

(b) $a \cdot 0 = 0$.

(c) $(-a)b = a(-b) = -(ab)$ and $(-a)(-b) = ab$.

(d) $(a+b)^2 = a^2 + 2ab + b^2$, where $x^2 := xx$ and $2 := 1+1$.

Idea. To show that something equals 0, or $-(ab)$, show that it satisfies the **defining equation** of that element and let uniqueness, part (a), finish.

Proof.

(a) Suppose 0 and $0'$ both satisfy (F3). Then $0' = 0' + 0 = 0 + 0' = 0$, by (F3) for 0, (F1), and (F3) for $0'$. Likewise, if 1 and $1'$ both satisfy (F7), then $1' = 1' \cdot 1 = 1 \cdot 1' = 1$, by (F7) and (F6). Next, suppose $a + b = 0$ and $a + c = 0$. Then

$$\begin{aligned} b &= b + 0 = b + (a + c) = (b + a) + c \\ &= (a + b) + c = 0 + c = c + 0 = c, \end{aligned}$$

by (F3), the second hypothesis, (F2), (F1), the first hypothesis, (F1) and (F3). So $-a$ is the **only** element whose sum with a is 0. Finally, $(-a) + a = a + (-a) = 0$ by (F1) and (F4), so a satisfies the defining equation of $-(-a)$, and uniqueness gives $-(-a) = a$.

(b) Let $z \coloneqq a \cdot 0$. By (F3) and (F9), $z = a(0 + 0) = a \cdot 0 + a \cdot 0 = z + z$. Adding $-z$ to both sides,

$$0 = z + (-z) = (z + z) + (-z) = z + \big(z + (-z)\big) = z + 0 = z,$$

by (F4), the equation $z = z + z$, (F2), (F4) and (F3). So $a \cdot 0 = 0$.

(c) By (F6) and (F9), $ab + (-a)b = b(a + (-a)) = b \cdot 0 = 0$, the last step by (b). So $(-a)b$ satisfies the defining equation of $-(ab)$, and (a) gives $(-a)b = -(ab)$. By (F6) also $a(-b) = (-b)a = -(ba) = -(ab)$. Hence $(-a)(-b) = -\big(a(-b)\big) = -\big(-(ab)\big) = ab$, using the first identity, the second, and (a).

(d) By (F9) and (F6), $(a + b)(a + b) = (a + b)a + (a + b)b = a^2 + ba + ab + b^2$, and $ba + ab = ab \cdot 1 + ab \cdot 1 = ab(1 + 1) = 2ab$ by (F6), (F7) and (F9). Regrouping by (F1) and (F2) gives $(a + b)^2 = a^2 + 2ab + b^2$.

■

The argument of (a) with multiplication shows that in a **field** a non-zero a has only one inverse: if $ab = 1$ and $ac = 1$, then $b = b \cdot 1 = b(ac) = (ba)c = (ab)c = 1 \cdot c = c$, by (F7), (F5) and (F6). So a^{-1}, and with it a/b, names a single element. And (b) explains the zero ring: if $1 = 0$ in a commutative ring, then $a = a \cdot 1 = a \cdot 0 = 0$ for every a, so the ring is $\{0\}$.

2.3.3 Ordered rings

The integers carry an order tied to the operations: adding the same number to both sides keeps an inequality, and a product of positives is positive. We want every rule for inequalities to follow from a short list of axioms, as for equations. The list is the order axioms (O1)–(O4) of *Analysis on the Real Line*, Chapter 1, with "field" relaxed to "commutative ring".

An ordered ring is a ring with a strict order that adding respects, in which a product of positives is positive.

Definition 2.3.2. Ordered Ring, Ordered Field

An **ordered commutative ring** is a commutative ring R (Definition 2.3.1) together with a relation $<$ on R such that, **for all** $a, b, c \in R$:

(O1) (trichotomy) **exactly one** of $a < b$, $a = b$, $b < a$ holds;

(O2) (transitivity) if $a < b$ and $b < c$, then $a < c$;

(O3) (addition) if $a < b$, then $a + c < b + c$;

(O4) (multiplication) if $0 < a$ and $0 < b$, then $0 < ab$.

An **ordered field** is a field that is an ordered commutative ring. We write $b > a$ for $a < b$, and $a \leq b$ for "$a < b$ or $a = b$" (likewise $b \geq a$). An element a is **positive** if $a > 0$, **negative** if $a < 0$, and **non-negative** if $a \geq 0$.

In words: any two elements compare in exactly one way, so $a < a$ never holds; comparisons chain; we may add the same element to both sides; and (O4), the only clause about multiplication, speaks only of **positive** elements.

The degenerate case: the zero ring $\{0\}$ with the **empty** relation is an ordered commutative ring. Only the middle option of (O1) holds, and (O2)–(O4) are vacuous. There $1 = 0$, so any rule concluding $1 > 0$ must assume $1 \neq 0$.

Checking (O1)–(O4) directly means handling pairs of elements. It is easier to say which elements are positive and let $a < b$ mean that $b - a$ is positive. The next proposition says which "positive sets" give an order, and that every order arises so. We call the move **order from a positive set**.

Proposition 2.3.3. Order from a Positive Set

Let R be a commutative ring and $P \subseteq R$ a subset such that:

(P1) if $a, b \in P$, then $a + b \in P$ and $ab \in P$;

(P2) **for every** $a \in R$, **exactly one** of $a \in P$, $a = 0$, $-a \in P$ holds.

Define $a < b$ to mean $b - a \in P$. Then $(R, <)$ is an ordered commutative ring (Definition 2.3.2), and $P = \{a \in R : a > 0\}$.

Conversely, if $(R, <)$ is an ordered commutative ring, then $P := \{a \in R : a > 0\}$ satisfies (P1) and (P2), and for all $a, b \in R$, $a < b$ holds if and only if $b - a \in P$.

The labels (P1), (P2) are local; they are not the clauses of a partition (*Sets, Relations, and Cardinality*, Section 2.5) or Peano's axioms (Chapter 1).

Idea. Each order axiom is a statement about one difference. (O1) is (P2) applied to $b - a$, since $-(b - a) = a - b$. (O2) is closure under sums, since $(c - b) + (b - a) = c - a$. (O3) holds since $(b + c) - (a + c) = b - a$. (O4) is closure under products. The algebra is uniqueness of negatives.

Proof. We first record three identities in R, for $a, b, c \in R$:

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

For the first, $(b-a)+(a-b) = (b+(-b))+((-a)+a) = 0$ by (F1), (F2), (F4), and Proposition 2.3.2 (a) gives the identity. The second is regrouping by (F1)–(F4). For the third, $(a+c)+((-a)+(-c)) = 0$ likewise, so $-(a+c) = (-a)+(-c)$ by Proposition 2.3.2 (a), and the c's cancel. Also $-0 = 0$ (as $0+0=0$), so $a-0 = a$; and $b-a = 0$ if and only if $b = a$: one direction is (F4), and $b-a = 0$ gives $b = (b-a)+a = 0+a = a$.

Now let P satisfy (P1), (P2), and let $a < b$ mean $b-a \in P$.

(O1) Apply (P2) to $b-a$: exactly one of $b-a \in P$, $b-a = 0$, $-(b-a) \in P$ holds. By the facts above, these say $a < b$, $a = b$ and $a-b \in P$, that is, $b < a$. So exactly one of the three holds.

(O2) If $b-a \in P$ and $c-b \in P$, then $c-a = (c-b)+(b-a) \in P$ by (P1).

(O3) If $b-a \in P$, then $(b+c)-(a+c) = b-a \in P$.

(O4) Since $a-0 = a$, we have $0 < a$ if and only if $a \in P$. So $0 < a$ and $0 < b$ give $a, b \in P$, hence $ab \in P$ by (P1), that is, $0 < ab$. The same equivalence says $P = \{a \in R : a > 0\}$.

Conversely, let $(R, <)$ be an ordered commutative ring and $P := \{a : a > 0\}$. If $a < b$, then $a+(-a) < b+(-a)$ by (O3), that is, $0 < b-a$. If $0 < b-a$, then $0+a < (b-a)+a = b$ by (O3), that is, $a < b$. This is the last claim. For (P1), let $a, b > 0$. Then $0+b < a+b$ by (O3), so $0 < b < a+b$ and $a+b > 0$ by (O2); and $ab > 0$ by (O4). For (P2), let $a \in R$. By (O1) exactly one of $0 < a$, $a = 0$, $a < 0$ holds, and by the last claim $a < 0$ holds exactly when $0-a = -a$ is in P. This proves the proposition. ∎

So an order on a commutative ring is the same thing as a positive set. For $\mathbb{Z}$, the natural one is the classes $[(a,b)]$ with $b < a$. Section 2.1 warned that "$b < a$" is a statement about **one** representative; we now check that it is a statement about the class.

2.3.4 The order of the integers

Definition 2.3.3. Order on the Integers

Let $S := \{(a,b) \in \mathbb{N} \times \mathbb{N} : b < a\}$, and

$$P_{\mathbb{Z}} := \{[(a,b)] : (a,b) \in S\} \subseteq \mathbb{Z}.$$

For $x, y \in \mathbb{Z}$, define $x < y$ to mean $y - x \in P_{\mathbb{Z}}$.

Well-definedness. By Theorem 2.2.1 (c), with $X = \mathbb{N} \times \mathbb{N}$, it suffices that $(a,b) \sim (a',b')$ and $b < a$ imply $b' < a'$. Write $a = b+k$ with $k \neq 0$ (Proposition 1.6.3 (a)). Then $a+b' = b+a'$ reads $(b'+k)+b = a'+b$ (Theorem 1.5.3 (b), (c)), and canceling b (Proposition 1.6.3 (b)) gives $a' = b'+k$, so $b' < a'$. Hence, for **every** pair,

$$[(a,b)] \in P_{\mathbb{Z}} \iff b < a.$$

So "positive" may be tested on any representative, as Section 2.1's warning required.

Examples. $[(5,2)] \in P_{\mathbb{Z}}$, since $2 < 5$; this is the integer 3. $[(2,5)] \notin P_{\mathbb{Z}}$ and $[(4,4)] \notin P_{\mathbb{Z}}$, since neither $5 < 2$ nor $4 < 4$. To compare $x = [(3,5)]$ and $y = [(4,1)]$: $y - x = [(4,1)] + [(5,3)] = [(9,4)]$, and $4 < 9$, so $x < y$, as "$-2 < 3$" leads us to expect. The degenerate pair $(0,0)$ is not in S, so $0 \notin P_{\mathbb{Z}}$: zero is not positive.

Quick check. Is $[(7,10)] < [(2,4)]$?

Solution. Compute $[(2,4)] - [(7,10)] = [(2,4)] + [(10,7)] = [(12,11)]$ by Definition 2.2.1. Since $11 < 12$, the difference lies in $P_{\mathbb{Z}}$, so yes: $[(7,10)] < [(2,4)]$, which is $-3 < -2$.

Recall from Section 2.2 that $\mathbb{Z}_{>0} \coloneqq \{n \in \mathbb{N} : n \neq 0\}$, a set of natural numbers, inside $\mathbb{Z}$ through the identification $n = \iota(n)$. The next theorem shows that it is exactly the set of positive integers.

Theorem 2.3.4. The Integers Form an Ordered Ring

$\mathbb{Z}$, with the operations of Definition 2.2.1 and the relation $<$ of Definition 2.3.3, is an ordered commutative ring, and its positive elements are exactly $P_{\mathbb{Z}}$. Moreover:

(a) for $m, n \in \mathbb{N}$, $\iota(m) < \iota(n)$ if and only if $m < n$;

(b) $\mathbb{N} = \{x \in \mathbb{Z} : x \geq 0\}$, and $\mathbb{Z}_{>0} = \{x \in \mathbb{Z} : x > 0\}$;

(c) (**discreteness**) there is no $k \in \mathbb{Z}$ with $0 < k < 1$; hence, for $m, n \in \mathbb{Z}$ with $m < n$, $m + 1 \leq n$.

Idea. By Proposition 2.3.3 it is enough to check (P1) and (P2) for $P_{\mathbb{Z}}$, each a fact about $\mathbb{N}$. For products, write $a = b + k$, $c = d + l$: the product's two coordinates differ by exactly $kl \neq 0$. (P2) is trichotomy in $\mathbb{N}$. Parts (a)–(c) follow from the positive classes being the $[(k,0)]$ with $k \neq 0$.

Proof. By Theorem 2.3.1, $\mathbb{Z}$ is a commutative ring. We check (P1) and (P2) of Proposition 2.3.3 for $P = P_{\mathbb{Z}}$, using the criterion $[(a,b)] \in P_{\mathbb{Z}} \iff b < a$.

(P1), sums. Let $x = [(a,b)]$ and $y = [(c,d)]$ with $b < a$ and $d < c$. Then $x + y = [(a+c, b+d)]$, and

$$b + d < a + d < a + c$$

by Proposition 1.6.3 (c) twice (with commutativity), so $b + d < a + c$ by transitivity (Theorem 1.6.1 (a)). Hence $x + y \in P_{\mathbb{Z}}$.

(P1), products. With the same x, y, write $a = b + k$ and $c = d + l$ with $k, l \neq 0$ (Proposition 1.6.3 (a)). Then $xy = [(ac + bd, ad + bc)]$, and expanding by Theorem 1.5.3,

$$ac + bd = (b + k)(d + l) + bd = bd + bl + kd + kl + bd,$$
$$ad + bc = (b + k)d + b(d + l) = bd + kd + bd + bl.$$

So $ac + bd = (ad + bc) + kl$. Since $k \neq 0$ and $l \neq 0$, $kl \neq 0$ by Proposition 1.6.3 (d), and

therefore $ad + bc < ac + bd$ by Proposition 1.6.3 (a). Hence $xy \in P_{\mathbb{Z}}$.

(P2). Let $x = [(a, b)]$. Then $x \in P_{\mathbb{Z}}$ if and only if $b < a$. Next, $x = [(0, 0)]$ if and only if $(a, b) \sim (0, 0)$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)), that is, $a+0 = b+0$, that is, $a = b$. Finally, $-x = [(b, a)] \in P_{\mathbb{Z}}$ if and only if $a < b$. By trichotomy in $\mathbb{N}$ (Theorem 1.6.1 (a)), exactly one of $b < a$, $a = b$, $a < b$ holds. So exactly one of $x \in P_{\mathbb{Z}}$, $x = 0$, $-x \in P_{\mathbb{Z}}$ holds.

By Proposition 2.3.3, $(\mathbb{Z}, <)$ is an ordered commutative ring with positive set $P_{\mathbb{Z}}$.

(a) By Definition 2.2.1, $\iota(n) - \iota(m) = [(n, 0)] + [(0, m)] = [(n, m)]$, which lies in $P_{\mathbb{Z}}$ if and only if $m < n$.

(b) If $x = [(a, b)] > 0$, then $a = b + k$ with $k \neq 0$ (Proposition 1.6.3 (a)), and $x = [(k, 0)] = \iota(k)$ by Proposition 2.1.2 (a). Conversely, for $k \neq 0$, $\iota(k) = [(k, 0)]$ is positive since $0 < k$ (Theorem 1.6.1 (d)). So the positive integers are the $\iota(k)$ with $k \in \mathbb{Z}_{>0}$; adding $0 = \iota(0)$ gives the first equation.

(c) If $0 < k$, then $k = \iota(m)$ with $m \neq 0$ by (b), so $1 \leq m$ (Theorem 1.6.1 (d)) and $1 \leq k$ by (a); then $k < 1$ would give $1 < 1$ by (O2), or contradict (O1) directly. Now let $m < n$. Then $n - m > 0$ (Proposition 2.3.3), so $n - m \geq 1$, and adding m by (O3) gives $n \geq 1 + m$. This proves the theorem.

■

With the identification of Section 2.2, (a) says that the order of $\mathbb{Z}$ **extends** the order of $\mathbb{N}$.

Two non-examples by minimal change, each breaking exactly one order axiom:

- Replace $<$ by $\leq$. (O2)–(O4) still hold, by the axioms for $<$ and the cases of equality (for (O4), $ab = 0$ when a factor is 0, by Proposition 2.3.2 (b)). But (O1) fails at $a = b = 0$: both "$0 \leq 0$" and "$0 = 0$" hold. The axioms describe a **strict** order.
- Reverse the order: $a \prec b$ means $b < a$. (O1)–(O3) hold, by the same axioms for $<$. But (O4) fails: $0 \prec -1$, since $0 - (-1) = 1 = [(1, 0)] \in P_{\mathbb{Z}}$ (Proposition 2.3.2 (a)), and $(-1)(-1) = 1$ (Proposition 2.3.2 (c)), yet $0 \prec 1$ is false, since $0 - 1 = [(0, 1)] \notin P_{\mathbb{Z}}$. So (O4) fixes which way round the order goes.

2.3.5 Rules for inequalities

The rules below hold in every ordered commutative ring: in $\mathbb{Z}$ now, and in $\mathbb{Q}$ once Section 2.5 orders it. One of them needs multiples first.

Multiples. Let R be a commutative ring and $a \in R$. By Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b), with $X = R$, starting value 0 and step $(n, y) \mapsto y + a$, there is exactly one function $\mathbb{N} \to R$, written $n \mapsto n \cdot a$, with

$$0 \cdot a = 0, \qquad (n + 1) \cdot a = n \cdot a + a \quad \text{for every } n \in \mathbb{N}.$$

So $1 \cdot a = a$ and $2 \cdot a = a + a$. From here to the end of the section, a dot with a natural number on its left means this multiple, and the product of two elements of R is written by juxtaposition, ab. When $R = \mathbb{Z}$, the two readings agree, by Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)): $0 \cdot a = 0 = 0a$ by Proposition 2.3.2 (b) and (F6), and $n \cdot a = na$ gives $(n + 1) \cdot a = na + a = (n + 1)a$, since $(n + 1)a = a(n + 1) = an + a = na + a$ by (F6), then (F9) with (F7), then (F6) (the sum $n + 1$ is the same in $\mathbb{N}$ and $\mathbb{Z}$ by Proposition 2.2.3 (a)).

Proposition 2.3.5. Rules for Inequalities in an Ordered Ring

Let R be an ordered commutative ring, and let $a, b, c, d \in R$.

(a) $a < b$ if and only if $b - a > 0$, if and only if $-b < -a$.

(b) If $a \leq b$ and $c \leq d$, then $a + c \leq b + d$, and the conclusion is strict if one of the hypotheses is strict.

(c) If $a < b$ and $c > 0$, then $ac < bc$.

(d) If $a < b$ and $c < 0$, then $ac > bc$.

(e) If $a \neq 0$, then $a^2 > 0$. If $1 \neq 0$, then $1 > 0$.

(f) If $ab = 0$, then $a = 0$ or $b = 0$.

(g) If $a \geq 0$ and $b \geq 0$, then $a < b$ if and only if $a^2 < b^2$.

(h) For every $m, k \in \mathbb{N}$, $(m + k) \cdot a = m \cdot a + k \cdot a$. If $1 \neq 0$, then $n \cdot 1 > 0$ for every $n \in \mathbb{N}$ with $n \geq 1$, and $m \cdot 1 < n \cdot 1$ whenever $m, n \in \mathbb{N}$ and $m < n$.

(i) $\leq$ is a total order on R (Sets, Relations, and Cardinality, Definition 2.6.1), and its strict part is $<$.

Idea. The method is **turn the comparison into one positive element**: by Proposition 2.3.3, $a < b$ says that $b - a$ is positive, and positive elements are closed under $+$ and $\cdot$. For instance (c) says: if $b - a$ and c are positive, so is $bc - ac = (b - a)c$. We prove (i) first, since (b) and (g) use it.

Proof.

(i) Reflexivity is $a = a$. Antisymmetry: $a \leq b$, $b \leq a$ and $a \neq b$ give $a < b < a$, so $a < a$ by (O2), against (O1). Transitivity: if $a \leq b \leq c$, either an equality holds and we substitute, or $a < c$ by (O2); and $a < c$ whenever one step is strict. Totality is (O1). Strict part: $a < b$ implies $a \neq b$ by (O1), and $a \leq b$, $a \neq b$ give $a < b$.

(a) The first equivalence is the last claim of Proposition 2.3.3. For the second, $(-a)-(-b) = (-a) + b = b - a$ by Proposition 2.3.2 (a) and (F1), so $-b < -a$ also says $b - a > 0$.

(b) By (O3) (or equality) and (F1), $a + c \leq b + c \leq b + d$; conclude by (i).

(c) By (a), $b - a > 0$. By (O4), $(b - a)c > 0$. By (F9), (F6) and Proposition 2.3.2 (c), $(b - a)c = bc + (-a)c = bc - ac$. So $bc - ac > 0$, and $ac < bc$ by (a).

(d) By (a), $c < 0$ gives $0 - c = -c > 0$. By (c) applied with $-c$, $a(-c) < b(-c)$, that is, $-(ac) < -(bc)$ by Proposition 2.3.2 (c), and then $bc < ac$ by (a) together with $-(-x) = x$.

(e) If $a > 0$, then $a^2 = aa > 0$ by (O4). If $a < 0$, then $-a > 0$ as in (d), and $a^2 = (-a)(-a) > 0$ by Proposition 2.3.2 (c) and (O4). By (O1) one of the two cases occurs when $a \neq 0$. If $1 \neq 0$, then $1 = 1^2 > 0$.

(f) Contrapositive: let $a, b \neq 0$. By (O1) and (a), some $a' \in \{a, -a\}$ and $b' \in \{b, -b\}$ are positive, so $a'b' > 0$ by (O4). By Proposition 2.3.2 (c), $a'b'$ is ab or $-(ab)$; if $ab = 0$, both are 0, against (O1). So $ab \neq 0$.

(g) Let $a, b \geq 0$. ($\Rightarrow$) If $a < b$, then $b > 0$ by (i), and (c) gives $ab < b^2$. Also $a^2 \leq ab$: by (c) if $a > 0$, and by Proposition 2.3.2 (b) if $a = 0$. So $a^2 < b^2$ by (i). ($\Leftarrow$) If $a^2 < b^2$ but not $a < b$, then $b = a$ or $b < a$ by (O1), giving $a^2 = b^2$ or, by ($\Rightarrow$), $b^2 < a^2$; both

contradict (O1).

(h) Fix $m \in \mathbb{N}$. We prove the sum law by Logic and Proof, Axiom 4.2.1 on k (through Corollary 1.7.3 (b)), for $P(k)$: "$(m+k)\cdot a = m\cdot a + k\cdot a$". The base case is $(m+0)\cdot a = m\cdot a + 0\cdot a$, by Definition 1.5.1 (a), (F3) and $0\cdot a = 0$. For the step, $P(k)$ gives $(m+(k+1))\cdot a = (m+k)\cdot a + a = (m\cdot a + k\cdot a) + a = m\cdot a + (k+1)\cdot a$, by associativity in $\mathbb{N}$ (Theorem 1.5.3 (c)) with the definition of multiples, $P(k)$, and (F2) with the definition of multiples again.

Now let $1 \neq 0$. By Logic and Proof, Axiom 4.2.1 on n, for $Q(n)$: "$(n+1)\cdot 1 > 0$": the base is $1\cdot 1 = 0+1 = 1 > 0$ by (e), and the step is $(n+2)\cdot 1 = (n+1)\cdot 1 + 1 > 0$ by (b) and (e). Every $n \geq 1$ in $\mathbb{N}$ is $j+1$ for some j (Proposition 1.6.3 (a)), so $n\cdot 1 > 0$. Finally, if $m < n$, then $n = m+k$ with $k \neq 0$ (Proposition 1.6.3 (a)), so $k \geq 1$ (Theorem 1.6.1 (d)), and by the sum law and (b),

$$n\cdot 1 = m\cdot 1 + k\cdot 1 > m\cdot 1 + 0 = m\cdot 1.$$

This proves the proposition.

■

The school rules for inequalities between integers are now theorems about $\mathbb{Z}$. The hypothesis $1 \neq 0$ in (e) and (h) cannot be dropped: in the zero ring with the empty order, neither $1 > 0$ nor $1\cdot 1 > 0$ holds.

Warning. (O4) and (c) are about multiplying by something **positive**. The tempting rule "if $a < b$ and $c < d$, then $ac < bd$" is false in $\mathbb{Z}$: $-3 < -1$ and $-3 < -1$, yet $(-3)(-3) = 9$ and $(-1)(-1) = 1$, and $9 > 1$. What is true is the version with $0 \leq a$ and $0 \leq c$: then $b > 0$ by (i), so $ac \leq bc < bd$, the first step by (c) if $c > 0$ and by Proposition 2.3.2 (b) if $c = 0$, the second by (c) and (F6); conclude by (i).

Quick check. Is $\mathbb{Z}$ a field?

Solution. No: $1+1$ is non-zero (it is $> 1 > 0$) and has no inverse. Suppose $(1+1)x = 1$. Since $(1+1)x = x(1+1) = x+x$ by (F6), (F9) and (F7), this says $x+x = 1$. If $x \leq 0$, then $x+x \leq 0 < 1$ by Proposition 2.3.5 (b), (e). If $x > 0$, then $x \geq 1$ by discreteness (Theorem 2.3.4 (c)), so $x+x \geq 1+1 > 1$ by Proposition 2.3.5 (b), (e). Both contradict $x+x = 1$ by (O1).

2.3.6 Least elements, and a ring that cannot be ordered

The natural numbers are well ordered (Theorem 1.6.2); the integers are not, since $x-1 < x$ for every x. But a set of integers bounded below can be shifted into $\mathbb{N}$.

Proposition 2.3.6. Least and Greatest Elements of Sets of Integers

Let $S \subseteq \mathbb{Z}$ be **non-empty**. If S has a lower bound $b \in \mathbb{Z}$ (that is, $b \leq s$ for every $s \in S$), then S has a least element. If S has an upper bound $c \in \mathbb{Z}$, then S has a greatest element.

Proof. Let b be a lower bound and $T := \{s - b : s \in S\}$, a non-empty set. For $s \in S$, $s \geq b$ gives $s - b \geq 0$ by (O3), so $T \subseteq \{x \in \mathbb{Z} : x \geq 0\} = \mathbb{N}$ by Theorem 2.3.4 (b). By Theorem 1.6.2, T has an element m with $m \leq t$ for every $t \in T$, in the order of $\mathbb{N}$ and so in that of $\mathbb{Z}$ (Theorem 2.3.4 (a)). Then $m = s_0 - b$ for some $s_0 \in S$, so $m + b = s_0 \in S$; and for $s \in S$, $s - b \geq m$ gives $s \geq m + b$ by (O3). So $m + b$ is least in S.

If c is an upper bound, then $-S := \{-s : s \in S\}$ has the lower bound $-c$ (Proposition 2.3.5 (a)), so by the first part it has a least element, which is $-s_1$ for some $s_1 \in S$, by the definition of $-S$. For $s \in S$, $-s_1 \leq -s$ gives $s \leq s_1$, by Proposition 2.3.5 (a), or by $-(-s) = s$ (Proposition 2.3.2 (a)) when $-s_1 = -s$. So s_1 is greatest in S. This proves the proposition.

■

Both hypotheses are needed: $\mathbb{Z}$ has no least element, and neither has $\emptyset$, which every integer bounds. The rules of Proposition 2.3.5 also show that some rings carry no order at all.

Example. A field with two elements that cannot be ordered

Let $R := \{0, 1\}$, two distinct elements, with addition and multiplication given by the tables

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Check that R satisfies (F1)–(F9), so that it is a field, and show that no relation $<$ on R makes it an ordered commutative ring.

Solution. The axioms. (F1), (F6): both tables are symmetric. (F3), (F7): the row of 0 in the first table and of 1 in the second repeat the headings. (F4): $-0 = 0$ and $-1 = 1$, since $1 + 1 = 0$. (F8): 1 has inverse 1, since the second table gives the product of 1 and 1 as 1. And $1 \neq 0$.

The three laws with three variables can be checked over the eight triples (a, b, c), or by cases. From the tables, $0x = 0$ and $1x = x$, and $x + y = 1$ exactly when exactly one of x, y is 1. (F2): both sides are 1 exactly when an odd number of a, b, c (one or three) are 1. (F5): both sides are 1 exactly when $a = b = c = 1$. (F9): if $a = 0$, both sides are 0; if $a = 1$, both sides are $b + c$. So R is a field.

No order. If $<$ made R an ordered commutative ring, then, as $1 \neq 0$, Proposition 2.3.5 (h) would give $2 \cdot 1 > 0$. But $2 \cdot 1 = 1 \cdot 1 + 1 = (0 + 1) + 1 = 0$ by the definition of multiples, and $0 > 0$ is forbidden by (O1).

The argument uses only $1 \neq 0$ and $n \cdot 1 = 0$ for some $n \geq 1$. So **no commutative ring with $1 \neq 0$ in which $n \cdot 1 = 0$ for some $n \in \mathbb{N}$, $n \geq 1$, can be ordered**. The integers modulo $n \geq 2$ of *Sets, Relations, and Cardinality*, Section 2.5 are such rings (their ring laws taken for

granted, as in the warning after Definition 2.3.1), since $n \cdot [1] = [n] = [0]$; a later chapter of this volume rebuilds them on the constructed $\mathbb{Z}$.

2.3.7 Exercises

2.3.7.1 A. Check your understanding

Exercise. A1

Let R be an ordered commutative ring.

(a) Here is a proof that if $a < 0 < b$, then $ab < a^2 + b^2$. Name the axiom, rule or result that justifies each of the five steps. (1) $ab < 0b$. (2) $0b = 0$. (3) $a^2 > 0$ and $b^2 > 0$. (4) $a^2 + b^2 > 0 + 0 = 0$. (5) $ab < a^2 + b^2$.

(b) True or false: "R has no zero divisors." Give a reason.

(c) True or false: "$1 > 0$ in every ordered commutative ring." Give a reason.

(d) Name the move by which Theorem 2.3.1 was proved, and say which row of its table needed the relation $\sim$ rather than equality of pairs.

Solution.

(a) (a) Proposition 2.3.5 (c), with $a < 0$ and $b > 0$. (2) Proposition 2.3.2 (b) and (F6). (3) Proposition 2.3.5 (e), as $a, b \neq 0$ by (O1). (4) Proposition 2.3.5 (b), strict, and (F3). (5) (O2).

(b) True, by Proposition 2.3.5 (f); a general commutative ring may have zero divisors (the warning after Definition 2.3.1).

(c) False: in the zero ring with the empty relation, $1 = 0$. It is true when $1 \neq 0$ (Proposition 2.3.5 (e)).

(d) **Reduce to the level below**: on representatives, each axiom becomes an identity in $\mathbb{N}$. Only (F4) needed $\sim$: $(a + b, b + a)$ is related to $(0, 0)$ but is not that pair.

2.3.7.2 B. Practice

Exercise. B1: Three rows in detail

Let $x = [(a, b)]$, $y = [(c, d)]$, $z = [(e, f)]$ be integers. Verify in detail, naming the law of $\mathbb{N}$ used at each step, the rows (F9), (F4) and (F6) of the table in the proof of Theorem 2.3.1.

Solution. All operations are those of Definition 2.2.1, and "Law" refers to Theorem 1.5.3.

(F9). Since $y + z = [(c + e, d + f)]$, $x(y + z) = [(a(c + e) + b(d + f),\ a(d + f) + b(c + e))]$, whose coordinates are $ac + ae + bd + bf$ and $ad + af + bc + be$ by Law (d). And $xy + xz = [((ac + bd) + (ae + bf),\ (ad + bc) + (af + be))]$, whose coordinates are the same sums by Laws (b), (c). So both sides are the class of one pair.

(F4). $x + (-x) = [(a + b, b + a)]$, and $(a + b, b + a) \sim (0, 0)$ means $(a + b) + 0 = (b + a) + 0$ (Definition 2.1.1), true by Definition 1.5.1 (a) and Law (b). So $x + (-x) = [(0, 0)]$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

(F6). $yx = [(ca + db, cb + da)]$. By Law (f), $ca = ac$, $db = bd$, $cb = bc$, $da = ad$, and by Law (b), $bc + ad = ad + bc$; so the pair is $(ac + bd, ad + bc)$, the pair of xy.

Exercise. B2: The wrong positive set

Let $P' := \{[(a, b)] : a < b\} \subseteq \mathbb{Z}$. Show that membership in P' does not depend on the representative, that P' satisfies (P2) of Proposition 2.3.3, and that it does not satisfy (P1). Which part of (P1) fails?

Solution. Independence of representatives. If $(a, b) \sim (a', b')$ and $b = a + k$ with $k \neq 0$, then $b' + a = a + b' = b + a' = (a + k) + a' = (a' + k) + a$ (Theorem 1.5.3 (b), (c)), so $b' = a' + k$ by cancellation (Proposition 1.6.3 (b)), and $a' < b'$ by Proposition 1.6.3 (a). By Theorem 2.2.1 (c), $[(a, b)] \in P'$ if and only if $a < b$.

(P2). For $x = [(a, b)]$: $x \in P'$ iff $a < b$; $x = 0$ iff $a = b$ (as in the proof of Theorem 2.3.4); $-x \in P'$ iff $b < a$. By trichotomy in $\mathbb{N}$ (Theorem 1.6.1 (a)), exactly one holds.

(P1) fails for products. $[(0, 1)] \in P'$, but, by Definition 2.2.1 with $a = c = 0$ and $b = d = 1$ in $[(ac + bd,\ ad + bc)]$,

$$[(0, 1)][(0, 1)] = [(1, 0)],$$

and $[(1, 0)] \notin P'$, since $1 < 0$ is false. The sum half of (P1) holds, as in the proof of Theorem 2.3.4. So P', the negative integers, fails only closure under products; it would define the reversed order, which breaks exactly (O4).

2.3.7.3 C. Going deeper

Exercise. C1: Ordered rings with $1 \neq 0$ are infinite

(a) Prove the sum law of Proposition 2.3.5 (h) in full: for every ordered commutative ring R, every $a \in R$ and all $m, k \in \mathbb{N}$, $(m + k) \cdot a = m \cdot a + k \cdot a$. State the property you induct on and justify every equality.

(b) Deduce that every ordered commutative ring with $1 \neq 0$ is infinite (Sets, Relations, and Cardinality, Definition 3.1.3).

(c) The set $\{0, 1, 2\}$ with addition and multiplication "modulo 3", given by the tables below, is a commutative ring with $1 \neq 0$ (you may take this for granted). Show that it cannot be ordered.

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Solution (Hint 1). For (b): what does (h) say about $k \mapsto k \cdot 1$?

Solution (Hint 2). Suppose R has m elements; consider $k \mapsto k \cdot 1$ on $\{1, \dots, m+1\}$.

Solution.

(a) Fix R, $a \in R$, $m \in \mathbb{N}$, and use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k, for $P(k)$: "$(m+k) \cdot a = m \cdot a + k \cdot a$". *Base:* $(m+0) \cdot a = m \cdot a = m \cdot a + 0 = m \cdot a + 0 \cdot a$, by Definition 1.5.1 (a), (F3) and $0 \cdot a = 0$. *Step:* assume $P(k)$. Then

$$\begin{aligned}(m + (k+1)) \cdot a = ((m+k)+1) \cdot a &= (m+k) \cdot a + a \\ &= (m \cdot a + k \cdot a) + a = m \cdot a + (k \cdot a + a) \\ &= m \cdot a + (k+1) \cdot a,\end{aligned}$$

by associativity in $\mathbb{N}$ (Theorem 1.5.3 (c)), the recursion for multiples, $P(k)$, (F2), and the recursion again. So $P(k+1)$ holds, and $P(k)$ holds for every k.

(b) Let $1 \neq 0$ in R, and suppose R has m elements for some $m \in \mathbb{N}$ (Sets, Relations, and Cardinality, Definition 3.1.2). Let $f\colon \{1, \dots, m+1\} \to R$, $f(k) := k \cdot 1$. If $k < l$, then $k \cdot 1 < l \cdot 1$ by Proposition 2.3.5 (h), so $k \cdot 1 \neq l \cdot 1$ by (O1): f is injective. But $\{1, \dots, m+1\}$ has $m+1 > m$ elements (by its identity map), so Sets, Relations, and Cardinality, Corollary 3.1.3 (a) forbids an injection into R. This contradiction shows that R is infinite (Sets, Relations, and Cardinality, Definition 3.1.3).

(c) By the definition of multiples and the tables, $1 \cdot 1 = 0 + 1 = 1$, $2 \cdot 1 = 1 + 1 = 2$ and $3 \cdot 1 = 2 + 1 = 0$. An order would give $3 \cdot 1 > 0$ by Proposition 2.3.5 (h), since $1 \neq 0$, that is, $0 > 0$, against (O1). (Alternatively, (b) applies to this finite ring.)

Marking scheme. (10 points) - 1: (a) states the induction property with m and a fixed and k the induction variable. - 2: (a) base case, citing $m + 0 = m$, $0 \cdot a = 0$ and (F3). - 2: (a) inductive step, with associativity in $\mathbb{N}$, the recursion, the hypothesis and (F2) each named. - 2: (b) shows that $k \mapsto k \cdot 1$ is injective, citing (h) and (O1). - 2: (b) derives the contradiction from Sets, Relations, and Cardinality, Corollary 3.1.3 (a) with $m+1 > m$ elements (or any correct count). - 1: (c) computes $3 \cdot 1 = 0$ and concludes. Fatal errors (0 for the step): inducting on m and k at once without fixing one; using $1 > 0$ without $1 \neq 0$; in (b), "the multiples are all different, so there are infinitely many" with no finiteness result cited (Sets, Relations, and Cardinality, Corollary 3.1.3, Sets, Relations, and Cardinality, Proposition 3.1.6 with $\mathbb{N}$ infinite, or equivalent). A route for (b) composing the inverse of a bijection $\{1, \dots, m\} \to R$ with f earns the same 4 points. So does a route that shows $k \mapsto k \cdot 1$ is an injection $\mathbb{N} \to R$ (2 points, as above), so that its image is an infinite subset of R, and concludes with Sets, Relations, and Cardinality, Proposition 3.1.6 (2 points).

2.4 The Rationals as Classes of Fractions

Section 2.3 made $\mathbb{Z}$ an ordered commutative ring and showed, in a Quick check, that $\mathbb{Z}$ is not a field, because $2x = 1$ has no solution in it. This section repairs that defect the way Section 2.1 repaired the missing subtraction in $\mathbb{N}$. There, a pair (a, b) recorded "$a - b$" without performing it; here a pair (p, q) records "p divided by q" without performing it. We define $\mathbb{Q}$ as a set of classes of such pairs, give it addition and multiplication through representatives, check that they are well defined, and prove that $\mathbb{Q}$ is a field containing a copy of $\mathbb{Z}$.

2.4.1 Fractions as pairs

At school, $\frac{1}{2}$, $\frac{2}{4}$ and $\frac{3}{6}$ are "the same number", and the test for sameness is cross-multiplication: $\frac{p}{q} = \frac{r}{s}$ when $ps = rq$. The test is the useful part. It mentions only multiplication of integers, which we have, and not division, which we do not. So we keep the test and let a rational number **be** what the equivalent fractions have in common: the class of all of them.

A rational number is a quotient you have not computed yet: the set of all fractions that would compute to it.

Signs need a decision. The fractions $\frac{1}{-2}$ and $\frac{-1}{2}$ name the same number, and allowing both would put a sign in two places. We require the **denominator to be positive**, so that a sign can sit only in the numerator. This is the convention *Sets, Relations, and Cardinality*, Section 2.5 already used for its fractions, and Section 2.5 below will profit from it: there, the order of $\mathbb{Q}$ is read off a single inequality between integers. Recall that $\mathbb{Z}_{>0} = \{n \in \mathbb{Z} : n > 0\}$, by Theorem 2.3.4 (b).

On $\mathbb{Z} \times \mathbb{Z}_{>0}$, let $(p, q) \sim (r, s)$ mean $ps = rq$. We reuse the symbol $\sim$ of Section 2.1; which relation is meant is always clear from the pairs, natural numbers there and an integer with a positive integer here.

Proposition 2.4.1. Cross-Multiplication Is an Equivalence Relation

The relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}_{>0}$ given by

$$(p, q) \sim (r, s) \iff ps = rq$$

is an equivalence relation (Sets, Relations, and Cardinality, Definition 2.5.1).

Idea. Reflexivity and symmetry are just reflexivity and symmetry of $=$: with this form of the test, $(p, q) \sim (p, q)$ reads $pq = pq$. For transitivity we are given $ps = rq$ and $ru = ts$, and want $pu = tq$. Multiply the first equation by u and use the second to replace ru: this gives $(pu)s = (tq)s$. We cannot divide by s, but $\mathbb{Z}$ has no zero divisors, so a factor $s \neq 0$ can be canceled.

Proof. Let $(p, q), (r, s), (t, u) \in \mathbb{Z} \times \mathbb{Z}_{>0}$. All computations are in the commutative ring $\mathbb{Z}$ (Theorem 2.3.1).

(Eq1) $pq = pq$, so $(p, q) \sim (p, q)$.

(Eq2) If $ps = rq$, then $rq = ps$, so $(r, s) \sim (p, q)$.

(Eq3) Let $ps = rq$ and $ru = ts$. Then, using (F5), (F6) and the two hypotheses,

$$(pu)s = (ps)u = (rq)u = (ru)q = (ts)q = (tq)s.$$

Hence $(pu - tq)s = (pu)s - (tq)s = 0$, by (F9) and Proposition 2.3.2 (c). Since $s \in \mathbb{Z}_{>0}$, $s > 0$, so $s \neq 0$ by (O1). By Proposition 2.3.5 (f), $pu - tq = 0$, that is, $pu = tq$. So $(p, q) \sim (t, u)$. This proves the proposition.

■

The hypothesis that denominators are non-zero is used once, for the **middle** pair (r, s), and it is needed there. On $\mathbb{Z} \times \mathbb{N}$, with the same formula, $(1, 0) \sim (0, 0)$ (since $1 \cdot 0 = 0 \cdot 0$) and $(0, 0) \sim (1, 1)$ (since $0 \cdot 1 = 1 \cdot 0$), but $(1, 0) \not\sim (1, 1)$ (since $1 \cdot 1 \neq 1 \cdot 0$). The pair $(0, 0)$ is related to everything, and transitivity collapses through it.

Definition 2.4.1. The Rational Numbers

Let $\sim$ be the equivalence relation of Proposition 2.4.1 on $\mathbb{Z} \times \mathbb{Z}_{>0}$. The set of **rational numbers** is the quotient set (Sets, Relations, and Cardinality, Definition 2.5.4)

$$\mathbb{Q} \coloneqq (\mathbb{Z} \times \mathbb{Z}_{>0})/\!\sim.$$

A **rational number** is an element of $\mathbb{Q}$: a class $[(p, q)]$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$. We call p a **numerator** and q a **denominator** of the representative (p, q).

In words: a rational number is a **set of pairs**, all the pairs (r, s) with positive s that cross-multiply with (p, q). It is not a pair, and it is not a number with a sign and a bar. Numerators and denominators belong to representatives, not to the class: $(1, 2)$ and $(2, 4)$ represent the same rational number with different denominators. Nothing in the definition needs checking beyond Proposition 2.4.1, which makes the quotient set meaningful. By Sets, Relations, and Cardinality, Theorem 2.5.1, every pair lies in exactly one class, and $[(p, q)] = [(r, s)]$ holds exactly when $ps = rq$.

Examples.

- $[(1, 2)] = \{(k, 2k) : k \in \mathbb{Z}_{>0}\}$. Indeed, $(r, s) \sim (1, 2)$ means $2r = s$. Then $2r > 0$, and $r \leq 0$ would give $2r = r + r \leq 0$ (Proposition 2.3.5 (b)), so $r > 0$. Conversely each $(k, 2k)$ with $k > 0$ has $2k > 0$ and $k \cdot 2 = 2k \cdot 1$. Its elements with denominator at most 6 are $(1, 2)$, $(2, 4)$, $(3, 6)$.
- $[(-2, 1)] = \{(-2s, s) : s \in \mathbb{Z}_{>0}\}$, since $(r, s) \sim (-2, 1)$ means $r = -2s$. This is the class that will be the integer -2.
- The degenerate case: $[(0, 1)] = \{(0, s) : s \in \mathbb{Z}_{>0}\}$, since $(r, s) \sim (0, 1)$ means $r \cdot 1 = 0 \cdot s = 0$. Every fraction with numerator 0 lies in this one class, whatever its denominator; it will be the zero of $\mathbb{Q}$.
- $[(2, 3)]$ contains $(4, 6)$ and $(6, 9)$, since $2 \cdot 6 = 4 \cdot 3$ and $2 \cdot 9 = 6 \cdot 3$. It does not contain $(-4, -6)$, which is not even an element of $\mathbb{Z} \times \mathbb{Z}_{>0}$. Describing the whole class needs divisibility, which a later chapter of this volume develops on the constructed $\mathbb{Z}$.

A non-example by minimal change. Allow the denominator 0: take $\mathbb{Z} \times \mathbb{N}$ instead of $\mathbb{Z} \times \mathbb{Z}_{>0}$, with the same formula. Reflexivity and symmetry survive, by the same lines. Transitivity fails, as the triple $(1, 0)$, $(0, 0)$, $(1, 1)$ above shows, so there is no quotient set. *Sets, Relations, and Cardinality*, Section 2.5 met the same failure in an exercise; it is the reason denominators must be non-zero.

Why positive and not merely non-zero denominators? Non-zero denominators would also give an equivalence relation, and, as the last example of this section shows, essentially the same set $\mathbb{Q}$. The difference appears with the order. With positive denominators, "$[(p, q)]$ is positive" can mean "$p > 0$", one inequality in $\mathbb{Z}$; with non-zero denominators, it would have to say "p and q have the same sign", and a naive comparison of two fractions breaks, as the warning at the end of this section shows.

Warning. Do not read $[(p, q)]$ as the number p/q. There is no division in $\mathbb{Z}$, and $\mathbb{Q}$ has no multiplication yet, so "p/q" means nothing at this point. The class is a set of pairs, and any statement made with one representative, such as "the denominator of $[(1, 2)]$ is 2", must be checked against all of them (here it fails: $(2, 4)$ is also in the class). The fraction notation becomes honest once $\mathbb{Z}$ is placed inside $\mathbb{Q}$, in Proposition 2.4.4 and the identification that follows it.

Quick check. Is $(3, 4) \sim (-6, -8)$ in $\mathbb{Z} \times \mathbb{Z}_{>0}$?

Solution. The question is malformed: $-8 \notin \mathbb{Z}_{>0}$, so $(-6, -8)$ is not an element of $\mathbb{Z} \times \mathbb{Z}_{>0}$, and $\sim$ says nothing about it. In the variant with non-zero denominators (the last example of this section), the answer is yes: $3 \cdot (-8) = -24 = (-6) \cdot 4$.

This construction pays a debt. *Sets, Relations, and Cardinality*, Section 2.5 described $\mathbb{Q}$ as the quotient of $\mathbb{Z} \times \mathbb{Z}_{>0}$ by cross-multiplication while $\mathbb{Z}$ was still the school integers; now $\mathbb{Z}$ is the ring of Section 2.3, and the quotient is the definition of $\mathbb{Q}$.

2.4.2 Adding and multiplying classes

The school rules $\frac{p}{q} + \frac{r}{s} = \frac{ps+rq}{qs}$ and $\frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs}$ are a **heuristic** for the definition, nothing more; the definition below does not depend on them, and the proofs never use them.

Definition 2.4.2. Operations on the Rationals

For $[(p, q)], [(r, s)] \in \mathbb{Q}$, define

$$\begin{aligned} [(p, q)] + [(r, s)] &\coloneqq [(ps + rq,\ qs)], \\ [(p, q)] \cdot [(r, s)] &\coloneqq [(pr,\ qs)], \\ -[(p, q)] &\coloneqq [(-p,\ q)]. \end{aligned}$$

The right-hand sides are classes of elements of $\mathbb{Z} \times \mathbb{Z}_{>0}$: since $q > 0$ and $s > 0$, $qs > 0$ by (O4), so $qs \in \mathbb{Z}_{>0}$ by Theorem 2.3.4 (b). Each formula is computed from **chosen representatives**, so the definition is legitimate only once we know that other representatives give the same class. That is the next proposition.

Proposition 2.4.2. The Operations on the Rationals Are Well Defined

Let $X \coloneqq \mathbb{Z} \times \mathbb{Z}_{>0}$, with $\sim$ as in Proposition 2.4.1. The maps

$$\begin{aligned} g_+((p, q), (r, s)) &\coloneqq (ps + rq,\ qs), \\ g_\cdot((p, q), (r, s)) &\coloneqq (pr,\ qs) \end{aligned}$$

from $X \times X$ to X satisfy the hypotheses of Theorem 2.2.1 (b), and $f(p, q) \coloneqq (-p, q)$ satisfies the hypothesis of Theorem 2.2.1 (a). Hence $+$, $\cdot$ and $-$ of Definition 2.4.2 are functions $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ and $\mathbb{Q} \to \mathbb{Q}$.

Idea. This is the move of Section 2.2, **check on representatives, one argument at a time**. Change the first pair to an equivalent one with the second fixed, then the second with the first fixed; Theorem 2.2.1 (b) joins the two changes by transitivity. Each change is one identity in $\mathbb{Z}$: expand both cross-products and use the hypothesis $pq' = p'q$ (or $rs' = r's$) once.

Proof. All computations are in the commutative ring $\mathbb{Z}$ (Theorem 2.3.1), and we reorder products by (F5), (F6) without comment.

Multiplication, first argument. Let $(p, q) \sim (p', q')$, so $pq' = p'q$, and let $(r, s) \in X$. We need $(pr, qs) \sim (p'r, q's)$, that is, $(pr)(q's) = (p'r)(qs)$. Indeed

$$(pr)(q's) = (pq')(rs) = (p'q)(rs) = (p'r)(qs).$$

Multiplication, second argument. Let $(r, s) \sim (r', s')$, so $rs' = r's$, and let $(p, q) \in X$. We need $(pr)(qs') = (pr')(qs)$. Indeed

$$(pr)(qs') = (pq)(rs') = (pq)(r's) = (pr')(qs).$$

Addition, first argument. Let $pq' = p'q$. We need $(ps + rq, qs) \sim (p's + rq', q's)$, that is, $(ps + rq)(q's) = (p's + rq')(qs)$. By (F9),

$$(ps + rq)(q's) = (pq')s^2 + rqq's,$$
$$(p's + rq')(qs) = (p'q)s^2 + rqq's,$$

and the right-hand sides are equal because $pq' = p'q$.

Addition, second argument. Let $rs' = r's$. We need $(ps + rq, qs) \sim (ps' + r'q, qs')$, that is, $(ps + rq)(qs') = (ps' + r'q)(qs)$. By (F9),

$$(ps + rq)(qs') = pqss' + (rs')q^2,$$
$$(ps' + r'q)(qs) = pqss' + (r's)q^2,$$

and the right-hand sides are equal because $rs' = r's$.

Negation. If $pq' = p'q$, then $(-p)q' = -(pq') = -(p'q) = (-p')q$ by Proposition 2.3.2 (c), so $(-p, q) \sim (-p', q')$.

So g_+ and $g_\cdot$ satisfy the hypotheses of Theorem 2.2.1 (b): for a change of both arguments, $g(x, y) \sim g(x', y) \sim g(x', y')$, and transitivity joins the two steps, which is what the proof of that lemma does. By its conclusion, there is exactly one function $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ with $[x] + [y] = [g_+(x, y)]$, and likewise for $\cdot$. By Theorem 2.2.1 (a), $-[(p, q)] \coloneqq [(-p, q)]$ is a function $\mathbb{Q} \to \mathbb{Q}$. This proves the proposition.

■

Not every formula in representatives passes this test, and the failures are instructive.

Example. Formulas that ignore the class

Decide which of the following rules define a function on $\mathbb{Q}$.

(a) $[(p, q)] \mapsto p + q \in \mathbb{Z}$.
(b) $[(p, q)] \mapsto p \in \mathbb{Z}$.
(c) $[(p, q)] \mapsto [(p^2, q^2)] \in \mathbb{Q}$.

Solution.

(a) Not a function. $[(1, 2)] = [(2, 4)]$, since $1 \cdot 4 = 2 \cdot 2$, yet the rule gives $1 + 2 = 3$ for one representative and $2 + 4 = 6$ for the other. A function has one value at each argument (Sets, Relations, and Cardinality, Definition 1.4.1), and this rule offers two at the same class.

(b) Not a function, for the same pair: the outputs are 1 and 2.

(c) A function. First, $q^2 > 0$ by Proposition 2.3.5 (e), as $q \neq 0$, so $(p^2, q^2) \in \mathbb{Z} \times \mathbb{Z}_{>0}$. If $pq' = p'q$, then

$$p^2(q')^2 = (pq')^2 = (p'q)^2 = (p')^2q^2,$$

by (F5), (F6), so $(p^2, q^2) \sim (p'^2, q'^2)$, and Theorem 2.2.1 (a) applies. The function is $x \mapsto x \cdot x$, since $[(p, q)] \cdot [(p, q)] = [(p^2, q^2)]$ by Definition 2.4.2.

The tempting argument for (a) runs: "a rational number x **is** p/q, so p and q are determined by x, and so is $p+q$". Its conclusion is false, and the flaw is the word "is": (p, q) is **a** representative of x, and $(2p, 2q)$ is another, just as legitimate. A rule defined through representatives is a function on classes only when every representative gives the same value, and that has to be proved, as in (c).

2.4.3 The rationals form a field

The ring axioms for $\mathbb{Q}$ are checked by the move of Section 2.3, **reduce to the level below**: each side of an axiom, written on representatives, is the class of an explicit pair in $\mathbb{Z} \times \mathbb{Z}_{>0}$, and the two pairs are compared by an identity in $\mathbb{Z}$. One small observation recurs: for every $(p, q) \in \mathbb{Z} \times \mathbb{Z}_{>0}$ and $k \in \mathbb{Z}_{>0}$,

$$(p, q) \sim (kp, kq), \quad \text{since } p(kq) = (kp)q;$$

here $kq > 0$ by (O4), so $(kp, kq) \in \mathbb{Z} \times \mathbb{Z}_{>0}$ by Theorem 2.3.4 (b), and the relation makes sense. Scaling both entries by the same positive integer does not change the class.

Theorem 2.4.3. The Rationals Form a Field

$\mathbb{Q}$, with the operations of Definition 2.4.2, is a field (Definition 2.3.1), with zero $[(0, 1)]$ and one $[(1, 1)]$. A rational number $[(p, q)]$ is zero if and only if $p = 0$. If $p > 0$, its inverse is $[(q, p)]$; if $p < 0$, its inverse is $[(-q, -p)]$.

Idea. Eight laws are identities between pairs of integers, checked in a table as for $\mathbb{Z}$. As there, (F4) gives pairs that are related rather than equal; here (F9) does too, and the scaling observation is what relates them. The genuinely new axiom is (F8): inverting $[(p, q)]$ should

swap numerator and denominator, and the sign of p decides whether the swapped pair is allowed.

Proof. By Proposition 2.4.2, the operations are functions. Let $x, y, z \in \mathbb{Q}$; by Sets, Relations, and Cardinality, Definition 2.5.4, $x = [(p, q)]$, $y = [(r, s)]$, $z = [(t, u)]$ with $p, r, t \in \mathbb{Z}$ and $q, s, u \in \mathbb{Z}_{>0}$. As in Section 2.3, "equal" means that the two pairs are the same pair of integers, by the laws (F1)–(F9) of $\mathbb{Z}$ (Theorem 2.3.1); "related" means that they are related by $\sim$, so that the classes agree by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

Axiom	left side, as a pair	right side, as a pair	why the classes agree
(F1)	$(ps + rq, qs)$	$(rq + ps, sq)$	equal
(F2)	$((ps+rq)u+tqs,\ qsu)$	$(psu + (ru + ts)q,\ qsu)$	equal: both first entries are $psu + qru + qst$
(F3) $x + [(0,1)] = x$	$(p \cdot 1 + 0 \cdot q,\ q \cdot 1)$	(p, q)	equal, by Proposition 2.3.2 (b)
(F4) $x + (-x) = [(0,1)]$	$(pq + (-p)q,\ q^2) = (0, q^2)$	$(0, 1)$	related: $0 \cdot 1 = 0 \cdot q^2$
(F5)	$((pr)t,\ (qs)u)$	$(p(rt),\ q(su))$	equal
(F6)	(pr, qs)	(rp, sq)	equal
(F7) $x \cdot [(1,1)] = x$	$(p \cdot 1,\ q \cdot 1)$	(p, q)	equal
(F9)	$(p(ru + ts),\ q(su))$	$((pr)(qu) + (pt)(qs),\ (qs)(qu))$	related: the right pair is the left pair scaled by q

In (F4), $pq + (-p)q = pq - pq = 0$ by Proposition 2.3.2 (c), and $0 \cdot 1 = 0 = 0 \cdot q^2$ by Proposition 2.3.2 (b). In (F9), the right side is $xy + xz = [(pr, qs)] + [(pt, qu)]$; its first entry is $q(pru + pts) = q \cdot p(ru + ts)$ and its second is $q \cdot q(su)$, so the scaling observation above, with $k = q$, relates it to the left pair.

The zero. $[(p, q)] = [(0, 1)]$ if and only if $p \cdot 1 = 0 \cdot q$, that is, $p = 0$.

$1 \neq 0$. $[(1, 1)] = [(0, 1)]$ would need $1 = 0$ in $\mathbb{Z}$, which Theorem 2.3.1 excludes.

(F8). Let $x = [(p, q)] \neq 0$, so $p \neq 0$. If $p > 0$, then $p \in \mathbb{Z}_{>0}$, so $(q, p) \in \mathbb{Z} \times \mathbb{Z}_{>0}$, and

$$[(p, q)] \cdot [(q, p)] = [(pq,\ qp)] = [(1, 1)],$$

since $(pq) \cdot 1 = 1 \cdot (qp)$. If $p < 0$, then $-p > 0$ by Proposition 2.3.5 (a), so $(-q, -p) \in \mathbb{Z} \times \mathbb{Z}_{>0}$, and

$$\begin{aligned}[(p, q)] \cdot [(-q, -p)] &= [(p(-q),\ q(-p))] \\ &= [(-(pq),\ -(pq))] = [(1, 1)],\end{aligned}$$

by Proposition 2.3.2 (c) and $(-(pq)) \cdot 1 = 1 \cdot (-(pq))$. By (O1), one of the two cases occurs. So every non-zero element has an inverse, and $\mathbb{Q}$ is a field. This proves the theorem.

■

Six rows of the table are "equal" and two, (F4) and (F9), are "related"; the second needs the

scaling observation because the sum formula multiplies the denominators qs and qu, so their common factor q appears twice. The case split in (F8) is the price of the convention: the obvious inverse $[(q,p)]$ of a negative $[(p,q)]$ is not a class at all, since its "denominator" p is negative, and we must move the sign to the numerator.

Quick check. Compute $[(1,2)]+[(1,3)]$ and $[(-3,4)]^{-1}$, and check the second answer by multiplying.

Solution. $[(1,2)]+[(1,3)] = [(1\cdot3+1\cdot2,\ 2\cdot3)] = [(5,6)]$. For $[(-3,4)]$ the numerator is negative, so the inverse is $[(-4,3)]$. Check: $[(-3,4)]\cdot[(-4,3)] = [(12,12)]$, and $(12,12)\sim(1,1)$ since $12\cdot1 = 1\cdot12$.

2.4.4 The integers inside the rationals

As $\mathbb{N}$ sat inside $\mathbb{Z}$ through ι (Proposition 2.2.3), $\mathbb{Z}$ sits inside $\mathbb{Q}$ through the classes with denominator 1.

Proposition 2.4.4. The Integers Embed in the Rationals

Let $j\colon \mathbb{Z}\to\mathbb{Q}$, $j(m) \coloneqq [(m,1)]$. Then j is injective, and for all $m,n\in\mathbb{Z}$,

$$j(m+n) = j(m)+j(n), \qquad j(mn) = j(m)j(n), \qquad j(-m) = -j(m),$$

with $j(0)=0$ and $j(1)=1$. Moreover, writing $a/b \coloneqq ab^{-1}$ in the field $\mathbb{Q}$:

(a) for $p\in\mathbb{Z}$ and $q\in\mathbb{Z}_{>0}$, $[(p,q)] = j(p)\cdot j(q)^{-1} = j(p)/j(q)$;

(b) every rational number is $j(p)/j(q)$ for some $p\in\mathbb{Z}$ and $q\in\mathbb{Z}_{>0}$, and for $p,r\in\mathbb{Z}$ and $q,s\in\mathbb{Z}_{>0}$, $j(p)/j(q) = j(r)/j(s)$ if and only if $ps = rq$;

(c) $\mathbb{Q} = \{j(m)/j(n) : m\in\mathbb{Z},\ n\in\mathbb{Z}_{>0}\}$.

Idea. The classes with denominator 1 add and multiply exactly as their numerators do, by the formulas of Definition 2.4.2 with $q=s=1$. For (a), $[(1,q)]$ is the inverse of $[(q,1)]$, and multiplying it by $[(p,1)]$ gives $[(p,q)]$; (b) and (c) are (a) read through Sets, Relations, and Cardinality, Theorem 2.5.1.

Proof. Injective. If $j(m)=j(n)$, then $(m,1)\sim(n,1)$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b), that is, $m\cdot1 = n\cdot1$, so $m=n$.

Operations. By Definition 2.4.2, $j(m)+j(n) = [(m\cdot1+n\cdot1,\ 1\cdot1)] = [(m+n,1)] = j(m+n)$, $j(m)j(n) = [(mn,1)] = j(mn)$, and $-j(m) = [(-m,1)] = j(-m)$. Also $j(0) = [(0,1)]$ and $j(1) = [(1,1)]$ are the zero and one of $\mathbb{Q}$ (Theorem 2.4.3).

(a) Since $q>0$, $j(q) = [(q,1)] \neq 0$, and by Theorem 2.4.3 its inverse is $[(1,q)]$ (the case of

a positive numerator; inverses are unique, Section 2.3). Hence

$$\begin{aligned} j(p)\cdot j(q)^{-1} &= [(p,1)]\cdot[(1,q)] \\ &= [(p\cdot 1,\ 1\cdot q)] = [(p,q)]. \end{aligned}$$

(b) Every element of $\mathbb{Q}$ is $[(p,q)]$ for some $(p,q)\in\mathbb{Z}\times\mathbb{Z}_{>0}$ (Sets, Relations, and Cardinality, Definition 2.5.4), which is $j(p)/j(q)$ by (a). By (a) again, $j(p)/j(q)=j(r)/j(s)$ means $[(p,q)]=[(r,s)]$, which holds if and only if $ps=rq$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)).

(c) Each $j(m)/j(n)$ lies in $\mathbb{Q}$, and every element of $\mathbb{Q}$ has this form by (b). This proves the proposition.

■

Identification. From now on we write m for $j(m)$, as Section 2.2 wrote n for $\iota(n)$, and regard $\mathbb{Z}$ as a subset of $\mathbb{Q}$; together with Section 2.2, $\mathbb{N}\subseteq\mathbb{Z}\subseteq\mathbb{Q}$. This is a convenient abuse, made honest by j: it is injective and carries 0, 1, $+$, $\cdot$ and $-$ of $\mathbb{Z}$ to those of $\mathbb{Q}$, so a computation with integers gives the same result read in $\mathbb{Z}$ or in $\mathbb{Q}$. (One could replace $j(\mathbb{Z})$ by $\mathbb{Z}$ itself inside $\mathbb{Q}$, but nothing needs it.) With this convention, (a)–(c) read:

$$[(p,q)]=\frac{p}{q},\qquad \frac{p}{q}=\frac{r}{s}\iff ps=rq,$$
$$\mathbb{Q}=\left\{\frac{m}{n}:m\in\mathbb{Z},\ n\in\mathbb{Z}_{>0}\right\}.$$

The last equation is the description of $\mathbb{Q}$ that *Sets, Relations, and Cardinality*, Section 2.1 took as a definition; it is now a theorem about the constructed set. The fraction bar is division in the field $\mathbb{Q}$, and the school rules for fractions, such as $\frac{p}{q}+\frac{r}{s}=\frac{ps+rq}{qs}$, are now the formulas of Definition 2.4.2 read through (a). The order of $\mathbb{Q}$, and the fact that j respects it, are Section 2.5's business.

2.4.5 Other denominators

The convention of positive denominators was a choice. The next example shows that the other natural choice gives the same set of rational numbers, and where it would cost us.

Example. Non-zero denominators give the same rationals

Let $Y:=\mathbb{Z}\times(\mathbb{Z}\setminus\{0\})$, and for $(p,q),(r,s)\in Y$ let $(p,q)\approx(r,s)$ mean $ps=rq$. Show that $\approx$ is an equivalence relation on Y, and that

$$I\colon\mathbb{Q}\to Y/{\approx},\qquad I\big([(p,q)]\big):=[(p,q)]_{\approx},$$

is a well-defined bijection.

Solution. Equivalence relation. The proof of Proposition 2.4.1 works on Y with two changes: the pairs are taken in Y instead of $\mathbb{Z}\times\mathbb{Z}_{>0}$, and the line deriving $s\neq 0$ from $s>0$ is dropped, since in Y the middle denominator s is non-zero by definition. That is the only property of the denominators the proof uses.

Well defined. Apply Theorem 2.2.1 (a) to the inclusion $(p,q)\mapsto(p,q)$ from $\mathbb{Z}\times\mathbb{Z}_{>0}$ to Y: if $(p,q)\sim(p',q')$, then $pq'=p'q$, which says $(p,q)\approx(p',q')$.

Injective. If $[(p,q)]_{\approx} = [(p',q')]_{\approx}$, then $(p,q) \approx (p',q')$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b), so $pq' = p'q$, so $(p,q) \sim (p',q')$ and $[(p,q)] = [(p',q')]$.

Surjective. Let $[(p,q)]_{\approx} \in Y/\!\approx$, with $q \neq 0$. By (O1), $q > 0$ or $q < 0$. If $q > 0$, the class is $I([(p,q)])$. If $q < 0$, then $-q > 0$ (Proposition 2.3.5 (a)), and $(p,q) \approx (-p,-q)$, since $p(-q) = -(pq) = (-p)q$ by Proposition 2.3.2 (c); so the class is $I([(-p,-q)])$.

So I is a bijection: the two constructions have "the same" rational numbers, matched by I.

Warning. The rule "$[(p,q)]_{\approx} < [(r,s)]_{\approx}$ if and only if $ps < rq$" is **not** well defined on $Y/\!\approx$. Compare $(0,1)$ with $(1,1)$: $0 \cdot 1 < 1 \cdot 1$, that is, $0 < 1$, true. Now replace $(1,1)$ by the equivalent $(-1,-1)$: the rule asks whether $0 \cdot (-1) < (-1) \cdot 1$, that is, $0 < -1$, which is false. Multiplying an inequality by a negative denominator reverses it (Proposition 2.3.5 (d)), so the rule's verdict depends on the representative. With positive denominators this cannot happen, which is why Definition 2.4.1 uses them.

2.4.6 Exercises

2.4.6.1 A. Check your understanding

Exercise. A1

(a) In Definition 2.4.2, why is $qs \in \mathbb{Z}_{>0}$?

(b) Where does the proof of transitivity in Proposition 2.4.1 use that a denominator is non-zero, and which denominator is it?

(c) Find the inverse of $[(-3,5)]$ in $\mathbb{Q}$, as a class with positive denominator.

(d) True or false: $j(2)^{-1} = [(1,2)]$. Give a reason.

Solution.

(a) $q, s \in \mathbb{Z}_{>0}$ means $q > 0$ and $s > 0$ (Theorem 2.3.4 (b)); then $qs > 0$ by (O4), so $qs \in \mathbb{Z}_{>0}$, again by Theorem 2.3.4 (b).

(b) In the last step: from $(pu - tq)s = 0$ it concludes $pu - tq = 0$ by Proposition 2.3.5 (f), which needs $s \neq 0$. Here s is the denominator of the **middle** pair (r,s).

(c) The numerator -3 is negative, so by Theorem 2.4.3 the inverse is $[(-5,3)]$. Check: $[(-3,5)] \cdot [(-5,3)] = [(15,15)] = [(1,1)]$, since $15 \cdot 1 = 1 \cdot 15$.

(d) True. $j(2) = [(2,1)]$ has positive numerator, so its inverse is $[(1,2)]$ by Theorem 2.4.3; directly, $[(2,1)] \cdot [(1,2)] = [(2,2)] = [(1,1)]$.

2.4.6.2 B. Practice

Exercise. B1: Three rows in detail

Let $x = [(p,q)]$, $y = [(r,s)]$, $z = [(t,u)]$ be rational numbers. Verify in detail, naming

the law of $\mathbb{Z}$ used at each step, the rows (F9), (F4) and (F2) of the table in the proof of Theorem 2.4.3.

Solution. All operations are those of Definition 2.4.2; (F1)–(F9) refer to the laws of $\mathbb{Z}$ (Theorem 2.3.1).

(F9). $y+z=[(ru+ts,\ su)]$, so $x(y+z)=[(p(ru+ts),\ q(su))]=[(pru+pts,\ qsu)]$ by (F9) and (F5). On the other side, $xy=[(pr,qs)]$ and $xz=[(pt,qu)]$, so

$$\begin{aligned} xy+xz &= [((pr)(qu)+(pt)(qs),\ (qs)(qu))] \\ &= [(q(pru+pts),\ q(qsu))], \end{aligned}$$

by (F5), (F6) and (F9). The pairs $(A,B):=(pru+pts,qsu)$ and (qA,qB) are related, since $A(qB)=(qA)B$ by (F5), (F6). By Sets, Relations, and Cardinality, Theorem 2.5.1 (b), $x(y+z)=xy+xz$.

(F4). $-x=[(-p,q)]$, so $x+(-x)=[(pq+(-p)q,\ qq)]$. By Proposition 2.3.2 (c), $(-p)q=-(pq)$, so the numerator is $pq+(-(pq))=0$ by (F4). The pair $(0,q^2)$ is related to $(0,1)$, since $0\cdot 1=0=0\cdot q^2$ (Proposition 2.3.2 (b)). So $x+(-x)=[(0,1)]$.

(F2). $x+y=[(ps+rq,qs)]$, so

$$\begin{aligned} (x+y)+z &= [((ps+rq)u+t(qs),\ (qs)u)] \\ &= [(psu+qru+qst,\ qsu)], \end{aligned}$$

by (F9), (F5), (F6). And $y+z=[(ru+ts,su)]$, so

$$\begin{aligned} x+(y+z) &= [(p(su)+(ru+ts)q,\ q(su))] \\ &= [(psu+qru+qst,\ qsu)], \end{aligned}$$

by the same laws. The pairs are equal, so $(x+y)+z=x+(y+z)$.

Exercise. B2: Addition is well defined

Prove, without looking back at Proposition 2.4.2, that the addition of Definition 2.4.2 is well defined. The first half is done for you.

Change the first argument. Let $(p,q)\sim(p',q')$, that is, $pq'=p'q$, and fix (r,s). Then

$$\begin{aligned} (ps+rq)(q's) &= (pq')s^2+rqq's \\ &= (p'q)s^2+rqq's=(p's+rq')(qs), \end{aligned}$$

so $(ps+rq,qs)\sim(p's+rq',q's)$.

Name the move, then complete it.

Solution. The move is **check on representatives, one argument at a time**: having changed the first argument with the second fixed, change the second with the first fixed, and join the two changes by transitivity.

Change the second argument. Let $(r,s)\sim(r',s')$, that is, $rs'=r's$, and fix (p,q). Then, by

(F5), (F6), (F9) in $\mathbb{Z}$,

$$\begin{aligned}(ps + rq)(qs') &= pqss' + (rs')q^2 \\ &= pqss' + (r's)q^2 = (ps' + r'q)(qs),\end{aligned}$$

so $(ps + rq, qs) \sim (ps' + r'q, qs')$.

Join. Let $(p, q) \sim (p', q')$ and $(r, s) \sim (r', s')$. Write g for the sum formula on representatives. The first step, with (r, s) fixed, gives $g((p, q), (r, s)) \sim g((p', q'), (r, s))$; the second, with (p', q') fixed, gives $g((p', q'), (r, s)) \sim g((p', q'), (r', s'))$. By transitivity (Proposition 2.4.1), $g((p, q), (r, s)) \sim g((p', q'), (r', s'))$, so the classes $[g((p, q), (r, s))]$ and $[g((p', q'), (r', s'))]$ are equal (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)). These are exactly the hypotheses, and the conclusion, of Theorem 2.2.1 (b), which then gives addition as a function $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$.

2.4.6.3 C. Going deeper

Exercise. C1: The least denominator

(a) Show that the rule "$[(p, q)] \mapsto q$" does not define a function $\mathbb{Q} \to \mathbb{Z}_{>0}$.

(b) Prove that there is exactly one function $N \colon \mathbb{Q} \to \mathbb{Z}_{>0}$ such that, for every $x \in \mathbb{Q}$, $N(x)$ is the least $q \in \mathbb{Z}_{>0}$ for which $x = [(p, q)]$ for some $p \in \mathbb{Z}$.

(c) Compute $N(6/4)$, justifying that no smaller denominator is possible.

Solution (Hint 1). For (b), fix x and look at the set of all denominators of representatives of x. Which result of Section 2.3 gives a least element?

Solution (Hint 2). Let $D_x := \{q \in \mathbb{Z}_{>0} : (p, q) \in x \text{ for some } p \in \mathbb{Z}\}$. It is non-empty because a class contains its representative.

Solution.

(a) $[(1, 2)] = [(2, 4)]$, since $1 \cdot 4 = 2 \cdot 2$, but the rule gives 2 for one representative and 4 for the other. A function has one value at each argument (Sets, Relations, and Cardinality, Definition 1.4.1).

(b) For $x \in \mathbb{Q}$, let $D_x := \{q \in \mathbb{Z}_{>0} : (p, q) \in x \text{ for some } p \in \mathbb{Z}\}$. Note that $x = [(p, q)]$ holds if and only if $(p, q) \in x$. If $x = [(p, q)]$, then $(p, q) \in x$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (a). Conversely, write $x = [(p_0, q_0)]$ (Sets, Relations, and Cardinality, Definition 2.5.4); then $(p, q) \in x$ means $(p, q) \sim (p_0, q_0)$ (Sets, Relations, and Cardinality, Definition 2.5.2), so $[(p, q)] = [(p_0, q_0)] = x$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). So D_x is the set of denominators allowed in the definition of $N(x)$.
D_x has a least element. By Sets, Relations, and Cardinality, Definition 2.5.4, $x = [(p_0, q_0)]$ for some pair, and $(p_0, q_0) \in x$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (a), so $q_0 \in D_x$ and D_x is non-empty. Every element of D_x is positive, so 0 is an integer lower bound. By Proposition 2.3.6, D_x has a least element, and it is unique,

by antisymmetry of $\leq$ (Proposition 2.3.5 (i)).
Existence and uniqueness of N. Let N be the function whose graph is $\{(x, \min D_x) : x \in \mathbb{Q}\} \subseteq \mathbb{Q} \times \mathbb{Z}_{>0}$. Each x has exactly one partner $\min D_x$, so N is a function (Sets, Relations, and Cardinality, Definition 1.4.1), and it has the required property. Any function with the property must take the value $\min D_x$ at every x, so it equals N (Sets, Relations, and Cardinality, Proposition 1.4.1). Unlike the rule in (a), N is defined from the class x itself, not from a chosen representative, so there is nothing to check for well-definedness.

(c) $6/4 = [(6, 4)]$ by Proposition 2.4.4 (a). A pair (p, q) lies in this class if and only if $4p = 6q$.

- $q = 1$ is impossible: it needs $4p = 6$. By discreteness (Theorem 2.3.4 (c)), $p \leq 1$ or $p \geq 2$. If $p \leq 1$, then $4p \leq 4 < 6$; if $p \geq 2$, then $4p \geq 8 > 6$, by Proposition 2.3.5 (c) (or equality). Either way $4p \neq 6$, by (O1).
- $q = 2$ works: $4 \cdot 3 = 12 = 6 \cdot 2$, so $(3, 2) \in [(6, 4)]$.

Since 1 is the only element of $\mathbb{Z}_{>0}$ below 2 (discreteness), $N(6/4) = 2$.

Marking scheme. (10 points) - 2: (a) exhibits two representatives of one class with different denominators. - 2: (b) defines the set of denominators of x from the class, and shows it non-empty. - 2: (b) gets a least element from Proposition 2.3.6, with the lower bound named. - 2: (b) explains why this defines exactly one function (one value per class; any such function has those values). - 2: (c) shows $4p = 6$ has no integer solution and exhibits $(3, 2)$. Fatal errors (0 for the step): defining $N([(p, q)])$ as "q divided by a common factor" (divisibility is not available, and the formula uses a representative); in (c), "$4p = 6$ has no solution since $p = 3/2$" (uses division in $\mathbb{Q}$ to decide a question about $\mathbb{Z}$ without the discreteness argument).

Exercise. C2: Find the error

Here is a proof that multiplication on $\mathbb{Q}$ is well defined. Is the conclusion true? Is the proof correct? If not, name the first line that fails, the error, and give a counterexample to the step.

(1) Let $(p, q) \sim (p', q')$, and let $(r, s) \in \mathbb{Z} \times \mathbb{Z}_{>0}$. (2) Then $[(p, q)] = [(p', q')]$. (3) So $[(p, q)] \cdot [(r, s)] = [(p', q')] \cdot [(r, s)]$. (4) That is, $[(pr, qs)] = [(p'r, q's)]$, so the product does not depend on the first representative. (5) The second argument is handled in the same way.

Solution. The conclusion is **true**: it is Proposition 2.4.2. The proof is **not** correct.

The failing line is (3). Line (2) is correct: it is (1) restated by Sets, Relations, and Cardinality, Theorem 2.5.1 (b), and says that (p, q) and (p', q') name the same class. Line (3) then multiplies that class by $[(r, s)]$, using the multiplication of $\mathbb{Q}$. But that multiplication is exactly what is being shown to exist: until it is known to be well defined, "$[(p, q)] \cdot [(r, s)]$" has no meaning, or rather means "$[(pr, qs)]$, computed from the representative (p, q)", and whether this depends on the representative is the claim. The line cites the goal, which is circular: entry (N9) of Logic and Proof, Example (What is not allowed).

Counterexample to the step. The inference "$[(p, q)] = [(p', q')]$, so $F(p, q) = F(p', q')$" fails for

formulas F in representatives. With $F(p,q) \coloneqq p+q$, the same lines "prove" that $[(p,q)] \mapsto p+q$ is a function on $\mathbb{Q}$, since "$p+q$ is a function of the class $[(p,q)]$". It is not: $(1,2) \sim (2,4)$ and $1+2 \neq 2+4$ (Example (Formulas that ignore the class) (a)).

Repair. Replace (2)–(3) by an identity in $\mathbb{Z}$: from $pq' = p'q$, $(pr)(q's) = (pq')(rs) = (p'q)(rs) = (p'r)(qs)$, so $(pr, qs) \sim (p'r, q's)$. Do the same for the second argument, and conclude with Theorem 2.2.1 (b). Lines (1), (4) and (5) are correct as stated, once (3) is replaced.

Marking scheme. (10 points) - 2: says that the conclusion is true and that the proof is not correct. - 3: names line (3) as the failing line and explains that it uses the multiplication of $\mathbb{Q}$, which is being defined. - 2: names the error as circular, entry (N9). - 2: gives a counterexample to the step (any formula in representatives that is not constant on classes, such as $p+q$). - 1: gives the repair through the integer identity $(pr)(q's) = (p'r)(qs)$. Fatal errors (0 for the step): declaring the conclusion false; blaming line (1) or (2), which only restate the hypothesis.

2.5 The Rationals as an Ordered Field; the Archimedean Property

Section 2.4 built $\mathbb{Q}$ as a set of classes of fractions and proved that it is a field (Theorem 2.4.3): we can add, subtract, multiply, and divide by anything non-zero. What it did not give us is an order. In school one "sees" that $-3/5 < -7/12$ by placing both on a line, but the line is the school picture this volume has set aside, so the order must be defined on the classes and its rules proved. This section does that, by the method of Section 2.3: name the positive rationals, check two closure properties, and let Proposition 2.3.3 produce the rest. Then we prove the first fact about $\mathbb{Q}$ that later chapters of this volume lean on at every turn: every rational is exceeded by a natural number, and so $1/n$ eventually drops below any positive rational. Finally every rational gets an integer part.

2.5.1 Which fractions are positive

Throughout, a rational is written as in Section 2.4. By Proposition 2.4.4 (b), every $x \in \mathbb{Q}$ is p/q for some $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$, where p/q is the class $[(p,q)]$, and $p/q = r/s$ if and only if $ps = rq$. The integers sit inside $\mathbb{Q}$ as $m = m/1$, and the natural numbers inside the integers.

The order of $\mathbb{Z}$ came from its positive set $P_{\mathbb{Z}}$ (Definition 2.3.3), and Proposition 2.3.3 turned closure under $+$ and $\cdot$, together with a trichotomy, into all of (O1)–(O4). We do the same for $\mathbb{Q}$, so the whole question is: which fractions should count as positive? A first guess from school is "numerator and denominator have the same sign". But in $\mathbb{Q}$ the denominator is always positive. That was the decision recorded with Definition 2.4.1, and here it pays: the sign of a fraction is the sign of its numerator.

A fraction with a positive denominator is positive exactly when its numerator is.

Definition 2.5.1. The Order on $\mathbb{Q}$

Let

$$P_{\mathbb{Q}} \coloneqq \{[(p,q)] : (p,q) \in \mathbb{Z} \times \mathbb{Z}_{>0} \text{ and } p > 0\}.$$

For $x, y \in \mathbb{Q}$, we write $x < y$ if $y - x \in P_{\mathbb{Q}}$. As in Definition 2.3.2, $x > y$ means $y < x$, and $x \leq y$ means $x < y$ or $x = y$.

In words: a rational is in $P_{\mathbb{Q}}$ if it **has a** representative (p, q) with $p > 0$, the inequality being the order of $\mathbb{Z}$; and $x < y$ means that the difference $y - x$, computed in the field $\mathbb{Q}$, lies in $P_{\mathbb{Q}}$.

Well-definedness. The definition says "has a representative with $p > 0$". To use it we want more: that **every** representative of a class in $P_{\mathbb{Q}}$ has a positive numerator, so that the test can be run on whichever representative is at hand. This is the situation of Theorem 2.2.1 (c), with $X = \mathbb{Z} \times \mathbb{Z}_{>0}$ and $S = \{(p, q) \in X : p > 0\}$. Its hypothesis asks: if $(p, q) \sim (p', q')$ and $p > 0$, is $p' > 0$? It is, and the reason is the sign of the denominators.

Suppose $pq' = p'q$ with $q, q' \in \mathbb{Z}_{>0}$ and $p > 0$. Since $q, q' > 0$ (Theorem 2.3.4 (b)), (O4) gives $pq' > 0$, so $p'q > 0$. If $p' = 0$, then $p'q = 0$ (Proposition 2.3.2 (b)); if $p' < 0$, then $p'q < 0 \cdot q = 0$ by Proposition 2.3.5 (c), since $q > 0$. Both contradict $p'q > 0$ by trichotomy, so $p' > 0$. Now Theorem 2.2.1 (c) gives, for every $(p, q) \in \mathbb{Z} \times \mathbb{Z}_{>0}$,

$$[(p, q)] \in P_{\mathbb{Q}} \iff p > 0. \tag{2.5.1}$$

Examples. The class $3/4$ lies in $P_{\mathbb{Q}}$, and so does $6/8$, the same class: both numerators are positive, as Eq. 2.5.1 says they must be. The class $-6/8$ does not, since $-6 > 0$ fails. The degenerate case is zero: $0 = 0/1 = 0/7$, and $0 > 0$ fails, so $0 \notin P_{\mathbb{Q}}$. That matters, because (P2) of Proposition 2.3.3 will ask for exactly one of $x \in P_{\mathbb{Q}}$, $x = 0$, $-x \in P_{\mathbb{Q}}$.

Non-example by minimal change. Allow negative denominators, as in Example (Non-zero denominators give the same rationals), and keep the test "the numerator is positive". The relation is the same and still an equivalence relation, but now $(1, 1) \approx (-1, -1)$: the class of $(1, 1)$ passes the test with one representative and fails it with another. The clause that fails is the hypothesis of Theorem 2.2.1 (c), and the step of the check above that breaks is "$q > 0$".

Why this definition. With non-zero denominators of either sign one could test "$pq > 0$" instead, which is well defined; the price is a product in every sign test and two sign cases in every proof. Positive denominators put all the sign information in one integer. The order is defined through the positive set rather than through a formula for $x < y$, so that Proposition 2.3.3 does the work of (O1)–(O4) once and for all.

Warning. Comparing numerators alone is not comparing fractions. The rationals $1/2$ and $2/5$ have $1 < 2$, yet $1/2 > 2/5$, since $1/2 - 2/5 = 1/10 \in P_{\mathbb{Q}}$. Only after bringing both to the same positive denominator do the numerators decide: that is the comparison formula below.

Quick check. Which of $[(-2, 3)]$, $[(0, 5)]$, $[(4, 6)]$ lie in $P_{\mathbb{Q}}$? Why can one not ask whether $[(2, -3)]$ lies in $P_{\mathbb{Q}}$?

Solution. By Eq. 2.5.1, only $[(4, 6)]$ does: its numerator 4 is positive, while $-2 > 0$ and $0 > 0$ both fail. The symbol $[(2, -3)]$ names no element of $\mathbb{Q}$: the denominator -3 is not in $\mathbb{Z}_{>0}$, so $(2, -3)$ is not in $\mathbb{Z} \times \mathbb{Z}_{>0}$ and has no class. (The rational one might mean is $-2/3 = [(-2, 3)]$, which is not in $P_{\mathbb{Q}}$.)

2.5.2 The rationals form an ordered field

The payoff is that the order of $\mathbb{Q}$ has every property of the order of $\mathbb{Z}$, together with a formula that compares two fractions with one integer inequality.

Theorem 2.5.1. The Rationals Form an Ordered Field

With the order of Definition 2.5.1, $\mathbb{Q}$ is an ordered field (Definition 2.3.2), and $P_{\mathbb{Q}} = \{x \in \mathbb{Q} : x > 0\}$. Moreover:

(a) *(The comparison formula.)* For $p, r \in \mathbb{Z}$ and $q, s \in \mathbb{Z}_{>0}$,

$$\frac{p}{q} < \frac{r}{s} \iff ps < rq;$$

in particular $p/q > 0$ if and only if $p > 0$.

(b) For $m, n \in \mathbb{Z}$, $j(m) < j(n)$ if and only if $m < n$.

Idea. By Proposition 2.3.3, a commutative ring with a positive set satisfying (P1) and (P2) is an ordered commutative ring, and $\mathbb{Q}$ is already a field. So everything reduces to the two properties of $P_{\mathbb{Q}}$, and Eq. 2.5.1 turns each into a statement about numerators in $\mathbb{Z}$: reduce to the level below. A sum $p/q + r/s$ has numerator $ps + rq$, and with $p, r > 0$ this is positive **because** s **and** q **are**. The comparison formula is (P1)'s computation run on a difference. *Before reading on: where is $q > 0$ used?*

Proof. By Theorem 2.4.3, $\mathbb{Q}$ is a field, in particular a commutative ring. We check (P1) and (P2) of Proposition 2.3.3 for $P = P_{\mathbb{Q}}$.

(P1) Let $x, y \in P_{\mathbb{Q}}$, say $x = [(p, q)]$ and $y = [(r, s)]$ with $q, s \in \mathbb{Z}_{>0}$. By Eq. 2.5.1, $p > 0$ and $r > 0$; and $q, s > 0$ by Theorem 2.3.4 (b). By Definition 2.4.2,

$$x + y = [(ps + rq,\ qs)], \qquad xy = [(pr,\ qs)],$$

with $qs \in \mathbb{Z}_{>0}$. By (O4) in $\mathbb{Z}$, $ps > 0$ and $rq > 0$, so $ps + rq > 0 + 0 = 0$ by Proposition 2.3.5 (b); and $pr > 0$ by (O4). By Eq. 2.5.1, $x + y \in P_{\mathbb{Q}}$ and $xy \in P_{\mathbb{Q}}$.

(P2) Let $x = [(p, q)] \in \mathbb{Q}$. Then $-x = [(-p, q)]$ (Definition 2.4.2). By Theorem 2.4.3, the zero of $\mathbb{Q}$ is $[(0, 1)]$, and $x = 0$ if and only if $p = 0$. By Eq. 2.5.1, $x \in P_{\mathbb{Q}}$ if and only if $p > 0$, and $-x \in P_{\mathbb{Q}}$ if and only if $-p > 0$, which by Proposition 2.3.5 (a) means $p < 0$. By (O1) in $\mathbb{Z}$, exactly one of $p > 0$, $p = 0$, $p < 0$ holds, so exactly one of $x \in P_{\mathbb{Q}}$, $x = 0$, $-x \in P_{\mathbb{Q}}$ holds.

By Proposition 2.3.3, the relation $x < y \iff y - x \in P_{\mathbb{Q}}$, which is the relation of Definition 2.5.1, satisfies (O1)–(O4), and $P_{\mathbb{Q}} = \{x : x > 0\}$. A field that is an ordered commutative ring is an ordered field (Definition 2.3.2).

(a) Let $p, r \in \mathbb{Z}$ and $q, s \in \mathbb{Z}_{>0}$. By Definition 2.4.2 and Proposition 2.3.2 (c),

$$\frac{r}{s} - \frac{p}{q} = [(r, s)] + [(-p, q)] = [(rq + (-p)s,\ sq)] = [(rq - ps,\ qs)].$$

So $p/q < r/s$ holds if and only if this class lies in $P_{\mathbb{Q}}$, that is (by Eq. 2.5.1, since $qs \in \mathbb{Z}_{>0}$), if and only if $rq - ps > 0$, which by Proposition 2.3.5 (a) means $ps < rq$.

For the special case, compare $0 = 0/1$ with p/q: the formula gives $0/1 < p/q$ if and only if $0 \cdot q < p \cdot 1$, that is, $p > 0$.

(b) $j(m) = m/1$ and $j(n) = n/1$ (Proposition 2.4.4), so by (a), $j(m) < j(n)$ if and only if $m \cdot 1 < n \cdot 1$, that is, $m < n$.

This proves the theorem.

■

The answer to the prompt: $q > 0$ is used three times. In the well-definedness check before Eq. 2.5.1, the positive q and q' carry the sign from p to p'. In (P1), ps and rq are positive because s and q are. In (a), the difference has denominator qs, a positive integer, so that Eq. 2.5.1 can read its sign off the numerator. Without positive denominators the first of these fails outright (the non-example above), and so does the comparison formula: see the warning after Example (Non-zero denominators give the same rationals).

Since $p/q = r/s$ exactly when $ps = rq$ (Proposition 2.4.4 (b)), the comparison formula has a version with $\leq$: for $q, s \in \mathbb{Z}_{>0}$,

$$\frac{p}{q} \leq \frac{r}{s} \iff ps \leq rq. \tag{2.5.2}$$

By (b), the order of $\mathbb{Q}$ restricted to $\mathbb{Z}$ is the order of $\mathbb{Z}$, which in turn restricts to the order of $\mathbb{N}$ (Theorem 2.3.4 (a)). So after the identifications $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$, a comparison of two integers means the same thing in all three sets, and we never need to say where it is made. Every rule of Proposition 2.3.5 now holds in $\mathbb{Q}$, and we write

$$\mathbb{Q}_{>0} := \{x \in \mathbb{Q} : x > 0\}$$

for the set of positive rationals, which is $P_{\mathbb{Q}}$.

One rule is new with fields, because rings have no inverses: taking reciprocals reverses the order of positive elements.

Proposition 2.5.2. Reciprocals Reverse Order

Let F be an ordered field and $a, b \in F$.

(a) If $a > 0$, then $a^{-1} > 0$.

(b) If $0 < a < b$, then $0 < b^{-1} < a^{-1}$, that is, $0 < 1/b < 1/a$.

Idea. For (a), turn a^{-1} into a product we can sign: $a^{-1} = a \cdot (a^{-1})^2$, a positive element times a non-zero square. For (b), multiply $a < b$ by the positive element $a^{-1}b^{-1}$: it cancels a on the left and b on the right, and swaps which reciprocal stands on which side.

Proof.

(a) Let $a > 0$. Then $a \neq 0$, so a^{-1} exists, and $a^{-1} \neq 0$: otherwise $1 = aa^{-1} = a \cdot 0 = 0$ (Proposition 2.3.2 (b)), against $1 \neq 0$ in a field. By Proposition 2.3.5 (e), $(a^{-1})^2 > 0$. Now

$$a \cdot (a^{-1})^2 = (aa^{-1})a^{-1} = 1 \cdot a^{-1} = a^{-1},$$

and the left side is a product of two positive elements, so $a^{-1} > 0$ by (O4).

(b) Let $0 < a < b$. Then $b > 0$ by (O2), so $a^{-1} > 0$ and $b^{-1} > 0$ by (a), and $c := a^{-1}b^{-1} > 0$ by (O4). Multiplying $a < b$ by $c > 0$ (Proposition 2.3.5 (c)) gives

$$b^{-1} = (aa^{-1})b^{-1} = ac < bc = (bb^{-1})a^{-1} = a^{-1}.$$

Since $1/b = 1 \cdot b^{-1} = b^{-1}$ (Definition 2.3.1), this is $0 < 1/b < 1/a$.

■

2.5.3 Between two rationals

In $\mathbb{Z}$ every integer has a next one: there is no integer strictly between m and $m+1$ (Theorem 2.3.4 (c)). In $\mathbb{Q}$ the opposite holds. Let $2 := 1+1$. By Proposition 2.3.5 (e) and (b), $1 > 0$ and $2 > 0$, so $1/2 > 0$ by Proposition 2.5.2 (a). Now let $a < b$ in $\mathbb{Q}$. Adding a, and then b, by (O3),

$$a + a < a + b < b + b,$$

and $a+a = (1+1)a = 2a$, $b+b = 2b$ by distributivity. Multiplying by $1/2 > 0$ (Proposition 2.3.5 (c)) gives

$$a < \frac{a+b}{2} < b.$$

So between two distinct rationals there is always a third, their **midpoint**. Applied with $a = 0$: for every $x > 0$, also $0 < x/2 < x$. Hence $\mathbb{Q}_{>0}$ has **no least element**. In particular, well-ordering is a property of $\mathbb{N}$ (Theorem 1.6.2) that $\mathbb{Q}_{>0}$ does not share.

Warning. Between any two rationals lies another, but that does not mean every rational has a neighbor "right next to it". It means the opposite: no rational x has a next one, since any candidate $y > x$ is beaten by the midpoint of x and y. Phrases like "the rational just after 0" name nothing.

Quick check. Find the midpoint of $1/3$ and $1/2$, and confirm with the comparison formula that it lies strictly between them.

Solution. $1/3 + 1/2 = 5/6$ by Definition 2.4.2 ($[(1 \cdot 2 + 1 \cdot 3, 6)]$), and $(5/6) \cdot (1/2) = 5/12$. By Theorem 2.5.1 (a): $1/3 < 5/12$ because $1 \cdot 12 = 12 < 15 = 5 \cdot 3$, and $5/12 < 1/2$ because $5 \cdot 2 = 10 < 12 = 1 \cdot 12$.

2.5.4 The Archimedean property

Here is a statement so familiar that it seems not to need proof: every rational number is smaller than some natural number. For the real numbers it is a genuine theorem. *Analysis on the Real Line*, Chapter 1 obtains it from the least upper bound axiom, and there are ordered fields, built from rational functions, in which the natural numbers are bounded above. For the constructed $\mathbb{Q}$ it is true for a simpler reason: every rational is a quotient of integers, and integers are as discrete as natural numbers.

Theorem 2.5.3. The Rationals Are Archimedean

For every $x \in \mathbb{Q}$ there is $n \in \mathbb{N}$ with $x < n$.

Try it first (ten minutes): given $x = p/q$ with $q \in \mathbb{Z}_{>0}$, which natural number beats it? Do not forget negative p.

Idea. The first candidate is $p + 1$, on the grounds that dividing by $q \geq 1$ "makes things smaller", so $p/q \leq p < p + 1$. An attempt meets two obstacles. Dividing by $q \geq 1$ makes a **negative** number larger: $-3/2 > -3$. And $p+1$ is not a natural number when $p \leq -2$. Both disappear if we split on the sign of p. A fraction with $p \leq 0$ is at most 0, and 1 beats it. For $p > 0$ the school intuition is right and provable: $q \geq 1$ gives $p \leq pq$, which the comparison formula reads as $p/q \leq p$.

Proof. Let $x \in \mathbb{Q}$. By Proposition 2.4.4 (b), $x = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$.

Case 1: $p \leq 0$. By Eq. 2.5.2, $p/q \leq 0/1$, since $p \cdot 1 \leq 0 \cdot q$. So $x \leq 0 < 1$ (Proposition 2.3.5 (e)), and $n \coloneqq 1 \in \mathbb{N}$ works.

Case 2: $p > 0$. Then $p \in \mathbb{N}$ by Theorem 2.3.4 (b). Since $q > 0$, discreteness (Theorem 2.3.4 (c), with $0 < q$) gives $1 \leq q$. Multiplying by $p > 0$ (Proposition 2.3.5 (c), or equality if $q = 1$) gives $p \cdot 1 \leq p \cdot q$. By Eq. 2.5.2 this says $p/q \leq p/1 = p$. Finally $p < p + 1$ by (O3), since $0 < 1$. So $x \leq p < p + 1$, and $n \coloneqq p + 1$ lies in $\mathbb{N}$, because $\mathbb{N}$ is closed under addition. Hence $x < n$.

In both cases $x < n$ for some $n \in \mathbb{N}$, as claimed.

■

The n found depends on x, through the numerator of the representative used; a different representative may give a different n, and any one will do. The proof uses only the representation $x = p/q$, the comparison formula, and elementary facts about the ordered ring $\mathbb{Z}$ (discreteness, $\mathbb{N} = \{x \in \mathbb{Z} : x \geq 0\}$, the rules for inequalities) and about $\mathbb{N}$. No supremum is mentioned in it. That is the contrast with the real numbers, where the same conclusion needs the completeness axiom.

The form in which later chapters of this volume use the property is about reciprocals: $1/n$ eventually falls below any positive rational.

Corollary 2.5.4. Small Reciprocals in $\mathbb{Q}$

For every $\varepsilon \in \mathbb{Q}_{>0}$ there is $n \in \mathbb{Z}_{>0}$ with $1/n < \varepsilon$. Moreover every $m \in \mathbb{Z}$ with $m \geq n$ then also satisfies $1/m < \varepsilon$. In symbols,

$$\forall \varepsilon \in \mathbb{Q}_{>0} \;\; \exists n \in \mathbb{Z}_{>0} \;\; \forall m \in \mathbb{Z}_{>0} \;\; (m \geq n \Rightarrow 1/m < \varepsilon).$$

Proof. Let $\varepsilon \in \mathbb{Q}_{>0}$. By Proposition 2.5.2 (a), $1/\varepsilon > 0$. By Theorem 2.5.3 there is $n \in \mathbb{N}$ with $1/\varepsilon < n$. Then $n > 0$, so $n \in \mathbb{Z}_{>0}$ (Theorem 2.3.4 (b)). By Proposition 2.5.2 (b) with

$0 < 1/\varepsilon < n$,

$$\frac{1}{n} < \frac{1}{1/\varepsilon} = \varepsilon,$$

where $1/(1/\varepsilon) = \varepsilon$ because ε is a multiplicative inverse of ε^{-1}, and inverses are unique (if $bc = 1 = bc'$, then $c = c(bc') = (cb)c' = c'$). If $m \geq n$, then either $m = n$, or $0 < n < m$ and $1/m < 1/n$ by Proposition 2.5.2 (b); in both cases $1/m \leq 1/n < \varepsilon$.

■

The order of the quantifiers is the point: ε is given first, and n is chosen after it and **depends on** ε. Exercise A1 asks what happens if the order is swapped. Here ε is a **rational** number, because the real numbers have not yet been constructed, and the corollary is the form in which a later chapter of this volume uses the Archimedean property, with $\varepsilon \in \mathbb{Q}_{>0}$.

2.5.5 The integer part

By Theorem 2.5.3, every rational lies below an integer. With the least-element property of $\mathbb{Z}$, it pins every rational between two consecutive integers.

Proposition 2.5.5. Integer Part of a Rational

For every $x \in \mathbb{Q}$ there is **exactly one** $m \in \mathbb{Z}$ with $m \leq x < m + 1$. It is called the **integer part** of x and written $\lfloor x \rfloor$.

Idea. The integer just above x should be the least integer exceeding x. So look at the set of integers that exceed x: Theorem 2.5.3 makes it non-empty, and applied to $-x$ it bounds the set below, so Proposition 2.3.6 hands us its least element. One less than that is the integer part. Uniqueness is discreteness: two candidates one unit apart cannot both fit below x and above $x - 1$.

Proof. Let $x \in \mathbb{Q}$, and let $S := \{k \in \mathbb{Z} : x < k\}$. By Theorem 2.5.3 there is $n \in \mathbb{N} \subseteq \mathbb{Z}$ with $x < n$, so $S \neq \emptyset$. Applying Theorem 2.5.3 to $-x$ gives $n' \in \mathbb{N}$ with $-x < n'$, that is, $-n' < x$ (Proposition 2.3.5 (a)). Every $k \in S$ then has $-n' < x < k$, so the integer $-n'$ is a lower bound of S. By Proposition 2.3.6, S has a least element l. Let $m := l - 1 \in \mathbb{Z}$, so that $l = m + 1$.

Existence. $x < l = m + 1$, since $l \in S$. And $m \notin S$: otherwise $l \leq m = l - 1 < l$, so $l < l$, against (O1). So $x < m$ fails, and $m \leq x$ by totality (Proposition 2.3.5 (i)).

Uniqueness. Let $m, m' \in \mathbb{Z}$ both satisfy the condition, and suppose $m < m'$. By discreteness (Theorem 2.3.4 (c)), $m + 1 \leq m'$. Then $m' \leq x < m + 1 \leq m'$, so $m' < m'$, against (O1). By symmetry $m' < m$ is impossible too, so $m = m'$ by (O1).

■

The comparisons in this proof mix integers and rationals ($x < k$), and the least element is taken in $\mathbb{Z}$; by Theorem 2.5.1 (b) the two orders agree on $\mathbb{Z}$, so nothing is lost. A later chapter of this volume reads $\lfloor p/q \rfloor$, for $q \in \mathbb{Z}_{>0}$, as the quotient in the division of p by q.

Quick check. Compute $\lfloor -7/3 \rfloor$. Then find the least $n \in \mathbb{Z}_{>0}$ with $1/n < 3/1000$.

Solution. $\lfloor -7/3 \rfloor = -3$: by Eq. 2.5.2, $-3 = -3/1 \leq -7/3$ because $(-3) \cdot 3 = -9 \leq -7$, and $-7/3 < -2$ because $-7 < -6 = (-2) \cdot 3$ (Theorem 2.5.1 (a)). Not -2: the rule "drop the fraction" gives $\lfloor x \rfloor$ for $x \geq 0$, but not for negative non-integers such as $-7/3$.

For $n \in \mathbb{Z}_{>0}$, the comparison formula gives $1/n < 3/1000$ if and only if $1 \cdot 1000 < 3n$. Since $3 \cdot 333 = 999 < 1000$ and $3 \cdot 334 = 1002 > 1000$, the least such n is 334 (for $n \leq 333$, $3n \leq 999$).

2.5.6 Exercises

2.5.6.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 2.5.1 and about Corollary 2.5.4.

(a) Where in the proof (including the well-definedness check before it) is $q > 0$ used? What goes wrong without it?

(b) Why does (P2) need "**exactly** one" of $x \in P_{\mathbb{Q}}$, $x = 0$, $-x \in P_{\mathbb{Q}}$, and not just "at least one"?

(c) Run the comparison formula on $-2/3$ and $-3/5$: which is larger?

(d) State the idea of the proof in two sentences.

(e) Write Corollary 2.5.4 (its first sentence) in symbols. Then write the version with the two quantifiers swapped, negate it, and decide which of the two versions is true.

Solution.

(a) Three times: in the well-definedness check, $q, q' > 0$ turn $p > 0$ into $p' > 0$; in (P1), $ps > 0$ and $rq > 0$ need $s > 0$ and $q > 0$; in (a), the difference $r/s - p/q$ has denominator $qs > 0$, so its sign is the sign of its numerator. Without it, the test "numerator positive" is not well defined: in the non-example after Definition 2.5.1, $(1, 1) \approx (-1, -1)$ gives two verdicts for one class, and, as the warning after Example (Non-zero denominators give the same rationals) shows, "$ps < rq$" gives $0 < 1$ for $(0, 1), (1, 1)$ but $0 < -1$ for $(0, 1), (-1, -1)$.

(b) "At least one" would allow a non-zero x with both x and $-x$ positive, and then $x < 0$ and $0 < x$ at once, against (O1). The "at most one" half is what makes the order antisymmetric; Proposition 2.3.3 uses it to get trichotomy (O1).

(c) By Theorem 2.5.1 (a), $-2/3 < -3/5$ if and only if $(-2) \cdot 5 < (-3) \cdot 3$, that is, $-10 < -9$, which is true. So $-3/5$ is the larger.

(d) By Proposition 2.3.3 it is enough to show that the positive set is closed under $+$ and $\cdot$ and satisfies the trichotomy (P2). Each of these, run on representatives with positive denominators, is a statement about integer numerators, which holds because $\mathbb{Z}$ is an ordered ring.

(e) The corollary: $\forall \varepsilon \in \mathbb{Q}_{>0}\ \exists n \in \mathbb{Z}_{>0}\ \ 1/n < \varepsilon$. Swapped: $\exists n \in \mathbb{Z}_{>0}\ \forall \varepsilon \in \mathbb{Q}_{>0}\ \ 1/n < \varepsilon$,

"one n works for every ε". By Logic and Proof, Proposition 6.2.3 (the bounded form of Logic and Proof, Proposition 6.2.1), applied twice, its negation is

$$\forall n \in \mathbb{Z}_{>0} \;\; \exists \varepsilon \in \mathbb{Q}_{>0} \;\; \neg(1/n < \varepsilon).$$

This negation is true: given $n \in \mathbb{Z}_{>0}$, take $\varepsilon := 1/n$, which lies in $\mathbb{Q}_{>0}$ by Proposition 2.5.2 (a), and $1/n < 1/n$ fails by (O1). So the swapped version is false, and the corollary (proved above) is true. The n must be allowed to depend on ε.

2.5.6.2 B. Practice

Exercise. B1

Put $-7/12$, $-3/5$ and $-4/7$ in increasing order, using only the comparison formula.

Solution. By Theorem 2.5.1 (a), pair by pair:

- $-3/5 < -7/12$, since $(-3)\cdot 12 = -36 < -35 = (-7)\cdot 5$;
- $-7/12 < -4/7$, since $(-7)\cdot 7 = -49 < -48 = (-4)\cdot 12$;
- $-3/5 < -4/7$, since $(-3)\cdot 7 = -21 < -20 = (-4)\cdot 5$.

Hence $-3/5 < -7/12 < -4/7$. (The third comparison follows from the first two by (O2); computing it is a check.)

Exercise. B2: A rational of the form a + 1/n

(a) Let $a, b \in \mathbb{Q}$ with $a < b$. Prove that there is $n \in \mathbb{Z}_{>0}$ with $a < a + 1/n < b$, and that every $m \in \mathbb{Z}$ with $m \geq n$ works too.

(b) Find the least such n for $a = 1/3$ and $b = 1/2$.

Solution.

(a) Let $\varepsilon := b - a$. Then $\varepsilon > 0$ by Proposition 2.3.5 (a), so Corollary 2.5.4 gives $n \in \mathbb{Z}_{>0}$ with $1/n < \varepsilon$, and $1/m < \varepsilon$ for every $m \geq n$. Fix such an m (for instance $m = n$). Adding a (O3): $a + 1/m < a + (b - a) = b$. And $1/m > 0$ by Proposition 2.5.2 (a), since $m > 0$; adding a gives $a < a + 1/m$.

(b) Here $b - a = 1/2 - 1/3 = 1/6$, and for $n \in \mathbb{Z}_{>0}$ the comparison formula gives $1/n < 1/6$ if and only if $6 < n$. So the least n is 7, and $a + 1/7 = 1/3 + 1/7 = 10/21$. Check: $1/3 < 10/21$ since $21 < 30$, and $10/21 < 1/2$ since $20 < 21$. For $n = 6$, $a + 1/6 = 1/2 = b$, which is not strictly below b.

Exercise. B3

Compute $\lfloor 22/7 \rfloor$, $\lfloor -22/7 \rfloor$, $\lfloor 0 \rfloor$ and $\lfloor -1/100 \rfloor$, and check each against the defining inequalities of Proposition 2.5.5.

Solution. We use Eq. 2.5.2 and the comparison formula with denominators 1 for the integers.

- $\lfloor 22/7 \rfloor = 3$: $3 \le 22/7$ since $3 \cdot 7 = 21 \le 22$, and $22/7 < 4$ since $22 < 28 = 4 \cdot 7$.
- $\lfloor -22/7 \rfloor = -4$: $-4 \le -22/7$ since $-28 \le -22$, and $-22/7 < -3$ since $-22 < -21$.
- $\lfloor 0 \rfloor = 0$: $0 \le 0 < 1$.
- $\lfloor -1/100 \rfloor = -1$: $-1 \le -1/100$ since $-100 \le -1$, and $-1/100 < 0$ since $-1 < 0$.

The first two show that $\lfloor -x \rfloor \neq -\lfloor x \rfloor$ in general: $\lfloor 22/7 \rfloor = 3$, but $\lfloor -22/7 \rfloor = -4$, not -3.

2.5.6.3 C. Going deeper

Exercise. C1: Exit check: $\mathbb{Q}$ is Archimedean

Prove, from the construction of $\mathbb{Q}$ and the results of this chapter before Theorem 2.5.3 (and not from any fact about the real numbers), that for every $x \in \mathbb{Q}$ there is $n \in \mathbb{N}$ with $x < n$; and deduce that for every $\varepsilon \in \mathbb{Q}_{>0}$ there is $n \in \mathbb{Z}_{>0}$ with $1/n < \varepsilon$.

Solution. The theorem. Let $x \in \mathbb{Q}$. By Proposition 2.4.4 (b), $x = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$.

If $p \le 0$, then $p \cdot 1 \le 0 \cdot q$, so $x = p/q \le 0/1 = 0$ by the comparison formula (Theorem 2.5.1 (a), with equality by Proposition 2.4.4 (b)). Hence $x \le 0 < 1$, and $n = 1$ works.

If $p > 0$, then $p \in \mathbb{N}$ (Theorem 2.3.4 (b)), and $q \ge 1$ by discreteness (Theorem 2.3.4 (c)), since $q > 0$. Multiplying $1 \le q$ by $p > 0$ (Proposition 2.3.5 (c), or equality if $q = 1$) gives $p \le pq$, which by the comparison formula says $p/q \le p/1$. So $x \le p < p + 1$, and $n \coloneqq p + 1 \in \mathbb{N}$ works.

The corollary. Let $\varepsilon \in \mathbb{Q}_{>0}$ be given. Then $1/\varepsilon > 0$ (Proposition 2.5.2 (a)), and the theorem gives $n \in \mathbb{N}$ with $1/\varepsilon < n$; in particular $n > 0$, so $n \in \mathbb{Z}_{>0}$. By Proposition 2.5.2 (b), $1/n < 1/(1/\varepsilon) = \varepsilon$. The n depends on ε, which was fixed first.

Marking scheme. (10 points) - 1: writes $x = p/q$ with $p \in \mathbb{Z}$, $q \in \mathbb{Z}_{>0}$, citing the representation of rationals. - 2: the case $p \le 0$: $x \le 0 < 1$, justified by the comparison formula. - 3: the case $p > 0$: $q \ge 1$ by discreteness of $\mathbb{Z}$, and $p/q \le p$ by the comparison formula (or an equivalent inequality argument from the ordered-field rules). - 1: exhibits $n \in \mathbb{N}$ (for instance $p + 1$) and checks $x < n$. - 3: the corollary: ε fixed first, $n > 1/\varepsilon$ from the theorem, reciprocals reverse order. Fatal errors (0 for the step): no separate treatment of $p < 0$ (there "$p/q \le p$" fails once $q \ge 2$, as $-3/2 > -3$ shows, and $p + 1 \notin \mathbb{N}$ once $p \le -2$); using a supremum, the real numbers, or "the decimal expansion of x". A route through Proposition 2.5.5 is circular here (its proof uses the theorem) and earns nothing for the theorem's 7 points; the corollary's 3 points, deduced from the theorem as stated, can still be earned.

Exercise. C2: The comparison, on representatives

The comparison formula of Theorem 2.5.1 (a) was proved through the positive set. Prove directly that the relation "$ps < rq$" on pairs in $\mathbb{Z} \times \mathbb{Z}_{>0}$ does not depend on the representatives: if $(p, q) \sim (p', q')$ and $(r, s) \sim (r', s')$, then $ps < rq$ if and only if $p's' < r'q'$.

Here is the start.

> We change one argument at a time. *First argument:* let $(p, q) \sim (p', q')$ and $ps < rq$; we must show $p's < rq'$.
>
> ...

Name the move, then complete it.

Solution. The move is **check on representatives, one argument at a time**: change the first pair with the second fixed, then the second with the first fixed, and join the two by transitivity of $\iff$.

First argument. Let $pq' = p'q$ and $ps < rq$. Multiply by $q' > 0$ (Proposition 2.3.5 (c)): $psq' < rqq'$. The left side is $(pq')s = (p'q)s = (p's)q$, and the right side is $(rq')q$. So $(p's)q < (rq')q$. If $p's \geq rq'$, then multiplying by $q > 0$ would give $(p's)q \geq (rq')q$ (Proposition 2.3.5 (c), or equality), against (O1). So $p's < rq'$. Since $\sim$ is symmetric, the same argument with the roles of (p, q) and (p', q') exchanged gives the converse: $ps < rq \iff p's < rq'$.

Second argument. Let $rs' = r's$, and fix (p', q'). If $p's < rq'$, multiply by $s' > 0$: $p'ss' < rq's'$, and $rs' = r's$ turns the right side into $(r's)q' = (r'q')s$. So $(p's')s < (r'q')s$, and canceling $s > 0$ as above gives $p's' < r'q'$. By symmetry again, $p's < rq' \iff p's' < r'q'$.

Join. Chaining the two equivalences,

$$ps < rq \iff p's < rq' \iff p's' < r'q'.$$

This proves the claim. By Theorem 2.5.1 (a) both sides also mean $p/q < r/s$; the point of the exercise is that the integer relation alone is already independent of the representatives, and positivity of q, q', s, s' is used at every multiplication and cancellation.

Marking scheme. (10 points) - 2: names the move (check on representatives, one argument at a time, joined by transitivity of $\iff$). - 3: first argument: multiplies $ps < rq$ by $q' > 0$, uses $pq' = p'q$ to reach $(p's)q < (rq')q$, cancels $q > 0$ (by contradiction via trichotomy), and gets the converse from symmetry of $\sim$. - 3: second argument: the same with $s' > 0$, $rs' = r's$ and cancellation of $s > 0$, with the converse by symmetry. - 2: chains $ps < rq \iff p's < rq' \iff p's' < r'q'$ and says where positivity of q, q', s, s' is used. Fatal errors (0 for the step): multiplying or canceling by a denominator without using its sign; proving only one direction of an equivalence without invoking symmetry; using $p/q < r/s$ (the comparison formula of Theorem 2.5.1 (a)) instead of working on the integer relation.

Exercise. C3: Find the error

Here is an argument for Theorem 2.5.3. Is the conclusion true? Is the argument correct? If not, name the first line that fails, say what makes it illegal, and give a counterexample to the step.

> (1) Let $x \in \mathbb{Q}$, and write $x = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$. (2) Then $q \geq 1$, by discreteness of $\mathbb{Z}$. (3) So $x = p/q \leq p$, since dividing by a number at least 1 cannot increase it. (4) Hence $n := p + 1$ is a natural number with $n > x$.

Solution. The conclusion is **true**: it is Theorem 2.5.3. The argument is **not** correct.

The failing line is (3). It cites the general "fact" that $p/q \leq p$ whenever $q \geq 1$. That is neither proved in this chapter nor true, so the line rests on nothing: entry (N9) of Logic and Proof, Example (What is not allowed). What is true, and what the proof of the theorem uses, is the special case $p \geq 0$.

Counterexample to the step. $p = -3$, $q = 2$: then $q \geq 1$, but $x = -3/2 > -3 = p$. Indeed the comparison formula gives $-3/1 < -3/2$, because $(-3) \cdot 2 = -6 < -3 = (-3) \cdot 1$. There is no counterexample to the conclusion.

A second failing line. Line (4) also fails for $p \leq -2$: then $p + 1 \leq -1$, which is not in $\mathbb{N} = \{n \in \mathbb{Z} : n \geq 0\}$ (Theorem 2.3.4 (b)). For $p = -1$, $n = 0$ happens to work, because $x < 0$; but the stated reason "$x \leq p$" still fails there once $q \geq 2$.

Repair. Split on the sign of p, as the proof of Theorem 2.5.3 does: for $p \leq 0$, $x \leq 0 < 1$; for $p > 0$, line (3) is correct, with the comparison formula as its reason ($p \cdot 1 \leq p \cdot q$).

Marking scheme. (10 points) - 2: says that the conclusion is true and the argument is not correct. - 3: names line (3), and the unproved general claim "$p/q \leq p$ for $q \geq 1$" (entry (N9)) as what makes it illegal. - 2: gives a counterexample to the step with $p < 0$ and $q \geq 2$, checked by the comparison formula. - 1: notes that line (4) fails as well for $p \leq -2$. - 2: repairs the argument by a case split on the sign of p. Fatal errors (0 for the step): declaring the conclusion false; offering a counterexample to the conclusion.

2.6 The Gaps in the Rationals

The rationals now have everything Section 2.5 asked of them: they form an ordered field (Theorem 2.5.1), every rational lies below a natural number (Theorem 2.5.3), and no two rationals are neighbors, since their midpoint lies between them. Is anything missing? School geometry suggests an answer. The diagonal of a square with side 1 has a length whose square is 2, and no fraction seems to measure it. That picture is motivation only. This section proves the two facts behind it from the construction: no rational squares to 2, and the set of rationals whose square is below 2 has no least upper bound in $\mathbb{Q}$. We also build a sequence of rationals that closes in on the gap from above without ever landing in it. A later chapter of this volume fills the gap.

2.6.1 No rational squares to two

Throughout, $x^2 := x \cdot x$ and $2 := 1 + 1$, with $4 = 2 \cdot 2$ and so on; by the identifications of Sections 2.2 and 2.4 these are the natural numbers of the same names. Their sums and products may be computed in $\mathbb{N}$, since the maps ι and j preserve $+$ and $\cdot$ (Proposition 2.2.3, Proposition 2.4.4), and so may their comparisons (Theorem 2.3.4 (a), Theorem 2.5.1 (b)).

Logic and Proof, Chapter 2 proved that "$\sqrt{2}$ is irrational" (Logic and Proof, Theorem 2.2.1), about the school rationals, by parity: an argument about even and odd integers. Parity is available here too, from Proposition 1.6.4 (d) for positive integers or from the integer part (Proposition 2.5.5; see Exercise R4). We prove the statement for the constructed $\mathbb{Q}$ by a different route, one that needs only the order and the least-element property of $\mathbb{Z}$, and that adapts to other non-squares (Exercise C2).

Theorem 2.6.1. No Rational Squares to Two

There is no $x \in \mathbb{Q}$ with $x^2 = 2$.

Idea. Work backwards from a supposed square root $x = p/q$. From $x^2 = 2$ we get $2 - x = x^2 - x = x(x-1)$, so

$$x = \frac{2-x}{x-1} = \frac{2-p/q}{p/q-1} = \frac{2q-p}{p-q}.$$

Since $1 < x < 2$ (as $1^2 < 2 < 2^2$), the new denominator $p - q$ is positive and smaller than q. So no fraction for x can have a **least** positive denominator, while Proposition 2.3.6 says some fraction must. The written proof checks the descent with integers only: it never divides, and it verifies $(2q-p)^2 = 2(p-q)^2$ by expanding. *Before reading on: using only integers, why is the new denominator $q' := p - q$ positive, and why is $q' < q$?*

Proof. Suppose, for a contradiction, that $x \in \mathbb{Q}$ and $x^2 = 2$. Then $x \neq 0$, since $0^2 = 0$ (Proposition 2.3.2 (b)) and $2 > 0$. By (O1), $x > 0$ or $x < 0$; let $y := x$ in the first case and $y := -x$ in the second. Then $y > 0$ (Proposition 2.3.5 (a)) and $y^2 = x^2 = 2$, because $(-x)(-x) = xx$ (Proposition 2.3.2 (c)).

By Proposition 2.4.4 (b), $y = a/b$ with $a \in \mathbb{Z}$ and $b \in \mathbb{Z}_{>0}$, and $a > 0$ by Theorem 2.5.1 (a). By Definition 2.4.2, $y^2 = a^2/b^2$, and $2 = 2/1$, so $y^2 = 2$ says $a^2 \cdot 1 = 2b^2$ (Proposition 2.4.4 (b)). Hence the set

$$D := \{q \in \mathbb{Z}_{>0} : p^2 = 2q^2 \text{ for some } p \in \mathbb{Z}_{>0}\}$$

contains b, so it is non-empty, and 0 is a lower bound of it. By Proposition 2.3.6, D has a least element q. Choose $p \in \mathbb{Z}_{>0}$ with $p^2 = 2q^2$.

$q < p$. Suppose $p \leq q$. Since $p, q \geq 0$, Proposition 2.3.5 (g) gives $p^2 < q^2$ if $p < q$, and $p^2 = q^2$ if $p = q$; so $p^2 \leq q^2$, that is, $q^2 + q^2 \leq q^2$, and adding $-q^2$ (O3) gives $q^2 \leq 0$. But $q \neq 0$, so $q^2 > 0$ by Proposition 2.3.5 (e). This contradicts (O1), so $p \leq q$ fails, and $q < p$ by totality (Proposition 2.3.5 (i)).

$p < 2q$. Suppose $2q \leq p$. Since $2q, p \geq 0$, (g) (or equality, if $2q = p$) gives $(2q)^2 \leq p^2$, that is, $2q^2 + 2q^2 = 4q^2 \leq 2q^2$, so $2q^2 \leq 0$. But $2q^2 > 0$ by (O4). As before, $p < 2q$.

Descent. Let $p' := 2q - p$ and $q' := p - q$, integers. By Proposition 2.3.5 (a), $p < 2q$ gives $p' > 0$, and $q < p$ gives $q' > 0$; so $p', q' \in \mathbb{Z}_{>0}$ (Theorem 2.3.4 (b)). Adding $-q$ to $p < 2q$ (O3) gives $q' = p - q < q$. Expanding with Proposition 2.3.2 (c) and (d),

$$\begin{aligned} p'^2 - 2q'^2 &= (4q^2 - 4pq + p^2) - (2p^2 - 4pq + 2q^2) \\ &= 2q^2 - p^2 = 0. \end{aligned}$$

So $p'^2 = 2q'^2$ with $p' \in \mathbb{Z}_{>0}$, that is, $q' \in D$. But $q' < q$, and q is the least element of D. This contradiction shows that no $x \in \mathbb{Q}$ has $x^2 = 2$.

■

The answer to the prompt: $q' = p - q > 0$ because $q < p$, and $q' < q$ because $p < 2q$. These are the two inequalities "$1 < x < 2$" of the Idea, written for integers. The hypothesis $p^2 = 2q^2$ is

used in each of the three steps, and the minimality of q only once, in the final contradiction.

Remark. The proof assumes $x^2 = 2$, reaches a contradiction, and concludes $\neg(\exists x \in \mathbb{Q}\ x^2 = 2)$: its last step is ¬I (Logic and Proof, Definition 2.2.1), not RAA (Logic and Proof, Definition 2.5.1), since what is refuted is the assumption and the conclusion is its negation. The steps "$q < p$" and "$p < 2q$" refute $p \leq q$ and $2q \leq p$ and then read the remaining case off trichotomy.

2.6.2 The set below the gap

Sets, Relations, and Cardinality, Section 2.6 stated, without proof, that the set of positive rationals with square below 2 has no least upper bound in $\mathbb{Q}$, and pointed out that the ambient set decides whether a supremum exists. We now prove it. By Proposition 2.3.5 (i) and Theorem 2.5.1, $\leq$ is a total order on $\mathbb{Q}$, so $(\mathbb{Q}, \leq)$ is a partially ordered set and the words of *Sets, Relations, and Cardinality*, Section 2.6 apply: upper bounds (Sets, Relations, and Cardinality, Definition 2.6.2) and a supremum, an upper bound s with (S1) $x \leq s$ for every x in the set and (S2) $s \leq u$ for every upper bound u (Sets, Relations, and Cardinality, Definition 2.6.5).

Theorem 2.6.2. A Bounded Set of Rationals with No Rational Supremum

Let

$$A \coloneqq \{x \in \mathbb{Q} : x^2 < 2\}, \qquad A^+ \coloneqq \{x \in A : x > 0\} = \{x \in \mathbb{Q} : x > 0,\ x^2 < 2\}.$$

(a) A is non-empty and bounded above in $(\mathbb{Q}, \leq)$, but A has no supremum in $\mathbb{Q}$.

(b) A and A^+ have the same upper bounds in $\mathbb{Q}$. Hence A^+ is also non-empty and bounded above, and has no supremum in $\mathbb{Q}$.

Try it first (ten minutes): a rational $s > 0$ is offered as $\sup A$. Find a rational that proves the offer wrong, once in the case $s^2 < 2$ and once in the case $s^2 > 2$.

Idea. If $s^2 < 2$, we need an element of A above s. The obvious try is $s + h$ with $h > 0$ small, but $(s+h)^2 = s^2 + 2sh + h^2$ has an h^2 in it, and taming that needs a separate size estimate on h. If $s^2 > 2$, we need a smaller upper bound, and a guess $s - h$ runs into the same algebra.

A cleaner idea finds the new rational in one formula. If $s^2 > 2$, then $2/s < s$ and $(2/s)^2 = 4/s^2 < 2$: the gap lies between $2/s$ and s. Their average

$$t(s) \coloneqq \frac{1}{2}\Big(s + \frac{2}{s}\Big) = \frac{s^2 + 2}{2s}$$

is smaller than s, and its square is still above 2; that is the smaller upper bound. If $s^2 < 2$, apply the same trick from the other side: $t(s)$ overshoots, so $u(s) \coloneqq 2/t(s) = 4s/(s^2+2)$ undershoots, and it lands above s. Each claim is an identity in the field $\mathbb{Q}$, checked by clearing denominators. The proof uses Theorem 2.6.1 exactly once, to rule out the third case $s^2 = 2$.

Proof.

(a) *Non-empty:* $1^2 = 1 < 2$, so $1 \in A$. *Bounded above:* 2 is an upper bound. Indeed, let $x \in A$, and suppose $x > 2$. Then $0 \leq 2 < x$, so $4 = 2^2 < x^2$ by Proposition 2.3.5 (g), and $2 < 4$, so $2 < x^2$, against $x^2 < 2$ by (O1). So $x \leq 2$, by totality.
Two maps. For $s \in \mathbb{Q}_{>0}$, we have $2s > 0$ by (O4), and $s^2 + 2 > 0$ by Proposition 2.3.5 (e) and (b); both are non-zero. Let

$$t(s) \coloneqq \frac{s^2+2}{2s}, \qquad u(s) \coloneqq \frac{4s}{s^2+2}.$$

Both are rationals, and both are positive, by Proposition 2.5.2 (a) and (O4). They satisfy

$$\begin{aligned} t(s)^2 - 2 &= \frac{(s^2-2)^2}{4s^2}, & s - t(s) &= \frac{s^2-2}{2s}, \\ u(s) - s &= \frac{s(2-s^2)}{s^2+2}, & 2 - u(s)^2 &= \frac{2(s^2-2)^2}{(s^2+2)^2}. \end{aligned} \tag{2.6.1}$$

Each identity is checked by multiplying both sides by its denominator, which is non-zero, so that the multiplication can be undone with the inverse. For the first, $2s \cdot t(s) = s^2 + 2$, so $4s^2(t(s)^2 - 2) = (s^2+2)^2 - 8s^2 = s^4 - 4s^2 + 4 = (s^2-2)^2$. For the second, $2s(s - t(s)) = 2s^2 - (s^2+2) = s^2 - 2$. For the third, $(s^2+2)(u(s) - s) = 4s - s^3 - 2s = s(2 - s^2)$. For the fourth, $(s^2+2)^2(2 - u(s)^2) = 2(s^2+2)^2 - 16s^2 = 2(s^2-2)^2$.
No supremum. Suppose, for a contradiction, that $s \in \mathbb{Q}$ is a supremum of A. By (S1) and $1 \in A$, $s \geq 1 > 0$, so $s \in \mathbb{Q}_{>0}$. By Theorem 2.6.1, $s^2 \neq 2$, so by (O1), $s^2 < 2$ or $s^2 > 2$. In both cases $s^2 - 2 \neq 0$, so $(s^2-2)^2 > 0$ by Proposition 2.3.5 (e).
Case 1: $s^2 < 2$. Then $2 - s^2 > 0$, and the third identity of Eq. 2.6.1 exhibits $u(s) - s$ as a product of positive elements (with Proposition 2.5.2 (a) for the reciprocal), so $u(s) > s$. The fourth gives $2 - u(s)^2 > 0$, so $u(s)^2 < 2$ and $u(s) \in A$. So A has an element $u(s) > s$, and s is not an upper bound of A, against (S1).
Case 2: $s^2 > 2$. The first identity gives $t(s)^2 - 2 > 0$, so $t(s)^2 > 2$. Then $t(s)$ is an upper bound of A: if some $x \in A$ had $x > t(s)$, then $0 < t(s) < x$ would give $2 < t(s)^2 < x^2$ by Proposition 2.3.5 (g), against $x^2 < 2$; so $x \leq t(s)$ for every $x \in A$. By (S2), $s \leq t(s)$. But the second identity, with $s^2 - 2 > 0$ and $2s > 0$, gives $s - t(s) > 0$, that is, $t(s) < s$. Then $t(s) < s \leq t(s)$, so $t(s) < t(s)$, against (O1).
Both cases are impossible, so A has no supremum in $\mathbb{Q}$.

(b) Since $A^+ \subseteq A$, every upper bound of A is an upper bound of A^+. Conversely, let u be an upper bound of A^+. Since $1 > 0$ and $1^2 < 2$, $1 \in A^+$, so $u \geq 1 > 0$. Let $x \in A$. If $x > 0$, then $x \in A^+$ and $x \leq u$. Otherwise $x \leq 0 < u$ (Proposition 2.3.5 (i)). So u is an upper bound of A.

Conditions (S1) and (S2) for a set mention only the set of its upper bounds: (S1) says that s is one, and (S2) that it lies below all of them. So s is a supremum of A^+ if and only if it is a supremum of A, and by (a), A^+ has none. It is non-empty, since $1 \in A^+$, and bounded above, by 2.

■

The set A^+ is the set of *Sets, Relations, and Cardinality*, Section 2.6, so clause (b) is the statement left unproved there. The two descriptions differ only by the non-positive elements of A, which lie below every upper bound anyway.

Remark. *Analysis on the Real Line*, Chapter 1 proves the same conclusion for A^+ in $\mathbb{Q}$, as part of its motivation for the axioms of the real numbers, and then shows that A^+ does have a supremum in the real numbers. That is the sense of "gap": a bounded set of rationals whose least upper bound, if it is anywhere, is not a rational.

Warning. "A has no supremum in $\mathbb{Q}$, so A is unbounded" mixes up the two conditions a supremum needs. A is bounded above, by 2 for instance; what fails is (S2): among the rational upper bounds there is no least one, because every upper bound s is beaten by the smaller upper bound $t(s)$.

2.6.3 A rational sequence aimed at the gap

The map t does more than refute a candidate. Started at an upper bound, it produces a smaller one, and it can be applied again. Iterating it gives a sequence of rationals that approaches the gap from above.

Example. A Rational Sequence Aimed at the Gap

For $s \in \mathbb{Q}_{>0}$ let $t(s) := (s^2+2)/(2s)$. Define $x_0 := 2$ and $x_{n+1} := t(x_n)$ for $n \in \mathbb{N}$, and let $e_n := x_n^2 - 2$. Show that this defines a sequence with $x_n \in \mathbb{Q}_{>0}$ for every $n \in \mathbb{N}$, and that for every $n \in \mathbb{N}$:

(i) $e_n > 0$, that is, $x_n^2 > 2$;
(ii) $e_{n+1} = e_n^2/(4x_n^2) \le e_n^2/8$;
(iii) $x_n - x_{n+1} = e_n/(2x_n) > 0$, so $x_{n+1} < x_n$.

Compute $x_0, \dots, x_3$ and $e_0, \dots, e_3$.

Solution. *The sequence.* For $s \in \mathbb{Q}_{>0}$, $s^2+2 > 0$ and $2s > 0$, so $t(s) \in \mathbb{Q}_{>0}$ by (O4) and Proposition 2.5.2 (a): t maps $\mathbb{Q}_{>0}$ to $\mathbb{Q}_{>0}$. The recursion theorem (Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ as Theorem 1.5.1 by Corollary 1.7.3 (b)), with $X = \mathbb{Q}_{>0}$, starting value 2 and $g(n,s) := t(s)$, gives exactly one function $n \mapsto x_n$ from $\mathbb{N}$ to $\mathbb{Q}_{>0}$ with $x_0 = 2$ and $x_{n+1} = t(x_n)$. So every x_n is a positive rational, with no further argument.

Two identities. For $s \in \mathbb{Q}_{>0}$, $2s \cdot t(s) = s^2+2$. Hence $4s^2(t(s)^2-2) = (s^2+2)^2 - 8s^2 = (s^2-2)^2$ and $2s(s-t(s)) = 2s^2 - s^2 - 2 = s^2 - 2$. Multiplying by the inverses of $4s^2 \neq 0$ and $2s \neq 0$,

$$t(s)^2 - 2 = \frac{(s^2-2)^2}{4s^2}, \qquad s - t(s) = \frac{s^2-2}{2s}. \tag{2.6.2}$$

(i). We use Logic and Proof, Axiom 4.2.1 (read on ω as Theorem 1.3.3 (b), by Corollary 1.7.3 (b)) for $P(n)$: $e_n > 0$. Base $P(0)$: $e_0 = 2^2 - 2 = 2 > 0$. Step: suppose $e_n > 0$. By Eq. 2.6.2 with $s = x_n$,

$$e_{n+1} = t(x_n)^2 - 2 = \frac{e_n^2}{4x_n^2}.$$

Here $e_n^2 > 0$ by Proposition 2.3.5 (e), since $e_n \neq 0$, and $1/(4x_n^2) > 0$ by Proposition 2.5.2 (a); so $e_{n+1} > 0$ by (O4). This proves (i) for every n.

(ii). The equation was just shown, for every n. By (i), $x_n^2 > 2$, so $4x_n^2 > 4 \cdot 2 = 8$ by Proposition 2.3.5 (c). From $0 < 8 < 4x_n^2$, Proposition 2.5.2 (b) gives $1/(4x_n^2) < 1/8$, and multiplying by $e_n^2 > 0$ (Proposition 2.3.5 (c)) gives $e_{n+1} < e_n^2/8$; in particular $e_{n+1} \leq e_n^2/8$.

(iii). By Eq. 2.6.2 with $s = x_n$, $x_n - x_{n+1} = e_n/(2x_n)$, a product of the positive elements e_n and $1/(2x_n)$; so it is positive, and $x_{n+1} < x_n$ by Proposition 2.3.5 (a).

The first terms. By the definition of t,

$$x_1 = \frac{4+2}{4} = \frac{3}{2}, \quad x_2 = \frac{9/4+2}{3} = \frac{17}{12}, \quad x_3 = \frac{289/144+2}{17/6} = \frac{577}{408}.$$

The errors are $e_0 = 2$, $e_1 = 9/4 - 2 = 1/4$, $e_2 = 289/144 - 2 = 1/144$ and $e_3 = 577^2/408^2 - 2 = 1/166464$, since $577^2 = 332929 = 2 \cdot 166464 + 1$ and $408^2 = 166464$. As a check on (ii): $e_1 = 2^2/(4 \cdot 4) = 1/4$, $e_2 = (1/16)/9 = 1/144$, and $e_3 = (1/144)^2/(4 \cdot 289/144) = 1/(144 \cdot 1156) = 1/166464$.

For orientation only (decimals are not part of this volume's toolkit), here are the terms in school decimals:

n	x_n	$e_n = x_n^2 - 2$	x_n, roughly
0	2	2	2
1	3/2	1/4	1.5
2	17/12	1/144	1.41667
3	577/408	1/166464	1.414216

The terms decrease, every one of them is an upper bound of A (a positive rational whose square exceeds 2, as in Case 2 of the proof above), and by (ii) the error is roughly squared at each step: after three steps, $e_3 = 1/166464$ is already below $1/100000$. The terms crowd together. Yet no rational L can be "the number they approach", for if they approached L, their squares would approach L^2, and the errors e_n force $L^2 = 2$, which Theorem 2.6.1 forbids. This paragraph is a picture, not a proof: limits in $\mathbb{Q}$ have not been defined. A later chapter of this volume defines them, with $\varepsilon \in \mathbb{Q}_{>0}$, and proves that this sequence is Cauchy in $\mathbb{Q}$ yet has no limit in $\mathbb{Q}$, citing (i)–(iii) from this example.

Quick check. Is $3/2$ an upper bound of $A = \{x \in \mathbb{Q} : x^2 < 2\}$? Is $17/12$? Is $4/3$?

Solution. The first two are. Each is positive, and $(3/2)^2 = 9/4 > 2$ because $9 > 8$, while $(17/12)^2 = 289/144 > 2$ because $289 > 288$ (the comparison formula, Theorem 2.5.1 (a)). A positive rational b with $b^2 > 2$ bounds A: if $x \in A$ had $x > b$, then $x^2 > b^2 > 2$ by Proposition 2.3.5 (g), against $x^2 < 2$. (These are x_1 and x_2 of Example (A Rational Sequence Aimed at the Gap), and $17/12 = t(3/2)$ is the smaller upper bound that Case 2 of the proof of Theorem 2.6.2 produces from $s = 3/2$.)

But $4/3$ is not, although it is positive: $(4/3)^2 = 16/9 < 2$ because $16 < 18$, so the hypothesis $b^2 > 2$ fails, and indeed $7/5 \in A$, as $(7/5)^2 = 49/25 < 2$ ($49 < 50$), while $7/5 > 4/3$

$(21 > 20)$.

This chapter built two number systems by one method. From $\mathbb{N}$ to $\mathbb{Z}$: pairs (a, b), identified when $a + d = b + c$, so that subtraction becomes possible. From $\mathbb{Z}$ to $\mathbb{Q}$: pairs (p, q) with $q > 0$, identified when $ps = rq$, so that division becomes possible. Each time an operation was defined on representatives and checked to be well defined (Theorem 2.2.1), and each axiom was reduced to the level below. The third step, from $\mathbb{Q}$ to the real numbers, takes the same route in a later chapter of this volume: the representatives are sequences of rationals like the one above, identified when their difference tends to 0, so that the gap found in this section, and every other gap of $\mathbb{Q}$, is filled.

2.6.4 Exercises

2.6.4.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 2.6.2.

(a) Where is Theorem 2.6.1 used, and what would go wrong without it?

(b) Run the proof on $s = 3/2$ and on $s = 7/5$: which case applies, which rational does the proof produce, and why does it refute s?

(c) Let F be an ordered field containing an element $r > 0$ with $r^2 = 2$, and let $A_F := \{x \in F : x^2 < 2\}$. Which step of the proof fails for A_F, and what happens to $t(r)$ and $u(r)$?

(d) State the idea of the proof in two sentences.

Solution.

(a) Once, right after a supposed supremum s is fixed: it excludes $s^2 = 2$, so that trichotomy leaves the two cases $s^2 < 2$ and $s^2 > 2$. Without it there is a third case, in which neither map helps: $(s^2 - 2)^2 = 0$, so the identities give $u(s) = s$ and $t(s) = s$.

(b) For $s = 3/2$: $s^2 = 9/4 > 2$ (as $9 > 8$), so Case 2. The proof produces $t(3/2) = (9/4 + 2)/3 = 17/12$, an upper bound of A, since $(17/12)^2 = 289/144 > 2$; and $17/12 < 3/2$, since $17 \cdot 2 = 34 < 36 = 3 \cdot 12$. So $3/2$ is not the **least** upper bound. For $s = 7/5$: $s^2 = 49/25 < 2$ (as $49 < 50$), so Case 1. The proof produces

$$u(7/5) = \frac{28/5}{49/25 + 2} = \frac{28/5}{99/25} = \frac{140}{99},$$

which lies in A, since $140^2 = 19600 < 19602 = 2 \cdot 99^2$, and exceeds $7/5$, since $140 \cdot 5 = 700 > 693 = 7 \cdot 99$. So $7/5$ is not an upper bound.

(c) The step "$s^2 \neq 2$": in F the candidate $s = r$ satisfies $r^2 = 2$, and it falls in neither case. Indeed r is the supremum of A_F in F: it is an upper bound (if $x > r > 0$, then $x^2 > 2$), and an upper bound $v < r$ would have $v > 0$ (as $1 \in A_F$) and $v^2 < 2$, and then $u(v) \in A_F$ with $u(v) > v$, as in Case 1. The maps fix r: $t(r) = (2 + 2)/(2r) = 2/r = r$, since $r \cdot r = 2$, and $u(r) = 4r/4 = r$.

(d) Suppose a rational s were the supremum; since no rational squares to 2, either $s^2 < 2$ or $s^2 > 2$. The explicit rationals $u(s)$ and $t(s)$ then give, respectively, an element of the set above s or an upper bound below s, so s fails (S1) or (S2).

2.6.4.2 B. Practice

Exercise. B1: Put the proof in order

Here are lines of a proof of Theorem 2.6.1, shuffled, together with two lines that do not belong. Write the lines of a correct proof in a correct order, and say what is wrong with each line you leave out.

(i) Set $p' := 2q - p$ and $q' := p - q$; both are positive integers, by the two inequalities between p and q.

(ii) Write $x = a/b$ with $a, b \in \mathbb{Z}_{>0}$; then $a^2 = 2b^2$.

(iii) Since $p^2 = 2q^2$, we have $p < q$.

(iv) Since $q' = p - q < q$, this contradicts the choice of q; so no rational squares to 2.

(v) Suppose $x \in \mathbb{Q}$ has $x^2 = 2$; replacing x by $-x$ if necessary, take $x > 0$.

(vi) $p < 2q$, since $p \geq 2q$ would give $p^2 \geq 4q^2 > 2q^2$.

(vii) Choose such an x with the least positive value.

(viii) Among all pairs (p, q) of positive integers with $p^2 = 2q^2$, choose one with q least (Proposition 2.3.6; b is such a q).

(ix) Expanding, $p'^2 - 2q'^2 = 2q^2 - p^2 = 0$, so $p'^2 = 2q'^2$.

(x) $q < p$, since $p \leq q$ would give $2q^2 = p^2 \leq q^2$, that is, $q^2 \leq 0$.

Solution. An admissible order is

$$\text{(v)},\ \text{(ii)},\ \text{(viii)},\ \text{(x)},\ \text{(vi)},\ \text{(i)},\ \text{(ix)},\ \text{(iv)}.$$

Lines (x) and (vi) may be exchanged, since neither uses the other. Otherwise the order is forced by what each line uses: (ii) needs the $x > 0$ of (v); (viii) needs the pair (a, b) of (ii), which makes the set of such q non-empty; (x) and (vi) are about the pair fixed in (viii), and if they came before it, the pair might change afterwards; (i) needs both inequalities; (ix) needs the definitions in (i); and (iv) needs $q' \in \mathbb{Z}_{>0}$ with $p'^2 = 2q'^2$, from (i) and (ix), and $q' < q$, which comes from (vi).

Line (vii) is legal but does no work. There is only one positive rational x with $x^2 = 2$: if $0 < x < y$ had $x^2 = y^2 = 2$, then Proposition 2.3.5 (g) would give $2 = x^2 < y^2 = 2$, against (O1). So "the least" just picks out the x of (v) and gives nothing new. The descent does not produce a smaller x: it keeps the number and shrinks its denominator, $p'/q' = p/q$, so choosing x least never leads to a contradiction. What must be least is the denominator q, as in (viii).

Line (iii) fails. Its conclusion is false: from $p^2 = 2q^2$ and $q > 0$ we get $p^2 > q^2$, hence $q < p$ by Proposition 2.3.5 (g), the opposite of (iii). It also clashes with (x).

Exercise. B2

Let $A := \{x \in \mathbb{Q} : x^2 < 2\}$ and $A^+ := \{x \in \mathbb{Q} : x > 0,\ x^2 < 2\}$. Without looking back, prove that A and A^+ have the same upper bounds in $\mathbb{Q}$, and deduce from Theorem 2.6.2 (a) that A^+ has no supremum in $\mathbb{Q}$.

Solution. ($\subseteq$) for upper bounds: if u bounds A, it bounds the subset $A^+ \subseteq A$. Conversely, let u bound A^+. Since $1 > 0$ and $1^2 = 1 < 2$, $1 \in A^+$, so $1 \leq u$ and $u > 0$. Let $x \in A$. If $x > 0$, then $x \in A^+$, so $x \leq u$. If $x \leq 0$, then $x \leq 0 < u$. So u bounds A.

Now suppose s were a supremum of A^+. Then s is an upper bound of A^+, hence of A (S1), and $s \leq u$ for every upper bound u of A^+, which are exactly the upper bounds of A (S2). So s would be a supremum of A, which Theorem 2.6.2 (a) excludes. Hence A^+ has no supremum in $\mathbb{Q}$.

Exercise. B3

Compute x_4 of Example (A Rational Sequence Aimed at the Gap), as a fraction of integers, and $e_4 = x_4^2 - 2$. Check your value of e_4 against (ii) of the example. Hence x_4 is an upper bound of $\{x \in \mathbb{Q} : x^2 < 2\}$ below x_3.

Solution. With $x_3 = 577/408$,

$$x_4 = t(x_3) = \frac{577^2 + 2 \cdot 408^2}{2 \cdot 577 \cdot 408} = \frac{332929 + 332928}{470832} = \frac{665857}{470832}.$$

Then $665857^2 = 443365544449$ and $470832^2 = 221682772224$, and $443365544449 - 2 \cdot 221682772224 = 1$, so

$$e_4 = \frac{1}{221682772224}.$$

By (ii), $e_4 = e_3^2/(4x_3^2) = 1/(166464 \cdot 4 \cdot 332929)$, and $4 \cdot 166464 \cdot 332929 = 221682772224$, which agrees. Since $x_4 > 0$ and $e_4 > 0$, x_4 bounds the set (the argument of the Quick check), and $x_4 < x_3$ by (iii).

2.6.4.3 C. Going deeper

Exercise. C1: Exit check: the gap at $\sqrt{2}$

Prove that no rational number squares to 2, and that the set $\{x \in \mathbb{Q} : x^2 < 2\}$ has no supremum in $\mathbb{Q}$.

Solution. No rational squares to 2. Suppose $x \in \mathbb{Q}$ with $x^2 = 2$. Then $x \neq 0$, and replacing x by $-x$ if needed ($(-x)^2 = x^2$), $x > 0$. Write $x = a/b$ with $a, b \in \mathbb{Z}_{>0}$ (Proposition 2.4.4 (b), Theorem 2.5.1 (a)); then $a^2 = 2b^2$. Among all $q \in \mathbb{Z}_{>0}$ for which some $p \in \mathbb{Z}_{>0}$ has $p^2 = 2q^2$, take the least (Proposition 2.3.6; the set contains b and is bounded below by 0), with its p. Then $q < p$ (else $p^2 \leq q^2$ by Proposition 2.3.5 (g), so $q^2 \leq 0$, against $q^2 > 0$) and $p < 2q$ (else $4q^2 \leq p^2 = 2q^2$, so $2q^2 \leq 0$). So $p' := 2q - p$ and $q' := p - q$ are positive integers with $q' < q$, and $p'^2 - 2q'^2 = 2q^2 - p^2 = 0$ by expanding. This contradicts the choice of q.

Bounded above. 2 bounds the set A: $x > 2$ would give $x^2 > 4 > 2$ (Proposition 2.3.5 (g)). And $1 \in A$.

No supremum. Suppose $s \in \mathbb{Q}$ is a supremum of A. As $1 \in A$, $s \geq 1 > 0$, and $s^2 \neq 2$ by the first part.

If $s^2 < 2$, let $u := 4s/(s^2+2) \in \mathbb{Q}_{>0}$. Clearing denominators, $u - s = s(2-s^2)/(s^2+2) > 0$ and $2 - u^2 = 2(s^2-2)^2/(s^2+2)^2 > 0$. So $u \in A$ and $u > s$: s is not an upper bound.

If $s^2 > 2$, let $t := (s^2+2)/(2s) \in \mathbb{Q}_{>0}$. Clearing denominators, $t^2 - 2 = (s^2-2)^2/(4s^2) > 0$ and $s - t = (s^2-2)/(2s) > 0$. So $t^2 > 2$, and t bounds A (an $x \in A$ with $x > t > 0$ would have $x^2 > t^2 > 2$); and $t < s$. So s is not the least upper bound.

By trichotomy one of the two cases holds, and each contradicts the choice of s. So A has no supremum in $\mathbb{Q}$.

Marking scheme. (10 points) - 4: no rational squares to 2: sets up a least denominator, or any other correct route (1); proves the key inequalities $q < p < 2q$, or the parity step of another route (2); reaches the contradiction (1). - 1: shows that A is bounded above, with an explicit upper bound, and non-empty. - 2: the case $s^2 < 2$: an explicit element of A larger than s, with both properties verified. - 2: the case $s^2 > 2$: an explicit smaller rational, verified to be an upper bound of A. - 1: concludes by trichotomy, with $s^2 = 2$ excluded by the first part. Fatal errors (0 for the step): assuming that $\sqrt{2}$ exists (in $\mathbb{Q}$ or in a larger field) and comparing s with it; "$s^2 < 2$, so $s + \varepsilon \in A$ for small ε" with no ε exhibited and checked. A second route for the first part, by parity (every integer is $2m$ or $2m+1$, with $m = \lfloor a/2 \rfloor$ from Proposition 2.5.5, and not both), earns the same 4 points: 1 for the least denominator, 2 for "p even, then q even", 1 for the contradiction.

Exercise. C2: Which integers have a rational square root?

For which $n \in \mathbb{Z}_{>0}$ does the set $A_n := \{x \in \mathbb{Q} : x^2 < n\}$ have a supremum in $\mathbb{Q}$? Prove your answer.

Solution (Hint 1). First find $k \in \mathbb{N}$ with $k^2 \le n < (k+1)^2$, and treat $n = k^2$ on its own. For $k^2 < n$, adapt both proofs of this section: the descent, and then the two maps.

Solution (Hint 2). Let $T := \{m \in \mathbb{Z} : m \ge 0,\ n < m^2\}$. It contains $n+1$, it is bounded below by 0, and $0 \notin T$; so its least element is $k+1$ for some $k \in \mathbb{N}$. In the descent for $k^2 < n$, try $q' := p - kq$.

Solution. **Answer:** exactly when $n = k^2$ for some $k \in \mathbb{N}$, and then the supremum is k.

Finding k. Let $T := \{m \in \mathbb{Z} : m \ge 0,\ n < m^2\}$. Since $n+1 \ge 1$, $(n+1)^2 \ge n+1 > n$ (Proposition 2.3.5 (c)), so $n+1 \in T$; 0 is a lower bound of T; and $0 \notin T$, as $0^2 = 0 < n$. By Proposition 2.3.6, T has a least element l. Since $0 \notin T$ and $l \ge 0$, we have $l > 0$, so $l \ge 1$ by discreteness (Theorem 2.3.4 (c)); then $l - 1 \ge 0$ by (O3), so $k := l - 1 \in \mathbb{N}$ by Theorem 2.3.4 (b). Then $n < (k+1)^2$, and $k \notin T$ with $k \ge 0$ gives $k^2 \le n$. So either $n = k^2$ or $k^2 < n < (k+1)^2$.

Case $n = k^2$. Here $k \ge 1$, since $n \ge 1$. We show $k = \sup A_n$. (S1): if $x \in A_n$ had $x > k$, then $x > k \ge 0$ would give $x^2 > k^2 = n$ (Proposition 2.3.5 (g)), against $x^2 < n$; so $x \le k$. (S2): let v be an upper bound of A_n, and suppose $v < k$. Let $w := v$ if $v \ge 0$ and $w := 0$ otherwise; then $v \le w$ and $0 \le w < k$. Let c be the midpoint of w and k (Section 2.5):

$w < c < k$. Then $c > 0$ and $c^2 < k^2 = n$ by (g), so $c \in A_n$, and $c > w \geq v$, so v is not an upper bound. Hence $k \leq v$ for every upper bound v, and k is the supremum.

Case $k^2 < n < (k+1)^2$. Then $n \geq 2$, since $n = 1 = 1^2$ is the other case; so $1 \in A_n$.

(i) *No rational squares to* n. Suppose $x^2 = n$ with $x \in \mathbb{Q}$; as in Theorem 2.6.1, we may take $x > 0$, and then $p^2 = nq^2$ for some $p, q \in \mathbb{Z}_{>0}$, with q least among all such denominators (Proposition 2.3.6). Multiplying $k^2 < n < (k+1)^2$ by $q^2 > 0$ gives $(kq)^2 < p^2 < ((k+1)q)^2$, so $kq < p < (k+1)q$ by (g). Let

$$p' \coloneqq nq - kp, \qquad q' \coloneqq p - kq.$$

Then $0 < q' < q$. For p', expand:

$$q \cdot p' = nq^2 - kpq = p^2 - kpq = p(p - kq) > 0$$

by (O4); since $q > 0$, $p' \leq 0$ would give $qp' \leq 0$ (Proposition 2.3.5 (c)), so $p' > 0$. (The bounds alone do not give this: for $n = 5$, $k = 2$, $q = 10$, $p = 28$, we have $kq < p < (k+1)q$, but $nq - kp = -6$. It is $p^2 = nq^2$ that forces the sign.) Finally, expanding,

$$p'^2 - nq'^2 = (k^2 - n)(p^2 - nq^2) = 0,$$

so q' is a smaller admissible denominator, a contradiction.

(ii) *No supremum.* The proof of Theorem 2.6.2 (a) goes through with n in place of 2, using (i) where it used Theorem 2.6.1. The set A_n contains 1 and is bounded above by n (if $x > n \geq 1$, then $x^2 > n^2 \geq n$). For $s \in \mathbb{Q}_{>0}$, let

$$t(s) \coloneqq \frac{s^2 + n}{2s}, \qquad u(s) \coloneqq \frac{2ns}{s^2 + n}.$$

Clearing denominators as before,

$$t(s)^2 - n = \frac{(s^2 - n)^2}{4s^2}, \qquad s - t(s) = \frac{s^2 - n}{2s},$$
$$u(s) - s = \frac{s(n - s^2)}{s^2 + n}, \quad n - u(s)^2 = \frac{n(s^2 - n)^2}{(s^2 + n)^2}.$$

A supremum s would satisfy $s \geq 1 > 0$ and $s^2 \neq n$ by (i). If $s^2 < n$, then $u(s) \in A_n$ and $u(s) > s$; if $s^2 > n$, then $t(s)$ is an upper bound of A_n with $t(s) < s$. Either way s is not the supremum.

So A_n has a supremum in $\mathbb{Q}$ exactly when n is the square of a natural number.

Marking scheme. (10 points) - 2: produces $k \in \mathbb{N}$ with $k^2 \leq n < (k+1)^2$, from a least element. - 2: the case $n = k^2$: checks (S1) and (S2) for k. - 3: the case $k^2 < n$: no rational squares to n, by a descent with a least denominator (or another correct route), including a proof that the new numerator is positive. - 2: the two maps (or any explicit rationals) and the two cases, as in the proof for 2. - 1: states the answer: exactly the squares of natural numbers. Fatal errors (0 for the step): using a square root of n, or the floor of one, as if it existed; deducing $nq - kp > 0$ from $kq < p < (k+1)q$ alone.

2.6.4.4 R. Review

Exercise. R1

Prove, by induction on m, that for all $m, n \in \mathbb{N}$: if $m + n = m$, then $n = 0$. Use only the equations of addition ($m+0 = m$, $m+(n+1) = (m+n)+1$, $0+n = n$, $(m+1)+n = (m+n)+1$) and the fact that $k + 1 = l + 1$ implies $k = l$. Do not use cancellation or any fact about the order: cancellation gives the claim in one line ($m + n = m + 0$), and it is itself proved from the facts this exercise asks you to use.

Solution. We use Theorem 1.3.3 (b) (the form of Logic and Proof, Axiom 4.2.1 on $\omega = \mathbb{N}$, by Corollary 1.7.3 (b)) on m, for the property $P(m)$: "for every $n \in \mathbb{N}$, if $m + n = m$, then $n = 0$". Keeping "for every n" inside $P(m)$ lets the step use the hypothesis for the same n.

Base $P(0)$. Let $0 + n = 0$. Since $0 + n = n$ (Theorem 1.5.3 (a)), $n = 0$.

Step. Assume $P(m)$, and let $(m + 1) + n = m + 1$. By Theorem 1.5.3 (a), $(m + 1) + n = (m + n) + 1$, so $(m + n) + 1 = m + 1$. By Corollary 1.5.2, $m + n = m$. By $P(m)$, $n = 0$. This proves $P(m + 1)$.

By induction, $P(m)$ holds for every $m \in \mathbb{N}$.

Cancellation (Proposition 1.6.3 (b)) and the order facts are deliberately not used: each gives a one-line proof, and cancellation is itself proved by peeling off one successor at a time with Corollary 1.5.2, the move of the step above.

Exercise. R2

Show that the set $\mathbb{Q}$, as constructed in this chapter, is countable.

Solution. The tools are the results on countable sets of *Sets, Relations, and Cardinality*, Chapter 3, read on $\mathbb{N} = \omega$ through the dictionary of Section 1.7 (Corollary 1.7.3 and the table after it).

$\mathbb{Z}$ is countable. $\mathbb{N} \times \mathbb{N}$ is countable (Sets, Relations, and Cardinality, Theorem 3.2.6). The map $a \mapsto [a]$, from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/\!\sim$, is surjective, since every class is the class of each of its elements (Sets, Relations, and Cardinality, Definition 2.5.4). By Sets, Relations, and Cardinality, Proposition 3.2.4, $\mathbb{Z}$ is countable.

$\mathbb{Z} \times \mathbb{Z}_{>0}$ is countable. $\mathbb{Z}_{>0} \subseteq \mathbb{Z}$ is countable by Sets, Relations, and Cardinality, Proposition 3.2.3, and the product of two countable sets is countable (Sets, Relations, and Cardinality, Corollary 3.2.7 (a)).

$\mathbb{Q}$ is countable. The map $(p, q) \mapsto [(p, q)]$, from $\mathbb{Z} \times \mathbb{Z}_{>0}$ to $\mathbb{Q}$, is surjective (Sets, Relations, and Cardinality, Definition 2.5.4), so $\mathbb{Q}$ is countable by Sets, Relations, and Cardinality, Proposition 3.2.4.

(Section 1.7 lets a review exercise cite Sets, Relations, and Cardinality, Theorem 3.2.6 and Sets, Relations, and Cardinality, Corollary 3.2.7 (a), which run on ω through Proposition 1.6.4.) The conclusion is Sets, Relations, and Cardinality, Theorem 3.2.8 for the constructed $\mathbb{Q}$; that theorem, about the school rationals, is not used.

Exercise. R3

Let $A := \{x \in \mathbb{Q} : x^2 < 2\}$ and $s \in \mathbb{Q}$. Write "s is a supremum of A in $\mathbb{Q}$" in symbols, with bounded quantifiers. Then write its negation, moving $\neg$ inward as far as it goes and naming the law used at each step. The proof that A has no supremum in $\mathbb{Q}$ has two cases; which half of the negation does each case establish, and with which witness?

Solution. By Sets, Relations, and Cardinality, Definition 2.6.5, "s is a supremum of A" is the conjunction of (S1) and (S2):

$$(\forall x \in A \;\; x \le s) \;\wedge\; \forall v \in \mathbb{Q}\; ((\forall x \in A \;\; x \le v) \to s \le v).$$

Its negation is, by Logic and Proof, Proposition 2.5.5 (d),

$$\neg(\forall x \in A \;\; x \le s) \;\vee\; \neg\forall v \in \mathbb{Q}\; ((\forall x \in A \;\; x \le v) \to s \le v).$$

By Logic and Proof, Proposition 6.2.3 (a), the first disjunct becomes $\exists x \in A\; \neg(x \le s)$, that is, $\exists x \in A \;\; s < x$ (totality of $\le$, Proposition 2.3.5 (i)). By Logic and Proof, Proposition 6.2.3 (a) and then Logic and Proof, Proposition 6.2.2, the second becomes

$$\exists v \in \mathbb{Q}\; ((\forall x \in A \;\; x \le v) \wedge \neg(s \le v)),$$

that is, there is an upper bound v of A with $v < s$. In words: s fails to be the supremum when **some element of A lies above s**, or **some upper bound lies below s**.

In the proof of Theorem 2.6.2, Case 1 ($s^2 < 2$) establishes the first disjunct, with witness $x = u(s) = 4s/(s^2+2)$. Case 2 ($s^2 > 2$) establishes the second, with witness $v = t(s) = (s^2+2)/(2s)$. (The proof itself needs no negation law: it derives a contradiction from (S1) in Case 1 and from (S2) in Case 2.)

Exercise. R4

Is there a rational number x with $x^3 = 2$, where $x^3 := x^2 \cdot x$? Prove your answer.

Solution. **No.** The tool is descent on the denominator, as in Theorem 2.6.1, but run with parity, which Proposition 2.5.5 supplies.

Parity. For every $a \in \mathbb{Z}$ there is $m \in \mathbb{Z}$ with $a = 2m$ or $a = 2m+1$, and not both forms occur. Indeed, let $m := \lfloor a/2 \rfloor$ (Proposition 2.5.5): $m \le a/2 < m+1$, so $2m \le a < 2m+2$ (Proposition 2.3.5 (c)). The integer $r := a - 2m$ has $0 \le r < 2$; if $r > 0$, then $r \ge 1$, and if $r > 1$, then $r \ge 2$, by discreteness (Theorem 2.3.4 (c)); so $r = 0$ or $r = 1$. If $2m = 2m' + 1$, then $d := m - m'$ has $2d = 1$; but $d \le 0$ gives $2d \le 0$, and $d \ge 1$ gives $2d \ge 2$, and by discreteness one of these holds. Moreover an odd cube is odd:

$$(2m+1)^3 = 2(4m^3 + 6m^2 + 3m) + 1.$$

Descent. Suppose $x \in \mathbb{Q}$ and $x^3 = 2$. Then $x > 0$: $x = 0$ gives $x^3 = 0$, and $x < 0$ gives $x^2 > 0$ and $x^3 = x^2 \cdot x < 0$ (Proposition 2.3.5 (e), (c)). So $x = p/q$ with $p, q \in \mathbb{Z}_{>0}$ and $p^3 = 2q^3$ (Proposition 2.4.4 (b), Definition 2.4.2). Take q least among such denominators

(Proposition 2.3.6), with its p. Since $p^3 = 2q^3$ has the even form, p is not odd (its cube would be odd), so $p = 2m$ with $m \in \mathbb{Z}$, and $m > 0$ because $p > 0$ ($m \leq 0$ would give $2m \leq 0$). Then $2q^3 = 8m^3$, and canceling the non-zero factor 2 (Proposition 2.3.5 (f)), $q^3 = 2(2m^3)$. So q is not odd either: $q = 2l$ with $l \in \mathbb{Z}_{>0}$. Then $8m^3 = 2 \cdot 8l^3$, so $m^3 = 2l^3$, and $l < l + l = q$. So l is a smaller admissible denominator, with $m \in \mathbb{Z}_{>0}$: a contradiction.

Why the look-alike fails. The substitution $p' \coloneqq 2q - p$, $q' \coloneqq p - q$ of the square-root proof came from $x = (2 - x)/(x - 1)$, which used $x^2 = 2$. For cubes the same algebra fails: expanding,

$$(2q - p)^3 - 2(p - q)^3 = -3p^3 + 12p^2q - 18pq^2 + 10q^3,$$

which is not a constant multiple of $p^3 - 2q^3$ (it has a p^2q term), so this identity gives no way to derive $p'^3 = 2q'^3$ from $p^3 = 2q^3$, and the descent has no smaller pair to move to. Parity, by contrast, works for every power: "even cube, so even base" is the same argument for squares and cubes.

Chapter 3

Divisibility, Division with Remainder, and Modular Arithmetic

Chapter 2 built $\mathbb{Z}$ out of pairs of natural numbers and proved it an ordered ring, but it said nothing about which integers divide which. *Sets, Relations, and Cardinality*, Section 2.2 did, for the school integers: it proved division with remainder, showed that every integer from 2 on has a prime factor, and set the uniqueness of prime factorization aside as a fact it neither proved nor used. This chapter proves that arithmetic again for the constructed $\mathbb{Z}$, and goes further, using the ordered-ring axioms and two properties of the order of $\mathbb{Z}$: there is no integer strictly between 0 and 1 (Theorem 2.3.4 (c)), and a non-empty set of integers with a lower bound has a least element (Proposition 2.3.6); the field $\mathbb{Q}$ enters chiefly for the integer part $\lfloor a/n \rfloor$ and for lowest terms. The second property gives division with remainder at once. Then one move does most of the work. In a set of integers closed under sums and under multiplication by any integer, the **least positive element** divides every element, and dividing by it with remainder is how one sees this. Applied to the combinations $ax + by$, it makes the greatest common divisor a combination of a and b (Bézout's theorem), and the Euclidean algorithm computes it. Multiplying that combination by a third integer gives Euclid's lemma, that a prime dividing a product divides one of the factors, and with it the uniqueness of prime factorization and the lowest terms of a rational number.

The second half returns to the clock of *Sets, Relations, and Cardinality*, Section 2.5. There, and again for $\mathbb{Z}$ and $\mathbb{Q}$ in Chapter 2, each sum and product of classes was checked separately to be well defined. Section 3.4 makes the check once, for the quotient of any commutative ring by an ideal, and $\mathbb{Z}/n\mathbb{Z}$ becomes a commutative ring with n elements on which, for $n \geq 2$, no order fits. In it the classes that can be divided by are exactly the classes of the integers coprime to n. So $\mathbb{Z}/p\mathbb{Z}$ is a field with p elements for every prime p, and two congruences with coprime moduli have exactly one class of solutions modulo the product of the moduli, the Chinese remainder theorem. Chapter 5 builds $\mathbb{R}$ in the same shape, as classes of rational Cauchy sequences whose difference is null (the null sequences play the part of the ideal), though it checks its operations directly; *Algebraic Preliminaries*, Chapter 5 runs division with remainder, gcd and unique factorization again for polynomials over a field.

Two integers, one theory. *Sets, Relations, and Cardinality*, Chapter 2 proved its facts about divisibility for the school integers, from school facts such as the closure of $\mathbb{Z}$ under $+$, $-$ and $\cdot$ (Sets, Relations, and Cardinality, Proposition 2.1.2). This chapter keeps their definitions of divisibility, prime and congruence (Sets, Relations, and Cardinality, Definition 2.2.1, Sets,

Relations, and Cardinality, Definition 2.2.2, Sets, Relations, and Cardinality, Definition 2.5.5), read word for word on the ring of Chapter 2, and proves the results again; no proof in it cites a result of *Sets, Relations, and Cardinality* about integers. Each such result has a twin here, named in the section that proves it: division with remainder (Sets, Relations, and Cardinality, Theorem 2.2.3) and the antisymmetry of divisibility on $\mathbb{N}$ (Sets, Relations, and Cardinality, Proposition 2.4.1) in Section 3.1; a prime factor of every $n \geq 2$ (Sets, Relations, and Cardinality, Theorem 2.2.5) and lowest terms (Sets, Relations, and Cardinality, Lemma 2.2.4) in Section 3.3; and the arithmetic of congruence classes (Sets, Relations, and Cardinality, Proposition 2.5.5) in Section 3.4. Nothing earlier was circular: the earlier chapters used these facts about the school integers as stated facts, and this chapter proves them for the constructed ones.

What you need. Chapters 1 and 2, chiefly Section 2.3, and *Sets, Relations, and Cardinality*, Chapters 1–3, chiefly Sections 2.2 and 2.5. From Chapter 2: the commutative ring $\mathbb{Z}$ and its algebra (Definition 2.3.1, Theorem 2.3.1, Proposition 2.3.2), its order, its positive elements, the rules for inequalities and discreteness (Proposition 2.3.3, Proposition 2.3.5, Theorem 2.3.4), the least element of a set of integers bounded below (Proposition 2.3.6), the copy of $\mathbb{N}$ inside $\mathbb{Z}$ (Proposition 2.2.3), the lemma that a rule on classes is a function when equivalent inputs give equivalent outputs (Theorem 2.2.1), the two-element field that cannot be ordered (Example (A field with two elements that cannot be ordered)), and the field $\mathbb{Q}$ with $\mathbb{Z}$ inside it, its order, reciprocals and integer part (Definition 2.4.2, Theorem 2.4.3, Proposition 2.4.4, Theorem 2.5.1, Proposition 2.5.2, Proposition 2.5.5). From Chapter 1, the citing rule of Section 1.7 (Corollary 1.7.3), by which induction (Logic and Proof, Axiom 4.2.1) and the recursion theorem (Sets, Relations, and Cardinality, Theorem 2.3.1, Theorem 1.5.1) apply to $\mathbb{N}$; and powers and parity on $\mathbb{N}$ (Proposition 1.6.4), compared with in Section 3.1 and used in review exercise R2. From *Sets, Relations, and Cardinality*, Section 2.2, the definitions of divisibility and prime (Sets, Relations, and Cardinality, Definition 2.2.1, Sets, Relations, and Cardinality, Definition 2.2.2); from its Section 2.3, finite products (Sets, Relations, and Cardinality, Definition 2.3.1), used in Section 3.3 as a pattern for products of lists; from its Section 2.5, equivalence relations, classes and quotient sets (Sets, Relations, and Cardinality, Definition 2.5.1, Sets, Relations, and Cardinality, Definition 2.5.4, Sets, Relations, and Cardinality, Theorem 2.5.1) and congruence (Sets, Relations, and Cardinality, Definition 2.5.5); from its Section 2.6, suprema and the divisibility order (Sets, Relations, and Cardinality, Definition 2.6.5, Sets, Relations, and Cardinality, Example (The supremum under divisibility is the least common multiple)), in a remark of Section 3.2 and in Section 3.3's B2. From *Sets, Relations, and Cardinality*, Chapter 1: products, equality of functions and bijections (Sets, Relations, and Cardinality, Definition 1.3.2, Sets, Relations, and Cardinality, Proposition 1.4.1, Sets, Relations, and Cardinality, Definition 1.4.5); and from its Section 3.1, finite sets and equinumerous sets (Sets, Relations, and Cardinality, Definition 3.1.2, Sets, Relations, and Cardinality, Proposition 3.1.1), to count $\mathbb{Z}/n\mathbb{Z}$, with more on finite sets (Sets, Relations, and Cardinality, Proposition 3.1.6, Sets, Relations, and Cardinality, Corollary 3.1.4, Sets, Relations, and Cardinality, Theorem 3.1.9, Sets, Relations, and Cardinality, Exercise (C2: Counting products and functions), Sets, Relations, and Cardinality, Corollary 3.1.3) in Section 3.5's B3 and C2. From *Logic and Proof*, Chapter 5, the catalogue of what a proof may not do (Logic and Proof, Example (What is not allowed)), for Section 3.3's B1 and C2 and Section 3.5's C2. A few results are compared with and not used: the twins listed above; the even square of *Logic and Proof*, Section 2.1 (Logic and Proof, Theorem 2.1.3) in Section 3.1; Euclid's argument in *Logic and Proof*, Section 5.1's C1 (Logic and Proof, Exercise (C1: Euclid without the contradiction)) in Section 3.3's C3; a rule on classes that is not a function (Sets, Relations, and Cardinality, Example (A rule on classes that is not a function)) in Section 3.4's C1; lowest terms in *Sets, Relations, and Cardinality*, Section 2.2's C2 (Sets, Relations, and Cardinality, Exercise (C2: Lowest terms)) in Section 3.3; and, in the review exercises, the countable product of *Sets, Relations, and Cardinality*, Section 3.2 (Sets, Relations, and Cardinality, Exercise (B2: Putting a proof in order), Sets, Relations, and Cardinality, Theorem 3.2.6) in R2,

the two proofs that $\sqrt{2}$ is not rational (Logic and Proof, Theorem 2.2.1, Theorem 2.6.1) in R3. Review exercise R1 also returns to the integers as classes of pairs (Definition 2.1.1).

Time. About 25 hours on the core track (all of A, one B and one C per section, the exit checks); about 30 hours for everything.

3.0.1 Roadmap

- **Divisibility and Division with Remainder.** Divisibility and primes read on the constructed $\mathbb{Z}$, the rules for divisibility proved from the ring and its order, division with a unique quotient and remainder from a least element, the quotient as the integer part $\lfloor a/n \rfloor$, even and odd integers, and the small-primes remark that 2, 3, 5 and 7 are prime.
- **The Greatest Common Divisor, Bézout, and the Euclidean Algorithm.** The combinations $a\mathbb{Z}+b\mathbb{Z}$ and the multiples $n\mathbb{Z}$, the greatest common divisor with $\gcd(0,0) \coloneqq 0$, Bézout's theorem (the least positive combination is the gcd, and every combination is a multiple of it), coprime integers, and the Euclidean algorithm as a recursion that keeps the set of combinations fixed, with back-substitution for the coefficients.
- **Euclid's Lemma and the Fundamental Theorem of Arithmetic.** Euclid's lemma from the Bézout identity, a prime dividing a product of primes equal to one of them, coprime divisors, the least divisor above 1 as a prime, existence and uniqueness of non-decreasing prime factorizations by comparing the largest primes, a monoid in which factorization is not unique, and lowest terms in $\mathbb{Q}$.
- **Ideals, Quotient Rings, and the Integers Modulo n.** Ideals as the subsets that a quotient needs, the quotient ring R/I of a commutative ring by an ideal, the failure for $\mathbb{Z} \subseteq \mathbb{Q}$, the ideals of $\mathbb{Z}$ as the sets $n\mathbb{Z}$, and $\mathbb{Z}/n\mathbb{Z}$ as a commutative ring with n elements that no order fits when $n \geq 2$.
- **Units, Prime Moduli, and the Chinese Remainder Theorem.** Units of a commutative ring, the units of $\mathbb{Z}/n\mathbb{Z}$ as the classes coprime to n with inverses from back-substitution, $\mathbb{Z}/n\mathbb{Z}$ a field exactly when n is prime, and the Chinese remainder theorem for two coprime moduli, with the failures when the moduli share a factor; the review exercises reach back to Chapters 1 and 2, *Logic and Proof*, Chapter 2, and *Sets, Relations, and Cardinality*, Chapter 3.

3.0.2 Named moves

- **Least positive element, then divide.** Take the least positive element d of a set of integers closed under integer combinations; divide any element of the set by d; the remainder lies in the set and is smaller than d, so it is 0. This proves Bézout's theorem in Section 3.2 and the description of the ideals of $\mathbb{Z}$ in Section 3.4, and Section 3.3's B2 and Section 3.4's C2 use it for common multiples and for sets closed under subtraction.
- **Multiply the Bézout identity.** From $1 = ax + by$, multiply by c: $c = acx + bcy$, and read off what divides c. It turns a statement about common divisors into an equation, and it gives Euclid's lemma in Section 3.3; in Section 3.5 the same identity, read modulo n, gives inverses, and read modulo m and modulo n at once, the Chinese remainder theorem.
- **Compute on representatives; the ideal absorbs the change.** An operation on classes $a + I$ is well defined because a change of representative is an element of I, and multiplying it by the other argument keeps it in I. This is Chapter 2's check one argument at a time, now made once for every commutative ring and every ideal.

3.1 Divisibility and Division with Remainder

Chapter 2 built $\mathbb{Z}$ as an ordered commutative ring, and a Quick check in Section 2.3 showed that it is not a field: $2x = 1$ has no solution in $\mathbb{Z}$. So "a/b" is not available among the integers. What remains is a relation, "b is a multiple of a", and the next best thing to division: dividing with a quotient and a remainder. *Sets, Relations, and Cardinality*, Section 2.2 defined the relation and proved division with remainder for the school integers. This section does both again for the integers of Chapter 2, where every step must come from the ordered-ring axioms and from the least-element property of $\mathbb{Z}$.

Throughout, $\mathbb{Z}$ is the ordered ring of Theorem 2.3.4, with $\mathbb{N} \subseteq \mathbb{Z}$ and $\mathbb{Z}_{>0} = \{n \in \mathbb{Z} : n > 0\}$ (Theorem 2.3.4 (b)), and $a - b \coloneqq a + (-b)$. Numerals $2 \coloneqq 1 + 1$, $3 \coloneqq 2 + 1$, and so on, are the natural numbers of those names, as in Section 2.6; a sum or product of numerals may be computed in $\mathbb{N}$, because the embedding of Proposition 2.2.3 preserves $+$ and $\cdot$, and a sign is moved by Proposition 2.3.2 (c), as in $(-4) \cdot 5 = -(4 \cdot 5) = -20$.

3.1.1 Divisibility on the constructed integers

The definition of divisibility needs nothing that the ring $\mathbb{Z}$ lacks. We restate it, and the definition of a prime, from *Sets, Relations, and Cardinality*, Section 2.2, with $\mathbb{Z}$ now the ring of Chapter 2.

Divisibility (Sets, Relations, and Cardinality, Definition 2.2.1). Let $a, b \in \mathbb{Z}$. Then a **divides** b, written $a \mid b$, if **there exists** $k \in \mathbb{Z}$ with $b = ka$. Then a is a **divisor** of b, and b is a **multiple** of a. We write $a \nmid b$ when a does not divide b.

Prime (Sets, Relations, and Cardinality, Definition 2.2.2). An integer p is **prime** if $p \geq 2$ and the only **positive** divisors of p are 1 and p. A **prime factor** of an integer n is a prime p with $p \mid n$.

Both definitions use only the multiplication of $\mathbb{Z}$, its order $\leq$, and the set $\mathbb{Z}_{>0}$, and Chapter 2 provides all three. They are the same definitions as before; what changes is that every fact about them must now be proved from Chapter 2.

Divisibility is a question about existence: is b some integer times a?

Some instances. $3 \mid -12$, since $-12 = (-4) \cdot 3$. $-4 \mid 0$, since $0 = 0 \cdot (-4)$. Even $0 \mid 0$ holds, since $0 = 1 \cdot 0$: the definition asks for **some** k, and here every k works. On the other hand $5 \nmid 12$. That is a statement that **no** $k \in \mathbb{Z}$ has $12 = 5k$, and to prove it we need the theorem on division with remainder below.

The relation "$a \mid b$" is not a fraction. It is a statement, true or false, about two integers. The expression b/a, for $a \neq 0$, is a rational number, and this chapter does not use it to decide divisibility.

Here are the rules for divisibility that the rest of the chapter uses. Each is a short consequence of the ring laws or of the order.

Proposition 3.1.1. Rules for Divisibility in $\mathbb{Z}$

Let $a, b, c, x, y \in \mathbb{Z}$.

(a) $a \mid 0$, $1 \mid a$ and $a \mid a$. If $0 \mid a$, then $a = 0$.

(b) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(c) If $a \mid b$ and $a \mid c$, then $a \mid bx + cy$.

(d) If $a \mid b$, then $(-a) \mid b$ and $a \mid (-b)$.

(e) If $a \mid b$, $a > 0$ and $b > 0$, then $a \leq b$.

(f) If $a, b \in \mathbb{N}$, $a \mid b$ and $b \mid a$, then $a = b$.

(g) If $a \mid 1$, then $a = 1$ or $a = -1$.

(h) If $a \neq 0$, then **exactly one** of a, $-a$ lies in $\mathbb{Z}_{>0}$.

Idea. Parts (a)–(d) are identities in the commutative ring $\mathbb{Z}$: write each hypothesis as an equation $b = ka$ and rearrange. Part (e) is the one that needs the order. If $b = ka$ with $a, b > 0$, the factor k cannot be zero or negative, so by discreteness it is at least 1, and multiplying a by something at least 1 cannot make it smaller. Parts (f) and (g) then follow from (e), and (h) is (P2) for the positives of $\mathbb{Z}$.

Proof. We use the ring laws (F1)–(F7) and (F9) of $\mathbb{Z}$ (Theorem 2.3.1) without further mention.

(a) $0 = a \cdot 0 = 0 \cdot a$ by Proposition 2.3.2 (b) and (F6), so $a \mid 0$ with $k = 0$. Also $a = a \cdot 1$ by (F7), so $1 \mid a$ with $k = a$; and $a = 1 \cdot a$, so $a \mid a$ with $k = 1$. If $0 \mid a$, then $a = k \cdot 0 = 0$ for some $k \in \mathbb{Z}$, by Proposition 2.3.2 (b).

(b) Let $b = ka$ and $c = lb$ with $k, l \in \mathbb{Z}$. Then $c = l(ka) = (lk)a$ by (F5), so $a \mid c$.

(c) Let $b = ka$ and $c = la$. Then

$$bx + cy = (ka)x + (la)y = (kx)a + (ly)a = (kx + ly)a$$

by (F5), (F6) and (F9), and $kx + ly \in \mathbb{Z}$. So $a \mid bx + cy$.

(d) Let $b = ka$. By Proposition 2.3.2 (c), $(-k)(-a) = ka = b$, so $(-a) \mid b$; and $(-k)a = -(ka) = -b$, so $a \mid (-b)$.

(e) Let $a \neq 0$. By the converse part of Proposition 2.3.3, applied to the ordered ring $\mathbb{Z}$, the set $\{x \in \mathbb{Z} : x > 0\}$ satisfies (P2), so exactly one of $a > 0$, $-a > 0$ holds. That set is $\mathbb{Z}_{>0}$ by Theorem 2.3.4 (b).

(f) Let $b = ka$ with $k \in \mathbb{Z}$, $a > 0$ and $b > 0$. We show first that $k > 0$. If $k = 0$, then $b = 0 \cdot a = 0$ by Proposition 2.3.2 (b), against $b > 0$. If $k < 0$, then $ka < 0 \cdot a = 0$ by Proposition 2.3.5 (c) (multiplying $k < 0$ by $a > 0$), again against $b > 0$. So $k > 0$, and by discreteness (Theorem 2.3.4 (c), with $0 < k$) $1 \leq k$. If $k = 1$, then $b = a$. If $1 < k$, then $k - 1 > 0$ by Proposition 2.3.5 (a), so $(k - 1)a > 0 \cdot a = 0$ by Proposition 2.3.5 (c), that is, $b - a = ka + (-1)a = (k - 1)a > 0$ (Proposition 2.3.2 (c), (F7) and (F9)), and $a < b$ by Proposition 2.3.5 (a). In both cases $a \leq b$.

(g) Let $a, b \in \mathbb{N}$ with $a \mid b$ and $b \mid a$. If $a = 0$, then $0 \mid b$, so $b = 0 = a$ by (a). If $b = 0$, then $a = 0$ in the same way. Otherwise $a, b \in \mathbb{N}$ are non-zero, hence positive (Theorem 2.3.4 (b)), and (e) gives $a \leq b$ and $b \leq a$. Since $\leq$ is a total order (Proposition 2.3.5 (i)), it is antisymmetric, and $a = b$.

(h) Let $a \mid 1$, say $1 = ka$. Then $a \neq 0$: otherwise $1 = k \cdot 0 = 0$, against $1 \neq 0$ in $\mathbb{Z}$ (Theorem 2.3.1). By (h), one of a, $-a$ is positive; call it e. Then $e \mid 1$, by (d) if $e = -a$. Since $1 > 0$ (Proposition 2.3.5 (e)), part (e) gives $e \leq 1$; and $e > 0$ gives $1 \leq e$ by discreteness. So $e = 1$. If $e = a$, then $a = 1$; if $e = -a$, then $a = -(-a) = -1$ by Proposition 2.3.2 (a). This proves the proposition.

■

Part (f) is the antisymmetry of divisibility on $\mathbb{N}$, Sets, Relations, and Cardinality, Proposition 2.4.1 of *Sets, Relations, and Cardinality*, Section 2.4, now proved on $\mathbb{N} \subseteq \mathbb{Z}$ for the constructed integers. On all of $\mathbb{Z}$ it fails: $2 \mid -2$ and $-2 \mid 2$, by (d), yet $2 \neq -2$. That is why (f) is stated for $a, b \in \mathbb{N}$, and why (e) needs $b > 0$: $3 \mid -3$, but $3 > -3$. (The hypothesis $a > 0$ is a convenience for the proof: $a = 0$ cannot divide $b > 0$, by (a), and if $a < 0 < b$, then $a \leq b$ holds anyway.)

Quick check. Which hypothesis of (e) fails for $a = 3$, $b = 0$, and is the conclusion $a \leq b$ true there?

Solution. $3 \mid 0$ by (a), and $a = 3 > 0$, but $b = 0$ is not positive. The conclusion $3 \leq 0$ is false. Every integer divides 0, so no bound of the form $a \leq b$ can hold for $b = 0$; that is the reason for the hypothesis $b > 0$.

3.1.2 Division with remainder

Now the substitute for division. Given $a \in \mathbb{Z}$ and $n \in \mathbb{Z}_{>0}$, we want to write a as a multiple of n plus something small, and we want "small" to make the answer unique. The obvious attempt subtracts n from a until what is left lies in $\{0, \ldots, n-1\}$, adding n instead when a is negative. That is a process; the theorem replaces it by a least element.

Theorem 3.1.2. Division with Remainder in $\mathbb{Z}$

Let $a \in \mathbb{Z}$ and $n \in \mathbb{Z}_{>0}$.

(a) There are **unique** $q, r \in \mathbb{Z}$ with

$$a = qn + r \quad \text{and} \quad 0 \leq r < n.$$

(b) Moreover $q = \lfloor a/n \rfloor$, the integer part of the rational number a/n (Proposition 2.5.5), so $r = a - n\lfloor a/n \rfloor$.

The integers q and r are the **quotient** and the **remainder** of a on division by n.

Idea. Among the integers $a - kn$, $k \in \mathbb{Z}$, look at the non-negative ones. If there is at least one, Proposition 2.3.6 hands us the least, r. It must be below n: otherwise $r - n$ would be a smaller non-negative one of the same form. Finding a non-negative $a - kn$ is easy for $a \geq 0$ (take $k = 0$) and needs one inequality for $a < 0$ (take $k = a$). For uniqueness, two remainders differ by a multiple of n and yet by less than n, and (e) of the rules above forbids a positive multiple of n below n. Part (b) moves the inequalities $qn \leq a < qn + n$ into $\mathbb{Q}$ and divides by n.

Proof.

(a) *Existence.* Let

$$R := \{a - kn : k \in \mathbb{Z}\} \cap \{x \in \mathbb{Z} : x \geq 0\}.$$

We show $R \neq \emptyset$. If $a \geq 0$, take $k = 0$: $a - 0 \cdot n = a - 0 = a \geq 0$, by (F6), Proposition 2.3.2 (b) and $-0 = 0$ ($0 + 0 = 0$ and Proposition 2.3.2 (a)). If $a < 0$, take $k = a$: then $a - an = a + a(-n) = a(1 - n)$ by Proposition 2.3.2 (c), (F7) and (F9). Since $n > 0$, discreteness (Theorem 2.3.4 (c)) gives $1 \leq n$. If $n = 1$, then $a(1 - n) = a \cdot 0 = 0$ by Proposition 2.3.2 (b). If $1 < n$, then $-n < -1$ by Proposition 2.3.5 (a), so $1 - n < 0$ by (O3); multiplying $a < 0$ by $1 - n < 0$ reverses the inequality (Proposition 2.3.5 (d)), so $a(1 - n) > 0 \cdot (1 - n) = 0$. In every case R has an element.
The set $R \subseteq \mathbb{Z}$ is non-empty and has the lower bound 0, so by Proposition 2.3.6 it has a least element r. Write $r = a - qn$ with $q \in \mathbb{Z}$. Then $a = qn + r$ and $r \geq 0$.
Suppose $r \geq n$. Then $r - n \geq 0$: if $r = n$, then $r - n = 0$ by (F4); if $r > n$, then $r - n > 0$ by Proposition 2.3.5 (a). Also $r - n = a - qn - n = a + (-1)(qn) + (-1)n = a + (-1)((q + 1)n) = a - (q + 1)n$, since $-x = (-1)x$ for every x (Proposition 2.3.2 (c) and (F7)) and by (F9), so $r - n \in R$. But $-n < -0 = 0$ (Proposition 2.3.5 (a), since $n > 0$), so $r - n < r$ by (O3). This contradicts the choice of r as least. Hence $r < n$.
Uniqueness. Let $a = qn + r = q'n + r'$ with $q, q', r, r' \in \mathbb{Z}$, $0 \leq r < n$ and $0 \leq r' < n$. Suppose $r \neq r'$. Exchanging (q, r) and (q', r') changes neither the hypotheses nor the goal, so we may suppose $r' < r$. Then

$$r - r' = (q' - q)n,$$

so $n \mid r - r'$. Here $r - r' > 0$ by Proposition 2.3.5 (a), and $n > 0$, so Proposition 3.1.1 (e) gives $n \leq r - r'$. But adding $r - r'$ to $0 \leq r'$ gives $r - r' \leq (r - r') + r' = r$ (Proposition 2.3.5 (b)), and $r < n$; so $r - r' < n$, a contradiction. Hence $r = r'$. Then $qn = q'n$, so $(q - q')n = 0$, and since $n \neq 0$, Proposition 2.3.5 (f) gives $q - q' = 0$, that is, $q = q'$.

(b) From (a), $0 \leq r < n$; adding qn (Proposition 2.3.5 (b)) gives

$$qn \leq a < qn + n.$$

These are inequalities between integers, and by Theorem 2.5.1 (b) and Proposition 2.4.4 they hold in $\mathbb{Q}$ as well, with the same sums and products. In $\mathbb{Q}$, $n > 0$, so $n^{-1} > 0$ (Proposition 2.5.2 (a)). Multiplying by n^{-1} preserves strict inequalities (Proposition 2.3.5 (c)) and equalities, so

$$q = qn \cdot n^{-1} \leq a \cdot n^{-1} < (qn + n) \cdot n^{-1} = q + 1.$$

That is, $q \leq a/n < q+1$, with $q \in \mathbb{Z}$. By the uniqueness in Proposition 2.5.5, $q = \lfloor a/n \rfloor$, and then $r = a - qn = a - n\lfloor a/n \rfloor$. This proves the theorem.

■

The hypothesis $n > 0$ is used four times in (a): to find an element of R when $a < 0$ (through $1 - n \leq 0$), to see that $r - n < r$, and in the uniqueness both to apply Proposition 3.1.1 (e) and, as $n \neq 0$, to cancel n from $(q - q')n = 0$. For $n = 0$ the statement fails outright: no r satisfies $0 \leq r < 0$. Exercise B1 handles $n < 0$.

Uniqueness is what turns the theorem into a test for divisibility: $n \mid a$ **if and only if the remainder of** a **on division by** n **is** 0. Indeed, if $a = kn$, then $a = kn + 0$ is a division with remainder, and by uniqueness it is the only one; conversely $a = qn + 0$ says $n \mid a$. For instance $12 = 2 \cdot 5 + 2$ with $0 \leq 2 < 5$, so the remainder is $2 \neq 0$, and $5 \nmid 12$.

Remark. The statement of (a) is Sets, Relations, and Cardinality, Theorem 2.2.3 of *Sets, Relations, and Cardinality*, Section 2.2, which was proved there for the school integers. The existence proof above follows the same route, with each school fact replaced by a result of Chapter 2: the closure of $\mathbb{Z}$ under $+$, $-$, $\cdot$ by Theorem 2.3.1; "a positive integer is at least 1" by Theorem 2.3.4 (b), (c); and the least element of a set of integers bounded below by Proposition 2.3.6. The uniqueness argument there is the one packaged here as Proposition 3.1.1 (e). For $n = 2$ and $a \in \mathbb{N}$, (a) is Proposition 1.6.4 (d) of Section 1.6, which we compare with but do not use.

Part (b) pays a debt of Section 2.5. There, after Proposition 2.5.5, the integer part $\lfloor p/q \rfloor$ was announced as a quotient; (b) says exactly that, and it gives a second way to compute q and r. Exercise C1 turns (b) around and proves the existence half of (a) from the integer part.

Quick check. Find the quotient and the remainder of $a = -17$ on division by $n = 5$, and check them against (b).

Solution. $q = -4$ and $r = 3$: $(-4) \cdot 5 + 3 = -20 + 3 = -17$, and $0 \leq 3 < 5$. For (b), $-4 \leq -17/5 < -3$, since $(-4) \cdot 5 = -20 \leq -17$ and $-17 < -15 = (-3) \cdot 5$ (Theorem 2.5.1 (a)); so $\lfloor -17/5 \rfloor = -4$.

Warning. "The remainder" means the r with $0 \leq r < n$. The equation $-17 = (-3) \cdot 5 - 2$ is true, but -2 is **not** the remainder of -17 on division by 5; the remainder is 3. If negative "remainders" were allowed, uniqueness would fail at once: $-17 = (-3)\cdot 5 + (-2) = (-4)\cdot 5 + 3$. For negative a, the quotient is not "a/n with the fraction dropped": that would give -3 here, while $\lfloor -17/5 \rfloor = -4$.

3.1.3 Even and odd

Recall that an integer a is **even** if $a = 2k$ for some $k \in \mathbb{Z}$, that is, $2 \mid a$, and **odd** if $a = 2k + 1$ for some $k \in \mathbb{Z}$. These are the words *Logic and Proof*, Chapter 1 used, now read on the constructed $\mathbb{Z}$. Section 2.6 had parity for natural numbers from Proposition 1.6.4 (d). With $n = 2$, the theorem gives it for every integer, negative ones included, with one citation.

Corollary 3.1.3. Even and Odd Integers

(a) Every $a \in \mathbb{Z}$ is even or odd, and **not both**. More precisely, either $a = 2q$ for exactly one $q \in \mathbb{Z}$ and a is not of the form $2k + 1$, or $a = 2q + 1$ for exactly one $q \in \mathbb{Z}$ and a is not of the form $2k$.

(b) If $a \in \mathbb{Z}$ and a^2 is even, then a is even.

Idea. Divide by 2. The remainder lies in $\{0, 1\}$, by discreteness, and it is unique, so each integer has exactly one of the two forms. For (b), square the odd form and read off the remainder.

Proof.

(a) Since $2 = 1 + 1 > 0$, the theorem with $n = 2$ gives unique $q, r \in \mathbb{Z}$ with $a = 2q + r$ and $0 \leq r < 2$. Then $r \in \{0, 1\}$: if $r \neq 0$, then $0 < r$, so $1 \leq r$ by discreteness (Theorem 2.3.4 (c)); and if moreover $1 < r$, then $2 = 1 + 1 \leq r$ by discreteness again, against $r < 2$. If $a = 2k$ and $a = 2k' + 1$ for some $k, k' \in \mathbb{Z}$, then $(k, 0)$ and $(k', 1)$ would be two pairs as in the theorem, and $0 \neq 1$ contradicts its uniqueness. The same uniqueness shows that q is the only integer with $a = 2q + r$.

(b) Suppose a is not even. By (a), $a = 2q + 1$ with $q \in \mathbb{Z}$. Then

$$a^2 = (2q + 1)^2 = 4q^2 + 4q + 1 = 2(2q^2 + 2q) + 1,$$

by Proposition 2.3.2 (d) and (F5), (F6), (F9), so a^2 is odd, and by (a) it is not even. This proves (b) by contraposition.

■

Part (b) is Logic and Proof, Theorem 2.1.3 of *Logic and Proof*, Section 2.1, proved there from the school facts (E1)–(E3) about the integers. Part (a) proves (E1) and (E2) for the constructed $\mathbb{Z}$, and the computation in (b) is (E3). We compare with that proof and do not use it. *Sets, Relations, and Cardinality*, Section 2.2 proved (E1) and (E2) for the school integers right after Sets, Relations, and Cardinality, Theorem 2.2.3.

Section 2.6 proved that no rational squares to 2 by a descent on denominators (Theorem 2.6.1), a route that needs only the order and the least-element property of $\mathbb{Z}$ and adapts to other non-squares. With (b), and with lowest terms from Section 3.3, the classical parity proof of *Logic and Proof*, Chapter 2 can also be run on the constructed $\mathbb{Q}$, every step citable; one of the review exercises at the end of this chapter does so.

3.1.4 Small primes

Later sections need to know that a few small integers are prime in the constructed $\mathbb{Z}$. *Sets, Relations, and Cardinality*, Section 2.2 checked some of these after Sets, Relations, and Cardinality, Definition 2.2.2, for the school integers; that check is not a proof about the ring of Chapter 2, so we redo it here.

Remark. **The small-primes remark.** The integers 2, 3, 5 and 7 are prime in $\mathbb{Z}$. Let $p \in \{2, 3, 5, 7\}$. Then $p \geq 2$. Let d be a positive divisor of p. Then $1 \leq d$ by discreteness (Theorem 2.3.4 (c)) and $d \leq p$ by Proposition 3.1.1 (e), so d is one of $1, 2, \ldots, p$: applying discreteness repeatedly (if $d \neq 1$ then $2 \leq d$; if $d \neq 2$ then $3 \leq d$; and so on up to p) leaves no other value. Each d with $1 < d < p$ leaves a non-zero remainder:

$$3 = 1 \cdot 2 + 1, \qquad 5 = 2 \cdot 2 + 1 = 1 \cdot 3 + 2 = 1 \cdot 4 + 1,$$
$$7 = 3 \cdot 2 + 1 = 2 \cdot 3 + 1 = 1 \cdot 4 + 3 = 1 \cdot 5 + 2 = 1 \cdot 6 + 1.$$

(For $p = 2$ there is no such d.) So $d \nmid p$ for these d, by the divisibility test after Theorem 3.1.2: $p = kd$ would give $p = kd + 0$, a second division of p by d, with remainder 0, against the uniqueness in Theorem 3.1.2 (a). Hence the positive divisors of p are 1 and p, and p is prime.

The later sections of this chapter cite this paragraph as "the small-primes remark". The same

procedure decides, for any given $p \geq 2$, whether p is prime; Section 3.3 shortens it.

3.1.5 Exercises

3.1.5.1 A. Check your understanding

Exercise. A1

(a) Find the quotient and the remainder on division by $n = 6$ of $a = 47$, of $a = -47$ and of $a = 0$.

(b) Where is $n > 0$ used in the existence half of Theorem 3.1.2 (a)?

(c) True or false, with a reason: $0 \mid 0$.

(d) True or false, with a reason: for all $a, b \in \mathbb{Z}$, if $a \mid b$ and $b \mid a$, then $a = b$.

Solution.

(a) $47 = 7 \cdot 6 + 5$, so $(q, r) = (7, 5)$. $-47 = (-8) \cdot 6 + 1$, since $-48 + 1 = -47$, so $(q, r) = (-8, 1)$. $0 = 0 \cdot 6 + 0$, so $(q, r) = (0, 0)$. In each case $0 \leq r < 6$, so by uniqueness these are the quotient and remainder. Note $(-8, 1)$, not $(-7, -5)$.

(b) Twice. For $a < 0$, the element $a - an = a(1 - n)$ of R is non-negative because $1 - n \leq 0$, which comes from $n \geq 1$, that is, from $n > 0$ and discreteness. And the contradiction "$r - n < r$" needs $-n < 0$, that is, $n > 0$.

(c) True: $0 = 1 \cdot 0$, so $k = 1$ witnesses $0 \mid 0$ (Proposition 3.1.1 (a)).

(d) False: $2 \mid -2$ and $-2 \mid 2$ (Proposition 3.1.1 (d)), but $2 \neq -2$. The statement is true for $a, b \in \mathbb{N}$ (Proposition 3.1.1 (f)).

3.1.5.2 B. Practice

Exercise. B1: Negative divisors

Let $a, n \in \mathbb{Z}$ with $n < 0$. Prove that there are unique $q, r \in \mathbb{Z}$ with $a = qn + r$ and $0 \leq r < -n$.

Solution. Since $n < 0$, $-n > 0$ by Proposition 2.3.5 (a), so $-n \in \mathbb{Z}_{>0}$.

Existence. By Theorem 3.1.2 (a) applied to a and $-n$, there are $q_0, r \in \mathbb{Z}$ with $a = q_0(-n) + r$ and $0 \leq r < -n$. Let $q \coloneqq -q_0$. Then $q_0(-n) = (-q_0)n = qn$ by Proposition 2.3.2 (c), so $a = qn + r$.

Uniqueness. Let $a = qn + r = q'n + r'$ with $0 \leq r, r' < -n$. By Proposition 2.3.2 (c), $qn = (-q)(-n)$ and $q'n = (-q')(-n)$, so $a = (-q)(-n) + r = (-q')(-n) + r'$ are two divisions of a by $-n$. The uniqueness in Theorem 3.1.2 (a) gives $r = r'$ and $-q = -q'$, hence $q = -(-q) = -(-q') = q'$ by Proposition 2.3.2 (a).

Exercise. B2: Three in a row

Let $a \in \mathbb{Z}$. Prove that **exactly one** of a, $a + 1$, $a + 2$ is divisible by 3.

Solution. At least one. By Theorem 3.1.2 (a), $a = 3q + r$ with $0 \le r < 3$, and $r \in \{0, 1, 2\}$ by discreteness, as in the proof of Corollary 3.1.3. If $r = 0$, then $3 \mid a$. If $r = 1$, then $a + 2 = 3q + 3 = 3(q + 1)$. If $r = 2$, then $a + 1 = 3q + 3 = 3(q + 1)$.

At most one. Suppose $3 \mid a + i$ and $3 \mid a + j$ with $0 \le i < j \le 2$. By Proposition 3.1.1 (c) (with $x = 1$, $y = -1$), $3 \mid (a + j) - (a + i) = j - i$, and $j - i \in \{1, 2\}$ is positive. By Proposition 3.1.1 (e), $3 \le j - i \le 2$, which is false. So no two of the three are divisible by 3.

(Equivalently: the remainders of a, $a + 1$, $a + 2$ on division by 3 are r, and then the other two of $0, 1, 2$, and uniqueness of remainders allows only one of them to be 0.)

3.1.5.3 C. Going deeper

Exercise. C1: Division from the integer part

Let $a \in \mathbb{Z}$ and $n \in \mathbb{Z}_{>0}$. Prove the existence half of Theorem 3.1.2 (a) directly from Proposition 2.5.5, without the set R: put $q := \lfloor a/n \rfloor$ and $r := a - qn$, and show that $0 \le r < n$. Say where $n > 0$ is used.

Solution. By Proposition 2.5.5, $q := \lfloor a/n \rfloor \in \mathbb{Z}$ satisfies

$$q \le a/n < q + 1$$

in $\mathbb{Q}$, where $a/n = a \cdot n^{-1}$ and $n \neq 0$ in $\mathbb{Q}$. Since $n > 0$, multiplying by n preserves the strict inequality (Proposition 2.3.5 (c)) and the equality case, so

$$qn \le (a \cdot n^{-1}) \cdot n = a < (q + 1)n = qn + n.$$

Subtracting qn (that is, adding $-qn$, (O3)) gives $0 \le a - qn < n$. Here $r := a - qn$ is an integer, and by Theorem 2.5.1 (b) the inequalities $0 \le r < n$, proved in $\mathbb{Q}$, hold in $\mathbb{Z}$. So $a = qn + r$ with $0 \le r < n$.

The hypothesis $n > 0$ is used when multiplying by n: for $n < 0$, multiplication reverses inequalities (Proposition 2.3.5 (d)), and the same q would give $qn \ge a > qn + n$, so $n < r \le 0$ instead. (This exercise is practice with a result already proved; nothing later rests on it.)

3.2 The Greatest Common Divisor, Bézout, and the Euclidean Algorithm

Section 3.1 gave the constructed integers a relation $a \mid b$ and a division with remainder (Theorem 3.1.2). This section asks what two integers have in common. Here is a concrete version. A machine accepts only coins worth 12 and 42 units, and gives change in the same coins. Which amounts can be paid exactly? Paying x coins of the first kind and y of the second, with negative numbers for coins received as change, pays $12x + 42y$. So $42 - 3 \cdot 12 = 6$ can be paid, and hence every multiple of 6; and since 6 divides both 12 and 42, nothing else can. Why 6, and why exactly the multiples of 6? The answer is a theorem about every pair of integers, and its proof is the main move of this chapter.

3.2.1 Integer combinations

The amounts $12x + 42y$ form a set that deserves a name.

Definition 3.2.1. Integer Combinations

Let $a, b, n \in \mathbb{Z}$. The set of **integer combinations** of a and b is

$$a\mathbb{Z} + b\mathbb{Z} := \{ax + by : x, y \in \mathbb{Z}\},$$

and the set of **multiples** of n is $n\mathbb{Z} := \{kn : k \in \mathbb{Z}\}$.

So $n\mathbb{Z}$ is the set of $m \in \mathbb{Z}$ with $n \mid m$ (Sets, Relations, and Cardinality, Definition 2.2.1), and $a\mathbb{Z} + b\mathbb{Z}$ is written as one symbol: it is defined directly, not as a sum of two sets. Three closure properties hold for every $a, b \in \mathbb{Z}$. For $x, y, x', y', m \in \mathbb{Z}$, the ring laws of Theorem 2.3.1 and Proposition 2.3.2 (c) give

$$\begin{aligned}(ax + by) + (ax' + by') &= a(x + x') + b(y + y'),\\ -(ax + by) &= a(-x) + b(-y),\\ m(ax + by) &= a(mx) + b(my),\end{aligned}$$

so $a\mathbb{Z} + b\mathbb{Z}$ is closed under $+$, under $-$, and under multiplication by **any** integer. It contains $0 = a \cdot 0 + b \cdot 0$, $a = a \cdot 1 + b \cdot 0$ and $b = a \cdot 0 + b \cdot 1$. Taking $b = 0$ gives $n\mathbb{Z} + 0\mathbb{Z} = \{nx + 0y\} = \{xn : x \in \mathbb{Z}\} = n\mathbb{Z}$, by Proposition 2.3.2 (b) and (F6); so $n\mathbb{Z}$ has the same three properties.

Examples:

- $12\mathbb{Z} + 42\mathbb{Z}$ contains 12, 42, $30 = 42 - 12$ and $6 = 12 \cdot (-3) + 42 \cdot 1$, so it contains every multiple of 6, by the third closure property.
- $2\mathbb{Z} + 3\mathbb{Z} = \mathbb{Z}$: it contains $1 = 2 \cdot (-1) + 3 \cdot 1$, and with it $m \cdot 1 = m$ for every $m \in \mathbb{Z}$.
- The degenerate case: $0\mathbb{Z}+0\mathbb{Z} = \{0\}$ (the case $n = 0$ above, since $0\mathbb{Z} = \{k \cdot 0 : k \in \mathbb{Z}\} = \{0\}$ by Proposition 2.3.2 (b)). It is the only case in which $a\mathbb{Z} + b\mathbb{Z}$ has no positive element (if $a \neq 0$, one of a, $-a$ is a positive element, by Proposition 3.1.1 (h); likewise for b), and the theorem below must exclude it.

Non-example by minimal change. Allow only paying, with no change given: $\{12x + 42y : x, y \in \mathbb{N}\}$. It contains 12 and 42 and is closed under $+$, but not under $-$: it contains no negative integer, since $x, y \geq 0$ give $12x + 42y \geq 0$. Nor does it contain 6: a non-zero element has $x \geq 1$ or $y \geq 1$, so it is at least 12. The coefficients must range over all of $\mathbb{Z}$.

3.2.2 The greatest common divisor

An integer d is a **common divisor** of a and b if $d \mid a$ and $d \mid b$. Every pair has the common divisor 1 (Proposition 3.1.1 (a)). For the coins, 6 is a common divisor of 12 and 42, and the question was why it is the one that matters.

The greatest common divisor is the largest integer that divides both.

Definition 3.2.2. Greatest Common Divisor

Let $a, b \in \mathbb{Z}$. If $(a, b) \neq (0, 0)$, the **greatest common divisor** $\gcd(a, b)$ is the greatest element, for $\leq$, of the set of common divisors of a and b. We set $\gcd(0, 0) := 0$.

In words: among all integers dividing both a and b, take the largest. The definition says "the greatest element", so it owes us a proof that there is one.

Well-definedness. Let $(a, b) \neq (0, 0)$, and let D be the set of common divisors of a and b. It is non-empty, since $1 \in D$. Say $a \neq 0$ (otherwise $b \neq 0$, and the same argument runs with b). By Proposition 3.1.1 (h), one of a, $-a$ is positive; call it c. Every $d \in D$ satisfies $d \leq c$: if $d > 0$, then $d \mid c$ by Proposition 3.1.1 (d), and $d \leq c$ by Proposition 3.1.1 (e); if $d \leq 0$, then $d \leq 0 < c$. So D is bounded above, and by Proposition 2.3.6 it has a greatest element. Since $1 \in D$, that element is at least 1. Hence $\gcd(a, b) \geq 1$ **whenever** $(a, b) \neq (0, 0)$, and $\gcd(a, b) = \gcd(b, a)$, since the condition "$d \mid a$ and $d \mid b$" is symmetric in a and b.

Examples:

- $\gcd(4, 6) = 2$. A positive common divisor d satisfies $d \leq 4$, by Proposition 3.1.1 (e), and $d \geq 1$, so $d \in \{1, 2, 3, 4\}$ by discreteness (Theorem 2.3.4 (c)), as in the small-primes remark of Section 3.1. Now $4 \nmid 6$ and $3 \nmid 4$, because $6 = 1 \cdot 4 + 2$ and $4 = 1 \cdot 3 + 1$ have non-zero remainders (the divisibility test after Theorem 3.1.2), while 2 divides both. So the greatest common divisor is 2.
- $\gcd(-5, 5) = 5$: the common divisors of -5 and 5 are the divisors of 5 (Proposition 3.1.1 (d) and Proposition 2.3.2 (a)), and the greatest is 5, by Proposition 3.1.1 (e).
- $\gcd(1, b) = 1$ for every $b \in \mathbb{Z}$: a common divisor d divides 1, so $d = \pm 1$ (Proposition 3.1.1 (g)), and the greater is 1.

Why the convention for $(0, 0)$. Every integer divides 0, so 0 and 0 have no greatest common divisor for $\leq$, and some value has to be chosen. The value 0 is the one that keeps clauses (c) and (e) of the theorem below true for $(0, 0)$, as the remark after it shows. It also matches Sets, Relations, and Cardinality, Example (The supremum under divisibility is the least common multiple) of *Sets, Relations, and Cardinality*, Section 2.6, where 0 is the top of $(\mathbb{N}, \mid)$, divisible by everything.

Listing divisors one by one, as in the first example, is slow. The Euclidean algorithm below replaces the list by one division per step, and Bézout's theorem is what makes it work.

3.2.3 Bézout's theorem

Here is the answer to the coin question for every pair of integers: the payable amounts are exactly the multiples of the greatest common divisor, and the greatest common divisor is itself payable.

Theorem 3.2.1. Bézout's Theorem

Let $a, b \in \mathbb{Z}$, not both 0, and let $S := (a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{Z}_{>0}$.

(a) S has a least element d, and $d = ax_0 + by_0$ for some $x_0, y_0 \in \mathbb{Z}$.
(b) $d \mid a$ and $d \mid b$.
(c) Every common divisor of a and b divides d.
(d) $d = \gcd(a, b)$. In particular, $\gcd(a, b) = ax + by$ for some $x, y \in \mathbb{Z}$.
(e) $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

Try it first (ten minutes): for $a = 12$ *and* $b = 42$*, list the positive numbers* $12x + 42y$ *you can reach. Which is the least, and why should the least one divide* 12*?*

Idea. The obvious plan starts from the greatest common divisor and tries to write it as $ax + by$. It stalls: the definition of $\gcd(a, b)$ says nothing about combinations, so there is nothing to build x and y from. The attempt on 12 and 42 suggests starting at the other end. The positive combinations $12, 42, 30, 18, 6, \ldots$ have a least element, and it was found as a combination. Why should it divide 12? Divide 12 by it. The remainder is 12 minus a multiple of a combination, so it is again a combination, and it is smaller than the least positive one. So it cannot be positive: it is 0. Once the least positive combination d is known to divide a and b, the rest is bookkeeping: every common divisor divides the combination d, so none exceeds it.

Before reading on: where is the minimality of d used, and where its positivity?

Proof.

(a) Suppose $a \neq 0$; if $a = 0$, then $b \neq 0$, and the argument is the same with b in place of a. Both $a = a \cdot 1 + b \cdot 0$ and $-a = a \cdot (-1) + b \cdot 0$ lie in $a\mathbb{Z} + b\mathbb{Z}$, and by Proposition 3.1.1 (h) one of them is positive. So $S \neq \emptyset$. Every element of S is positive, so 0 is a lower bound, and by Proposition 2.3.6 S has a least element d. Since $d \in a\mathbb{Z} + b\mathbb{Z}$, there are $x_0, y_0 \in \mathbb{Z}$ with $d = ax_0 + by_0$.

(b) Since $d \in \mathbb{Z}_{>0}$, Theorem 3.1.2 (a) gives $q, r \in \mathbb{Z}$ with $a = qd + r$ and $0 \leq r < d$. Then

$$r = a - qd = a - q(ax_0 + by_0) = a(1 - qx_0) + b(-qy_0),$$

so $r \in a\mathbb{Z} + b\mathbb{Z}$. If $r > 0$, then $r \in S$ and $r < d$, contradicting the minimality of d. Hence $r = 0$, and $a = qd$, that is, $d \mid a$. Dividing b by d in the same way gives $d \mid b$.

(c) Let $c \mid a$ and $c \mid b$. By Proposition 3.1.1 (c), $c \mid ax_0 + by_0 = d$.

(d) By (b), d is a common divisor of a and b. Let c be any common divisor. If $c > 0$, then $c \mid d$ by (c), and $c \leq d$ by Proposition 3.1.1 (e), since $d > 0$. If $c \leq 0$, then $c \leq 0 < d$. So d is the greatest common divisor: $d = \gcd(a, b)$ by Definition 3.2.2, and $\gcd(a, b) = ax_0 + by_0$.

(e) ($\supseteq$) Let $k \in \mathbb{Z}$. Then $kd = k(ax_0 + by_0) = a(kx_0) + b(ky_0) \in a\mathbb{Z} + b\mathbb{Z}$. ($\subseteq$) Let $x, y \in \mathbb{Z}$. By (b), $d \mid a$ and $d \mid b$, so $d \mid ax + by$ by Proposition 3.1.1 (c), that is, $ax + by \in d\mathbb{Z}$. This proves the theorem.

■

The answer to the prompt: the minimality of d is used once, in (b), to force the remainder r to be 0. The positivity of d is used twice: to divide by d at all (Theorem 3.1.2 needs a positive divisor), and in (d), where $c \leq d$ for positive c comes from Proposition 3.1.1 (e) and for the other c from $0 < d$. (The step $r \in S$ uses the positivity of r, not of d, which is why the case $r = 0$ is split off.)

The proof of (b) is the move this chapter is built on, and we give it a name: **least positive element, then divide**. Take the least positive element d of a set closed under integer combinations; divide any element of the set by d; the remainder lies in the set and is smaller than d, so it is 0. Section 3.3 uses the move again for common multiples, and Section 3.4 for sets closed under subtraction.

For the coins, the theorem says: the payable amounts form $12\mathbb{Z} + 42\mathbb{Z} = d\mathbb{Z}$, where d is the least

positive payable amount, and $d = \gcd(12, 42)$. The Try-it-first list finds $d = 6 = 12 \cdot (-3) + 42 \cdot 1$.

Remark. Clause (c) says that $\gcd(a, b)$ is greatest also for divisibility, not only for $\leq$. For $a, b \in \mathbb{N}$, not both 0, it is the infimum of $\{a, b\}$ in $(\mathbb{N}, |)$ (Sets, Relations, and Cardinality, Definition 2.6.5): it divides both, and every $c \in \mathbb{N}$ dividing both divides it. This is the mirror image of Sets, Relations, and Cardinality, Example (The supremum under divisibility is the least common multiple), where the supremum is the least common multiple. For general $a, b \in \mathbb{Z}$, let a' be the one of a, $-a$ that lies in $\mathbb{N}$, and likewise b' (Proposition 3.1.1 (h), with $a' = 0$ if $a = 0$). Since a and $-a$ have the same divisors (Proposition 3.1.1 (d) and Proposition 2.3.2 (a)), $\gcd(a, b) = \gcd(a', b')$. With the convention $\gcd(0, 0) = 0$, clauses (c) and (e) also hold for $(0, 0)$ when d is read as 0: every integer divides 0, and $0\mathbb{Z} + 0\mathbb{Z} = \{0\} = 0\mathbb{Z}$. So 0, the top of $(\mathbb{N}, |)$, is also the infimum of $\{0, 0\} = \{0\}$.

Pairs whose greatest common divisor is 1 get a name of their own.

Definition 3.2.3. Coprime Integers

Integers $a, b \in \mathbb{Z}$ are **coprime** if $\gcd(a, b) = 1$.

For instance, 2 and 3 are coprime, and so are 1 and any b; but 4 and 6 are not, since $\gcd(4, 6) = 2$, and neither are 0 and 0, since $\gcd(0, 0) = 0$. For coprime integers, (d) of the theorem produces 1 as a combination, and conversely:

Corollary 3.2.2. Coprime Means 1 Is a Combination

Let $a, b \in \mathbb{Z}$. Then a and b are coprime if and only if $ax + by = 1$ for some $x, y \in \mathbb{Z}$.

Proof. ($\Rightarrow$) If $\gcd(a, b) = 1$, then $(a, b) \neq (0, 0)$, since $\gcd(0, 0) = 0 \neq 1$, and Theorem 3.2.1 (d) gives $x, y \in \mathbb{Z}$ with $ax + by = \gcd(a, b) = 1$.

($\Leftarrow$) Let $ax + by = 1$. Then $(a, b) \neq (0, 0)$, since otherwise $ax + by = 0 \neq 1$. So $g := \gcd(a, b) \geq 1$, and g divides a and b, hence divides $ax + by = 1$ by Proposition 3.1.1 (c). By Proposition 3.1.1 (g), $g = 1$ or $g = -1$, and $g \geq 1$ rules out -1. So $\gcd(a, b) = 1$.

■

The direction ($\Leftarrow$) is how coprimality is usually proved: exhibit one combination equal to 1. Thus $2 \cdot (-1) + 3 \cdot 1 = 1$ shows again that 2 and 3 are coprime, with no list of divisors.

3.2.4 The Euclidean algorithm

Bézout's theorem says that $\gcd(a, b)$ is a combination of a and b, but its proof finds it as a least element, which is no way to compute it. Computing needs a step that makes the numbers smaller without changing the set of combinations. Division with remainder is that step.

Lemma 3.2.3. One Step of the Euclidean Algorithm

Let $a \in \mathbb{Z}$ and $b \in \mathbb{Z}_{>0}$, and let $a = qb + r$ with $q, r \in \mathbb{Z}$. Then $a\mathbb{Z} + b\mathbb{Z} = b\mathbb{Z} + r\mathbb{Z}$, and hence

$\gcd(a, b) = \gcd(b, r)$.

Proof. ($\subseteq$) For $x, y \in \mathbb{Z}$, $ax + by = (qb + r)x + by = b(qx + y) + rx \in b\mathbb{Z} + r\mathbb{Z}$. ($\supseteq$) For $x, y \in \mathbb{Z}$, $bx + ry = bx + (a - qb)y = ay + b(x - qy) \in a\mathbb{Z} + b\mathbb{Z}$. So the two sets are equal. Neither pair is $(0, 0)$, since $b \neq 0$. By Theorem 3.2.1 (a) and (d), $\gcd(a, b)$ is the least element of $(a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{Z}_{>0}$, and $\gcd(b, r)$ is the least element of $(b\mathbb{Z} + r\mathbb{Z}) \cap \mathbb{Z}_{>0}$, the same set. So $\gcd(a, b) = \gcd(b, r)$. ∎

The lemma does not need r to be the remainder; any q will do. Choosing the remainder makes the new pair smaller, and repeating the step gives a sequence. To define "repeat" without "and so on", we use the recursion theorem.

Theorem 3.2.4. The Euclidean Algorithm

Let $a \in \mathbb{Z}$ and $b \in \mathbb{Z}_{>0}$. Define $(x_k, y_k) \in \mathbb{Z} \times \mathbb{N}$ for $k \in \mathbb{N}$ by $(x_0, y_0) := (a, b)$ and

$$(x_{k+1}, y_{k+1}) := \begin{cases} (y_k, r_k) & \text{if } y_k > 0, \\ (x_k, y_k) & \text{if } y_k = 0, \end{cases}$$

where, when $y_k > 0$, r_k is the remainder of x_k on division by y_k. Then:

(a) $x_k\mathbb{Z} + y_k\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$ for every $k \in \mathbb{N}$;

(b) there is a least $K \in \mathbb{N}$ with $y_K = 0$; moreover $K \geq 1$, and $y_{k+1} < y_k$ for every $k < K$;

(c) $x_K = \gcd(a, b)$.

The sequence is well defined. Let $G\colon \mathbb{Z}\times\mathbb{N} \to \mathbb{Z}\times\mathbb{N}$ send (x, y) to (y, r), where r is the remainder of x on division by y, if $y > 0$, and to (x, y) if $y = 0$. This is a function: the remainder exists and is unique by Theorem 3.1.2 (a), and it lies in $\mathbb{N}$ because $r \geq 0$ (Theorem 2.3.4 (b)). The recursion theorem (Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ as Theorem 1.5.1 by Corollary 1.7.3 (b)), with $X = \mathbb{Z} \times \mathbb{N}$, starting value (a, b) and step $(k, (x, y)) \mapsto G(x, y)$, gives exactly one sequence with the two displayed properties.

Idea. Part (a) is Lemma 3.2.3, once per step, carried along by induction. For (b), the second entries are natural numbers that drop at each step while they are positive, and a set of natural numbers has a least element, which cannot be positive. For (c), at the first index K with $y_K = 0$, the invariant (a) reads $a\mathbb{Z} + b\mathbb{Z} = x_K\mathbb{Z}$, and Bézout's theorem identifies the gcd as the least positive element of this set.

Proof.

(a) We use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k, for $P(k)$: "$x_k\mathbb{Z} + y_k\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$". The base case $P(0)$ holds since $(x_0, y_0) = (a, b)$. Suppose $P(k)$. If $y_k > 0$, then $x_k = q_k y_k + r_k$ for some $q_k \in \mathbb{Z}$, and Lemma 3.2.3 with x_k, y_k in place of a, b gives

$$x_{k+1}\mathbb{Z} + y_{k+1}\mathbb{Z} = y_k\mathbb{Z} + r_k\mathbb{Z} = x_k\mathbb{Z} + y_k\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}.$$

If $y_k = 0$, then $(x_{k+1}, y_{k+1}) = (x_k, y_k)$, and $P(k+1)$ is $P(k)$.

(b) The set $Y := \{y_k : k \in \mathbb{N}\} \subseteq \mathbb{N}$ is non-empty and has the lower bound 0, so it has a least element y_m (Proposition 2.3.6). If $y_m > 0$, then $y_{m+1} = r_m < y_m$ by the definition of the step and Theorem 3.1.2 (a), contradicting the choice of y_m. So $y_m = 0$, and the set $\{k \in \mathbb{N} : y_k = 0\}$ is non-empty, with lower bound 0. Let K be its least element (Proposition 2.3.6). Then $K \neq 0$, since $y_0 = b > 0$; so $K \geq 1$. For $k < K$, $y_k \neq 0$ by the choice of K, so $y_k > 0$ (as $y_k \in \mathbb{N}$, Theorem 2.3.4 (b)), and $y_{k+1} = r_k < y_k$.

(c) Since $K \geq 1$, $K - 1 \in \mathbb{N}$ and $K - 1 < K$, so $y_{K-1} > 0$ by (b), and the definition of the step gives $x_K = y_{K-1} > 0$. By (a) with $k = K$, and since $y_K = 0$,

$$a\mathbb{Z} + b\mathbb{Z} = x_K\mathbb{Z} + 0\mathbb{Z} = x_K\mathbb{Z}.$$

The least positive element of $x_K\mathbb{Z}$ is x_K: it lies in the set $(x_K = 1 \cdot x_K)$, and every positive element mx_K is divisible by x_K, so $x_K \leq mx_K$ by Proposition 3.1.1 (e). Since $b > 0$, $(a, b) \neq (0, 0)$, and by Theorem 3.2.1 (a) and (d), the least element of $(a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{Z}_{>0}$ is $\gcd(a, b)$. Hence $x_K = \gcd(a, b)$.

■

In practice the sequence is written as a chain of divisions, each dividing the previous divisor by the previous remainder, until a remainder is 0; then $x_K = y_{K-1}$, the last non-zero second entry, is the gcd. If $K \geq 2$, it is the last non-zero remainder r_{K-2}; if $K = 1$, it is b, and then $b \mid a$. Exercise B3 shows that at most b divisions are needed.

To find coefficients x, y with $\gcd(a, b) = ax + by$, we run the identity of Lemma 3.2.3 backwards.

Algorithm. **Back-substitution.** Write out the divisions $x_k = q_k y_k + r_k$ for $k = 0, \ldots, K-1$, so that $x_{k+1} = y_k$ and $y_{k+1} = r_k$. Start from $\gcd(a, b) = x_K = 1 \cdot x_K + 0 \cdot y_K$, a combination of the last pair. Then, for $k = K-1, K-2, \ldots, 0$ in turn, rewrite a combination $ux_{k+1} + vy_{k+1} = uy_k + vr_k$ of the pair with index $k+1$ as one of the pair with index k, by substituting $r_k = x_k - q_k y_k$:

$$uy_k + v(x_k - q_k y_k) = vx_k + (u - q_k v)y_k,$$

the identity in the proof of Lemma 3.2.3. After $k = 0$, the combination is one of $x_0 = a$ and $y_0 = b$. Multiply it out to check it. (By hand, the step $k = K - 1$ changes nothing, since $r_{K-1} = 0$; so one starts from the last division with a non-zero remainder, as in the example below.)

The procedure only computes; nothing rests on it. That some x, y exist is Theorem 3.2.1 (d), and the final multiplication checks the particular x, y found.

Example. Two runs of the algorithm

Compute $\gcd(252, 198)$ and $\gcd(-35, 15)$, and write each as a combination of the two numbers.

Solution. *The pair* $(252, 198)$. The chain of divisions is

$$\begin{aligned} 252 &= 1 \cdot 198 + 54, \\ 198 &= 3 \cdot 54 + 36, \\ 54 &= 1 \cdot 36 + 18, \\ 36 &= 2 \cdot 18 + 0. \end{aligned}$$

In the notation of the theorem, (x_k, y_k) runs through $(252, 198)$, $(198, 54)$, $(54, 36)$, $(36, 18)$, $(18, 0)$, so $K = 4$ and $\gcd(252, 198) = x_4 = 18$. Back-substitution, from the third division up:

$$\begin{aligned} 18 &= 54 - 1 \cdot 36 = 54 - (198 - 3 \cdot 54) = 4 \cdot 54 - 198 \\ &= 4 \cdot (252 - 198) - 198 = 4 \cdot 252 - 5 \cdot 198. \end{aligned}$$

Check: $4 \cdot 252 = 1008$, $5 \cdot 198 = 990$, and $1008 - 990 = 18$.

The pair $(-35, 15)$. The first entry may be negative; only the second must be positive. The remainder of -35 on division by 15 is 10, not -5:

$$\begin{aligned} -35 &= (-3) \cdot 15 + 10, \\ 15 &= 1 \cdot 10 + 5, \\ 10 &= 2 \cdot 5 + 0. \end{aligned}$$

So $\gcd(-35, 15) = 5$, and

$$5 = 15 - 10 = 15 - (-35 + 3 \cdot 15) = (-1) \cdot (-35) + (-2) \cdot 15.$$

Check: $(-1) \cdot (-35) = 35$, $(-2) \cdot 15 = -30$, and $35 - 30 = 5$.

Warning. Bézout coefficients are **not** unique. From $18 = 4 \cdot 252 - 5 \cdot 198$ and $11 \cdot 252 = 2772 = 14 \cdot 198$,

$$18 = (4 - 11) \cdot 252 + (-5 + 14) \cdot 198 = (-7) \cdot 252 + 9 \cdot 198.$$

So "the" coefficients of $\gcd(a, b)$ means nothing by itself; when a computation needs particular ones, they are those produced by back-substitution, and any other pair works as well.

Quick check. Compute $\gcd(0, -6)$ and $\gcd(84, 35)$.

Solution. $\gcd(0, -6) = 6$. Every integer divides 0, so the common divisors of 0 and -6 are the divisors of -6, which are those of 6 (Proposition 3.1.1 (d) and Proposition 2.3.2 (a)). The positive ones are at most 6 (Proposition 3.1.1 (e)), and 6 is one of them. (The theorem does not apply directly to $(0, -6)$, since its second entry is negative; it does apply to $(0, 6)$, with the one division $0 = 0 \cdot 6 + 0$.)

$\gcd(84, 35) = 7$, by the chain $84 = 2 \cdot 35 + 14$, $35 = 2 \cdot 14 + 7$, $14 = 2 \cdot 7 + 0$.

3.2.5 Exercises

3.2.5.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 3.2.1.

(a) Where is the minimality of d used?

(b) Run the proof on $(a, b) = (12, 42)$: find d, x_0, y_0, and the divisions of (b).

(c) Which step fails for $(a, b) = (0, 0)$?

(d) State the idea of the proof in two sentences.

Solution.

(a) Once, in (b): the remainder r of a on division by d is a combination of a and b, and if it were positive it would be an element of S smaller than d. Minimality forces $r = 0$.

(b) The set S contains $6 = 12 \cdot (-3) + 42 \cdot 1$. Every element of $12\mathbb{Z} + 42\mathbb{Z}$ is divisible by 6, since $6 \mid 12$ and $6 \mid 42$ (Proposition 3.1.1 (c)), so every element of S is at least 6 (Proposition 3.1.1 (e)). Hence $d = 6$, with $x_0 = -3$ and $y_0 = 1$. The divisions of (b) are $12 = 2 \cdot 6 + 0$ and $42 = 7 \cdot 6 + 0$: both remainders are 0, as the proof predicts.

(c) Step (a): $0\mathbb{Z} + 0\mathbb{Z} = \{0\}$, so $S = \emptyset$, and Proposition 2.3.6 does not apply. There is no d to divide by.

(d) Take the least positive integer combination d of a and b. Dividing a (or b) by d leaves a remainder that is again a combination and smaller than d, so it is 0; thus d is a common divisor, and since every common divisor divides the combination d, it is the greatest.

3.2.5.2 B. Practice

Exercise. B1

Run the Euclidean algorithm on $(1001, 357)$, and use back-substitution to write $\gcd(1001, 357)$ as $1001x + 357y$ with $x, y \in \mathbb{Z}$.

Solution. The chain of divisions is

$$\begin{aligned} 1001 &= 2 \cdot 357 + 287, \\ 357 &= 1 \cdot 287 + 70, \\ 287 &= 4 \cdot 70 + 7, \\ 70 &= 10 \cdot 7 + 0, \end{aligned}$$

so $\gcd(1001, 357) = 7$ by Theorem 3.2.4 (c). Back-substitution:

$$\begin{aligned} 7 &= 287 - 4 \cdot 70 = 287 - 4 \cdot (357 - 287) = 5 \cdot 287 - 4 \cdot 357 \\ &= 5 \cdot (1001 - 2 \cdot 357) - 4 \cdot 357 = 5 \cdot 1001 - 14 \cdot 357. \end{aligned}$$

Check: $5 \cdot 1001 = 5005$, $14 \cdot 357 = 4998$, and $5005 - 4998 = 7$. Hence $x = 5$, $y = -14$ works.

Exercise. B2

Let $a, b \in \mathbb{Z}$. Prove that $\gcd(a, b) = \gcd(-a, b)$, and that $\gcd(a, 0) = a$ if $a \in \mathbb{Z}_{>0}$.

Solution. First claim. If $(a, b) = (0, 0)$, then also $(-a, b) = (0, 0)$, and both sides are 0 by convention. Otherwise $(-a, b) \neq (0, 0)$ as well, since $-a = 0$ would give $a = -(-a) = -0 = 0$ (Proposition 2.3.2 (a); $-0 = 0$ because $0 + 0 = 0$ and the inverse is unique). By Proposition 3.1.1 (d), $c \mid a$ implies $c \mid -a$, and $c \mid -a$ implies $c \mid -(-a) = a$ (Proposition 2.3.2 (a)). So a and $-a$ have the same divisors, the pairs (a, b) and $(-a, b)$ have the same common divisors, and the greatest is the same (Definition 3.2.2).

Second claim. Let $a > 0$. Every integer divides 0 (Proposition 3.1.1 (a)), so the common divisors of a and 0 are the divisors of a. Among them is a itself. A positive divisor c satisfies $c \leq a$ by Proposition 3.1.1 (e), and a non-positive one satisfies $c \leq 0 < a$. So a is the greatest, and $\gcd(a, 0) = a$.

Exercise. B3: The algorithm stops quickly

In the notation of Theorem 3.2.4, prove that $y_k \leq b - k$ for every $k \in \mathbb{N}$ with $k \leq K$. Deduce that $K \leq b$: the algorithm stops after at most b divisions.

Solution. We use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k, for $P(k)$: "if $k \leq K$, then $y_k \leq b - k$". The base case $P(0)$ is $y_0 = b = b - 0$. Suppose $P(k)$, and let $k + 1 \leq K$. Then $k < K$, so $k \leq K$ and $P(k)$ gives $y_k \leq b - k$. By Theorem 3.2.4 (b), $y_{k+1} < y_k$, so $y_{k+1} + 1 \leq y_k$ by discreteness (Theorem 2.3.4 (c)), that is, $y_{k+1} \leq y_k - 1$. Hence $y_{k+1} \leq b - k - 1 = b - (k + 1)$, which is $P(k + 1)$.

With $k = K$: $0 = y_K \leq b - K$, so $K \leq b$. Each index $k < K$ is one division, so there are $K \leq b$ of them.

3.2.5.3 C. Going deeper

Exercise. C1: Exit check: Bézout

Let $a, b \in \mathbb{Z}$, not both 0, and let $S := (a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{Z}_{>0}$. Prove that S has a least element d; that $d \mid a$ and $d \mid b$; that every common divisor of a and b divides d; and that $d = \gcd(a, b)$.

Solution. S has a least element. Say $a \neq 0$ (otherwise exchange the roles of a and b). Both $a = a \cdot 1 + b \cdot 0$ and $-a = a \cdot (-1) + b \cdot 0$ are combinations of a and b, and one of them is positive (Proposition 3.1.1 (h)); so $S \neq \emptyset$. It is bounded below by 0, so it has a least element d (Proposition 2.3.6), and $d = ax_0 + by_0$ for some $x_0, y_0 \in \mathbb{Z}$.

d divides a and b. Divide a by $d > 0$ (Theorem 3.1.2 (a)): $a = qd + r$ with $0 \leq r < d$. Then $r = a(1 - qx_0) + b(-qy_0)$ is a combination of a and b. If $r > 0$, then $r \in S$ and $r < d$, against the minimality of d. So $r = 0$ and $d \mid a$. The same argument with b gives $d \mid b$.

Common divisors divide d. If $c \mid a$ and $c \mid b$, then $c \mid ax_0 + by_0 = d$ (Proposition 3.1.1 (c)).

d is the gcd. d is a common divisor. A common divisor $c > 0$ divides $d > 0$, so $c \leq d$ (Proposition 3.1.1 (e)); a common divisor $c \leq 0$ satisfies $c < d$. So d is the greatest common divisor (Definition 3.2.2).

Marking scheme. (10 points) - 2: shows $S \neq \emptyset$, with a reason that covers $a < 0$ or $b < 0$ (for instance, one of a, $-a$ is a positive combination). - 1: obtains the least element by a cited well-ordering result for sets of integers bounded below. - 3: divides a by d and writes the remainder as a combination of a and b. - 2: uses the minimality of d to force the remainder to be 0, for a and for b. - 2: shows every common divisor divides d, and deduces that d is the greatest (with the case of non-positive divisors). Fatal errors (0 for the step): assuming that the gcd is a combination of a and b (circular); citing Theorem 3.2.4 (c) or the gcd clause of Lemma 3.2.3 (both are proved from Theorem 3.2.1, so this is circular); "d divides a because d is the least positive combination", with no division step. A second route goes through the Euclidean algorithm, with back-substitution, by induction on the number of steps. Citing Theorem 3.2.4 (a), (b) and the equality $a\mathbb{Z} + b\mathbb{Z} = b\mathbb{Z} + r\mathbb{Z}$ of Lemma 3.2.3 is allowed; if the invariant used is "the common divisors are unchanged", it must be proved directly from $r = a - qb$. The invariant earns the 3 points for the division, and termination earns the 2 points for minimality. The answer must still identify the gcd with the least element d of S ($d \leq \gcd(a, b)$ since $\gcd(a, b) \in S$; $\gcd(a, b) \mid d$ since the gcd divides every combination). The rest is marked as written.

Exercise. C2: Scaling the gcd

Let $a, b \in \mathbb{Z}$ and $k \in \mathbb{Z}_{>0}$. Prove that $\gcd(ka, kb) = k \gcd(a, b)$.

Solution (Hint 1). Use the description of $a\mathbb{Z} + b\mathbb{Z}$ in Theorem 3.2.1 (e), and the description of the gcd as a least element in (a) and (d).

Solution (Hint 2). If $(a, b) \neq (0, 0)$ and $d := \gcd(a, b)$, first show $(ka)\mathbb{Z} + (kb)\mathbb{Z} = (kd)\mathbb{Z}$.

Solution. If $(a, b) = (0, 0)$, then $(ka, kb) = (0, 0)$ by Proposition 2.3.2 (b), and both sides are 0: $\gcd(0, 0) = 0 = k \cdot 0$.

Otherwise let $d := \gcd(a, b)$, so $d \geq 1$. Say $a \neq 0$ (otherwise $b \neq 0$; exchange the roles of a and b); then $ka \neq 0$, since $k \neq 0$ and $\mathbb{Z}$ has no zero divisors (Proposition 2.3.5 (f)). So $(ka, kb) \neq (0, 0)$, and Theorem 3.2.1 applies to both pairs.

The combinations. For $x, y \in \mathbb{Z}$, $(ka)x + (kb)y = k(ax + by)$. As x, y range over $\mathbb{Z}$, $ax + by$ ranges over $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$ (Theorem 3.2.1 (e)), so

$$(ka)\mathbb{Z} + (kb)\mathbb{Z} = \{k(md) : m \in \mathbb{Z}\} = \{m(kd) : m \in \mathbb{Z}\} = (kd)\mathbb{Z}.$$

The least positive element. $kd > 0$ by (O4), and $kd = 1 \cdot kd \in (kd)\mathbb{Z}$. Every positive element $m(kd)$ of $(kd)\mathbb{Z}$ is divisible by kd, so $kd \leq m(kd)$ by Proposition 3.1.1 (e). So kd is the least element of $((ka)\mathbb{Z} + (kb)\mathbb{Z}) \cap \mathbb{Z}_{>0}$, and by Theorem 3.2.1 (a) and (d), $\gcd(ka, kb) = kd = k \gcd(a, b)$.

For instance, $\gcd(12, 42) = 6\gcd(2, 7) = 6$, since 2 and 7 are coprime ($2 \cdot (-3) + 7 \cdot 1 = 1$, Corollary 3.2.2).

Marking scheme. (10 points) - 2: treats $(a, b) = (0, 0)$ by the convention, and checks $(ka, kb) \neq (0, 0)$ otherwise, with a reason (no zero divisors). - 4: shows $(ka)\mathbb{Z} + (kb)\mathbb{Z} = (kd)\mathbb{Z}$, citing Theorem 3.2.1 (e) for (a, b). - 2: shows that kd is the least positive element of $(kd)\mathbb{Z}$. - 2: concludes $\gcd(ka, kb) = kd$ through Theorem 3.2.1 (a), (d) for (ka, kb). Fatal errors (0 for the step): asserting that every common divisor of ka and kb is k times a common divisor of a and b (false: 1 divides $6 = 3 \cdot 2$ and $9 = 3 \cdot 3$, and 1 is not 3 times an integer); using prime factorizations, which this section does not have. A second route, showing kd is a common divisor and every common divisor c divides kd because $c \mid (ka)x_0 + (kb)y_0 = kd$, earns the same total: 2 for treating $(a, b) = (0, 0)$ by the convention and checking $(ka, kb) \neq (0, 0)$ otherwise (no zero divisors); 3 for showing kd is a common divisor and a combination ($kd = (ka)x_0 + (kb)y_0$ with $ax_0 + by_0 = d$ from Theorem 3.2.1); 3 for every common divisor dividing kd; 2 for concluding kd is the greatest ($kd > 0$, rule (e) for positive divisors, and the non-positive divisors) via Definition 3.2.2.

3.3 Euclid's Lemma and the Fundamental Theorem of Arithmetic

Section 3.2 gave every pair of integers, not both zero, a greatest common divisor, and wrote it as a combination $ax + by$ (Theorem 3.2.1). This section spends that identity. Take 60. It splits as $2 \cdot 30$, as $6 \cdot 10$, as $4 \cdot 15$, and every way of splitting it further, until no factor splits again, ends with the same primes $2, 2, 3, 5$. *Sets, Relations, and Cardinality*, Section 2.2 proved, for the school integers, that every integer at least 2 has a prime factor, and set the uniqueness of the full splitting aside. Here we prove both, for the integers of Chapter 2. The key is a property of primes that the definition ("no positive divisors except 1 and itself") does not obviously give: a prime that divides a product divides one of the factors.

As in Section 3.1, numerals are the natural numbers of those names, and their sums and products may be computed in $\mathbb{N}$ (Proposition 2.2.3 (a)).

3.3.1 Euclid's lemma

The definition of a prime p speaks about the divisors **of** p. The property we want speaks about the numbers that p **divides**. Something has to connect the two, and the connection is the Bézout identity of Section 3.2. We first prove a version for any divisor that shares no factor with one of the two numbers multiplied.

Theorem 3.3.1. Euclid's Lemma

Let $a, b, c \in \mathbb{Z}$. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.

Idea. The hypothesis $\gcd(a, b) = 1$ is a statement about common divisors, which is hard to use directly. By Corollary 3.2.2 it is the same as an equation $1 = ax + by$, and an equation can be multiplied. Multiplying it by c writes c as $acx + bcy$. The first term is a multiple of a on sight, and the second is one because bc is.

Proof. Suppose $a \mid bc$ and $\gcd(a,b) = 1$. By Corollary 3.2.2 there are $x, y \in \mathbb{Z}$ with $ax + by = 1$. Since $a \mid bc$, there is $k \in \mathbb{Z}$ with $bc = ka$ (Sets, Relations, and Cardinality, Definition 2.2.1). Multiplying the identity by c, with (F7), (F9), (F5) and (F6),

$$c = c(ax + by) = a(cx) + (bc)y = a(cx) + a(ky) = a(cx + ky).$$

Since $cx + ky \in \mathbb{Z}$, this shows $a \mid c$.

■

We call the move of this proof **multiply the Bézout identity**: from $1 = ax + by$, multiply by c and read off what divides c. It turns a statement about divisors into an equation, which is where the work can be done.

Each hypothesis has a job. Without $\gcd(a,b) = 1$ the conclusion fails: $4 \mid 2 \cdot 6 = 12$, but $4 \nmid 6$, since $6 = 1 \cdot 4 + 2$ leaves a non-zero remainder (uniqueness in Theorem 3.1.2 (a)); here $\gcd(4,2) = 2$. The common factor 2 of 4 and 2 let 4 divide the product without dividing 6.

3.3.2 Lists, products and prime factorizations

To say that a factorization is unique, we must first say what a factorization is. The products $2 \cdot 2 \cdot 3 \cdot 5$ and $3 \cdot 2 \cdot 5 \cdot 2$ are equal and use the same primes, and we want to count them as one. We do so by writing the primes in order, smallest first. A factorization also has a length, and the length may be 0: the number 1 is the product of no primes at all.

Products of lists are defined by the recursion of Sets, Relations, and Cardinality, Definition 2.3.1. That definition was made for sequences of real numbers, and we use it as a pattern, as group (iii) of the citing rule of Section 1.7 allows, with the ordered field $\mathbb{Q}$ (Theorem 2.5.1) in place of $\mathbb{R}$. A product of integers is an integer (Proposition 2.4.4), so the products of lists below lie in $\mathbb{Z}$; the paragraph "Why the product exists" below checks directly that the recursion runs in $\mathbb{Z}$.

Definition 3.3.1. List, Product of a List, Prime Factorization

Let $k \in \mathbb{N}$. A **list** of length k in $\mathbb{Z}$ is a function $p\colon \{1, \dots, k\} \to \mathbb{Z}$, written $(p_1, \dots, p_k)$ with $p_i := p(i)$; for $k = 0$ it is the **empty list**. Its **product** $\prod_{i=1}^{k} p_i \in \mathbb{Z}$ is given by the recursion

$$\prod_{i=1}^{0} p_i = 1, \qquad \prod_{i=1}^{j+1} p_i = \Big(\prod_{i=1}^{j} p_i\Big) p_{j+1} \quad \text{for } j < k,$$

where $\prod_{i=1}^{j} p_i$ is the product of the shortened list $(p_1, \dots, p_j)$.

A **prime factorization** of $n \in \mathbb{Z}_{>0}$ is a list $(p_1, \dots, p_k)$ such that **every** entry p_i is prime (Sets, Relations, and Cardinality, Definition 2.2.2), the list is **non-decreasing** ($i \le j$ implies $p_i \le p_j$), and $\prod_{i=1}^{k} p_i = n$.

In words: a list is a finite sequence of integers, possibly empty; its product starts at 1 and multiplies by one more entry at each stage; and a prime factorization lists primes in non-decreasing order with the right product.

Why the product exists. Extend the list to a sequence $\bar{p}$ on $\mathbb{Z}_{>0}$ by $\bar{p}_i := p_i$ for $i \le k$ and $\bar{p}_i := 1$ for $i > k$. Then Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ through

Corollary 1.7.3 (b), with $X = \mathbb{Z}$, starting value 1 and step $g(j, x) := x\bar{p}_{j+1}$, gives exactly one function $P\colon \mathbb{N} \to \mathbb{Z}$ with $P_0 = 1$ and $P_{j+1} = P_j\bar{p}_{j+1}$; we set $\prod_{i=1}^{j} p_i := P_j$ for $j \leq k$. The "Finite lists" remark after Sets, Relations, and Cardinality, Definition 2.3.1 shows that P_j depends only on $p_1, \ldots, p_j$, not on how the list was extended. Its proof is an induction on j that uses only the recursion and the multiplication of the target set, so it runs unchanged in $\mathbb{Q}$, and, since $\mathbb{Z}$ is closed under multiplication, with target set $\mathbb{Z}$. In particular P_j is also the product of the shortened list $(p_1, \ldots, p_j)$, which is the displayed recursion. We cite it as **the recursion of Definition 3.3.1**: it peels the **last** entry off a product.

Examples.

- $(2, 2, 3, 5)$ is a prime factorization of 60. Its entries are prime by the small-primes remark of Section 3.1; it is non-decreasing; and the recursion gives the products $1, 2, 4, 12, 60$.
- (5) is a prime factorization of 5, with products $1, 5$. A prime p has the one-entry factorization (p).
- The degenerate case: the empty list is a prime factorization of 1. Its product is 1 by the first clause of the recursion, and the conditions "every entry is prime" and "non-decreasing" hold vacuously. Without this case the number 1 would have no factorization, and the theorem below would fail for $n = 1$; its existence proof uses this case to rule out $n = 1$.

Non-examples by minimal change. Exchange two entries of the first example: $(2, 3, 2, 5)$ has prime entries and product 60, but it is not non-decreasing, since positions $2 \leq 3$ carry $3 > 2$. Merge two entries: $(2, 2, 15)$ is non-decreasing with product 60, but $15 = 3 \cdot 5$ is not prime.

Why these clauses. Without "non-decreasing", 60 would have twelve factorizations, its four primes in every order, and uniqueness could only be stated up to rearrangement, which needs a theory of permutations. Without "prime", splitting would stop anywhere. And the requirement $p \geq 2$ in Sets, Relations, and Cardinality, Definition 2.2.2 is what keeps $(1, 2, 2, 3, 5)$ out: if 1 counted as a prime, every factorization could be padded with as many 1s as we like.

Quick check. Find the prime factorization of 360, and check each clause of Definition 3.3.1.

Solution. $(2, 2, 2, 3, 3, 5)$. Its entries are prime by the small-primes remark of Section 3.1. It is non-decreasing. Its products, by the recursion, are $1, 2, 4, 8, 24, 72, 360$. By the theorem below there is no other.

3.3.3 Primes dividing products

For a prime, the coprimality that Euclid's lemma asks for comes almost free: a prime shares no factor with a number unless it divides that number.

Corollary 3.3.2. A Prime Dividing a Product

Let p be prime.

(a) For every $a \in \mathbb{Z}$: $p \mid a$ or $\gcd(p, a) = 1$.

(b) For all $b, c \in \mathbb{Z}$: if $p \mid bc$, then $p \mid b$ or $p \mid c$.

(c) If $(q_1, \ldots, q_l)$ is a list of primes, with $l \in \mathbb{N}$, and $p \mid \prod_{j=1}^{l} q_j$, then $p = q_j$ for some $j \in \{1, \ldots, l\}$.

Idea. For (a), the gcd of p and a is a positive divisor of p, so it is 1 or p, and the second case says $p \mid a$. Then (b) is Euclid's lemma in the case $\gcd(p, b) = 1$ of (a). For (c), peel the last prime off the product and apply (b) to the two factors; an induction on the length handles the rest, and a prime that divides a prime is equal to it.

Proof. *(a)* Since $p \geq 2$, $(p, a) \neq (0, 0)$, so $g := \gcd(p, a) \geq 1$ (Definition 3.2.2). As a common divisor, $g \mid p$ and $g \mid a$. So g is a positive divisor of the prime p, and $g = 1$ or $g = p$ (Sets, Relations, and Cardinality, Definition 2.2.2). If $g = p$, then $p \mid a$.

(b) Let $p \mid bc$. By (a), $p \mid b$ or $\gcd(p, b) = 1$. In the second case Theorem 3.3.1, with $a = p$, gives $p \mid c$.

(c) We use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on l, for $P(l)$: "for every list of primes $(q_1, \ldots, q_l)$, if $p \mid \prod_{j=1}^{l} q_j$, then $p = q_j$ for some $j \leq l$".

Base case $P(0)$. The empty product is 1 (the recursion of Definition 3.3.1). If $p \mid 1$, then $p = 1$ or $p = -1$ (Proposition 3.1.1 (g)), and both contradict $p \geq 2$. So the hypothesis of $P(0)$ never holds, and $P(0)$ is true.

Step. Assume $P(l)$, and let $(q_1, \ldots, q_{l+1})$ be a list of primes with $p \mid \prod_{j=1}^{l+1} q_j$. By the recursion of Definition 3.3.1,

$$\prod_{j=1}^{l+1} q_j = \Big(\prod_{j=1}^{l} q_j\Big) q_{l+1},$$

where the first factor is the product of the list of primes $(q_1, \ldots, q_l)$. By (b), p divides that first factor or $p \mid q_{l+1}$. In the first case $P(l)$ gives $j \leq l$ with $p = q_j$. In the second, $p \geq 2 > 0$ is a positive divisor of the prime q_{l+1}, so $p \in \{1, q_{l+1}\}$, and $p \neq 1$; hence $p = q_{l+1}$. This proves $P(l + 1)$, and with it the corollary.

■

Part (b) is the property promised at the start of the section, and it is what makes primes the right atoms for multiplication. It fails for 6, which is not prime: $6 \mid 2 \cdot 3$, but $6 \nmid 2$ and $6 \nmid 3$, since $0 < 2 < 6$ and $0 < 3 < 6$ rule both out by Proposition 3.1.1 (e).

The second application of Euclid's lemma combines two coprime divisors into one.

Proposition 3.3.3. Coprime Divisors of a Number

Let $m, n, c \in \mathbb{Z}$ with $m \mid c$, $n \mid c$ and $\gcd(m, n) = 1$. Then $mn \mid c$.

Proof. Write $c = km$ with $k \in \mathbb{Z}$. Then $n \mid km$, and $\gcd(n, m) = \gcd(m, n) = 1$ (Definition 3.2.2), so Theorem 3.3.1 gives $n \mid k$, say $k = ln$ with $l \in \mathbb{Z}$. Hence $c = (ln)m = l(mn)$, by (F5) and (F6), and $mn \mid c$.

■

The coprimality is needed: $4 \mid 12$ and $6 \mid 12$, but $24 \nmid 12$, since a positive divisor of 12 is at most 12 (Proposition 3.1.1 (e)); here $\gcd(4, 6) = 2$.

3.3.4 The least divisor is prime

The existence half of unique factorization needs a supply of primes: every integer $n \geq 2$ must have a prime factor. *Sets, Relations, and Cardinality*, Section 2.2 proved this for the school integers by strong induction (Sets, Relations, and Cardinality, Theorem 2.2.5). On the constructed $\mathbb{Z}$ the least-element property gives a shorter route: the least divisor bigger than 1 is already prime.

Lemma 3.3.4. The Least Divisor Is Prime

Let $n \in \mathbb{Z}$ with $n \geq 2$. The set $D := \{d \in \mathbb{Z} : d \geq 2 \text{ and } d \mid n\}$ has a least element, and that element is prime. In particular, n has a prime factor.

Proof. Since $n \geq 2$ and $n \mid n$ (Proposition 3.1.1 (a)), $n \in D$, so D is non-empty; and 2 is a lower bound of D by its definition. By Proposition 2.3.6, D has a least element p.

Suppose p were not prime. Since $p \geq 2$, the clause of Sets, Relations, and Cardinality, Definition 2.2.2 that fails is the one about divisors: p has a positive divisor e with $e \neq 1$ and $e \neq p$. Then $e \geq 1$ by discreteness (Theorem 2.3.4 (c)), and $e \leq p$ by Proposition 3.1.1 (e). So $1 < e < p$, and $e \geq 2$ by discreteness again. Also $e \mid p$ and $p \mid n$ give $e \mid n$ (Proposition 3.1.1 (b)). So $e \in D$ and $e < p$, against the choice of p. Hence p is prime, and $p \mid n$ because $p \in D$.

■

Remark. This is Sets, Relations, and Cardinality, Theorem 2.2.5 on the constructed $\mathbb{Z}$, proved by a least element instead of strong induction. It also shortens the primality test of Section 3.1's small-primes remark. If $p \geq 2$ is not prime, its least divisor $e \geq 2$ is prime by the lemma, so it is not p itself, and $p = ef$ with $f \in \mathbb{Z}$; then $f > 0$ (as $e, p > 0$), $f \neq 1$, so $f \geq 2$ and $f \mid p$, and therefore $e \leq f$ by the choice of e. Multiplying by $e > 0$ gives $e \cdot e \leq ef = p$ (Proposition 2.3.5 (c), or equality if $e = f$). So to prove $p \geq 2$ prime it is enough to check that **no** d with $d \geq 2$ and $d \cdot d \leq p$ divides p. For 11 only $d = 2, 3$ need checking, since $4 \cdot 4 = 16 > 11$.

3.3.5 The fundamental theorem

Everything is now in place for the main result of the chapter's first half.

Theorem 3.3.5. Fundamental Theorem of Arithmetic

Every $n \in \mathbb{Z}_{>0}$ has **exactly one** prime factorization.

Idea. Existence. Take the least n with no factorization. It is at least 2, so it has prime factors, and a largest one, p. Then $n = pm$ with $m < n$, so m has a factorization. Every entry of it divides n, so it is at most p, and appending p at the end keeps the list non-decreasing.

Uniqueness. Take two factorizations with the same product, $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_l)$. The largest entry p_k divides the product of the q's, so by Corollary 3.3.2 (c) it **is** one of the q_j, and that is at most the largest, q_l. By symmetry $q_l \leq p_k$. So the two last entries are equal; cancel them and induct on the length.

Before reading on: why compare the largest primes, p_k and q_l, and not the first ones?

Proof. Existence. Let $B \coloneqq \{n \in \mathbb{Z}_{>0} : n \text{ has no prime factorization}\}$, and suppose $B \neq \emptyset$. It has the lower bound 0, so it has a least element n (Proposition 2.3.6). The empty list is a prime factorization of 1, so $n \neq 1$; as $n \geq 1$ (discreteness, Theorem 2.3.4 (c)), $n > 1$, and so $n \geq 2$ by discreteness.

The set F of prime factors of n is non-empty by Lemma 3.3.4. Each $q \in F$ is positive and divides $n > 0$, so $q \leq n$ (Proposition 3.1.1 (e)). By Proposition 2.3.6, F has a greatest element p. Write $n = mp$ with $m \in \mathbb{Z}$. If $m \leq 0$, then $mp \leq 0$ ($m = 0$: Proposition 2.3.2 (b); $m < 0$: $mp < 0 \cdot p = 0$ by Proposition 2.3.5 (c)), against $n > 0$. So $m > 0$, that is, $m \in \mathbb{Z}_{>0}$ (Theorem 2.3.4 (b)). Moreover $m < n$: since $p \geq 2$ and $m > 0$, $n = pm \geq 2m = m + m > m$, by Proposition 2.3.5 (c) (or equality if $p = 2$), (F9) with (F7), and (O3).

By the minimality of n, $m \notin B$: let $(q_1, \ldots, q_l)$ be a prime factorization of m. If $l \geq 1$, the recursion of Definition 3.3.1 gives $m = (\prod_{j=1}^{l-1} q_j)q_l$, so $q_l \mid m$; with $m \mid n$ this gives $q_l \mid n$ (Proposition 3.1.1 (b)), so $q_l \in F$ and $q_l \leq p$. As the list is non-decreasing, $q_i \leq q_l \leq p$ for every $i \leq l$. Now let $(q_1, \ldots, q_l, p)$ be the list of length $l + 1$ whose first l entries are those of $(q_1, \ldots, q_l)$ and whose last entry is p. Its entries are prime. It is non-decreasing: two positions among the first l compare as before, and every entry is at most p (for $l = 0$ it is the one-entry list (p), and there is nothing to check). By the recursion its product is

$$\Big(\prod_{j=1}^{l} q_j\Big)p = mp = n.$$

So n has a prime factorization, against $n \in B$. Hence $B = \emptyset$.

Uniqueness. We prove, by Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k, the statement $P(k)$: "for every $l \in \mathbb{N}$, if $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_l)$ are non-decreasing lists of primes with $\prod_{i=1}^{k} p_i = \prod_{j=1}^{l} q_j$, then $l = k$ and $p_i = q_i$ for every $i \leq k$". Two prime factorizations of the same n are such lists, so $P(k)$ for every k gives the uniqueness.

First, a non-empty list of primes $(r_1, \ldots, r_h)$ does not have product 1: by the recursion its product is $(\prod_{i=1}^{h-1} r_i)r_h$, so r_h would divide 1, and $r_h = \pm 1$ (Proposition 3.1.1 (g)) contradicts $r_h \geq 2$.

Base case $P(0)$. If $k = 0$, then $\prod_{j=1}^{l} q_j = 1$, so $l = 0$ by the previous paragraph, and both lists are empty.

Step. Assume $P(k)$. Let $(p_1, \ldots, p_{k+1})$ and $(q_1, \ldots, q_l)$ be non-decreasing lists of primes with the same product N. Since $k+1 \geq 1$, $N \neq 1$, so $l \geq 1$. By the recursion, $p_{k+1} \mid N = \prod_{j=1}^{l} q_j$, so Corollary 3.3.2 (c) gives $p_{k+1} = q_j$ for some $j \leq l$, and $q_j \leq q_l$ because the q-list is non-decreasing. In the same way $q_l \mid N = \prod_{i=1}^{k+1} p_i$ gives $q_l = p_i \leq p_{k+1}$ for some $i \leq k + 1$. So $p_{k+1} \leq q_l \leq p_{k+1}$, and $p_{k+1} = q_l$ (Proposition 2.3.5 (i)). By the recursion,

$$\Big(\prod_{i=1}^{k} p_i\Big)p_{k+1} = N = \Big(\prod_{j=1}^{l-1} q_j\Big)p_{k+1}.$$

Subtracting, $(\prod_{i=1}^{k} p_i - \prod_{j=1}^{l-1} q_j)p_{k+1} = 0$ by (F9) and Proposition 2.3.2 (c), and $p_{k+1} \neq 0$, so the bracket is 0 (Proposition 2.3.5 (f)). The shortened lists $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_{l-1})$ are non-decreasing lists of primes with equal products, so $P(k)$, applied with $l - 1 \in \mathbb{N}$,

gives $l - 1 = k$ and $p_i = q_i$ for $i \leq k$. With $p_{k+1} = q_l = q_{k+1}$, the two lists have the same length $k+1$ and the same entries, so they are equal (Sets, Relations, and Cardinality, Proposition 1.4.1). This proves $P(k+1)$, and the theorem.

■

The answer to the prompt lies in the recursion of Definition 3.3.1. It peels off the **last** factor, so once the two last entries are known to be equal they can be canceled at once, leaving two shortened lists to which $P(k)$ applies. And the comparison "largest with largest" needs only that both lists are non-decreasing. Comparing the first entries would also show $p_1 = q_1$, but removing a first entry shifts every index, and the recursion says nothing directly about that.

Remark. *Sets, Relations, and Cardinality*, Section 2.2 named unique factorization as a fact it did not prove or use, and Sets, Relations, and Cardinality, Exercise (B2: Putting a proof in order) lists, as line (vii), a step that cites it and is therefore not available there. On the constructed $\mathbb{Z}$ the fact is now proved, and Exercise R2 of Section 3.5 uses it.

Warning. **Uniqueness is not automatic.** Let $H := \{4k+1 : k \in \mathbb{N}\} = \{1, 5, 9, 13, 17, 21, \dots\}$. It is closed under multiplication, since $(4k+1)(4l+1) = 4(4kl+k+l)+1$. Call $h \in H$ with $h > 1$ an **H-prime** if it is not a product of two elements of H both greater than 1. Then 9, 21 and 49 are H-primes, and

$$441 = 9 \cdot 49 = 21 \cdot 21$$

gives two different factorizations into H-primes. What breaks is Euclid's lemma: 9 divides $21 \cdot 21$ inside H, but $9 \nmid 21$. The proof of Theorem 3.3.1 wrote $1 = ax + by$, with **sums** and with negative coefficients, and H is closed under neither. The proof of uniqueness above is only as good as Euclid's lemma.

3.3.6 Lowest terms

Every rational number is a fraction in many ways: $6/4 = 3/2 = 9/6$. Section 2.4 found, in its Exercise C1, a least positive denominator by the least-element property, and *Sets, Relations, and Cardinality*, Section 2.2 removed common factors of 2 only (Sets, Relations, and Cardinality, Lemma 2.2.4). With the gcd we can remove every common factor at once, and Euclid's lemma shows that the result is unique. Recall that after the identification of Section 2.4, $p/q := pq^{-1}$ in $\mathbb{Q}$ for $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$, and $p/q = r/s$ holds if and only if $ps = rq$ (Proposition 2.4.4 (b)).

Proposition 3.3.6. Lowest Terms

Let $x \in \mathbb{Q}$.

(a) There are $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$ with $x = p/q$ and $\gcd(p, q) = 1$, and the pair (p, q) is **unique**.

(b) If $x = r/s$ with $r \in \mathbb{Z}$ and $s \in \mathbb{Z}_{>0}$, then $r = tp$ and $s = tq$ for some $t \in \mathbb{Z}_{>0}$. In particular $q \mid s$ and $q \leq s$: q is the least positive denominator of x.

Idea. For existence, divide numerator and denominator by their gcd d; the Bézout identity for d, divided by d, is a Bézout identity $1 = px_0 + qy_0$ for the new pair. For (b), cross-multiplying gives $ps = rq$, so $q \mid ps$, and Euclid's lemma removes p. Uniqueness is (b) applied to a second coprime pair.

Proof. Existence in (a). By Proposition 2.4.4 (b), $x = r/s$ for some $r \in \mathbb{Z}$, $s \in \mathbb{Z}_{>0}$. Since $s \neq 0$, $d \coloneqq \gcd(r, s) \geq 1$, and by Theorem 3.2.1 (d) there are $x_0, y_0 \in \mathbb{Z}$ with $d = rx_0 + sy_0$. As $d \mid r$ and $d \mid s$, write $r = dp$ and $s = dq$ with $p, q \in \mathbb{Z}$. If $q \leq 0$, then $dq \leq 0$ (as in the existence proof of Theorem 3.3.5), against $s > 0$; so $q \in \mathbb{Z}_{>0}$. Now

$$d \cdot 1 = d = (dp)x_0 + (dq)y_0 = d(px_0 + qy_0),$$

so $d(px_0 + qy_0 - 1) = 0$, and since $d \neq 0$, $px_0 + qy_0 = 1$ (Proposition 2.3.5 (f)). Hence $\gcd(p, q) = 1$ by Corollary 3.2.2. Finally $rq = dpq = p(dq) = ps$, so $r/s = p/q$ by Proposition 2.4.4 (b), and $x = p/q$.

(b), for any such pair. Let $x = p/q$ with $\gcd(p, q) = 1$, and $x = r/s$ with $s > 0$. By Proposition 2.4.4 (b), $ps = rq$. So $q \mid ps$, and $\gcd(q, p) = 1$, hence $q \mid s$ by Theorem 3.3.1; write $s = tq$ with $t \in \mathbb{Z}$. Since $s > 0$ and $q > 0$, $t > 0$, by the sign argument used for q above. Then $q(pt) = ps = rq$, so $q(pt - r) = 0$, and $r = tp$ because $q \neq 0$ (Proposition 2.3.5 (f)). As q and s are positive, $q \mid s$ gives $q \leq s$ (Proposition 3.1.1 (e)).

Uniqueness in (a). Let (r, s) be a second pair with $x = r/s$, $s > 0$ and $\gcd(r, s) = 1$. By (b), $r = tp$ and $s = tq$ with $t > 0$. So t is a common divisor of r and s, and it divides $\gcd(r, s) = 1$ (Theorem 3.2.1 (c), (d)). Hence $t = 1$ (Proposition 3.1.1 (g), with $t > 0$), and $(r, s) = (p, q)$.

■

So the least denominator $N(x)$ of Section 2.4's Exercise C1 is the q of (a), and every other denominator of x is a multiple of it. For example $252/198$ has $\gcd(252, 198) = 18$ (Example (Two runs of the algorithm)), and $252 = 18 \cdot 14$, $198 = 18 \cdot 11$, so its lowest terms are $14/11$.

Remark. The twins in *Sets, Relations, and Cardinality*, Section 2.2 are Sets, Relations, and Cardinality, Lemma 2.2.4, which removes only factors of 2, and Sets, Relations, and Cardinality, Exercise (C2: Lowest terms), which asks for lowest terms in the full sense, for the school $\mathbb{Q}$. Proposition 3.3.6 is the full statement, with uniqueness, for the constructed $\mathbb{Q}$.

3.3.7 Exercises

3.3.7.1 A. Check your understanding

Exercise. A1

These questions are about the uniqueness half of the proof of Theorem 3.3.5.

(a) Why must the empty list count as a prime factorization?

(b) Run the induction step on $(p_1, p_2, p_3, p_4) = (2, 2, 3, 5)$ and a second prime factorization

$(q_1, \ldots, q_l)$ of 60. Which equalities does the proof derive, and to which two lists is $P(3)$ applied?

(c) Which line of the proof uses Euclid's lemma, and through which result?

(d) Where is "non-decreasing" used?

(e) Give the idea of the uniqueness proof in two sentences.

Solution.

(a) Otherwise the number 1 would have no prime factorization, and the theorem would be false for $n = 1$. The existence proof uses it to exclude $n = 1$ from B. (The base case $P(0)$ of the uniqueness proof concerns the empty list as a list of primes with product 1, which it is in any case.)

(b) The step first gets $l \geq 1$, since the product is $60 \neq 1$. Then $p_4 = 5$ divides $\prod q_j$, so $5 = q_j \leq q_l$ for some j; and q_l divides $60 = 2 \cdot 2 \cdot 3 \cdot 5$, so $q_l = p_i \leq 5$ for some i. Hence $q_l = 5$. Canceling 5 gives $\prod_{j=1}^{l-1} q_j = 2 \cdot 2 \cdot 3 = 12$, and $P(3)$ is applied to $(2, 2, 3)$ and $(q_1, \ldots, q_{l-1})$, which gives $l - 1 = 3$ and $(q_1, q_2, q_3) = (2, 2, 3)$.

(c) The lines "$p_{k+1} = q_j$ for some j" and "$q_l = p_i$ for some i". They cite Corollary 3.3.2 (c), whose step uses (b), which is Theorem 3.3.1 applied with $a = p$.

(d) In $q_j \leq q_l$ and in $p_i \leq p_{k+1}$: the entry found in the other list is at most that list's last entry.

(e) The largest prime of one factorization divides the other product, so by Euclid's lemma it is an entry there, hence at most the other's largest; by symmetry the two largest entries are equal. Canceling them leaves two shorter factorizations with equal products, and induction on the length finishes.

3.3.7.2 B. Practice

Exercise. B1: Put the proof in order

The following lines contain a proof of Theorem 3.3.1, shuffled, together with two lines that do not belong. Write the seven correct lines in an order that makes a proof, say which other orders are also admissible, and say why each of the other two lines fails.

(i) Suppose $a \mid bc$ and $\gcd(a, b) = 1$.

(ii) There are $x, y \in \mathbb{Z}$ with $ax + by = 1$.
(ii′) Since $\gcd(a, b) = 1$ and $a \mid bc$, we have $a \mid b$ or $a \mid c$.

(iii) Multiplying by c: $c = acx + bcy$.
(iii′) Dividing $1 = ax + by$ by a: $1/a = x + (b/a)y$.

(iv) $a \mid acx$, since $acx = a(cx)$.

(v) Write $bc = ak$; then $bcy = a(ky)$, so $a \mid bcy$.

(vi) Hence $a \mid acx + bcy$.

(vii) So $a \mid c$.

Solution. An admissible order is (i), (ii), (iii), (iv), (v), (vi), (vii):

Suppose $a \mid bc$ and $\gcd(a, b) = 1$. There are $x, y \in \mathbb{Z}$ with $ax + by = 1$ (Corol-

lary 3.2.2). Multiplying by c: $c = acx + bcy$. Then $a \mid acx$, since $acx = a(cx)$. Write $bc = ak$; then $bcy = a(ky)$, so $a \mid bcy$. Hence $a \mid acx+bcy$ (Proposition 3.1.1 (c)). So $a \mid c$.

Line (i) comes first, and (ii) needs its hypothesis $\gcd(a, b) = 1$. Lines (iii), (iv) and (v) each need only (ii): (iii) multiplies its equation, and (iv) and (v) mention its x and y; (v) also uses the hypothesis $a \mid bc$ of (i). None of these three uses another of them. Line (vi) needs both (iv) and (v), but not (iii), since $acx + bcy$ is just an expression in a, b, c, x, y. Line (vii) rewrites (vi) through the equation of (iii), so it needs both and comes last. The admissible orders are therefore those that start (i), (ii), end with (vii), and place (vi) after (iv) and (v); there are eight. For example, (i), (ii), (iv), (v), (vi), (iii), (vii) is admissible: (iii) may come anywhere before (vii).

The distractor (ii′) cites a "fact" that is not available as a step for arbitrary a. For a prime a it is Corollary 3.3.2 (b), which is proved **from** this lemma, so citing it here is circular. For a general integer a the line is the lemma itself in a weaker form, so citing it is circular again; only special cases are available (for $a = \pm 1$, $a \mid b$ by Proposition 3.1.1 (a), with (d) for -1), and the proof must work for every a. Either way the line uses what has not been established, which is (N9) of Logic and Proof, Example (What is not allowed).

The distractor (iii′) divides in $\mathbb{Z}$, where there is no division: a may be 0 ($\gcd(0, 1) = 1$), and then the line divides by 0; for $a \neq 0$, $1/a$ is not an integer unless $a = \pm 1$. Divisibility is defined by multiplication (Sets, Relations, and Cardinality, Definition 2.2.1), not by fractions.

Exercise. B2: The least common multiple

Let $a, b \in \mathbb{Z}_{>0}$, and let $M := \{m \in \mathbb{Z}_{>0} : a \mid m \text{ and } b \mid m\}$. Prove that M has a least element L, and that $L \mid c$ for **every** $c \in \mathbb{Z}$ with $a \mid c$ and $b \mid c$.

The first part of a proof is given.

> $ab > 0$ by (O4), and $a \mid ab$, $b \mid ab$, so $ab \in M$; and 0 is a lower bound of M. By Proposition 2.3.6, M has a least element L. Let $c \in \mathbb{Z}$ with $a \mid c$ and $b \mid c$. By Theorem 3.1.2 (a), $c = qL + r$ with $q, r \in \mathbb{Z}$ and $0 \leq r < L$. Since $r = c + (-q)L$, and a divides c and L, $a \mid r$ (Proposition 3.1.1 (c)); likewise $b \mid r$.

The last step is blank. *Name the move, then complete it.*

Solution. The move is **least positive element, then divide**: the remainder lies in the set and is smaller than its least positive element, so it is not positive.

Completion: if $r > 0$, then $r \in \mathbb{Z}_{>0}$ is a common multiple of a and b, so $r \in M$, and $r < L$, against the choice of L. Hence $r = 0$ (as $r \geq 0$), and $c = qL$, that is, $L \mid c$.

The same argument proved Theorem 3.2.1 (b), with the set of positive combinations in place of M. This is the twin of Sets, Relations, and Cardinality, Example (The supremum under divisibility is the least common multiple), now on the constructed $\mathbb{Z}$.

Exercise. B3: Two factorizations and a gcd

Find the prime factorizations of 1001 and of $2^{10} - 1$, proving that each entry is prime. Then show that $\gcd(1001, 1023) = 11$ in two ways. First, use the factorizations to write $1001 = 11 \cdot 91$ and $1023 = 11 \cdot 93$: show $\gcd(91, 93) = 1$, and deduce $\gcd(1001, 1023) = 11$ from a Bézout identity for 91 and 93 (Corollary 3.2.2). Second, use the Euclidean algorithm.

Hint: the remark after Lemma 3.3.4 shortens each primality check. Do not use the school rule "the gcd is the product of the shared primes": this chapter does not prove it.

Solution. Factorizations. $1001 = 7 \cdot 143 = 7 \cdot 11 \cdot 13$, and $2^{10} - 1 = 1024 - 1 = 1023 = 3 \cdot 341 = 3 \cdot 11 \cdot 31$. So the prime factorizations are $(7, 11, 13)$ and $(3, 11, 31)$, both non-decreasing, once the entries are known to be prime. The primes 3 and 7 are covered by the small-primes remark of Section 3.1. For 11, 13 and 31 we use the remark after Lemma 3.3.4: it suffices that no $d \geq 2$ with $d \cdot d \leq p$ divides p.

- $p = 11$ and $p = 13$: $d \in \{2, 3\}$, since $d \geq 4$ gives $d \cdot d \geq 16$. The divisions $11 = 5 \cdot 2 + 1 = 3 \cdot 3 + 2$ and $13 = 6 \cdot 2 + 1 = 4 \cdot 3 + 1$ leave non-zero remainders.
- $p = 31$: $d \in \{2, 3, 4, 5\}$, since $d \geq 6$ gives $d \cdot d \geq 36$. The divisions $31 = 15 \cdot 2 + 1 = 10 \cdot 3 + 1 = 7 \cdot 4 + 3 = 6 \cdot 5 + 1$ leave non-zero remainders.

A non-zero remainder means $d \nmid p$, by the uniqueness in Theorem 3.1.2 (a). By Theorem 3.3.5 these are the only prime factorizations.

The gcd from the factorizations. We do not use a rule "the gcd is the product of the shared primes", which this chapter does not prove. We have $1001 = 11 \cdot 91$ and $1023 = 11 \cdot 93$. First, $\gcd(91, 93) = 1$: a common divisor of 91 and 93 divides $93 - 91 = 2$ (Proposition 3.1.1 (c)), so a positive one is 1 or 2 (Proposition 3.1.1 (e) and discreteness), and $2 \nmid 91 = 45 \cdot 2 + 1$. By Corollary 3.2.2, $91x + 93y = 1$ for some $x, y \in \mathbb{Z}$; indeed $91 \cdot 46 - 93 \cdot 45 = 4186 - 4185 = 1$. Multiplying by 11,

$$1001 \cdot 46 - 1023 \cdot 45 = 11.$$

So every common divisor of 1001 and 1023 divides 11 (Proposition 3.1.1 (c)), and a positive one is at most 11 (Proposition 3.1.1 (e)). Since 11 is itself a common divisor, $\gcd(1001, 1023) = 11$ (Definition 3.2.2).

The gcd by the algorithm (Theorem 3.2.4): $1023 = 1 \cdot 1001 + 22$, $1001 = 45 \cdot 22 + 11$, $22 = 2 \cdot 11 + 0$. The last non-zero remainder is 11.

3.3.7.3 C. Going deeper

Exercise. C1: Exit check: unique factorization

Prove that two prime factorizations of the same positive integer are equal: the uniqueness half of Theorem 3.3.5. You may cite Theorem 3.3.1 and Corollary 3.3.2.

Solution. The proof of uniqueness in Theorem 3.3.5 is a complete answer. Its steps, in order: set up $P(k)$ for non-decreasing lists of primes of length k and arbitrary length l, with equal products; base case $k = 0$, where a non-empty list of primes cannot have product 1; in the step, $l \geq 1$, then $p_{k+1} = q_j \leq q_l$ and $q_l = p_i \leq p_{k+1}$ by Corollary 3.3.2 (c) (proved from Euclid's lemma by induction on the length of the product) and the order of the lists; so $p_{k+1} = q_l$; cancel this non-zero factor (Proposition 2.3.5 (f)); apply $P(k)$ to the shortened lists.

Marking scheme. (10 points) - 2: sets up an induction on the length of one list (or a least counterexample), with the statement quantified over all lengths of the other list, and proves the base case. - 3: shows that the largest (or any chosen) prime of one list divides the other product, and identifies it with an entry of the other list by Euclid's lemma for products of several factors (cited as Corollary 3.3.2 (c), or proved by induction from Euclid's lemma). - 2: uses the symmetric step and the non-decreasing order to get equality of the two chosen extreme entries (both largest, or both smallest). - 1: justifies the cancellation by the absence of zero divisors. - 2: applies the induction hypothesis (or minimality) to the shortened lists and concludes equality of length and entries. Fatal errors (0 for the step): using the uniqueness of the factorization of a smaller number without an induction or a least counterexample set up; applying Euclid's lemma for two factors to a product of many with no induction. Canceling a prime without noting that it is non-zero costs 1 point and is not fatal. A least-counterexample proof on n (the least positive integer with two different factorizations) maps to the same points: its set-up and the case $n = 1$ take the first 2, and "the cofactor m in $n = mp_k$ has only one factorization" takes the last 2. A proof that compares the smallest entries $p_1 = q_1$ earns the same points, provided it proves (for example by induction on the length) that a product equals its first entry times the product of the re-indexed list $(p_2, \dots, p_{k+1})$, and states the re-indexing of the shortened lists $(p_2, \dots, p_{k+1})$ and $(q_2, \dots, q_l)$ before applying the induction hypothesis; asserting that identity without proof costs the 1 cancellation point.

Exercise. C2: Find the error

Here is a statement and an argument for it.

> **Claim.** Let $m, n, c \in \mathbb{Z}$. If $m \mid c$, $n \mid c$ and $\gcd(m, n) = 1$, then $mn \mid c$.
>
> **Argument.** (1) Write $c = mk$ and $c = nl$ with $k, l \in \mathbb{Z}$. (2) Then $c^2 = c \cdot c = (mk)(nl) = mn \cdot kl$. (3) So $mn \mid c^2$. (4) Hence $mn \mid c$.

Is the claim true? Does the argument prove it? If not, name the line that fails, explain why, and give a counterexample to the step.

Solution. The claim is **true**: it is Proposition 3.3.3. The argument does **not** prove it.

Lines (1)–(3) are correct. The failing line is (4). It rests on a general "fact": $x \mid c^2$ implies $x \mid c$. That has not been proved, and it is false, so the line cites nothing: entry (N9) of Logic and Proof, Example (What is not allowed). A counterexample to the step: $x = 4$, $c = 2$, with $4 \mid 4 = c^2$ and $4 \nmid 2$ (a positive divisor of 2 is at most 2, Proposition 3.1.1 (e)). There is no counterexample to the conclusion, which is true.

A sign that something is wrong: the argument never uses $\gcd(m, n) = 1$, and without that hypothesis the claim is false ($4 \mid 12$, $6 \mid 12$, $24 \nmid 12$). The repair is the proof of Proposition 3.3.3, which spends the coprimality in Euclid's lemma.

Exercise. C3: There is no last prime

Prove that no finite list contains every prime: for every list of primes $(p_1, \dots, p_k)$, with $k \in \mathbb{N}$, there is a prime that is not an entry of the list.

Solution (Hint 1). Consider $N := \prod_{i=1}^{k} p_i + 1$, and ask which primes can divide it.

Solution (Hint 2). First prove, by induction on k, that every entry of a list of primes divides its product, and that the product is at least 1.

Solution. A lemma on products. For a list of primes $(p_1, \ldots, p_k)$, write $P_j := \prod_{i=1}^{j} p_i$ for $j \leq k$. We prove, by Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on j, the statement "if $j \leq k$, then $P_j \geq 1$ and $p_i \mid P_j$ for every $i \leq j$". Base $j = 0$: $P_0 = 1$ (the recursion of Definition 3.3.1), and there are no entries to check. Step: let $j + 1 \leq k$; then $j \leq k$, so the statement holds for j, and $P_{j+1} = P_j p_{j+1}$ by the recursion. Each p_i with $i \leq j$ divides P_j, which divides P_{j+1}, so $p_i \mid P_{j+1}$ (Proposition 3.1.1 (b)); and $p_{j+1} \mid P_{j+1}$ by Sets, Relations, and Cardinality, Definition 2.2.1. Since $P_j \geq 1 > 0$ and $p_{j+1} \geq 2 > 0$, $P_{j+1} > 0 \cdot p_{j+1} = 0$ (Proposition 2.3.5 (c), Proposition 2.3.2 (b)), so $P_{j+1} \geq 1$ by discreteness (Theorem 2.3.4 (c)).

The new prime. Let $N := P_k + 1$. Then $N \geq 2$, so its least divisor $q \geq 2$ is prime (Lemma 3.3.4), and $q \mid N$. If $q = p_i$ for some i, then $q \mid P_k$ by the lemma, so q divides $N - P_k = 1$ (Proposition 3.1.1 (c)), and $q = \pm 1$ (Proposition 3.1.1 (g)), against $q \geq 2$. So q is a prime that is not an entry of the list.

For the empty list, $N = 2$ and $q = 2$. For $(2, 3, 5, 7)$, $N = 211$, which happens to be prime; for $(2, 3, 5, 7, 11, 13)$, $N = 30031 = 59 \cdot 509$ is not, and the new prime is 59. The proof promises a new prime factor of N, not that N itself is prime.

This is the direct proof asked for in part (c) of Logic and Proof, Exercise (C1: Euclid without the contradiction), now on the constructed $\mathbb{Z}$ and including the empty list. Its two school facts are replaced by Lemma 3.3.4 and Proposition 3.1.1 (c).

Marking scheme. (10 points) - 2: defines N as the product plus 1 and shows $N \geq 2$, with a reason that the product is positive. - 2: shows that each p_i divides the product, by an induction or a cited result. - 2: produces a prime factor q of N, citing Lemma 3.3.4 or an equivalent. - 3: shows q is not any p_i, since it would divide 1. - 1: concludes with the quantifiers in order: for every list there is a prime outside it. Fatal errors (0 for the step): claiming that N is prime; assuming "the set of primes is finite" and taking its product without a list (no list, no product).

3.4 Ideals, Quotient Rings, and the Integers Modulo n

Sets, Relations, and Cardinality, Section 2.5 turned the school integers into a clock, $\mathbb{Z}/n$, and checked by hand that its sum and product are well defined (Sets, Relations, and Cardinality, Proposition 2.5.5). Chapter 2 made the same kind of check twice more, for $\mathbb{Z}$ and for $\mathbb{Q}$, each time through Theorem 2.2.1 and a computation special to the formula. This section makes the check **once**, for every commutative ring and every subset of the right kind, called an ideal. The integers modulo n, now a ring, are the first example. Chapter 5 builds $\mathbb{R}$ in the same shape, from

classes of Cauchy sequences of rationals that differ by a null sequence, and checks its operations there directly; *Algebraic Preliminaries*, Chapter 5 runs division with remainder, gcd and unique factorization again for polynomials over a field.

Throughout, R is a commutative ring (Definition 2.3.1), with $a - b := a + (-b)$.

3.4.1 What a quotient needs

Let $I \subseteq R$, and call a and b related when they differ by an element of I:

$$a \sim_I b \quad :\Longleftrightarrow \quad a - b \in I.$$

For $R = \mathbb{Z}$ and I the multiples of n, this is congruence modulo n. We want two things: $\sim_I$ should be an equivalence relation, and the sum and product of classes should be computable on representatives. Each wish makes a demand on I.

- **Reflexivity**, $a - a \in I$, needs $0 \in I$.
- **Symmetry**: $b - a = -(a - b)$, so it needs I to be closed under negatives.
- **Transitivity**: $a - c = (a - b) + (b - c)$, so it needs I to be closed under sums.
- **The sum on classes.** Changing a to a' changes $a + b$ to $a' + b$, and the difference is $a - a'$ again: no new demand.
- **The product on classes.** Changing a to a' changes ab to $a'b$, and the difference is $(a - a')b$: an element of I multiplied by an **arbitrary** element of R. So I must swallow such products.

The last demand contains the second: $-x = (-1)x$. Three clauses remain, and they get a name.

An ideal is a set of allowed errors: it contains 0*, it is closed under adding errors, and an error multiplied by anything is still an error.*

Definition 3.4.1. Ideal

Let R be a commutative ring. A subset $I \subseteq R$ is an **ideal** of R if

(I1) $0 \in I$;

(I2) **for all** $a, b \in I$, $a + b \in I$;

(I3) **for all** $a \in I$ and **all** $r \in R$, $ra \in I$.

In words: (I1) and (I2) say that I contains 0 and is closed under addition; (I3), **absorption**, says that multiplying an element of I by any element of the ring, inside I or not, lands in I. The ring element r in (I3) ranges over all of R, and that is what separates (I3) from mere closure under multiplication, where r would have to lie in I.

Two consequences are used constantly. First, an ideal is closed under negatives and differences: for $a \in I$, $-a = (-1)a \in I$ by (I3), since $(-1)a = -(1 \cdot a) = -a$ by Proposition 2.3.2 (c) and (F7) with (F6); then $a - b = a + (-b) \in I$ for $a, b \in I$, by (I2). Second, if $1 \in I$, then $I = R$, since every $r = r \cdot 1 \in I$ by (I3).

Examples.

- The degenerate cases: $\{0\}$ and R itself are ideals of R. For $\{0\}$: $0 + 0 = 0$ and $r \cdot 0 = 0$ (Proposition 2.3.2 (b)). For R, every clause asks for an element of R. They are the two extremes: "no error allowed" and "everything allowed".
- For $n \in \mathbb{Z}$, the multiples $n\mathbb{Z} = \{kn : k \in \mathbb{Z}\}$ form an ideal of $\mathbb{Z}$ (Definition 3.2.1): $0 = 0 \cdot n$; $kn + ln = (k + l)n$; $r(kn) = (rk)n$. In particular the even integers $2\mathbb{Z}$ form an ideal.

- For $a, b \in \mathbb{Z}$, the combinations $a\mathbb{Z} + b\mathbb{Z}$ form an ideal of $\mathbb{Z}$, by the closure properties in Definition 3.2.1; Theorem 3.4.2 below records this.

Non-examples by minimal change. $\mathbb{N} \subseteq \mathbb{Z}$ satisfies (I1) and (I2), but (I3) fails at $a = 1$, $r = -1$: $(-1) \cdot 1 = -1 \notin \mathbb{N}$, since $-1 < 0$ ($1 > 0$ by Proposition 2.3.5 (e), as $1 \neq 0$ in $\mathbb{Z}$ by Theorem 2.3.1, then (a)) while $\mathbb{N} = \{x \in \mathbb{Z} : x \geq 0\}$ (Theorem 2.3.4 (b)). And $\mathbb{Z} \subseteq \mathbb{Q}$ satisfies (I1), (I2) and is even closed under negatives and products, but (I3) fails at $a = 1$, $r = 1/2$: $(1/2) \cdot 1 = 1/2 \notin \mathbb{Z}$. (If $1/2 = m \in \mathbb{Z}$, then $2m = 1$, but $1 = 2 \cdot 0 + 1$ is odd and so not of the form $2m$, by Corollary 3.1.3 (a).) In both cases the ring element r comes from **outside** the subset, which is exactly what (I3) allows.

Why these clauses. Each clause was forced by one of the five wishes above, and dropping one loses its wish: without (I1) or (I2), $\sim_I$ is not reflexive or not transitive; without (I3), the product on classes can fail, as the warning after the next theorem shows for $\mathbb{Z} \subseteq \mathbb{Q}$.

3.4.2 The quotient ring

The payoff is a single theorem that builds a ring of classes from any ideal. We write $[a]$, or $[a]_I$ when the ideal must be named, for the class of a under $\sim_I$, and R/I for the quotient set $R/\sim_I$ (Sets, Relations, and Cardinality, Definition 2.5.4).

Theorem 3.4.1. Quotient Ring by an Ideal

Let R be a commutative ring and I an ideal of R.

(a) Setting $a \sim_I b \iff a - b \in I$ defines an equivalence relation $\sim_I$ on R, and the class of a is $[a] = a + I := \{a + x : x \in I\}$.

(b) The rules $[a] + [b] := [a + b]$, $[a] \cdot [b] := [ab]$ and $-[a] := [-a]$ are well defined on $R/I := R/\sim_I$.

(c) R/I, with these operations, is a commutative ring with zero $[0]$ and one $[1]$. The **quotient map** $\pi \colon R \to R/I$, $\pi(a) := [a]$, satisfies $\pi(a + b) = \pi(a) + \pi(b)$, $\pi(ab) = \pi(a)\pi(b)$ and $\pi(1) = [1]$.

(d) $[a] = [0]$ if and only if $a \in I$. In particular $[1] = [0]$ if and only if $I = R$.

In words: (a) and (b) say that classes modulo an ideal can be added, multiplied and negated through any representatives; (c) says that the result obeys every ring law, and that passing to classes respects the operations; (d) says that the elements of I are exactly those that become 0, and that the quotient collapses to the zero ring exactly when I is everything.

Idea. A change of representative is an element of I. The equivalence-relation clauses ask that such changes can be reversed and chained, which is closure under negatives and sums. The operations ask that a change in one argument produce a change in the output: for the sum the change is the same element of I, and for the product it is that element multiplied by the other argument, which (I3) keeps in I. The ring laws are then laws of R read on classes.

Before reading on: where is "$ra \in I$ for every $r \in R$" used, and why does closure under $+$ not suffice for the product?

Proof. *(a)* Let $a, b, c \in R$. (Eq1): $a - a = 0 \in I$, by (F4) and (I1). (Eq2): if $a - b \in I$, then $b - a = -(a - b) \in I$, since I is closed under negatives; here $(a - b) + (b - a) = 0$ by (F1)–(F4), so $b - a = -(a - b)$ by the uniqueness of negatives (Proposition 2.3.2 (a)). (Eq3): if $a - b \in I$ and $b - c \in I$, then $a - c = (a - b) + (b - c) \in I$, by (F2)–(F4) and (I2). So $\sim_I$ is an equivalence relation (Sets, Relations, and Cardinality, Definition 2.5.1). For the class: if $x \sim_I a$, then $x - a \in I$ and $x = a + (x - a) \in a + I$; conversely, for $y \in I$, $(a + y) - a = y \in I$, so $a + y \sim_I a$. Hence $[a] = a + I$.

(b) We check the hypotheses of Theorem 2.2.1 (b), with $X = R$ and the relation $\sim_I$, for the maps $(a, b) \mapsto a + b$ and $(a, b) \mapsto ab$. Let $a \sim_I a'$, so $a - a' \in I$, and let $b \in R$. Then

$$(a + b) - (a' + b) = a - a' \in I,$$
$$ab - a'b = (a - a')b = b(a - a') \in I,$$

by (F1)–(F4) in the first line, and by (F9), (F6) and Proposition 2.3.2 (c) in the second, where the membership is (I3) with $r = b$: **the ideal absorbs** b. This is the first hypothesis. Since $a + b = b + a$ and $ab = ba$ ((F1), (F6)), the second hypothesis is the first with the two arguments exchanged. For the negative we check the hypothesis of Theorem 2.2.1 (a) for $a \mapsto -a$: $(-a) - (-a') = -(a - a') \in I$, by Proposition 2.3.2 (a) and closure under negatives. Hence the three rules define functions.

(c) Every element of R/I is a class $[a]$ (Sets, Relations, and Cardinality, Definition 2.5.4). Each axiom of Definition 2.3.1 other than (F8) is an equation, and by (b) both of its sides, read on classes, are the classes of the two sides of the same equation in R. For example, for (F9),

$$[a]([b] + [c]) = [a][b + c] = [a(b + c)] = [ab + ac] = [ab] + [ac] = [a][b] + [a][c],$$

using (F9) in R in the middle; for (F3), $[a] + [0] = [a + 0] = [a]$; for (F4), $[a] + (-[a]) = [a] + [-a] = [a + (-a)] = [0]$; for (F7), $[a][1] = [a \cdot 1] = [a]$. The laws (F1), (F2), (F5) and (F6) go the same way. So R/I is a commutative ring with zero $[0]$ and one $[1]$. The equations for π are the definitions in (b) read from right to left, and $\pi(1) = [1]$ by definition of π.

(d) $[a] = [0]$ holds if and only if $a \sim_I 0$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)), that is, $a - 0 \in I$; and $a - 0 = a + 0 = a$, since $-0 = 0$ ($0 + 0 = 0$ and Proposition 2.3.2 (a)). So $[a] = [0] \iff a \in I$. In particular $[1] = [0] \iff 1 \in I$. If $1 \in I$, then $I = R$, since $r = r \cdot 1 \in I$ for every $r \in R$ by (I3); and if $I = R$, then $1 \in I$. This proves the theorem. ∎

The answer to the prompt. Absorption with an arbitrary $r \in R$, the full force of (I3), is used exactly once in the proof of (a) and (b): in the product, at $ab - a'b = (a - a')b$, where the change of representative $a - a' \in I$ is multiplied by an arbitrary $b \in R$. Elsewhere (I3) enters only with $r = -1$, through closure under negatives ($-x = (-1)x$), and otherwise the proof uses (I1) and (I2). Closure under $+$ does not suffice for the product because b need not lie in I: the product changes by a multiple of the error, not by a sum of errors.

We call this move **compute on representatives; the ideal absorbs the change**. It is the move of Chapter 2's "one argument at a time", now with the change of representative recorded as one element of I.

Warning. **Closure under $+$ and $-$ is not enough for the product.** $\mathbb{Z} \subseteq \mathbb{Q}$ contains 0 and is closed under sums and negatives, and these are all that the proof of (a) and of the sum and negative in (b) used. So "$x - y \in \mathbb{Z}$" relates rationals by an equivalence relation, and $[x] + [y] := [x + y]$ is well defined on its classes. But $[x][y] := [xy]$ is **not**: $[1/2] = [3/2]$, since $3/2 - 1/2 = 1 \in \mathbb{Z}$, while

$$\tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \qquad \tfrac{3}{2} \cdot \tfrac{1}{2} = \tfrac{3}{4}, \qquad \tfrac{3}{4} - \tfrac{1}{4} = \tfrac{1}{2} \notin \mathbb{Z}.$$

One class, two products in different classes. The missing clause is (I3): $\mathbb{Z}$ is not an ideal of $\mathbb{Q}$.

3.4.3 The ideals of the integers

In $\mathbb{Z}$ every ideal is as simple as possible: the multiples of one number. The proof is the move that gave the gcd in Section 3.2.

Theorem 3.4.2. Ideals of the Integers

(a) For all $a, b \in \mathbb{Z}$, $a\mathbb{Z} + b\mathbb{Z}$ is an ideal of $\mathbb{Z}$. In particular $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$ for every $n \in \mathbb{Z}$.

(b) Every ideal I of $\mathbb{Z}$ equals $n\mathbb{Z}$ for **exactly one** $n \in \mathbb{N}$: $n = 0$ if $I = \{0\}$, and otherwise n is the least positive element of I.

Idea. For (b), take the least positive element n of I. Its multiples lie in I by absorption. Conversely, divide any $a \in I$ by n: the remainder $a - qn$ lies in I and is smaller than n, so it is not positive, and it is 0. This is **least positive element, then divide** once more.

Proof. *(a)* $0 = a \cdot 0 + b \cdot 0 \in a\mathbb{Z} + b\mathbb{Z}$ (Proposition 2.3.2 (b)), and Definition 3.2.1 shows that $a\mathbb{Z} + b\mathbb{Z}$ is closed under sums and under multiplication by any integer. These are (I1)–(I3). Since $n\mathbb{Z} = n\mathbb{Z} + 0\mathbb{Z}$ (Definition 3.2.1), $n\mathbb{Z}$ is an ideal.

(b) Existence. If $I = \{0\}$, then $I = 0\mathbb{Z}$, since $k \cdot 0 = 0$ for every k (Proposition 2.3.2 (b)). Otherwise I has an element $a \neq 0$, and $-a \in I$ as well; one of a, $-a$ is positive (Proposition 3.1.1 (h)). So $I \cap \mathbb{Z}_{>0}$ is non-empty, with lower bound 0, and has a least element n (Proposition 2.3.6). By (I3), $kn \in I$ for every $k \in \mathbb{Z}$, so $n\mathbb{Z} \subseteq I$. Conversely, let $a \in I$. By Theorem 3.1.2 (a), $a = qn + r$ with $q, r \in \mathbb{Z}$ and $0 \leq r < n$. Then $r = a + (-q)n \in I$, by (I3) and (I2). If $r > 0$, then r is a positive element of I smaller than n, against the choice of n. So $r = 0$, and $a = qn \in n\mathbb{Z}$. Hence $I = n\mathbb{Z}$.

Uniqueness. Let $n\mathbb{Z} = m\mathbb{Z}$ with $n, m \in \mathbb{N}$. Then $n = 1 \cdot n \in m\mathbb{Z}$, so $m \mid n$; likewise $n \mid m$. By Proposition 3.1.1 (f), $n = m$.

■

Remark. For $I = a\mathbb{Z} + b\mathbb{Z}$ with $(a, b) \neq (0, 0)$, (b) says $a\mathbb{Z} + b\mathbb{Z} = n\mathbb{Z}$ where n is the least positive combination. That is Theorem 3.2.1 (a), (d), (e), with $n = \gcd(a, b)$: the theorem

of Section 3.2 was this one in disguise.

3.4.4 The ring of integers modulo n

Now the clock of *Sets, Relations, and Cardinality*, Section 2.5 can be rebuilt on the constructed $\mathbb{Z}$, as an instance of the general theorem.

Definition 3.4.2. The Integers Modulo n

Let $n \in \mathbb{Z}_{>0}$. The ring $\mathbb{Z}/n\mathbb{Z}$ is the quotient ring of $\mathbb{Z}$ by the ideal $n\mathbb{Z}$ (Theorem 3.4.1, Theorem 3.4.2 (a)).

Its relation is congruence modulo n of Sets, Relations, and Cardinality, Definition 2.5.5, read on the constructed $\mathbb{Z}$: $a - b \in n\mathbb{Z}$ means $a - b = kn$ for some $k \in \mathbb{Z}$, that is, $n \mid a - b$, that is, $a \equiv b \pmod{n}$. Its elements are the classes $[a] = a + n\mathbb{Z}$, written $[a]_n$ when two moduli are in play. Its operations, $[a] + [b] = [a + b]$ and $[a][b] = [ab]$, are those of Sets, Relations, and Cardinality, Proposition 2.5.5; there they were checked by hand for the school integers, and here they are a consequence of Theorem 3.4.1 (b). The set $\mathbb{Z}/n$ of *Sets, Relations, and Cardinality* is the same construction on the school $\mathbb{Z}$; from now on we write $\mathbb{Z}/n\mathbb{Z}$, to recall the ideal.

Theorem 3.4.3. The Ring of Integers Modulo n

Let $n \in \mathbb{Z}_{>0}$.

(a) $\mathbb{Z}/n\mathbb{Z}$ is a commutative ring with zero $[0]$ and one $[1]$.

(b) For each $a \in \mathbb{Z}$ there is **exactly one** $r \in \{0, \dots, n-1\}$ with $[a] = [r]$, namely the remainder of a on division by n. The map $\{0, \dots, n-1\} \to \mathbb{Z}/n\mathbb{Z}$, $r \mapsto [r]$, is a bijection, so $\mathbb{Z}/n\mathbb{Z}$ is finite with n elements (Sets, Relations, and Cardinality, Definition 3.1.2).

(c) $[1] \neq [0]$ if and only if $n \geq 2$.

(d) $k \cdot [1] = [k]$ for every $k \in \mathbb{N}$, where $k \cdot x$ is the k-th multiple in the commutative ring $\mathbb{Z}/n\mathbb{Z}$; in particular $n \cdot [1] = [0]$.

(e) If $n \geq 2$, no relation $<$ makes $\mathbb{Z}/n\mathbb{Z}$ an ordered commutative ring.

The multiples in (d) are those of Section 2.3, defined there by recursion for every commutative ring R ($0 \cdot x \coloneqq 0$, $(k+1) \cdot x \coloneqq k \cdot x + x$), before Proposition 2.3.5; the recursion uses no order, so it applies to $\mathbb{Z}/n\mathbb{Z}$, which for $n \geq 2$ carries no order compatible with its arithmetic, by (e).

Idea. Most of the work is done: (a) is the general theorem, and (b) is division with remainder read on classes, the remainder naming the class and its uniqueness making the naming one-to-one. Then (c) asks when 1 is a multiple of n, (d) is an induction along the recursion that defines multiples, and (e) puts (c) and (d) against the rule that $n \cdot 1 > 0$ in an ordered ring.

Proof. *(a)* Theorem 3.4.1 (c), since $n\mathbb{Z}$ is an ideal (Theorem 3.4.2 (a)).

(b) Existence. By Theorem 3.1.2 (a), $a = qn + r$ with $0 \leq r < n$. Then $a - r = qn \in n\mathbb{Z}$, so $[a] = [r]$. Also $r \in \mathbb{N}$ by Theorem 2.3.4 (b), and $r + 1 \leq n$ by discreteness (Theorem 2.3.4 (c)), so $r \leq n - 1$ by (O3); that is, $r \in \{0, \dots, n-1\}$.

Uniqueness. Let $r, s \in \{0, \ldots, n-1\}$ with $[r] = [s]$. Then $r - s = kn$ for some $k \in \mathbb{Z}$, so $r = kn + s$ and $r = 0 \cdot n + r$ are two divisions of r by n, with remainders s and r, each at least 0 and less than n. By the uniqueness in Theorem 3.1.2 (a), $s = r$.

So the map $\varphi\colon \{0, \ldots, n-1\} \to \mathbb{Z}/n\mathbb{Z}$, $r \mapsto [r]$, is surjective (existence: every class is some $[a]$, Sets, Relations, and Cardinality, Definition 2.5.4) and injective (uniqueness), hence a bijection (Sets, Relations, and Cardinality, Definition 1.4.5). Since $n \geq 1$, adding -1 gives $n - 1 \geq 0$ (O3), so $n - 1 \in \mathbb{N}$ by Theorem 2.3.4 (b), and the third example after Sets, Relations, and Cardinality, Definition 3.1.2, with $n - 1$ in place of its n, gives the bijection $\psi\colon \{1, \ldots, n\} \to \{0, \ldots, n-1\}$, $k \mapsto k - 1$. The composite $\varphi \circ \psi\colon \{1, \ldots, n\} \to \mathbb{Z}/n\mathbb{Z}$ is a bijection (Sets, Relations, and Cardinality, Theorem 1.4.4 (c)), so $\mathbb{Z}/n\mathbb{Z}$ has n elements.

(c) By Theorem 3.4.1 (d), $[1] = [0]$ if and only if $1 \in n\mathbb{Z}$, that is, $n \mid 1$. By Proposition 3.1.1 (g) and $n > 0$, this happens exactly when $n = 1$. As $n \geq 1$ by discreteness, $n \neq 1$ means $n \geq 2$.

(d) By Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k. Base: $0 \cdot [1] = [0]$, by the recursion. Step: if $k \cdot [1] = [k]$, then

$$(k+1) \cdot [1] = k \cdot [1] + [1] = [k] + [1] = [k+1],$$

by the recursion, the hypothesis, and the definition of $+$ in Theorem 3.4.1 (b) (the sum $k+1$ is the same in $\mathbb{N}$ and in $\mathbb{Z}$, Proposition 2.2.3 (a)). Finally $[n] = [0]$, since $n - 0 = 1 \cdot n \in n\mathbb{Z}$.

(e) Let $n \geq 2$, and suppose $<$ made $\mathbb{Z}/n\mathbb{Z}$ an ordered commutative ring. By (c), $[1] \neq [0]$, so Proposition 2.3.5 (h), applied with $n \geq 1$, gives $n \cdot [1] > [0]$. By (d), $n \cdot [1] = [0]$, and $[0] > [0]$ contradicts (O1). This proves the theorem.

■

Part (e) is the general clause after Example (A field with two elements that cannot be ordered), now applied to a ring that this volume has built: $\mathbb{Z}/2\mathbb{Z}$ has the tables of that example, with $[0]$, $[1]$ for 0, 1. Part (b) says that the elements of $\mathbb{Z}/n\mathbb{Z}$ can be listed as $[0], [1], \ldots, [n-1]$, and that to compute with classes we may always reduce to remainders.

Example. Multiplication modulo six and modulo five

Write out the multiplication table of $\mathbb{Z}/6\mathbb{Z}$. Find the pairs of non-zero classes whose product is $[0]$, and the classes $[a]$ for which some $[b]$ has $[a][b] = [1]$. Then do the same for $\mathbb{Z}/5\mathbb{Z}$.

Solution. By Theorem 3.4.3 (b), the classes of $\mathbb{Z}/6\mathbb{Z}$ are $[0], \ldots, [5]$, and $[a][b] = [ab] = [r]$, where r is the remainder of ab on division by 6. For instance $[4][5] = [20] = [2]$, since $20 = 3 \cdot 6 + 2$. Writing r for $[r]$, the table is

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

The non-zero classes with product [0] are $[2][3] = [6] = [0]$ and $[3][4] = [12] = [0]$, with their mirror images $[3][2]$, $[4][3]$. So $\mathbb{Z}/6\mathbb{Z}$ has zero divisors, unlike $\mathbb{Z}$ (Proposition 2.3.5 (f)). Also $[4][4] = [16] = [4]$: a class can equal its own square without being [0] or [1]. The entry 1 occurs only in the rows of [1] and [5]: $[1][1] = [1]$ and $[5][5] = [25] = [1]$. So exactly [1] and [5] have a partner with product [1].

For $\mathbb{Z}/5\mathbb{Z}$ the same method gives

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

No product of non-zero classes is [0], and every non-zero class has a partner with product [1]: $[1][1] = [1]$, $[2][3] = [6] = [1]$, $[4][4] = [16] = [1]$. Moreover $5 \cdot [1] = [0]$ by Theorem 3.4.3 (d), so $\mathbb{Z}/5\mathbb{Z}$ cannot be ordered, by (e).

The classes with a partner of product [1] are the **units** of Section 3.5, which finds them for every n and shows that for a prime n, as for 5, every non-zero class is one. Chapter 6 uses the ring $\mathbb{Z}/5\mathbb{Z}$, with $5 \cdot [1] = [0]$, as a field that no order can fit.

Quick check. In $\mathbb{Z}/12\mathbb{Z}$, compute $[7] + [8]$ and $[7][8]$, each as $[r]$ with $0 \leq r \leq 11$. How many elements has $\mathbb{Z}/1\mathbb{Z}$?

Solution. $[7]+[8] = [15] = [3]$, since $15 = 1 \cdot 12 + 3$; and $[7][8] = [56] = [8]$, since $56 = 4 \cdot 12 + 8$. The ring $\mathbb{Z}/1\mathbb{Z}$ has one element, by Theorem 3.4.3 (b): every integer is congruent to 0 modulo 1, so $[0] = [1]$, and it is a zero ring, like the ring $\{0\}$ of Section 2.3.

Warning. **A class is not its remainder.** [3] in $\mathbb{Z}/12\mathbb{Z}$ is the set $3+12\mathbb{Z} = \{\ldots, -9, 3, 15, \ldots\}$, not the integer 3, and the same symbol [3] in $\mathbb{Z}/5\mathbb{Z}$ is a different set. A comparison such as "$[3] < [8]$" cannot be read off the remainders and respect the arithmetic: by Theorem 3.4.3 (e) no order makes $\mathbb{Z}/12\mathbb{Z}$ an ordered ring. Reducing to remainders is a way of **naming** classes, not a way of turning them back into integers.

3.4.5 Exercises

3.4.5.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 3.4.1.

(a) Where is the clause "$ra \in I$ for every $r \in R$" used?

(b) Run the product check of (b) with $R = \mathbb{Z}$, $I = 6\mathbb{Z}$, $a = 2$, $a' = 8$ and $b = 5$: verify $a \sim_I a'$, compute $ab - a'b$, and say why it lies in I.

(c) For $\mathbb{Z} \subseteq \mathbb{Q}$, which step of the proof fails, and which clause of Definition 3.4.1 is missing?

(d) Give the idea of the proof in two sentences.

Solution.

(a) In (b), for the product in the first argument: $ab - a'b = (a - a')b \in I$, with $a - a' \in I$ and $r = b$. (It also gives closure under negatives, which (a) and (b) use through $-x = (-1)x$; and it is used in (d), to get $I = R$ from $1 \in I$.)

(b) $a - a' = 2 - 8 = -6 = (-1) \cdot 6 \in 6\mathbb{Z}$, so $2 \sim_I 8$. Then $ab - a'b = 10 - 40 = -30$, which is $(a - a')b = (-6) \cdot 5$, and $-30 = (-5) \cdot 6 \in 6\mathbb{Z}$. It lies in I because I absorbs the factor 5.

(c) The product step of (b): $(a - a')b$ need not lie in $\mathbb{Z}$, as with $a = 3/2$, $a' = 1/2$, $b = 1/2$, where it is $1/2$. The missing clause is (I3): $(1/2) \cdot 1 \notin \mathbb{Z}$.

(d) A change of representative is an element of I; for the sum it passes through unchanged, and for the product it gets multiplied by the other argument, which absorption keeps inside I. The ring laws then hold in R/I because both sides of each are classes of the two sides of the same law in R.

3.4.5.2 B. Practice

Exercise. B1: The ring with four elements

Write the addition and multiplication tables of $\mathbb{Z}/4\mathbb{Z}$. Find its zero divisors (non-zero x with $xy = [0]$ for some non-zero y) and the classes $[a]$ with $[a][b] = [1]$ for some $[b]$.

Solution. By Theorem 3.4.3 (b) the elements are $[0], [1], [2], [3]$, and each entry is the class of the remainder of the sum or product on division by 4. Writing r for $[r]$:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

For instance $[3] + [3] = [6] = [2]$ and $[3][3] = [9] = [1]$, since $6 = 1 \cdot 4 + 2$ and $9 = 2 \cdot 4 + 1$. The only zero divisor is $[2]$: $[2][2] = [4] = [0]$, and in the rows of $[1]$ and $[3]$ the entry 0 occurs only in the column of $[0]$. The classes with a partner of product $[1]$ are $[1]$ ($[1][1] = [1]$) and $[3]$ ($[3][3] = [1]$); the rows of $[0]$ and $[2]$ contain no 1.

Exercise. B2: Determine which

For each subset of $\mathbb{Z}$ below, decide whether it is an ideal of $\mathbb{Z}$. Justify your answer.

(a) $\mathbb{N}$.

(b) $2\mathbb{Z} \cup 3\mathbb{Z}$.

(c) $4\mathbb{Z} + 6\mathbb{Z}$.

(d) $\{0\}$.

(e) $\{n \in \mathbb{Z} : 6 \mid n^2\}$, where $n^2 := nn$.

Solution.

(a) No. (I3) fails: $1 \in \mathbb{N}$, but $(-1) \cdot 1 = -1 \notin \mathbb{N}$, since $-1 < 0$ ($1 > 0$ by Proposition 2.3.5 (e), as $1 \neq 0$ in $\mathbb{Z}$ by Theorem 2.3.1, then (a)) while $\mathbb{N} = \{x \in \mathbb{Z} : x \geq 0\}$ (Theorem 2.3.4 (b)).

(b) No. (I2) fails: $2 \in 2\mathbb{Z}$ and $3 \in 3\mathbb{Z}$, but 5 is in neither, since $5 = 2 \cdot 2 + 1$ and $5 = 1 \cdot 3 + 2$ leave non-zero remainders (uniqueness in Theorem 3.1.2 (a)).

(c) Yes, by Theorem 3.4.2 (a). In fact $4\mathbb{Z} + 6\mathbb{Z} = 2\mathbb{Z}$: every $4x + 6y = 2(2x + 3y)$ is even, and every $2k = 4(-k) + 6k$ is a combination.

(d) Yes: the ideal $0\mathbb{Z}$, or directly as in the examples after Definition 3.4.1.

(e) Yes: the set is $6\mathbb{Z}$. If $n = 6k$, then $n^2 = 6(6k^2)$. Conversely, let $6 \mid n^2$, say $n^2 = 6m$. Then $n \cdot n = 2(3m) = 3(2m)$, so $2 \mid nn$ and $3 \mid nn$. Since 2 and 3 are prime (the small-primes remark of Section 3.1), Corollary 3.3.2 (b) gives $2 \mid n$ and $3 \mid n$. As $1 = (-1) \cdot 2 + 1 \cdot 3$, $\gcd(2, 3) = 1$ (Corollary 3.2.2), so $6 = 2 \cdot 3 \mid n$ by Proposition 3.3.3. Hence the set equals $6\mathbb{Z}$, an ideal by Theorem 3.4.2 (a).

3.4.5.3 C. Going deeper

Exercise. C1: Changing the modulus

Let $m, n \in \mathbb{Z}_{>0}$. Prove that there is a function $F\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ with $F([a]_n) = [a]_m$ for every $a \in \mathbb{Z}$ if and only if $m \mid n$. Show that, when it exists, F preserves sums, products and the one.

Solution. ($\Leftarrow$) Let $n = lm$. Then $n\mathbb{Z} \subseteq m\mathbb{Z}$, since $kn = (kl)m$. We apply Theorem 2.2.1 (a) with $X = Y = \mathbb{Z}$, congruence modulo n on X, congruence modulo m on Y, and $f(a) := a$: if $a - a' \in n\mathbb{Z}$, then $a - a' \in m\mathbb{Z}$. So F exists (and is unique).

($\Rightarrow$) Suppose F exists. Since $n - 0 = 1 \cdot n$, $[n]_n = [0]_n$, so

$$[n]_m = F([n]_n) = F([0]_n) = [0]_m,$$

so $n \in m\mathbb{Z}$ by Theorem 3.4.1 (d), that is, $m \mid n$.

Operations. $F([a]_n + [b]_n) = F([a + b]_n) = [a + b]_m = [a]_m + [b]_m$, by the definition of $+$ in Theorem 3.4.1 (b) on both sides; likewise for products, and $F([1]_n) = [1]_m$.

This is the general form of Sets, Relations, and Cardinality, Example (A rule on classes that is not a function): $[a]_6 \mapsto [a]_4$ fails because $4 \nmid 6$, and $[a]_6 \mapsto [a]_3$ works because $3 \mid 6$.

Marking scheme. (10 points) - 3: ($\Leftarrow$) writes $n = lm$, shows $n\mathbb{Z} \subseteq m\mathbb{Z}$, and invokes Theorem 2.2.1 (a) with $f(a) = a$. - 3: ($\Rightarrow$) uses $[n]_n = [0]_n$ and applies F to get $[n]_m = [0]_m$, hence $m \mid n$. - 3: computes F of a sum and of a product on representatives, citing the definition of the operations on both quotient rings. - 1: $F([1]_n) = [1]_m$. Fatal errors (0

for the step): defining F by a formula on representatives without checking that it does not depend on the representative; in ($\Rightarrow$), assuming $n\mathbb{Z} \subseteq m\mathbb{Z}$ (circular).

Exercise. C2: Closed under subtraction

Let $S \subseteq \mathbb{Z}$ be **non-empty**, with $a - b \in S$ for all $a, b \in S$. Prove that $S = n\mathbb{Z}$ for some $n \in \mathbb{N}$.

The first half of a proof is given.

> Pick $a \in S$; then $0 = a - a \in S$. If $S = \{0\}$, then $S = 0\mathbb{Z}$. Otherwise S contains some $a \neq 0$, and $-a = 0 - a \in S$; one of a, $-a$ is positive (Proposition 3.1.1 (h)). By Proposition 2.3.6, S has a least positive element n.

The second half is blank. *Name the move, then complete it.*

Solution. The move is **least positive element, then divide**.

$n\mathbb{Z} \subseteq S$. First, $kn \in S$ for every $k \in \mathbb{N}$, by Logic and Proof, Axiom 4.2.1 on k. Base: $0 \cdot n = 0 \in S$. Step: $-n = 0 - n \in S$, and if $kn \in S$, then $(k+1)n = kn + n = kn - (-n) \in S$, using (F9), (F7) and $-(-n) = n$ (Proposition 2.3.2 (a)). Next, every $k \in \mathbb{Z}$ is j or $-j$ for some $j \in \mathbb{N}$ (Proposition 2.2.3 (c)), and $(-j)n = -(jn) = 0 - jn \in S$ (Proposition 2.3.2 (c)).

$S \subseteq n\mathbb{Z}$. Let $a \in S$. By Theorem 3.1.2 (a), $a = qn + r$ with $0 \leq r < n$. Since $qn \in S$, $r = a - qn \in S$. If $r > 0$, it is a positive element of S smaller than n, against the choice of n. So $r = 0$ and $a = qn$.

Hence $S = n\mathbb{Z}$. No multiplication hypothesis was needed: closure under subtraction alone forces the multiples of n, and so S is automatically an ideal. The sets S of this exercise are the subgroups of $(\mathbb{Z}, +)$, which *Algebraic Preliminaries*, Chapter 3 studies.

3.5 Units, Prime Moduli, and the Chinese Remainder Theorem

Section 3.4 made $\mathbb{Z}/n\mathbb{Z}$ a commutative ring, and Example (Multiplication modulo six and modulo five) put two of these rings side by side. In $\mathbb{Z}/5\mathbb{Z}$ every non-zero class has a partner whose product with it is $[1]$, so we may divide by it. In $\mathbb{Z}/6\mathbb{Z}$ the class $[2]$ has no such partner: the products $[2][x]$ are only $[0], [2], [4]$. Which classes can be divided by, and for which moduli can all non-zero classes? This section answers both questions with the gcd of Section 3.2, and finds a field with p elements for every prime p. It then solves two congruences at once, the Chinese remainder theorem, by the move of Section 3.3, multiply the Bézout identity: write 1 as a combination and multiply.

Throughout, $n \in \mathbb{Z}_{>0}$ and classes are taken in $\mathbb{Z}/n\mathbb{Z}$ unless another modulus is named. Two classes are equal exactly when their representatives are congruent:

$$[u] = [v] \iff u - v \in n\mathbb{Z} \iff n \mid u - v.$$

The first equivalence is Sets, Relations, and Cardinality, Theorem 2.5.1 (b) for the relation of Theorem 3.4.1 (a); the second is the definition of $n\mathbb{Z}$ (Definition 3.2.1) read with Sets, Relations,

and Cardinality, Definition 2.2.1. We call it the **class test**. Products of classes are computed on representatives, $[u][v] = [uv]$ (Theorem 3.4.1 (b)).

3.5.1 Units of a commutative ring

In $\mathbb{Z}/6\mathbb{Z}$ the equation $[2][x] = [1]$ has no solution, while $[5][x] = [1]$ has the solution $[x] = [5]$. The elements that can be divided by are the ones that solve "u times something is 1", and they deserve a name in every commutative ring at once.

A unit is an element that has a multiplicative inverse in the ring itself.

Definition 3.5.1. Unit of a Commutative Ring

Let R be a commutative ring (Definition 2.3.1). An element $u \in R$ is a **unit** of R if **there exists** $v \in R$ with $uv = 1$. The set of units of R is written $R^\times$. For a unit u, the element v is **unique**; it is written u^{-1} and called the **inverse** of u.

In words: u is a unit when the equation $ux = 1$ can be solved **inside** R. The definition asks for one solution; it does not ask u to be non-zero.

Well-definedness of u^{-1}. Let $uv = 1$ and $uv' = 1$. Then, by (F7), (F5) and (F6),

$$v = v \cdot 1 = v(uv') = (vu)v' = (uv)v' = 1 \cdot v' = v'.$$

So the inverse is unique, and the notation u^{-1} names one element.

Closure. The element 1 is a unit, since $1 \cdot 1 = 1$. If u is a unit, so is u^{-1}, with $(u^{-1})^{-1} = u$, because $u^{-1}u = uu^{-1} = 1$ by (F6). If u and w are units, then

$$(uw)(u^{-1}w^{-1}) = (uu^{-1})(ww^{-1}) = 1 \cdot 1 = 1,$$

reordering the four factors by (F5) and (F6); so uw is a unit and $(uw)^{-1} = u^{-1}w^{-1}$, by uniqueness.

A unit is not a zero divisor. If u is a unit and $ux = 0$, then

$$x = (u^{-1}u)x = u^{-1}(ux) = u^{-1} \cdot 0 = 0,$$

by $u^{-1}u = 1$ (above) and $1 \cdot x = x \cdot 1 = x$ ((F6), (F7)) in the first equality, (F5) in the second, and Proposition 2.3.2 (b) in the third. So a unit times a non-zero element is never 0.

Examples.

- **The integers.** $\mathbb{Z}^\times = \{1, -1\}$. Indeed $1 \cdot 1 = 1$ and $(-1)(-1) = 1$ (Proposition 2.3.2 (c)). Conversely, $uv = 1$ says $u \mid 1$, so $u = 1$ or $u = -1$ by Proposition 3.1.1 (g).
- **The rationals.** $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\}$: every non-zero rational has an inverse by (F8) (Theorem 2.4.3), and 0 is not a unit, since $0 \cdot v = 0 \neq 1$ (Proposition 2.3.2 (b)).
- **The integers modulo six.** By the multiplication table of Example (Multiplication modulo six and modulo five), the units of $\mathbb{Z}/6\mathbb{Z}$ are $[1]$ and $[5]$, and $[5]^{-1} = [5]$.
- **The zero ring** $\{0\}$ of Section 2.3, where $1 = 0$. Here $0 \cdot 0 = 0 = 1$, so 0 is a unit and $R^\times = R$. This degenerate case is why "every non-zero element is a unit" is not by itself the definition of a field.

A non-example by minimal change. Change 5 to 2 in the third example. The class $[2] \in \mathbb{Z}/6\mathbb{Z}$ is non-zero, like $[5]$, but $[2][3] = [6] = [0]$: it is a zero divisor, so by the paragraph above it is

not a unit. Concretely, the clause "there exists v with $uv = 1$" fails: the products $[2][x]$ for $x = 0, \ldots, 5$ are $[0], [2], [4], [0], [2], [4]$, and $[1]$ is not among them.

Why this definition. With units, the field axiom (F8) reads: every non-zero element is a unit. So a field (Definition 2.3.1) is exactly a commutative ring with $1 \neq 0$ in which $R^\times = R \setminus \{0\}$; when $1 \neq 0$, the element 0 is never a unit, since $0 \cdot v = 0$ (Proposition 2.3.2 (b)). The definition asks for an inverse in R, not in some bigger ring: $2 \in \mathbb{Z}$ has the inverse $1/2$ in $\mathbb{Q}$, and it is still not a unit of $\mathbb{Z}$.

Warning. "Non-zero and not a zero divisor" does not make an element a unit. In $\mathbb{Z}$ there are no zero divisors at all (Proposition 2.3.5 (f)), yet 2 is not a unit, since $2 \nmid 1$. The paragraph above proves only one direction: a unit is never a zero divisor. So "not a unit" must be checked against the definition, or, in $\mathbb{Z}/n\mathbb{Z}$, by the next theorem.

The units of a commutative ring are closed under products and inverses and contain 1: they form what *Algebraic Preliminaries*, Chapter 3 calls a group. Here we only need to find them in $\mathbb{Z}/n\mathbb{Z}$.

3.5.2 Which classes are units

In $\mathbb{Z}/6\mathbb{Z}$ the units $[1]$ and $[5]$ are the classes of the integers coprime to 6, and the non-units $[0], [2], [3], [4]$ share a factor with 6. The equation $[a][x] = [1]$ says $n \mid ax - 1$, that is, $ax + ny = 1$ for some integer y: a Bézout identity. So the question is answered by Section 3.2.

Theorem 3.5.1. Units Modulo n

Let $n \in \mathbb{Z}_{>0}$ and $a \in \mathbb{Z}$. Then $[a]$ is a unit of $\mathbb{Z}/n\mathbb{Z}$ **if and only if** $\gcd(a, n) = 1$. In that case, for **any** $x, y \in \mathbb{Z}$ with $ax + ny = 1$, $[a]^{-1} = [x]$.

Idea. Both directions translate between one equation in $\mathbb{Z}/n\mathbb{Z}$ and one equation in $\mathbb{Z}$. A solution of $[a][x] = [1]$ is an integer combination $ax + ny = 1$, and Corollary 3.2.2 says that such a combination exists exactly when $\gcd(a, n) = 1$. The class test does the translation.

Proof. ($\Leftarrow$) Suppose $\gcd(a, n) = 1$. By Corollary 3.2.2 there are $x, y \in \mathbb{Z}$ with $ax + ny = 1$. Let x, y be any such pair. Then $ax - 1 = -ny = (-y)n$, so $n \mid ax - 1$, and by the class test $[ax] = [1]$. Hence $[a][x] = [ax] = [1]$, where the first equality is the definition of the product on classes (Theorem 3.4.1 (b)). So $[a]$ is a unit, and $[x]$ is its inverse, which is unique (Definition 3.5.1); thus $[a]^{-1} = [x]$ for every such pair.

($\Rightarrow$) Suppose $[a]$ is a unit. Every element of $\mathbb{Z}/n\mathbb{Z}$ is the class of an integer (Sets, Relations, and Cardinality, Definition 2.5.4), so $[a][x] = [1]$ for some $x \in \mathbb{Z}$. Then $[ax] = [1]$, so $n \mid ax - 1$ by the class test: $ax - 1 = yn$ for some $y \in \mathbb{Z}$. Hence $ax + n(-y) = 1$, and $\gcd(a, n) = 1$ by Corollary 3.2.2. This proves the theorem. ∎

The theorem turns "is $[a]$ invertible?" into a gcd, and the Euclidean algorithm (Theorem 3.2.4)

computes the gcd. Back-substitution, as in Example (Two runs of the algorithm), then produces the x with $ax + ny = 1$, and with it the inverse. Whether $\gcd(a, n) = 1$ depends only on the class $[a]$, since by the theorem it is equivalent to a statement about $[a]$; so any representative may be used.

Example. Inverting a class modulo sixty

Decide whether $[17]$ and $[12]$ are units of $\mathbb{Z}/60\mathbb{Z}$, and find the inverse of each one that is.

Solution. *The class* $[17]$. Run the Euclidean algorithm on $(60, 17)$:

$$60 = 3 \cdot 17 + 9, \quad 17 = 1 \cdot 9 + 8,$$
$$9 = 1 \cdot 8 + 1, \qquad 8 = 8 \cdot 1 + 0.$$

The last non-zero remainder is 1, so $\gcd(60, 17) = 1$ (Theorem 3.2.4 (c)), and $[17]$ is a unit by Theorem 3.5.1. Back-substitute, using the lines from the bottom up:

$$1 = 9 - 8 = 9 - (17 - 9) = 2 \cdot 9 - 17$$
$$= 2 \cdot (60 - 3 \cdot 17) - 17 = 2 \cdot 60 - 7 \cdot 17.$$

So $17 \cdot (-7) + 60 \cdot 2 = 1$, and Theorem 3.5.1 gives $[17]^{-1} = [-7]$. The representative in $\{0, \ldots, 59\}$ is 53, since $-7 - 53 = -60$. Check: $17 \cdot 53 = 901 = 15 \cdot 60 + 1$, so $[17][53] = [1]$.

The class $[12]$. Here $12 \mid 60$, so 12 is a common divisor of 12 and 60, and every common divisor of them divides 12; hence $\gcd(12, 60) = 12 \neq 1$ (Definition 3.2.2, with Proposition 3.1.1 (e) for the comparison), and $[12]$ is not a unit by Theorem 3.5.1. A witness: $[12][5] = [60] = [0]$ with $[5] \neq [0]$, so $[12]$ is a zero divisor, and a unit never is (Definition 3.5.1 and the paragraph after it).

Quick check. Find the inverse of $[3]$ in $\mathbb{Z}/7\mathbb{Z}$.

Solution. $[3]^{-1} = [5]$: $3 \cdot 5 = 15 = 2 \cdot 7 + 1$, so $[3][5] = [1]$. In the form of Theorem 3.5.1, $3 \cdot 5 + 7 \cdot (-2) = 1$.

3.5.3 The field of integers modulo a prime

By Theorem 3.5.1, every non-zero class of $\mathbb{Z}/n\mathbb{Z}$ is a unit exactly when every a with $n \nmid a$ is coprime to n. For a prime, Corollary 3.3.2 (a) says precisely this. For a composite n, a proper factor of n is neither divisible by n nor coprime to it. So the field property sorts the moduli into primes and the rest.

Theorem 3.5.2. The Integers Modulo n Form a Field Exactly When n Is Prime

Let $n \in \mathbb{Z}$ with $n \geq 2$. Then $\mathbb{Z}/n\mathbb{Z}$ is a field in the sense of Definition 2.3.1, that is, a

commutative ring with $[1] \neq [0]$ in which every non-zero element is a unit, **if and only if** n is prime.

Idea. For a prime p, a non-zero class $[a]$ has $p \nmid a$, so the only common divisors left are ± 1, and Theorem 3.5.1 inverts $[a]$. For a composite $n = rs$ with $1 < r, s < n$, the classes $[r]$ and $[s]$ are non-zero with product $[n] = [0]$; a zero divisor cannot be a unit, so (F8) fails.

Proof. By Theorem 3.4.3 (a), $\mathbb{Z}/n\mathbb{Z}$ is a commutative ring, and by Theorem 3.4.3 (c), $[1] \neq [0]$ since $n \geq 2$. So it is a field if and only if every non-zero class is a unit (Definition 2.3.1, (F8), and Definition 3.5.1).

($\Leftarrow$) Let n be prime, and let $[a] \neq [0]$. By Theorem 3.4.1 (d), $a \notin n\mathbb{Z}$, that is, $n \nmid a$. By Corollary 3.3.2 (a), $\gcd(n, a) = 1$, and $\gcd(a, n) = \gcd(n, a)$ (Definition 3.2.2). So $[a]$ is a unit by Theorem 3.5.1.

($\Rightarrow$) Let $n \geq 2$ be not prime. By Sets, Relations, and Cardinality, Definition 2.2.2, n has a positive divisor $r \notin \{1, n\}$; write $n = rs$ with $s \in \mathbb{Z}$. Then $1 \leq r$ by discreteness (Theorem 2.3.4 (c)) and $r \leq n$ by Proposition 3.1.1 (e), so $1 < r < n$. Next, $s > 0$: if $s \leq 0$, then $rs \leq 0$ ($rs = 0$ if $s = 0$ by Proposition 2.3.2 (b), and $rs < r \cdot 0 = 0$ if $s < 0$ by Proposition 2.3.5 (c)), against $n > 0$. So s is a positive divisor of n, and $s \leq n$ by Proposition 3.1.1 (e). Also $s \neq n$: otherwise $rn = n$, so $(r - 1)n = 0$ and $r = 1$ by Proposition 2.3.5 (f), since $n \neq 0$. Hence $0 < s < n$ as well.

Now $[r] \neq [0]$: if $n \mid r$, then $n \leq r$ by Proposition 3.1.1 (e), since $n, r > 0$, against $r < n$. In the same way $[s] \neq [0]$. But $[r][s] = [rs] = [n] = [0]$, by the class test. Suppose $[r]$ were a unit. Then $[s] = [0]$, since a unit is not a zero divisor (the paragraph after Definition 3.5.1). This contradiction shows that the non-zero class $[r]$ is not a unit, so $\mathbb{Z}/n\mathbb{Z}$ is not a field. This proves the theorem.

■

So for every prime p there is a field with exactly p elements (Theorem 3.4.3 (b)). The smallest is $\mathbb{Z}/2\mathbb{Z}$, whose tables are those of the two-element field of Example (A field with two elements that cannot be ordered) with 0, 1 renamed $[0]$, $[1]$. *Algebraic Preliminaries*, Chapter 2 studies fields in general and gives these finite fields a name of their own.

Remark. In particular $\mathbb{Z}/5\mathbb{Z}$ is a field, since 5 is prime (the small-primes remark of Section 3.1). In it $5 \cdot [1] = [5] = [0]$ (Theorem 3.4.3 (d)), so $[1]+[1]+[1]+[1]+[1] = [0]$ with $[1] \neq [0]$, and no order makes it an ordered field (Theorem 3.4.3 (e)). Being a field does not help a finite ring to be ordered.

Warning. In $\mathbb{Z}/p\mathbb{Z}$, "$[a] \neq [0]$" means $p \nmid a$, not $a \neq 0$. The class $[7]$ is the zero of $\mathbb{Z}/7\mathbb{Z}$, and it has no inverse, although $7 \neq 0$. The proof above applies Corollary 3.3.2 (a) to the condition $p \nmid a$; with only $a \neq 0$ in hand, the case $p \mid a$, $\gcd(p, a) = p$, is left open.

3.5.4 Two congruences at once

A number leaves the remainder 2 on division by 3 and the remainder 3 on division by 5. What can it be? Each condition alone is a single class. The question is whether the two classes meet, and in how much. The next theorem says that for coprime moduli they always meet, and in exactly one class modulo the product.

Theorem 3.5.3. Chinese Remainder Theorem for Two Moduli

Let $m, n \in \mathbb{Z}_{>0}$ with $\gcd(m,n) = 1$, and choose $u, v \in \mathbb{Z}$ with $mu + nv = 1$ (Corollary 3.2.2).

(a) For all $a, b \in \mathbb{Z}$, the integer $x := a \cdot nv + b \cdot mu$ satisfies

$$x \equiv a \pmod{m} \qquad \text{and} \qquad x \equiv b \pmod{n},$$

and an integer x' satisfies both congruences **if and only if** $x' \equiv x \pmod{mn}$.

(b) The rule $\Phi([x]_{mn}) := ([x]_m, [x]_n)$ is a well-defined **bijection** $\mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. With $+$ and $\cdot$ on $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ taken componentwise, for all $x, y \in \mathbb{Z}$,

$$\Phi([x]_{mn} + [y]_{mn}) = \Phi([x]_{mn}) + \Phi([y]_{mn}),$$
$$\Phi([x]_{mn}[y]_{mn}) = \Phi([x]_{mn})\,\Phi([y]_{mn}),$$

and $\Phi([1]_{mn}) = ([1]_m, [1]_n)$.

Try it first (ten minutes): find an x with $x \equiv 2 \pmod 3$ and $x \equiv 3 \pmod 5$ by trying. Then use $1 = 2 \cdot 3 - 1 \cdot 5$ to find a formula that works for every pair of right-hand sides.

Idea. Trying works: among $3, 8, 13, \ldots$, the numbers that are 3 modulo 5, the second one is 2 modulo 3. But trial and error gives one answer and no reason, and it does not say why no answer can be missing. The identity $mu + nv = 1$ supplies two special integers:

$$e := nv = 1 - mu, \qquad f := mu = 1 - nv.$$

The first is $\equiv 1 \pmod m$ and $\equiv 0 \pmod n$; the second the other way round. So $ae + bf$ is a modulo m and b modulo n, like a vector built from two coordinate vectors. For 3 and 5, $e = -5$ and $f = 6$, and $2e + 3f = 8$. Uniqueness modulo mn is Proposition 3.3.3: a difference divisible by m and by n is divisible by mn.

Proof.

(a) Since $nv - 1 = -mu$ and $mu - 1 = -nv$,

$$x - a = a(nv - 1) + bmu = m(bu - au),$$
$$x - b = anv + b(mu - 1) = n(av - bv).$$

So $m \mid x - a$ and $n \mid x - b$: $x \equiv a \pmod m$ and $x \equiv b \pmod n$ (Sets, Relations, and Cardinality, Definition 2.5.5).
Let $x' \in \mathbb{Z}$. Suppose first that $x' \equiv a \pmod m$ and $x' \equiv b \pmod n$. Then $x' - x = (x' - a) - (x - a)$ is divisible by m, and $x' - x = (x' - b) - (x - b)$ by n (Proposition 3.1.1 (c)). Since $\gcd(m,n) = 1$, Proposition 3.3.3 gives $mn \mid x' - x$, that is, $x' \equiv x \pmod{mn}$.

Conversely, suppose $mn \mid x' - x$. Since $m \mid mn$, $m \mid x' - x$ (Proposition 3.1.1 (b)), and then $m \mid (x' - x) + (x - a) = x' - a$ by Proposition 3.1.1 (c). In the same way $n \mid x' - b$. So x' satisfies both congruences.

(b) *Well defined.* Let $k \in \{m, n\}$. If $x \equiv x' \pmod{mn}$, then $k \mid mn \mid x - x'$, so $x \equiv x' \pmod{k}$ (Proposition 3.1.1 (b)). By Sets, Relations, and Cardinality, Theorem 2.5.1 (b) and Theorem 2.2.1 (a), applied to the identity map of $\mathbb{Z}$, with congruence modulo mn on the source and modulo k on the target, there is exactly one function $\mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/k\mathbb{Z}$ sending $[x]_{mn}$ to $[x]_k$. Pairing the two values for $k = m$ and $k = n$ gives the function Φ into $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ (Sets, Relations, and Cardinality, Definition 1.3.2).

Surjective. Let $(\alpha, \beta) \in \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. By Sets, Relations, and Cardinality, Definition 2.5.4, $\alpha = [a]_m$ and $\beta = [b]_n$ for some $a, b \in \mathbb{Z}$. With x as in (a), $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$, so $\Phi([x]_{mn}) = ([x]_m, [x]_n) = ([a]_m, [b]_n)$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

Injective. Let $z, z' \in \mathbb{Z}$ with $\Phi([z]_{mn}) = \Phi([z']_{mn})$. Then $[z']_m = [z]_m$ and $[z']_n = [z]_n$, so $m \mid z' - z$ and $n \mid z' - z$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b). By Proposition 3.3.3, $mn \mid z' - z$, that is, $[z']_{mn} = [z]_{mn}$. So Φ is a bijection (Sets, Relations, and Cardinality, Definition 1.4.5).

The operations. By the definition of $+$ on each of the three rings (Theorem 3.4.1 (b)) and of the componentwise sum,

$$\begin{aligned}\Phi([x]_{mn} + [y]_{mn}) &= \Phi([x + y]_{mn}) = ([x + y]_m, [x + y]_n) \\ &= ([x]_m + [y]_m, [x]_n + [y]_n) = \Phi([x]_{mn}) + \Phi([y]_{mn}).\end{aligned}$$

The same lines with xy in place of $x + y$ give the product, and $\Phi([1]_{mn}) = ([1]_m, [1]_n)$ is the definition. This proves the theorem.

■

The theorem answers the opening question: the numbers with remainder 2 modulo 3 and remainder 3 modulo 5 form one class modulo 15, and the formula of (a) names a representative. Clause (b) says the same thing as a bijection. Each class modulo mn is the same information as a pair of classes, one modulo m and one modulo n, and the sums and products of classes can be computed in either form. Each coordinate of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ is a commutative ring, so with zero $([0]_m, [0]_n)$, one $([1]_m, [1]_n)$ and negatives taken componentwise, every axiom (F1)–(F7), (F9) holds in each coordinate and hence for the pairs; then Φ matches two commutative rings, element for element and operation for operation.

Example. Two remainders at once, and two failures

(i) Find every $x \in \mathbb{Z}$ with $x \equiv 2 \pmod{3}$ and $x \equiv 3 \pmod{5}$.

(ii) Show that $x \equiv 0 \pmod{2}$, $x \equiv 1 \pmod{4}$ has no solution, and that the solutions of $x \equiv 0 \pmod{2}$, $x \equiv 2 \pmod{4}$ are not one class modulo 8. Which hypothesis of Theorem 3.5.3 fails, and what happens to its map $\Phi\colon \mathbb{Z}/8\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$?

Solution.

(i) Here $m = 3$, $n = 5$, and $1 = 2 \cdot 3 + (-1) \cdot 5$, so $\gcd(3, 5) = 1$ (Corollary 3.2.2) with

$u = 2$, $v = -1$. By Theorem 3.5.3 (a),

$$x = 2 \cdot 5 \cdot (-1) + 3 \cdot 3 \cdot 2 = -10 + 18 = 8.$$

Check: $8 = 2 \cdot 3 + 2$ and $8 = 1 \cdot 5 + 3$. The solutions are exactly the x' with $x' \equiv 8 \pmod{15}$, the class $8 + 15\mathbb{Z} = \{\dots, -7, 8, 23, 38, \dots\}$.

(ii) Here $\gcd(2, 4) = 2 \neq 1$: 2 divides both, and every common divisor divides 2, since $2 = 1 \cdot 2 + 0 \cdot 4$ (Proposition 3.1.1 (c), (e)). The hypothesis $\gcd(m, n) = 1$ fails, and both halves of the conclusion fail with it.

Existence fails. If $x \equiv 1 \pmod 4$, then $x = 4k + 1 = 2(2k) + 1$ for some $k \in \mathbb{Z}$, so x has the form $2q + 1$. If also $x \equiv 0 \pmod 2$, then $x = 2q'$. No integer has both forms (Corollary 3.1.3 (a)), so there is no solution.

Uniqueness modulo mn fails. If $x \equiv 2 \pmod 4$, then $x = 4k + 2 = 2(2k + 1)$, so $x \equiv 0 \pmod 2$ holds automatically, and the solutions are exactly the class $2 + 4\mathbb{Z}$. It contains 2 and 6, but $6 - 2 = 4$ and $8 \nmid 4$ (Proposition 3.1.1 (e)), so $2 \not\equiv 6 \pmod 8$: the solutions form two classes modulo 8, $[2]_8$ and $[6]_8$.

The map. The rule $\Phi([x]_8) := ([x]_2, [x]_4)$ is still well defined, since $2 \mid 8$ and $4 \mid 8$ (the first step of the proof of (b) used only these divisibilities). It is not injective: $\Phi([0]_8) = ([0]_2, [0]_4) = \Phi([4]_8)$, since $4 \equiv 0$ modulo 2 and modulo 4, while $[4]_8 \neq [0]_8$. It is not surjective either: by the existence part, no x has $\Phi([x]_8) = ([0]_2, [1]_4)$.

Remark. For a map from a finite set to itself, injectivity alone gives surjectivity, by counting. Take $g\colon \mathbb{Z}/5\mathbb{Z} \to \mathbb{Z}/5\mathbb{Z}$, $g([x]) := [2][x]$. It is injective: if $[2][x] = [2][y]$, multiplying by $[3]$, with $[3][2] = [6] = [1]$, gives $[x] = [y]$ ((F5)–(F7)). Let $T := g(\mathbb{Z}/5\mathbb{Z})$. With codomain T, g is a bijection $\mathbb{Z}/5\mathbb{Z} \to T$, and $\mathbb{Z}/5\mathbb{Z}$ is finite (Theorem 3.4.3 (b)); so T is not a proper subset of $\mathbb{Z}/5\mathbb{Z}$ (Sets, Relations, and Cardinality, Theorem 3.1.9), that is, $T = \mathbb{Z}/5\mathbb{Z}$, and g is surjective. Could injectivity of Φ alone have given its surjectivity in the same way? Its source and target are different sets, so their sizes must be compared first; Exercise B3 asks for this. The formula of (a) is kept because it does more: it **computes** the solution.

Warning. The coprimality hypothesis is used twice in Theorem 3.5.3: for the identity $mu + nv = 1$, which builds x, and in Proposition 3.3.3, which makes the solution unique modulo mn. Part (ii) of Example (Two remainders at once, and two failures) shows each use failing on its own: the moduli $2, 4$ with right-hand sides $0, 1$ have no solution, and with $0, 2$ they have solutions spread over two classes modulo 8.

In this section the identity $1 = ax + by$ of Section 3.2 gave inverses (Theorem 3.5.1) and the solutions of two congruences (Theorem 3.5.3), inside the rings that Section 3.4's quotient construction built.

3.5.5 Exercises

3.5.5.1 A. Check your understanding

Exercise. A1

(a) Find $[5]^{-1}$ in $\mathbb{Z}/12\mathbb{Z}$.

(b) List the units of $\mathbb{Z}/12\mathbb{Z}$.

(c) True or false: $\mathbb{Z}/1\mathbb{Z}$ is a field. Justify your answer.

(d) Find every $x \in \mathbb{Z}$ with $x \equiv 1 \pmod 4$ and $x \equiv 2 \pmod 9$.

(e) In the proof of Theorem 3.5.2, where is it used that a unit is not a zero divisor?

Solution.

(a) $[5]^{-1} = [5]$, since $5 \cdot 5 = 25 = 2 \cdot 12 + 1$.

(b) By Theorem 3.5.1, $[a]$ is a unit exactly when $\gcd(a, 12) = 1$, and by Theorem 3.4.3 (b) each class is $[a]$ for exactly one $a \in \{0, \dots, 11\}$. Among these, $0, 2, 4, 6, 8, 10$ share the divisor 2 with 12, and $3, 9$ share 3. The rest are coprime to 12: $1 \cdot 1 = 1$, $5 \cdot 5 = 25$, $7 \cdot 7 = 49$ and $11 \cdot 11 = 121$ are each 1 more than a multiple of 12, so each class is its own inverse. The units are $[1], [5], [7], [11]$.

(c) **False.** In $\mathbb{Z}/1\mathbb{Z}$ every a has $1 \mid a - 0$, so $[1] = [0]$ (Theorem 3.4.3 (c)), and a field requires $1 \neq 0$ (Definition 2.3.1). It is the zero ring. (Every element of it is a unit, as in the zero ring of the examples after Definition 3.5.1; that is not enough.)

(d) $\gcd(4, 9) = 1$, from $4 \cdot (-2) + 9 \cdot 1 = 1$ (Corollary 3.2.2), so Theorem 3.5.3 applies with $m = 4$, $n = 9$, $u = -2$, $v = 1$:

$$x = 1 \cdot 9 \cdot 1 + 2 \cdot 4 \cdot (-2) = 9 - 16 = -7.$$

The solutions are the $x' \equiv -7 \pmod{36}$, that is, the class $29 + 36\mathbb{Z}$, since $-7 + 36 = 29$. Check: $29 = 7 \cdot 4 + 1 = 3 \cdot 9 + 2$.

(e) In the direction ($\Rightarrow$), at the end: $[r][s] = [0]$ with $[s] \neq [0]$, so $[r]$, a non-zero class, cannot be a unit. That is how the proof shows (F8) fails.

3.5.5.2 B. Practice

Exercise. B1

Let p be prime and $a \in \mathbb{Z}$. Prove that if $[a]^2 = [1]$ in $\mathbb{Z}/p\mathbb{Z}$, then $[a] = [1]$ or $[a] = [-1]$. Show by an example that this fails in $\mathbb{Z}/8\mathbb{Z}$.

Solution. $[a]^2 = [a][a] = [a^2]$, so $[a^2] = [1]$, and $p \mid a^2 - 1$ by the class test. Expanding (Proposition 2.3.2), $a^2 - 1 = (a - 1)(a + 1)$. By Corollary 3.3.2 (b), $p \mid a - 1$ or $p \mid a + 1$. In the first case $[a] = [1]$; in the second, $a + 1 = a - (-1)$, so $[a] = [-1]$, again by the class test.

In $\mathbb{Z}/8\mathbb{Z}$, $[3]^2 = [9] = [1]$, since $9 - 1 = 8$. But $[3] \neq [1]$, since $8 \nmid 2$, and $[3] \neq [-1]$, since $3 - (-1) = 4$ and $8 \nmid 4$ (Proposition 3.1.1 (e)). The proof breaks at Corollary 3.3.2 (b): $8 \mid 2 \cdot 4 = (3 - 1)(3 + 1)$, yet 8 divides neither factor. (So do [5] and [7]: all four units of $\mathbb{Z}/8\mathbb{Z}$ square to [1].)

Exercise. B2

(a) Let $m, n \in \mathbb{Z}_{>0}$ and $x \in \mathbb{Z}$. Show that $[x]_{mn}$ is a unit of $\mathbb{Z}/mn\mathbb{Z}$ if and only if $[x]_m$ and $[x]_n$ are units of $\mathbb{Z}/m\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$.

(b) Now let m and n be coprime. Deduce that the map Φ of Theorem 3.5.3 (b) maps $(\mathbb{Z}/mn\mathbb{Z})^\times$ bijectively onto $(\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$.

(c) Which of (a) and (b) uses coprimality? Illustrate with $m = n = 2$.

Solution.

(a) ($\Rightarrow$) Let $[x]_{mn}[s]_{mn} = [1]_{mn}$, with $s \in \mathbb{Z}$. Then $mn \mid xs-1$ by the class test, and $m \mid mn$, so $m \mid xs - 1$ (Proposition 3.1.1 (b)); thus $[x]_m[s]_m = [1]_m$. Likewise $[x]_n[s]_n = [1]_n$. So both are units.
($\Leftarrow$) Let $[x]_m[s]_m = [1]_m$ and $[x]_n[s']_n = [1]_n$. By the class test, $xs - 1 = -mt$ and $xs' - 1 = -nt'$ for some $t, t' \in \mathbb{Z}$, so $xs + mt = 1$ and $xs' + nt' = 1$. Multiplying the two equations and expanding,

$$x(xss' + snt' + s'mt) + mn \cdot tt' = 1.$$

So $\gcd(x, mn) = 1$ (Corollary 3.2.2), and $[x]_{mn}$ is a unit (Theorem 3.5.1).

(b) By (a), $\Phi([x]_{mn}) = ([x]_m, [x]_n)$ lies in $(\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$ exactly when $[x]_{mn}$ is a unit. So Φ maps $(\mathbb{Z}/mn\mathbb{Z})^\times$ into that set, and the restriction is injective, since Φ is. Each element of that set is $\Phi(c)$ for some $c \in \mathbb{Z}/mn\mathbb{Z}$, since Φ is surjective, and that c is a unit by (a). So the restriction is a bijection.

(c) Only (b), through Theorem 3.5.3; (a) holds for all m, n. For $m = n = 2$, (a) says that $[x]_4$ is a unit exactly when $[x]_2$ is, and both hold exactly for odd x: the units of $\mathbb{Z}/4\mathbb{Z}$ are $[1], [3]$ and that of $\mathbb{Z}/2\mathbb{Z}$ is $[1]$. Yet $\Phi([x]_4) := ([x]_2, [x]_2)$ sends both units $[1]_4$ and $[3]_4$ to $([1]_2, [1]_2)$, so it is not injective on the units, and it is not a bijection $\mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ either (compare Example (Two remainders at once, and two failures) (ii)).

Exercise. B3

Let $m, n \in \mathbb{Z}_{>0}$ be coprime. Show by counting that injectivity of the map Φ of Theorem 3.5.3 (b) already forces its surjectivity, without using the formula of part (a).

Solution. By Theorem 3.4.3 (b), applied to the modulus mn, the set $\mathbb{Z}/mn\mathbb{Z}$ has mn elements: there is a bijection $\varphi\colon \{1, \dots, mn\} \to \mathbb{Z}/mn\mathbb{Z}$. Let $T := \Phi(\mathbb{Z}/mn\mathbb{Z})$ be the image of Φ. Since Φ is injective, Φ with codomain T is a bijection, and so is its composite with φ (Sets, Relations, and Cardinality, Proposition 3.1.1 (c)): T has mn elements.

The target $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ also has mn elements, by Theorem 3.4.3 (b) for m and for n, and the count of a product, part (a) of Sets, Relations, and Cardinality, Exercise (C2: Counting products and functions). Suppose T were a proper subset of the target. Then Sets, Relations, and Cardinality, Proposition 3.1.6 gives it k elements for some $k < mn$, and Sets, Relations, and Cardinality, Corollary 3.1.4 gives $k = mn$: a contradiction. So T is the whole target, and Φ is surjective. (Equivalently: composing the two bijections from $\{1, \dots, mn\}$ gives a bijection from the target onto T, which Sets, Relations, and Cardinality, Theorem 3.1.9

forbids when T is a proper subset.)

3.5.5.3 C. Going deeper

Exercise. C1: Exit check: units and the field $\mathbb{Z}/p\mathbb{Z}$

Let $n \in \mathbb{Z}_{>0}$ and $a \in \mathbb{Z}$. Prove that $[a] \in \mathbb{Z}/n\mathbb{Z}$ is a unit if and only if $\gcd(a, n) = 1$. Deduce that $\mathbb{Z}/p\mathbb{Z}$ is a field for every prime p.

Solution. Units, ($\Leftarrow$). If $\gcd(a, n) = 1$, then $ax + ny = 1$ for some $x, y \in \mathbb{Z}$ (Corollary 3.2.2). The product on $\mathbb{Z}/n\mathbb{Z}$ is well defined and computed on representatives (Theorem 3.4.1 (b)), so $[a][x] = [ax]$. Since $ax - 1 = (-y)n$, $n \mid ax - 1$, and $[ax] = [1]$ by the class test. So $[a]$ is a unit with inverse $[x]$.

Units, ($\Rightarrow$). If $[a][x] = [1]$ for some $x \in \mathbb{Z}$, then $[ax] = [1]$, so $ax - 1 = yn$ with $y \in \mathbb{Z}$, and $ax + n(-y) = 1$. By Corollary 3.2.2, $\gcd(a, n) = 1$.

The field. Let p be prime. $\mathbb{Z}/p\mathbb{Z}$ is a commutative ring (Theorem 3.4.3 (a)), and $[1] \neq [0]$, since $[1] = [0]$ would give $p \mid 1$, so $p \leq 1$ by Proposition 3.1.1 (e), while $p \geq 2$. Let $[a] \neq [0]$. Then $p \nmid a$. The integer $g := \gcd(p, a) \geq 1$ divides p, so $g \in \{1, p\}$ by Sets, Relations, and Cardinality, Definition 2.2.2; and $g = p$ would give $p \mid a$. So $g = 1$, and $[a]$ is a unit by the first part. Every non-zero class is a unit, which is (F8), so $\mathbb{Z}/p\mathbb{Z}$ is a field.

Marking scheme. (10 points) - 3: ($\Leftarrow$) of units: a Bézout identity $ax + ny = 1$ and the class computation $[a][x] = [ax] = [1]$, naming that the product is well defined (computed on representatives). - 3: ($\Rightarrow$) of units: from $[a][x] = [1]$ to an integer combination of a and n equal to 1, then $\gcd(a, n) = 1$ with a cited reason. - 2: $[a] \neq [0]$ gives $p \nmid a$, and $\gcd(a, p) = 1$ because the gcd divides p and is not p. - 1: $[1] \neq [0]$ for $p \geq 2$. - 1: concludes with (F8): every non-zero class is a unit. Fatal errors (0 for the step): canceling $[a]$ in $\mathbb{Z}/p\mathbb{Z}$ without justification (circular, since cancellation is what inverses give); "$\gcd(a, p) = 1$ because p is prime" with no treatment of the case $p \mid a$. A proof of the field part that cites Corollary 3.3.2 (a) instead of the divisor argument earns the same 2 points.

Exercise. C2: Find the error

Here is an argument.

> **Claim.** For every prime p, $\mathbb{Z}/p\mathbb{Z}$ is a field.
>
> **Proof.** (0) By Theorem 3.4.3 (a) and (c), $\mathbb{Z}/p\mathbb{Z}$ is a commutative ring with $[1] \neq [0]$, since $p \geq 2$. (1) Let $[a] \neq [0]$ in $\mathbb{Z}/p\mathbb{Z}$. (2) If $[a][x] = [a][y]$, then canceling $[a]$ gives $[x] = [y]$. (3) So the map $f\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$, $f([x]) := [a][x]$, is injective. (4) With codomain its image, f is a bijection from $\mathbb{Z}/p\mathbb{Z}$ onto $f(\mathbb{Z}/p\mathbb{Z})$, a subset of the finite set $\mathbb{Z}/p\mathbb{Z}$ (Theorem 3.4.3 (b)). By Sets, Relations, and Cardinality, Theorem 3.1.9 the image is not a proper subset, so f is surjective. (5) So some $[x]$ has $[a][x] = [1]$, and $[a]$ is a unit. So (F8) holds, and $\mathbb{Z}/p\mathbb{Z}$ is a field.

Is the claim true? Does the argument prove it? Name the line that fails and the reason, give a counterexample to the illegal step, and repair the argument.

Solution (Hint 1). List the properties of $\mathbb{Z}/p\mathbb{Z}$ that each line uses. Which of them have been proved in this chapter, and which are things a field has?

Solution (Hint 2). Write $[a][x] = [a][y]$ as a divisibility statement about integers.

Solution. **The claim is true:** it is Theorem 3.5.2. **The argument does not prove it as written.**

Line (2) fails. "Cancel a non-zero factor" is a rule of fields (multiply by the inverse) or of rings without zero divisors, and neither has been established for $\mathbb{Z}/p\mathbb{Z}$: it is close to what the argument is trying to prove, and it is false in general. So the line cites a "fact" that is neither proved nor true, entry (N9) of Logic and Proof, Example (What is not allowed). Counterexample to the step: in $\mathbb{Z}/6\mathbb{Z}$, $[2][1] = [2]$ and $[2][4] = [8] = [2]$, so $[2][1] = [2][4]$ with $[2] \neq [0]$, yet $[1] \neq [4]$. There is no counterexample to the conclusion.

Repair of line (2). If $[a][x] = [a][y]$, then $[a(x-y)] = [0]$, so $p \mid a(x-y)$ by the class test. Since $[a] \neq [0]$, $p \nmid a$, so $p \mid x - y$ by Corollary 3.3.2 (b), and $[x] = [y]$. This uses that p is prime, the hypothesis the flawed line never touched. With this repair, the argument is a correct second proof.

Marking scheme. (10 points) - 2: says the claim is true and the argument, as written, does not prove it. - 3: names line (2) and the reason: cancellation by a non-zero class is unproved here and false in general, (N9), or an equivalent statement of the reason. - 2: a counterexample to the step, such as $[2][1] = [2][4]$ in $\mathbb{Z}/6\mathbb{Z}$. - 3: repairs line (2) without assuming the field property: $p \mid a(x-y)$ by the class test and Euclid's lemma for a prime (Corollary 3.3.2 (b), from Theorem 3.3.1); or $\gcd(a,p) = 1$ (Corollary 3.3.2 (a)) and an inverse of $[a]$ from Theorem 3.5.1, which earns full points even though it makes lines (3)–(5) redundant. Fatal errors (0 for the step): a counterexample to the conclusion (there is none); a repair that cancels $[a]$ by appeal to the field property being proved (citing Theorem 3.5.1 is not such an appeal).

Exercise. C3

Find every $x \in \mathbb{Z}$ with

$$x \equiv 1 \pmod 3, \qquad x \equiv 2 \pmod 5, \qquad x \equiv 3 \pmod 7,$$

by applying Theorem 3.5.3 twice.

Solution. The first two congruences. $1 = 2 \cdot 3 + (-1) \cdot 5$, so $\gcd(3,5) = 1$ (Corollary 3.2.2), with $m = 3$, $n = 5$, $u = 2$, $v = -1$. By Theorem 3.5.3 (a),

$$x_1 = 1 \cdot 5 \cdot (-1) + 2 \cdot 3 \cdot 2 = -5 + 12 = 7,$$

and an integer satisfies the first two congruences **if and only if** it is $\equiv 7 \pmod{15}$.

Adding the third. So the three congruences together say exactly $x \equiv 7 \pmod{15}$ and $x \equiv 3$

(mod 7). Now $1 = 1 \cdot 15 + (-2) \cdot 7$, so $\gcd(15, 7) = 1$, with $m = 15$, $n = 7$, $u = 1$, $v = -2$. By Theorem 3.5.3 (a) again,

$$x_2 = 7 \cdot 7 \cdot (-2) + 3 \cdot 15 \cdot 1 = -98 + 45 = -53,$$

and the solutions are the $x \equiv -53 \pmod{105}$, that is, the class $52 + 105\mathbb{Z}$.

Check: $52 = 17 \cdot 3 + 1 = 10 \cdot 5 + 2 = 7 \cdot 7 + 3$. (Coprimality of 3 and 5, and of 15 and 7, also follows from the small-primes remark of Section 3.1 with Corollary 3.3.2 (a).) The "if and only if" in the first step is what allows replacing two congruences by one; with "if" alone, the second step could lose solutions.

3.5.5.4 R. Review

Exercise. R1

Let $\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/\sim$ be the integers built as classes of pairs, with $(a, b) \sim (c, d) \iff a + d = b + c$. Show that $[(a, b)] \mapsto [a + b]_2$ is a well-defined function $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$, and identify it.

Solution. This is the method of Theorem 2.2.1 (a), from the construction Definition 2.1.1. Take $X = \mathbb{N} \times \mathbb{N}$ with $\sim$, $Y = \mathbb{Z}$ with congruence modulo 2, the equivalence relation $\sim_{2\mathbb{Z}}$ whose quotient set is $\mathbb{Z}/2\mathbb{Z}$ (Definition 3.4.2, Theorem 3.4.1 (a)), and $f(a, b) := a + b$. Let $(a, b) \sim (a', b')$, so $a + b' = a' + b$. Then, in $\mathbb{Z}$,

$$\begin{aligned}(a + b) - (a' + b') &= (a + b' - a' - b) + 2b - 2b' \\ &= 0 + 2(b - b'),\end{aligned}$$

so $a + b \equiv a' + b' \pmod 2$. By Theorem 2.2.1 (a) there is exactly one function $F\colon \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ with $F([(a, b)]) = [a + b]_2$.

Identification. By the remark after Theorem 2.3.1, $[(a, b)] = a - b$ in $\mathbb{Z}$. Since $(a+b)-(a-b) = 2b$, $[a + b]_2 = [a - b]_2$ by the class test. So $F(z) = [z]_2$ for every $z \in \mathbb{Z}$: F is the quotient map of $\mathbb{Z}/2\mathbb{Z}$, which sends an integer to its parity class.

Exercise. R2

Let $m, n, m', n' \in \mathbb{N}$ with $2^m 3^n = 2^{m'} 3^{n'}$. Prove that $m = m'$ and $n = n'$. Conclude that $(m, n) \mapsto 2^m 3^n$ is injective on $\mathbb{N} \times \mathbb{N}$.

Solution. The tool is unique factorization, Theorem 3.3.5. The powers are those of $\mathbb{N} = \omega$ (Proposition 1.6.4), carried into $\mathbb{Z}$, which preserves products (Proposition 2.2.3); $b^0 = 1$ and $b^{j+1} = b^j \cdot b$.

A factorization. For $m, n \in \mathbb{N}$, let $L_{m,n}$ be the list of length $m + n$ whose first m entries are 2 and whose last n entries are 3. Its entries are prime, by the small-primes remark of Section 3.1, and it is non-decreasing, since $2 \le 3$. Its product is $2^m 3^n$. Indeed, with P_j the product of the first j entries, the recursion of Definition 3.3.1 and induction on $j \le m$ give $P_j = 2^j$ ($P_0 = 1 = 2^0$; $P_{j+1} = P_j \cdot 2 = 2^{j+1}$), and then induction on $i \le n$ gives $P_{m+i} = 2^m 3^i$. So $L_{m,n}$ is a prime factorization of $2^m 3^n$, which is positive, being a product of positive integers

(Proposition 2.3.5 (c)).

Uniqueness. If $2^m3^n = 2^{m'}3^{n'}$, then $L_{m,n}$ and $L_{m',n'}$ are prime factorizations of the same positive integer, so they are equal (Theorem 3.3.5): $m+n = m'+n'$, with equal entries. Suppose $m < m'$. Then the entry of index m' is 2 in $L_{m',n'}$, and in $L_{m,n}$ it is 3, since $m < m' \leq m+n$: a contradiction. Exchanging the two lists rules out $m' < m$. So $m = m'$, and then $n = n'$ by canceling m in $m+n = m'+n'$.

This is line (vii) of Sets, Relations, and Cardinality, Exercise (B2: Putting a proof in order), which was not available when that exercise was set; Sets, Relations, and Cardinality, Theorem 3.2.6 proves the same injectivity by parity instead.

Exercise. R3

Prove that no $x \in \mathbb{Q}$ satisfies $x^2 = 2$, by the parity argument: write x as a fraction and show that numerator and denominator are both even. Then show that, for every prime p, no $x \in \mathbb{Q}$ satisfies $x^2 = p$.

Solution. This is the school argument of Logic and Proof, Theorem 2.2.1, now with every step citable.

The case 2. Suppose $x \in \mathbb{Q}$ and $x^2 = 2$. By Proposition 3.3.6, $x = a/b$ with $a \in \mathbb{Z}$, $b \in \mathbb{Z}_{>0}$ and $\gcd(a,b) = 1$. By Definition 2.4.2, $x^2 = a^2/b^2$, and $2 = 2/1$, so $a^2 = 2b^2$ (Proposition 2.4.4 (b)). Then a^2 is even, so a is even by Corollary 3.1.3 (b): $a = 2c$ with $c \in \mathbb{Z}$. Then $4c^2 = 2b^2$, and canceling the non-zero factor 2 (Proposition 2.3.5 (f), applied to $2(b^2 - 2c^2) = 0$) gives $b^2 = 2c^2$. So b is even as well. Now 2 divides a and b, hence divides $\gcd(a,b) = 1$ (Theorem 3.2.1 (c), (d)), so $2 \in \{1,-1\}$ (Proposition 3.1.1 (g)): a contradiction.

A prime p. The same argument works with "$p\mid$" in place of "even": from $a^2 = pb^2$, $p \mid a \cdot a$, so $p \mid a$ by Corollary 3.3.2 (b); $a = pc$ gives $p^2c^2 = pb^2$, so $b^2 = pc^2$ (cancel $p \neq 0$), and $p \mid b$ in the same way. Then $p \mid \gcd(a,b) = 1$, against $p \geq 2$.

Comparison. Chapter 2's proof, Theorem 2.6.1, needed neither parity nor gcd: it descended to a smaller denominator using order alone. The argument here is shorter, but it stands on lowest terms, and lowest terms stands on Bézout (Proposition 3.3.6 uses Corollary 3.2.2).

Exercise. R4

Find all $x \in \mathbb{Z}$ with $6x \equiv 4 \pmod{10}$.

Solution. **Answer:** the class $4 + 5\mathbb{Z}$.

The tool is to divide the congruence by $\gcd(6,10) = 2$, and then invert a unit (Theorem 3.5.1). Since $6x - 4 = 2(3x-2)$ and $10 = 2 \cdot 5$,

$$10 \mid 6x - 4 \iff 5 \mid 3x - 2 :$$

if $2(3x-2) = 10k = 2(5k)$, cancel $2 \neq 0$ (Proposition 2.3.5 (f)); conversely, multiply

$3x - 2 = 5k$ by 2. So the condition is $[3][x] = [2]$ in $\mathbb{Z}/5\mathbb{Z}$. There $[3]^{-1} = [2]$, since $3 \cdot 2 = 6 = 5 + 1$. Multiplying by $[2]$, $[x] = [2][2] = [4]$; and conversely $[3][4] = [12] = [2]$. So the solutions are exactly the $x \equiv 4 \pmod 5$. Check: $6 \cdot 4 - 4 = 20$ and $6 \cdot 9 - 4 = 50$ are multiples of 10.

Why the look-alikes fail. "Multiply by $[6]^{-1}$ in $\mathbb{Z}/10\mathbb{Z}$" fails: $\gcd(6, 10) = 2$, so $[6]$ is not a unit there (Theorem 3.5.1). Canceling the 2 from every term but keeping the modulus gives $3x \equiv 2 \pmod{10}$, and that is wrong: it keeps 4 and loses 9 ($6 \cdot 9 - 4 = 50$, while $3 \cdot 9 - 2 = 25$). The modulus must be divided too. The Chinese remainder theorem is not a look-alike that fails: splitting $10 = 2 \cdot 5$, Theorem 3.5.3 (b) turns the condition $[6]_{10}[x]_{10} = [4]_{10}$ into $[0]_2[x]_2 = [0]_2$, which always holds, and $[x]_5 = [4]_5$, since $[6]_5 = [1]_5$; so it is a second route to the same answer. Note also that the answer is a class modulo 5, not modulo 10: it is the two classes $[4]_{10}$ and $[9]_{10}$.

Chapter 4

Ordered Fields and Sequences in an Ordered Field

In $\mathbb{Q}$ the numbers $2, 3/2, 17/12, 577/408, \ldots$ crowd together, and their squares crowd toward 2; yet Chapter 2 proved that no rational squares to 2 (Theorem 2.6.1). Chapter 5 will make the missing number out of sequences like this one. Before it can, we must say what "crowd together" means when the only yardsticks are rationals, and prove the rules for such sequences without presupposing the number they are aiming at. This chapter does that once, for an **arbitrary ordered field** F, with every tolerance ε taken from F itself. It adds to the inequality rules of Chapter 2 what limits need (halving, fractions, the absolute value and the triangle inequality), places a copy of $\mathbb{N}$ and of $\mathbb{Q}$ inside every ordered field, and names the Archimedean property, which $\mathbb{Q}$ has and an ordered field need not have. Then come limits (uniqueness, boundedness, the limit laws, limits and order) and Cauchy sequences (convergent implies Cauchy, Cauchy implies bounded, sums and products, and the sign of a Cauchy sequence that is not null). Chapter 5 applies all of it with $F = \mathbb{Q}$.

The chapter ends at the gap it was written for. The sequence above is Cauchy in $\mathbb{Q}$ and has no limit in $\mathbb{Q}$, so in $\mathbb{Q}$ "Cauchy" does not imply "convergent". If you have met limits in a course on $\mathbb{R}$, most statements here will look familiar. What is new is that they are proved in any ordered field, from the axioms alone. Watch where the Archimedean property is needed (each use names it), and note that completeness never appears: no supremum, no monotone convergence theorem and no Bolzano–Weierstrass theorem is used, as Section 4.6 checks.

What you need. Chapters 1 and 2, chiefly Sections 2.3, 2.5 and 2.6, and Section 3.4 for review exercise R1 only; *Logic and Proof*, Chapters 2–5, chiefly Chapter 4; and *Sets, Relations, and Cardinality*, Chapters 1 and 2, with Chapter 3 for one exercise. From Chapter 2, Section 2.3: commutative rings and fields, ordered rings and fields and the rules for inequalities (Definition 2.3.1, Definition 2.3.2, Proposition 2.3.2, Proposition 2.3.5, Proposition 2.3.3), the integers as an ordered ring with least elements of sets bounded below (Theorem 2.3.1, Theorem 2.3.4, Proposition 2.3.6), and the two-element field that cannot be ordered (Example (A field with two elements that cannot be ordered)). From Sections 2.2, 2.4 and 2.5: the copies of $\mathbb{N}$ in $\mathbb{Z}$ and of $\mathbb{Z}$ in $\mathbb{Q}$ (Proposition 2.2.3, Proposition 2.4.4), $\mathbb{Q}$ as an ordered field with reciprocals reversing order (Theorem 2.5.1, Definition 2.5.1, Proposition 2.5.2), and the Archimedean property of $\mathbb{Q}$, its form $1/N < \varepsilon$, and the integer part (Theorem 2.5.3, Corollary 2.5.4, Proposition 2.5.5). From Section 2.6: no rational squares to 2, and the Newton sequence with the term-wise facts proved there (Theorem 2.6.1, Example (A Rational Sequence Aimed at the Gap)), which Sections 4.3, 4.5 and 4.6 use. From Chapter 1: the citing rule of Section 1.7 (Corollary 1.7.3), by which induction (Logic and Proof, Axiom 4.2.1) and the recursion theorem (Sets, Relations, and

Cardinality, Theorem 2.3.1, Theorem 1.5.1) apply to $\mathbb{N}$; the arithmetic and order of $\mathbb{N}$ (Definition 1.5.1, Theorem 1.5.3, Theorem 1.6.1, Proposition 1.6.3); and the powers of Section 1.6 (Proposition 1.6.4), which Section 4.2 compares with powers in F and Exercise B2 of Section 4.6 uses as indices. From *Logic and Proof*, Chapter 4: the negation laws (Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2), used to negate the definitions of limit, Cauchy sequence and Archimedean field, and the statement "eventually keeps one sign" in review exercise R3; the translation of a limit statement and the bounded quantifier rules (Logic and Proof, Example (Five definitions, translated), Logic and Proof, Proposition 6.1.1) in Section 4.3; and the argument that numbers arbitrarily close are equal (Logic and Proof, Proposition 6.2.5), which Section 4.3 reruns in F. From *Logic and Proof*, Chapter 2, the rule RAA (Logic and Proof, Definition 2.5.1), and De Morgan's laws (Logic and Proof, Proposition 2.5.5), which R3 also uses; from its Chapter 5, the catalogue of what a proof may not do (Logic and Proof, Example (What is not allowed)), for the find-the-error exercises C2 of Section 4.3 and C3 of Section 4.5. From *Sets, Relations, and Cardinality*: sequences as functions (Sets, Relations, and Cardinality, Definition 1.4.1), from Chapter 1; total orders (Sets, Relations, and Cardinality, Definition 2.6.1) and finite maxima (Sets, Relations, and Cardinality, Proposition 2.3.8), from Sections 2.6 and 2.3. A few items reach further. Exercise C1 of Section 4.2 uses finite and infinite sets (Sets, Relations, and Cardinality, Definition 3.1.2, Sets, Relations, and Cardinality, Definition 3.1.3, Sets, Relations, and Cardinality, Corollary 3.1.3) from *Sets, Relations, and Cardinality*, Section 3.1. Review exercise R1 uses Chapter 3's ideals and quotient ring (Definition 3.4.1, Theorem 3.4.1), from Section 3.4: the null sequences form an ideal of the ring of Cauchy sequences in $\mathbb{Q}$, and the quotient by it, which Chapter 3 built with Chapter 2's lemma on operations on classes (Theorem 2.2.1), gives the classes and their sum. Nothing else from Chapter 3 is used. Section 4.6 names suprema (Sets, Relations, and Cardinality, Definition 2.6.5) without using them. One example, the non-Archimedean field $\mathbb{Q}(t)$ of Section 4.2, uses school algebra of polynomials, which *Algebraic Preliminaries*, Chapter 4 builds; it appears only in examples, remarks, warnings and exercises, and no proof of a numbered result cites it. Nothing from *Analysis on the Real Line* is used, and no fact about $\mathbb{R}$.

Time. About 25 hours on the core track (all of A, one B and one C per section, the exit checks); about 40 hours for everything.

4.0.1 Roadmap

- **Working in an Ordered Field.** The standing setup of an arbitrary ordered field F, the fraction rules of a field and the absence of zero divisors, halving a tolerance and the midpoint, comparing reciprocals and multiplying non-negative inequalities, and the absolute value with the triangle and reverse triangle inequalities, ending in the form $|x - a| < \varepsilon \iff a - \varepsilon < x < a + \varepsilon$ that the rest of the chapter uses.
- **The Rationals inside an Ordered Field; the Archimedean Property.** The elements $n \cdot 1_F$ and powers in F, which add, multiply and compare as in $\mathbb{N}$; the unique embedding ι_F of $\mathbb{Q}$ into every ordered field, strictly increasing and order-reflecting; the Archimedean property as a property of F, with its reciprocal form $1/(n \cdot 1_F) < \varepsilon$; the integer part and the density of the rationals in an Archimedean field; and $\mathbb{Q}(t)$, an ordered field that is not Archimedean.
- **Sequences and Limits in an Ordered Field.** Sequences indexed from 0 and "eventually", the limit with tolerances from F and its negation, first limits in $\mathbb{Q}$ and the divergence of $((-1)^n)$ in every ordered field, null and bounded sequences, $1/(n+1) \to 0$ as a restatement of the Archimedean property, changing finitely many terms, uniqueness of limits, and the boundedness of convergent sequences.
- **The Algebra and Order of Limits.** Null times bounded is null, terms near a non-zero limit stay away from 0, the limit laws for sums, multiples, products and quotients, and

weak inequalities passing to the limit, with the squeeze; no result of the section uses the Archimedean property.

- **Cauchy Sequences.** Terms close to one another instead of to a point, convergent sequences Cauchy and Cauchy sequences bounded, sums and products of Cauchy sequences, the sign theorem (a Cauchy sequence that is not null eventually stays at least some $c > 0$ above 0, or at least c below it), sign stability under null differences, reciprocals, and a sequence of rationals whose steps shrink to 0 and which is not Cauchy.
- **A Cauchy Sequence with No Rational Limit.** The Newton sequence for $\sqrt{2}$ is decreasing with overshoot $e_n \leq 1/(4n)$, Cauchy in $\mathbb{Q}$, and convergent to no rational, since a limit L would satisfy $L^2 = 2$; the section then lists what the chapter did not prove (suprema, monotone convergence, Bolzano–Weierstrass), each of which fails in $\mathbb{Q}$, and the review exercises return to Chapter 3's quotient by an ideal, to induction, and to the negation laws of *Logic and Proof*, Chapter 4.

4.0.2 Named moves

- **Split ε.** To make a sum of k small terms smaller than ε, give each term the tolerance ε/k, and combine the finitely many thresholds by a maximum. It proves the sum law in Section 4.4 and the sum of two Cauchy sequences in Section 4.5; Exercise B3 of Section 4.4 and Exercise C1 of Section 4.5 leave part of it to you.
- **Bound, then shrink.** In a product, replace the factor you cannot control by a bound fixed in advance, then make the other factor small enough to absorb the bound. The bound comes from the whole sequence, before ε is named. It proves the product law in Section 4.4 and the product of two Cauchy sequences in Section 4.5.
- **Fix the bad ε.** "Not null" says: some ε is missed beyond every threshold. It hands you one ε, to keep for the rest of the proof; ask the Cauchy condition for a threshold with $\varepsilon/2$ first, and only then take a term beyond it that misses ε. This is the sign theorem of Section 4.5.
- **Locate the missing limit.** Suppose a limit L exists, and use the limit laws and uniqueness to find an equation that L must satisfy; if the field has no solution, there is no limit. In Section 4.6 the equation is $L^2 = 2$.

4.1 Working in an Ordered Field

Section 2.3 proved the rules for inequalities from the order axioms, and Section 2.5 added the one rule that needs division: reciprocals reverse order. A limit argument asks for a few more. It halves a tolerance, so it needs $0 < \varepsilon/2 < \varepsilon$; it subtracts fractions, so it needs $1/b - 1/d = (d-b)/(bd)$; and above all it measures how far apart two numbers are, so it needs the absolute value and the triangle inequality. Each of these is a few lines from the rules we have. This section proves them once, in an arbitrary ordered field, and the rest of the chapter uses them on nearly every page.

4.1.1 The standing setup

Throughout this chapter, F is an **ordered field** (Definition 2.3.2): a field with a relation $<$ satisfying (O1)–(O4). The rationals are one, by Theorem 2.5.1, and for most of the chapter they are the one that matters: the next chapter applies these results first with $F = \mathbb{Q}$, to build the real numbers from Cauchy sequences of rationals, and then again with $F = \mathbb{R}$, once $\mathbb{R}$ is shown to be an ordered field. We state results for a general F because the proofs never use anything about $\mathbb{Q}$ beyond (F1)–(F9) and (O1)–(O4), and saying so on the page makes it checkable.

We use the results of Sections 2.3 and 2.5 by their letters, without restating them: the algebra of a commutative ring (Proposition 2.3.2), the rules for inequalities (Proposition 2.3.5 (a)–(i)), and the reversal of order by reciprocals (Proposition 2.5.2). In a field $1 \neq 0$, so Proposition 2.3.5 (e) gives $1 > 0$ in every ordered field. We write

$$F_{>0} := \{x \in F : x > 0\}, \qquad F_{\geq 0} := \{x \in F : x \geq 0\},$$

so that $\mathbb{Q}_{>0}$ of Section 2.5 is the case $F = \mathbb{Q}$. Recall that $a/b := ab^{-1}$ for $b \neq 0$, so $1/b = b^{-1}$, and that $2 := 1 + 1$ and $x^2 := xx$ (Proposition 2.3.2 (d)).

4.1.2 Fractions

School algebra adds fractions by cross-multiplying and subtracts reciprocals by a common denominator. Both are identities in any field, and the second is how we will compare $1/b_n$ with $1/b$ in Section 4.4. The only point needing care is that a product of non-zero elements is non-zero, so that bd may stand in a denominator.

Lemma 4.1.1. Fraction Rules in a Field

Let F be a field (Definition 2.3.1), and let $a, b, c, d \in F$ with $b \neq 0$ and $d \neq 0$.

(a) b^{-1} is the **only** element y with $by = 1$, and $(b^{-1})^{-1} = b$. Moreover $bd \neq 0$ and $(bd)^{-1} = b^{-1}d^{-1}$.

(b) $\dfrac{a}{b} + \dfrac{c}{d} = \dfrac{ad + bc}{bd}$ and $\dfrac{a}{b} \cdot \dfrac{c}{d} = \dfrac{ac}{bd}$.

(c) $\dfrac{1}{b} - \dfrac{1}{d} = \dfrac{d - b}{bd}$.

(d) For **all** $x, y \in F$: if $xy = 0$, then $x = 0$ or $y = 0$.

Idea. Everything rests on one move from Section 2.3: to identify an element, show it satisfies the **defining equation** of the element you want and let uniqueness finish. The product $b^{-1}d^{-1}$ satisfies the defining equation of $(bd)^{-1}$; once bd is known to be a legal denominator, (b) and (c) are bookkeeping with (F5), (F6) and (F9).

Proof.

(a) Suppose $by = 1$. Then

$$y = 1 \cdot y = (b^{-1}b)y = b^{-1}(by) = b^{-1} \cdot 1 = b^{-1},$$

by (F7), (F8) with (F6), (F5), the hypothesis, and (F7); this is the uniqueness argument of Proposition 2.3.2 (a), run with multiplication. Next, $b^{-1} \neq 0$: otherwise $1 = bb^{-1} = b \cdot 0 = 0$ by Proposition 2.3.2 (b), against $1 \neq 0$. So $(b^{-1})^{-1}$ exists, and since $b^{-1}b = bb^{-1} = 1$ by (F6), the element b satisfies its defining equation; uniqueness gives $(b^{-1})^{-1} = b$. Now, by (F5) and (F6),

$$(bd)(b^{-1}d^{-1}) = (bb^{-1})(dd^{-1}) = 1 \cdot 1 = 1.$$

If $bd = 0$, the left side would be $0 \cdot (b^{-1}d^{-1}) = 0$ by Proposition 2.3.2 (b) and (F6), so $1 = 0$, which is false in a field. Hence $bd \neq 0$, and the display says that $b^{-1}d^{-1}$ satisfies the defining equation of $(bd)^{-1}$; by uniqueness, $(bd)^{-1} = b^{-1}d^{-1}$.

(b) By (a), $bd \neq 0$, so both right-hand sides are defined. For the product, by (F5), (F6) and (a),

$$\frac{a}{b} \cdot \frac{c}{d} = (ab^{-1})(cd^{-1}) = (ac)(b^{-1}d^{-1})$$
$$= (ac)(bd)^{-1} = \frac{ac}{bd}.$$

For the sum, first $\frac{ad}{bd} = (ad)(b^{-1}d^{-1}) = (ab^{-1})(dd^{-1}) = \frac{a}{b}$, and likewise $\frac{bc}{bd} = \frac{c}{d}$. Then (F9) with (F6) gives

$$\frac{a}{b} + \frac{c}{d} = (ad)(bd)^{-1} + (bc)(bd)^{-1}$$
$$= (ad + bc)(bd)^{-1} = \frac{ad + bc}{bd}.$$

(c) Apply the sum rule of (b) with $a = 1$ and $c = -1$. Since $(-1)d^{-1} = -(d^{-1})$ by Proposition 2.3.2 (c), the left side is $1/b + (-1)/d = 1/b - 1/d$, and the numerator on the right is $1 \cdot d + b(-1) = d - b$, again by Proposition 2.3.2 (c).

(d) Let $xy = 0$, and suppose $x \neq 0$. Then $y = 1 \cdot y = (x^{-1}x)y = x^{-1}(xy) = x^{-1} \cdot 0 = 0$, by (F7), (F8), (F5) and Proposition 2.3.2 (b). So $x = 0$ or $y = 0$. This proves the lemma.

■

In an ordered field, (d) is also Proposition 2.3.5 (f). The point of (d) here is that it needs **no** order: a field has no zero divisors. The commutative ring of integers modulo 6, in the warning after Definition 2.3.1, has them ($[2][3] = [0]$ with $[2] \neq [0] \neq [3]$), so by (d) it is not a field.

4.1.3 Halving and comparing

The commonest move in a limit proof is to give each of two terms half the tolerance. That needs $\varepsilon/2$ to be positive, smaller than ε, and to add up to ε with itself. The next proposition collects these facts with two comparisons we will need as often.

Proposition 4.1.2. Halves and Comparisons in an Ordered Field

Let F be an ordered field.

(a) $\leq$ is a total order on F (Sets, Relations, and Cardinality, Definition 2.6.1), and its strict part is $<$.

(b) $2 := 1 + 1 > 0$. For every $\varepsilon \in F_{>0}$,

$$0 < \frac{\varepsilon}{2} < \varepsilon \qquad \text{and} \qquad \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

For all $a, b \in F$ with $a < b$, $a < \frac{a+b}{2} < b$.

(c) For $a, b \in F_{>0}$: if $a \leq b$, then $\frac{1}{b} \leq \frac{1}{a}$.

(d) For $a, b, c, d \in F$: if $0 \leq a \leq b$ and $0 \leq c \leq d$, then $ac \leq bd$.

Idea. Nothing new happens: each part is one or two of the rules of Sections 2.3 and 2.5 in sequence. The only care needed is with $\leq$, which hides an "or": each non-strict hypothesis splits into an equality case and a strict case, and the strict rules apply only to the second.

Proof.

(a) This is Proposition 2.3.5 (i).

(b) Since $1 > 0$, (O3) gives $2 = 1 + 1 > 0 + 1 = 1$, and $2 > 0$ by (O2). So $2 \neq 0$, and $2^{-1} > 0$ by Proposition 2.5.2 (a). For $\varepsilon > 0$, $\varepsilon/2 = \varepsilon \cdot 2^{-1} > 0$ by (O4). By (F9) and (F7),

$$\begin{aligned}\frac{\varepsilon}{2} + \frac{\varepsilon}{2} &= \varepsilon \cdot 2^{-1} \cdot 1 + \varepsilon \cdot 2^{-1} \cdot 1 \\ &= (\varepsilon \cdot 2^{-1})(1 + 1) = \varepsilon(2^{-1} \cdot 2) = \varepsilon,\end{aligned}$$

the third equality by (F5) and (F6). Adding $\varepsilon/2$ to $0 < \varepsilon/2$ by (O3) gives $\varepsilon/2 < \varepsilon/2 + \varepsilon/2 = \varepsilon$. For the midpoint, let $a < b$. By (O3), $a + a < a + b$ and $a + b < b + b$, and $a + a = a(1 + 1) = 2a$, $b + b = 2b$ by (F7), (F9), (F6). Multiplying by $2^{-1} > 0$ (Proposition 2.3.5 (c)) gives $a < (a + b)/2 < b$, since $(2a)2^{-1} = a(2 \cdot 2^{-1}) = a$.

(c) If $a = b$, then $1/b = 1/a$. Otherwise $0 < a < b$, and $1/b < 1/a$ by Proposition 2.5.2 (b). In both cases $1/b \leq 1/a$.

(d) First, $ac \leq bc$: if $a = b$ or $c = 0$, the two sides are equal (for $c = 0$ both are 0 by Proposition 2.3.2 (b)); otherwise $a < b$ and $c > 0$, and $ac < bc$ by Proposition 2.3.5 (c). Next, $bc \leq bd$, by the same argument with the roles of the factors exchanged by (F6): $c \leq d$, and $b \geq 0$ since $0 \leq a \leq b$ (transitivity, part (a)). By transitivity of $\leq$ (part (a)), $ac \leq bd$.

■

Part (b) is used whenever a proof splits a tolerance: the sum of two quantities, each below $\varepsilon/2$, is below ε, by Proposition 2.3.5 (b). The midpoint is the same fact read for two points: between any two elements of F lies a third. Part (d) is how a product of two small, or two bounded, quantities is estimated.

Warning. The non-negativity in (d) cannot be dropped. In $\mathbb{Q}$, take $a = c = -2$ and $b = d = 0$: then $a \leq b$ and $c \leq d$, but $ac = 4$ and $bd = 0$, and $4 \leq 0$ is false. Section 2.3 gave a strict version of the same trap (the warning after Proposition 2.3.5). Before multiplying two inequalities, check that both sides are non-negative.

4.1.4 Absolute value

To say that x is close to a, we want one element of F measuring the gap between them, whichever of the two is larger. In school this is $|x - a|$, and $|x|$ is sometimes written $\sqrt{x^2}$. That formula needs a square-root function, and an ordered field has none in general: no rational squares to 2 (Theorem 2.6.1). The square x^2 does have the two square roots x and $-x$, and picking the non-negative one is already a definition by cases. So we define the absolute value by cases, using only the order.

The absolute value of x is x with its sign made non-negative.

Definition 4.1.1. Absolute Value in an Ordered Field

Let F be an ordered field. For $x \in F$, the **absolute value** of x is

$$|x| := \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

For $x, y \in F$, the element $|x - y|$ is the **distance** from x to y.

The two cases do not overlap and cover F, by trichotomy (O1), so $|x|$ is a well-defined element of F. The map $x \mapsto |x|$ is a function $F \to F$.

Examples. In $\mathbb{Q}$, $|3/4| = 3/4$ and $|-3/4| = 3/4$, the second because $-3/4 < 0$ (Theorem 2.5.1 (a): $(-3) \cdot 1 < 0 \cdot 4$) and $-(-3/4) = 3/4$ (Proposition 2.3.2 (a)). The distance from $1/3$ to $1/2$ is $|1/3 - 1/2| = |-1/6| = 1/6$. The degenerate case is $|0| = 0$, from the first clause. Part (a) below shows that 0 is the only element with absolute value 0, so the distance from x to y is 0 exactly when $x = y$.

A non-example by minimal change. Replace the definition by $x \mapsto x^2$. This map is also non-negative, zero only at 0, and multiplicative, so it passes (a) and the product rule of (b) below. It fails the triangle inequality (e): in $\mathbb{Q}$, with $x = y = 1$, the "size" of $x + y$ would be $2^2 = 4$, while the sizes of x and y add up to $1 + 1 = 2$. A measure of size must not grow faster than its argument, and that is the clause the square breaks.

Proposition 4.1.3. Rules for the Absolute Value

Let F be an ordered field, and let $x, y, z, r \in F$.

(a) $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$.

(b) $|-x| = |x|$, $|xy| = |x||y|$ and $|x|^2 = x^2$; if $x \neq 0$, then $|1/x| = 1/|x|$.

(c) $-|x| \leq x \leq |x|$.

(d) $|x| \leq r$ if and only if $-r \leq x \leq r$; and $|x| < r$ if and only if $-r < x < r$.

(e) *(Triangle inequality.)* $|x + y| \leq |x| + |y|$, and $|x - z| \leq |x - y| + |y - z|$.

(f) *(Reverse triangle inequality.)* $||x| - |y|| \leq |x - y|$.

Idea. Parts (a)–(d) are case checks on the sign of x (and of y). The triangle inequality then needs **no** cases: by (c), x and y are each trapped between $\pm$ their absolute values; adding the two traps traps $x + y$ between $\pm(|x| + |y|)$, and (d) reads that as the inequality. The reverse triangle inequality is the triangle inequality applied to $x = (x - y) + y$ and to $y = (y - x) + x$.

Proof.

(a) If $x \geq 0$, then $|x| = x \geq 0$. If $x < 0$, then $|x| = -x > 0$ by Proposition 2.3.5 (a), since $-0 = 0$. So $|x| \geq 0$ always, and $|x| = 0$ only in the first case with $x = 0$; conversely $|0| = 0$.

(b) *Negation.* If $x > 0$, then $-x < 0$ (Proposition 2.3.5 (a)), so $|-x| = -(-x) = x = |x|$ (Proposition 2.3.2 (a)). If $x = 0$, both sides are 0. If $x < 0$, then $-x > 0$, so $|-x| = -x = |x|$.
Product. If $x \geq 0$ and $y \geq 0$, then $xy \geq 0$ (by (O4), or Proposition 2.3.2 (b) if a factor is 0), so $|xy| = xy = |x||y|$. If $x \geq 0$ and $y < 0$, then $-y > 0$, and $x(-y) = -(xy)$

by Proposition 2.3.2 (c), so $|x||y| = x(-y) = -(xy)$; and $-(xy) \geq 0$ by the first case, so $xy \leq 0$ and $|xy| = -(xy)$ (if $xy = 0$, both readings give 0). The case $x < 0 \leq y$ is the same with the roles exchanged. If $x < 0$ and $y < 0$, then $xy = (-x)(-y) > 0$ by Proposition 2.3.2 (c) and (O4), so $|xy| = xy = (-x)(-y) = |x||y|$.
Square. $|x|^2$ is xx or $(-x)(-x)$, and both are x^2 by Proposition 2.3.2 (c).
Reciprocal. Let $x \neq 0$. By the product rule, $|x|\,|1/x| = |x \cdot x^{-1}| = |1| = 1$, and $|x| \neq 0$ by (a). So $|1/x|$ satisfies the defining equation of $|x|^{-1}$, and Lemma 4.1.1 (a) gives $|1/x| = 1/|x|$.

(c) If $x \geq 0$, then $-|x| = -x \leq 0 \leq x = |x|$, using Proposition 2.3.5 (a) for $-x \leq 0$. If $x < 0$, then $-|x| = -(-x) = x$, and $x < 0 < -x = |x|$. In both cases $-|x| \leq x \leq |x|$, by transitivity (Proposition 2.3.5 (i)).

(d) ($\Rightarrow$) Let $|x| \leq r$. By (c), $x \leq |x| \leq r$. By (c) and (b) applied to $-x$, $-x \leq |-x| = |x| \leq r$, and negating reverses the inequality (Proposition 2.3.5 (a), or equality), so $-r \leq x$. ($\Leftarrow$) Let $-r \leq x \leq r$. If $x \geq 0$, then $|x| = x \leq r$. If $x < 0$, then $|x| = -x \leq r$, by negating $-r \leq x$. The strict version is the same argument with $<$ in place of $\leq$ at the steps that involve r: for instance $x \leq |x| < r$ gives $x < r$ by Proposition 2.3.5 (i).

(e) By (c), $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. Adding (Proposition 2.3.5 (b)),

$$-(|x| + |y|) \leq x + y \leq |x| + |y|,$$

where $(-|x|) + (-|y|) = -(|x| + |y|)$ by uniqueness of negatives (Proposition 2.3.2 (a)). By (d) with $r := |x| + |y|$, $|x + y| \leq |x| + |y|$. For the second form, apply the first to $x - z = (x - y) + (y - z)$.

(f) By (e), $|x| = |(x - y) + y| \leq |x - y| + |y|$, so $|x| - |y| \leq |x - y|$ by (O3). Exchanging x and y, and using $|y - x| = |-(x - y)| = |x - y|$ from (b), gives $|y| - |x| \leq |x - y|$. So $-|x-y| \leq -(|y|-|x|)$ by Proposition 2.3.5 (a) (or equality), and $-(|y|-|x|) = |x|-|y|$ by uniqueness of negatives (Proposition 2.3.2 (a)), since $(|y| - |x|) + (|x| - |y|) = 0$; that is, $-|x - y| \leq |x| - |y|$. By (d), $||x| - |y|| \leq |x - y|$. This proves the proposition.

■

The form of (d) that the rest of the chapter uses is about a distance. Apply the strict half of (d) to $x - a$ with $r = \varepsilon$, and add a throughout by (O3); the step is reversible, since adding $-a$ (by (O3) again) goes back, and $(x - a) + a = x$:

$$|x - a| < \varepsilon \iff a - \varepsilon < x < a + \varepsilon.$$

In words: x lies within ε of a exactly when it lies strictly between $a - \varepsilon$ and $a + \varepsilon$. In Section 4.3 the role of x is played by the terms a_n of a sequence and a is a candidate limit.

Example. In $\mathbb{Q}$, which x satisfy $|x - 1/2| < 1/4$? By the display above, exactly those with $1/2 - 1/4 < x < 1/2 + 1/4$, that is, $1/4 < x < 3/4$. So $2/3$ qualifies ($1 \cdot 3 = 3 < 8 = 2 \cdot 4$ and $2 \cdot 4 = 8 < 9 = 3 \cdot 3$, by the comparison formula of Theorem 2.5.1 (a)) and $3/4$ does not: the inequality is strict, and $|3/4 - 1/2| = 1/4$.

Quick check. Is $|a| - |b| \leq |a - b|$ for all $a, b \in F$? Is $|a - b| \leq |a| - |b|$ for all $a, b \in F$?

Solution. The first: yes. By (f) and (c), $|a| - |b| \leq ||a| - |b|| \leq |a - b|$. The second: no. In

$\mathbb{Q}$, take $a = 0$ and $b = 1$: then $|a - b| = |-1| = 1$, while $|a| - |b| = -1$, and $1 \leq -1$ is false.

The triangle inequality is the tool of the whole chapter. In its second form it says that a detour through a third point y cannot shorten the way from x to z. Run backwards, it turns two small distances, from x to y and from y to z, into one small distance from x to z. That is the move of almost every proof in Sections 4.3–4.5.

Remark. The same two-case formula appears in *Logic and Proof*, Chapter 1, for the school numbers, and in *Analysis on the Real Line*, Chapter 1, for $\mathbb{R}$. Here it is defined, and its rules are proved, for an arbitrary ordered field, from (F1)–(F9) and (O1)–(O4) alone.

4.1.5 Exercises

4.1.5.1 A. Check your understanding

Exercise. A1

Let F be an ordered field.

(a) Here is a proof that $0 < 1/2 < 1$ in F: (1) $1 > 0$; (2) $2 = 1 + 1 > 1 > 0$; (3) $0 < 1/2 < 1/1 = 1$. Name the rule behind each of the three steps.

(b) True or false in every ordered field: "if $a < b$, then $a^2 < b^2$". Give a reason, or a counterexample in $\mathbb{Q}$.

(c) True or false in every ordered field: "if $|a| < |b|$, then $a < b$". Give a reason, or a counterexample in $\mathbb{Q}$.

(d) True or false in every ordered field: "$x^2 + 1 > 0$ for every x". Give a reason, or a counterexample in $\mathbb{Q}$.

Solution.

(a) (a) Proposition 2.3.5 (e), since $1 \neq 0$ in a field. (2) $1 + 1 > 0 + 1 = 1$ by (O3) (or Proposition 2.3.5 (b), strict), then $2 > 0$ by (O2). (3) Proposition 2.5.2 (b), applied to $0 < 1 < 2$; and $1/1 = 1 \cdot 1^{-1} = 1$, since 1 is its own inverse ($1 \cdot 1 = 1$, and Lemma 4.1.1 (a)).

(b) False. In $\mathbb{Q}$, $-2 < 1$, but $(-2)^2 = 4 > 1 = 1^2$. The true statement needs $a, b \geq 0$: Proposition 2.3.5 (g).

(c) False. In $\mathbb{Q}$, $a = 1$ and $b = -2$ give $|a| = 1 < 2 = |b|$, but $a = 1 > -2 = b$.

(d) True. By Proposition 2.3.5 (e), $x^2 > 0$ if $x \neq 0$, and $x^2 = 0$ if $x = 0$ (Proposition 2.3.2 (b)); so $x^2 \geq 0$. Also $1 > 0$. Hence $x^2 + 1 > 0 + 0 = 0$ by Proposition 2.3.5 (b), strict in the second summand.

4.1.5.2 B. Practice

Exercise. B1: Two companions of (c) and (d)

Let F be an ordered field.

(a) Prove that if $0 \le a < b$ and $0 < c \le d$, then $ac < bd$.

(b) Prove that for $a, b \in F_{>0}$, $a \le b$ **if and only if** $1/b \le 1/a$.

Name the rule behind each step.

Solution.

(a) Since $0 \le a < b$, $b > 0$ by Proposition 2.3.5 (i). First, $ac < bc$: $a < b$ and $c > 0$, so Proposition 2.3.5 (c) applies. Next, $bc \le bd$: if $c = d$ the two sides are equal; otherwise $c < d$ and $b > 0$, so $cb < db$ by Proposition 2.3.5 (c), that is, $bc < bd$ by (F6). Combining $ac < bc \le bd$ by Proposition 2.3.5 (i) gives $ac < bd$.

(b) ($\Rightarrow$) is Proposition 4.1.2 (c). ($\Leftarrow$) Let $1/b \le 1/a$. Both are positive, by Proposition 2.5.2 (a). Applying Proposition 4.1.2 (c) to the positive elements $1/b \le 1/a$ gives $1/(1/a) \le 1/(1/b)$, and $1/(1/a) = (a^{-1})^{-1} = a$, $1/(1/b) = b$ by Lemma 4.1.1 (a). So $a \le b$.

Exercise. B2: Solving an inequality

In an ordered field F, find all x with $|2x - 3| < 5$, naming the rule behind each step. Here 3, 5, 8 and so on denote the multiples $3 \cdot 1$, $5 \cdot 1$, $8 \cdot 1$, ... of Section 2.3 (defined before Proposition 2.3.5), and $2 = 1 + 1 = 2 \cdot 1$.

Solution. We use the sum law of Proposition 2.3.5 (h), $(m+k)\cdot 1 = m\cdot 1 + k\cdot 1$; in particular $5 + 3 = 8$, $2 + 3 = 5$ and $4 + 4 = 8$ hold in F because they hold in $\mathbb{N}$.

By Proposition 4.1.3 (d), strict form, $|2x - 3| < 5$ if and only if

$$-5 < 2x - 3 < 5.$$

Adding 3 throughout is reversible (add -3 to go back), so by (O3) this holds if and only if $-5 + 3 < 2x < 5 + 3$. Here $5 + 3 = 8$, and $-5 + 3 = -(2 + 3) + 3 = -2$, using $-(2+3) = (-2) + (-3)$ (uniqueness of negatives, Proposition 2.3.2 (a)). So the condition is $-2 < 2x < 8$.

Multiplying by $1/2 > 0$ (Proposition 4.1.2 (b)) is reversible too (multiply by $2 > 0$ to go back), so by Proposition 2.3.5 (c) the condition is equivalent to $(-2)/2 < x < 8/2$. Now $2/2 = 1$, so $(-2)/2 = -(2/2) = -1$ by Proposition 2.3.2 (c). And $8 = 4 + 4 = 4(1 + 1)$ by (F7) and (F9), the field product of the elements 4 and 2, so $8/2 = 4(2/2) = 4$ by (F5). Hence

$$|2x - 3| < 5 \iff -1 < x < 4.$$

Check at an endpoint: $x = 4$ gives $|8 - 3| = 5$, not < 5, as the strict inequality predicts.

Exercise. B3: Two more triangle inequalities

Prove that for all a, b, c in an ordered field,

(a) $|a + b + c| \le |a| + |b| + |c|$;

(b) $||a| - |b|| \le |a + b|$.

Solution.

(a) By Proposition 4.1.3 (e) twice, $|(a+b)+c| \le |a+b|+|c| \le (|a|+|b|)+|c|$, the second step by adding $|c|$ to $|a+b| \le |a|+|b|$ (Proposition 2.3.5 (b)). Conclude by transitivity (Proposition 2.3.5 (i)).

(b) Apply Proposition 4.1.3 (f) to a and $-b$: $\big||a|-|-b|\big| \le |a-(-b)|$. By Proposition 4.1.3 (b), $|-b| = |b|$, and $a-(-b) = a+b$ by Proposition 2.3.2 (a). This is the claim.

4.1.5.3 C. Going deeper

Exercise. C1: A field that cannot be ordered

(a) Prove that in an ordered field no x satisfies $x^2 = -1$.

(b) Deduce that if a field contains an element i with $i^2 = -1$, then no relation $<$ makes it an ordered field.

Solution (Hint 1). Which sign does a square have, and which sign does -1 have?

Solution (Hint 2). Show first that $x^2 \ge 0$ for every x, splitting on whether $x = 0$.

Solution.

(a) Let F be an ordered field and $x \in F$. If $x = 0$, then $x^2 = 0$ (Proposition 2.3.2 (b)); if $x \neq 0$, then $x^2 > 0$ (Proposition 2.3.5 (e)). So $x^2 \ge 0$. On the other hand $1 > 0$ (Proposition 2.3.5 (e), as $1 \neq 0$ in a field), so $-1 < 0$ by Proposition 2.3.5 (a). If $x^2 = -1$, then $0 \le x^2 = -1 < 0$, so $0 < 0$ by Proposition 2.3.5 (i), against (O1).

(b) If some relation $<$ made the field an ordered field, then (a), applied to that ordered field, would forbid $i^2 = -1$. So no such relation exists.

The complex numbers of *Algebraic Preliminaries*, Chapter 1 contain such an i, so they form a field that cannot be ordered. Compare Example (A field with two elements that cannot be ordered), where the obstacle was $1+1=0$; here it is that -1 is a square.

Marking scheme. (10 points) - 4: shows $x^2 \ge 0$ for every x (cases $x = 0$, $x \neq 0$), or rules out $x = 0$ for the x with $x^2 = -1$ and then shows $x^2 > 0$, giving the rule for each step. - 2: shows $-1 < 0$ from $1 > 0$ (justified by $1 \neq 0$ and Proposition 2.3.5 (e), or by citing the standing remark of Section 4.1) and from negation reversing order. - 2: derives the contradiction with trichotomy (O1) or irreflexivity. - 2: (b) states that any order would make (a) apply, so none exists. Fatal errors (0 for the step): asserting $x^2 > 0$ for all x (false at 0) and then using strictness; asserting $1 > 0$ with no justification (accept $1 \neq 0$ with Proposition 2.3.5 (e), or the standing remark of Section 4.1); using properties of $\mathbb{R}$ or of $\sqrt{\ }$.

Exercise. C2: Maximum and minimum through the absolute value

For a, b in an ordered field F, let $\max\{a,b\}$ be b if $a \le b$ and a otherwise, and $\min\{a,b\}$ be

a if $a \le b$ and b otherwise.

(a) Prove that

$$\max\{a,b\} = \frac{a+b+|a-b|}{2}, \qquad \min\{a,b\} = \frac{a+b-|a-b|}{2}.$$

(b) Deduce that $|x| = \max\{x,-x\}$ for every $x \in F$.

Solution.

(a) *Case* $a \le b$. Then $a-b \le 0$ (Proposition 2.3.5 (a), or equality), so $|a-b| = b-a$: by the second clause of Definition 4.1.1 if $a-b<0$, and because both sides are 0 if $a=b$. Hence

$$\frac{a+b+|a-b|}{2} = \frac{a+b+b-a}{2} = \frac{2b}{2} = b = \max\{a,b\},$$

and likewise $\dfrac{a+b-(b-a)}{2} = \dfrac{2a}{2} = a = \min\{a,b\}$, using $b+b=2b$ and $(2b)2^{-1}=b$ as in the proof of Proposition 4.1.2 (b).
Case $b<a$. Then $a-b>0$, so $|a-b| = a-b$, and the two formulas give $(2a)/2 = a = \max\{a,b\}$ and $(2b)/2 = b = \min\{a,b\}$.

(b) By (a) with $a=x$ and $b=-x$: $\max\{x,-x\} = (x+(-x)+|2x|)/2 = |2x|/2$. By Proposition 4.1.3 (b), $|2x| = |2||x| = 2|x|$, since $2>0$. So $\max\{x,-x\} = |x|$.

4.2 The Rationals inside an Ordered Field; the Archimedean Property

In an ordered field F we can form 1, $1+1$, $1+1+1$, and so on. They look like the natural numbers $1,2,3,\ldots$, but they are elements of F, and two questions need answers before the next section can use them as thresholds. Do they behave like natural numbers, adding, multiplying and comparing as $\mathbb{N}$ does? And does every element of F lie below one of them? For $\mathbb{Q}$, Section 2.5 answered the second question with yes (Theorem 2.5.3). This section answers the first for every ordered field, and goes further: a whole copy of $\mathbb{Q}$ sits inside F, in exactly one way. The second question has the answer no in general. Fields where the answer is yes are called **Archimedean**, and in them every element has an integer part and the rationals are dense.

4.2.1 The natural numbers in an ordered field

Section 2.3 defined, in any commutative ring R and for any $a \in R$, the multiples $n \cdot a$ for $n \in \mathbb{N}$ (just before Proposition 2.3.5): $0 \cdot a = 0$ and $(n+1)\cdot a = n \cdot a + a$. With $a = 1$ they are the elements $0, 1, 1+1, \ldots$ of the opening paragraph. We give them a name, together with powers, which this chapter needs for estimates such as $2^k \ge k+1$.

Definition 4.2.1. Natural Numbers and Powers in an Ordered Field

Let F be an ordered field, with unit element 1_F.

(a) For $n \in \mathbb{N}$, $n \cdot 1_F \in F$ is the multiple of 1_F of Section 2.3: the map $\mathbb{N} \to F$, $n \mapsto n \cdot 1_F$,

is the unique function with

$$0 \cdot 1_F = 0, \qquad (n+1) \cdot 1_F = n \cdot 1_F + 1_F \quad \text{for every } n \in \mathbb{N}.$$

(b) For $x \in F$, the **powers** x^k ($k \in \mathbb{N}$) are given by Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b), with $X = F$, starting value 1_F and step $(k, y) \mapsto yx$:

$$x^0 := 1_F, \qquad x^{k+1} := x^k x \quad \text{for every } k \in \mathbb{N}.$$

The subscript in 1_F is there because two fields will often be in play, F and $\mathbb{Q}$, each with its own 1; where only one field is in play the subscript is dropped. For natural numbers n, k, the power $n^k \in \mathbb{N}$ is the one of Section 1.6, defined (just before Proposition 1.6.4) by the same recursion with $X = \mathbb{N}$: $n^0 = 1$ and $n^{k+1} = n^k n$. So 2^k may mean a natural number or an element of F; part (c) of the next proposition says the two readings match.

Examples. $1 \cdot 1_F = 0 \cdot 1_F + 1_F = 0 + 1_F = 1_F$, and $2 \cdot 1_F = 1_F + 1_F$, the element written 2 in Section 4.1. Also $x^1 = 1_F x = x$ and $x^2 = xx$, which is the square of Proposition 2.3.2 (d). The degenerate case: $0 \cdot 1_F = 0$, and $0^0 = 1_F$ by the first equation, a convention the recursion imposes and nothing below questions.

A non-example by minimal change. Keep the field axioms and drop the order. In the two-element field of Example (A field with two elements that cannot be ordered), the same recursion gives $2 \cdot 1 = 1 + 1 = 0 = 0 \cdot 1$. So the map $n \mapsto n \cdot 1$ is not injective there, and the elements $0 \cdot 1, 1 \cdot 1, 2 \cdot 1, \ldots$ are not a copy of $\mathbb{N}$. The next proposition shows that the order axioms prevent this.

Proposition 4.2.1. The Natural Numbers inside an Ordered Field

Let F be an ordered field, and let $m, n, k \in \mathbb{N}$.

(a) $(m+n) \cdot 1_F = m \cdot 1_F + n \cdot 1_F$ and $(mn) \cdot 1_F = (m \cdot 1_F)(n \cdot 1_F)$.

(b) If $m < n$, then $m \cdot 1_F < n \cdot 1_F$. Hence $n \mapsto n \cdot 1_F$ is injective, and $n \cdot 1_F \geq 1_F > 0$ whenever $n \geq 1$.

(c) $(n^k) \cdot 1_F = (n \cdot 1_F)^k$. In particular $(2^k) \cdot 1_F = (2 \cdot 1_F)^k$, the k-th power in F of $2 = 1+1$.

(d) In $\mathbb{Q}$, $n \cdot 1_{\mathbb{Q}} = j(\iota(n))$, the image of n under the embeddings Proposition 2.2.3 and Proposition 2.4.4. With the identifications $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$, this says $n \cdot 1_{\mathbb{Q}} = n$.

Idea. The sum law in (a) and part (b) are rules for multiples (Proposition 2.3.5 (h)); the product law and (c) are inductions following the recursion defining the left side. The product law is the only step with content: multiplying by $n+1$ is multiplying by n and adding one more copy, and the sum law and distributivity turn that into the same step on the right. Part (d) compares two functions on $\mathbb{N}$ that satisfy the same recursion, and the uniqueness half of the recursion theorem makes them equal.

Proof.

(a) The sum law is Proposition 2.3.5 (h) with $R = F$ and $a = 1_F$. For the product law, fix m and use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on n, for $P(n)$: "$(mn) \cdot 1_F = (m \cdot 1_F)(n \cdot 1_F)$". *Base:* $m \cdot 0 = 0$ by Definition 1.5.1 (b), so the left side is

$0 \cdot 1_F = 0$, and the right side is $(m \cdot 1_F) \cdot 0 = 0$ by Proposition 2.3.2 (b). *Step:* assume $P(n)$. By Definition 1.5.1 (b) and (c), $m(n+1) = mn + m$. Hence

$$\begin{aligned}(m(n+1)) \cdot 1_F &= (mn) \cdot 1_F + m \cdot 1_F = (m \cdot 1_F)(n \cdot 1_F) + (m \cdot 1_F)1_F \\ &= (m \cdot 1_F)(n \cdot 1_F + 1_F) = (m \cdot 1_F)((n+1) \cdot 1_F),\end{aligned}$$

by the sum law, $P(n)$ with (F7), (F9), and the recursion of (a) in Definition 4.2.1. So $P(n+1)$ holds, and $P(n)$ holds for every n.

(b) Since $1_F \neq 0$ in a field, Proposition 2.3.5 (h) gives $m \cdot 1_F < n \cdot 1_F$ whenever $m < n$. If $m \neq n$, then $m < n$ or $n < m$ by trichotomy in $\mathbb{N}$ (Theorem 1.6.1 (a)), so $m \cdot 1_F \neq n \cdot 1_F$ by (O1): the map is injective. Finally $1 \cdot 1_F = 1_F$ (the examples above), and for $n > 1$, $1_F = 1 \cdot 1_F < n \cdot 1_F$; so $n \cdot 1_F \geq 1_F$ for every $n \geq 1$, and $1_F > 0$ by Proposition 2.3.5 (e).

(c) Fix n and induct on k. *Base:* $n^0 = 1$, so $(n^0) \cdot 1_F = 1 \cdot 1_F = 1_F = (n \cdot 1_F)^0$. *Step:* if $(n^k) \cdot 1_F = (n \cdot 1_F)^k$, then, since $n^{k+1} = n^k n$,

$$(n^{k+1}) \cdot 1_F = ((n^k) \cdot 1_F)(n \cdot 1_F) = (n \cdot 1_F)^k (n \cdot 1_F) = (n \cdot 1_F)^{k+1},$$

by the product law of (a), the hypothesis, and Definition 4.2.1 (b). The special case is $n = 2$, with $2 \cdot 1_F = 1_F + 1_F$.

(d) Let $f(n) \coloneqq j(\iota(n)) \in \mathbb{Q}$. The zero and one of $\mathbb{Z}$ are $\iota(0)$ and $\iota(1)$ (Theorem 2.3.1), and j sends them to the zero and one of $\mathbb{Q}$ (Proposition 2.4.4). So $f(0) = 0$, and, since ι and j preserve sums (Proposition 2.2.3 (a), Proposition 2.4.4),

$$f(n+1) = j(\iota(n) + \iota(1)) = j(\iota(n)) + j(\iota(1)) = f(n) + 1_{\mathbb{Q}}.$$

So f and $n \mapsto n \cdot 1_{\mathbb{Q}}$ both satisfy the recursion of Definition 4.2.1 (a) with $F = \mathbb{Q}$, and by the uniqueness in Sets, Relations, and Cardinality, Theorem 2.3.1 (through Corollary 1.7.3 (b)) they are equal. This proves the proposition.

■

So the order axioms do what the non-example above lacked: by (b), the elements $n \cdot 1_F$ are all different and increase with n, and by (a) they add and multiply as natural numbers do. In $\mathbb{Q}$, (d) says that nothing new has happened: $n \cdot 1_{\mathbb{Q}}$ is n. When only one field is in play, we will sometimes write n for $n \cdot 1_F$ in a general ordered field too; the statements of this section keep $n \cdot 1_F$, because there the difference between $n \in \mathbb{N}$ and its image in F is the point.

4.2.2 A copy of the rationals in every ordered field

The copy of $\mathbb{N}$ extends. Negatives exist in F, so the integers have images too, and inverses of non-zero elements exist, so fractions have images. The next theorem says that the result is a copy of $\mathbb{Q}$, compatible with $+$, $\cdot$ and $<$, and that there is no other way to place $\mathbb{Q}$ inside F. Later chapters of this volume use this map to compare an arbitrary ordered field with the rationals.

Every ordered field contains a copy of $\mathbb{Q}$, and only one.

Theorem 4.2.2. The Rationals Embed in Every Ordered Field

Let F be an ordered field. There is **exactly one** map $\iota_F\colon \mathbb{Q} \to F$ such that, for all $x, y \in \mathbb{Q}$,

$$\iota_F(1) = 1_F,$$
$$\iota_F(x+y) = \iota_F(x) + \iota_F(y), \qquad \iota_F(xy) = \iota_F(x)\,\iota_F(y).$$

It satisfies $\iota_F(n) = n \cdot 1_F$ for every $n \in \mathbb{N}$. It is injective and **strictly increasing** ($x < y$ implies $\iota_F(x) < \iota_F(y)$), and it **reflects order**: $\iota_F(x) < \iota_F(y)$ implies $x < y$.

Idea. Uniqueness tells us how to build ι_F. A map respecting 1 and + must send $n = 1+\cdots+1$ to $n \cdot 1_F$; respecting negatives, it must send $-n$ to $-(n \cdot 1_F)$; and respecting products, it must send p/q to $(p \cdot 1_F)(q \cdot 1_F)^{-1}$. So we take that formula as the definition, and the work is to check it. **Step roadmap.** ① Extend $n \mapsto n \cdot 1_F$ to $\mathbb{Z}$ and check that it respects $+$, $\cdot$ and negation. ② Define ι_F on fractions and check that the value does not depend on the fraction chosen. ③ Check $+$ and $\cdot$ with the fraction rules. ④ Check the order. ⑤ Uniqueness: run the opening argument as a proof.

Proof. **Step 1: the integers.** By Proposition 2.2.3 (c), read through the identification $\mathbb{N} \subseteq \mathbb{Z}$, every $m \in \mathbb{Z}$ is n or $-n$ for some $n \in \mathbb{N}$. Define

$$m \cdot 1_F := \begin{cases} n \cdot 1_F & \text{if } m = n \text{ with } n \in \mathbb{N}, \\ -(n \cdot 1_F) & \text{if } m = -n \text{ with } n \in \mathbb{N}. \end{cases}$$

This is well defined. Within one case, n is determined by m: directly in the first case, and in the second because $-n = -n'$ gives $n = n'$ by Proposition 2.3.2 (a). The cases overlap only at 0: if $n = -n'$ with $n, n' \in \mathbb{N}$, then $n + n' = 0$, and $n, n' \geq 0$ (Theorem 2.3.4 (b)); if $n' > 0$, then $n + n' > n + 0 \geq 0$ by Proposition 2.3.5 (b), against $n + n' = 0$, so $n' = 0$ and then $n = 0$. At $m = 0$ both cases give 0, since $-0 = 0$.

We check three rules, for $m, m' \in \mathbb{Z}$.

Negation: $(-m) \cdot 1_F = -(m \cdot 1_F)$. If $m = n \in \mathbb{N}$, this is the second case of the definition. If $m = -n$, then $-m = n$ by Proposition 2.3.2 (a), and $(-m) \cdot 1_F = n \cdot 1_F = -(-(n \cdot 1_F)) = -(m \cdot 1_F)$.

Sums: $(m + m') \cdot 1_F = m \cdot 1_F + m' \cdot 1_F$. If $m, m' \in \mathbb{N}$, this is Proposition 4.2.1 (a). We write out the mixed case $m = n$, $m' = -n'$ with $n, n' \in \mathbb{N}$. If $n' \leq n$, then $n = n' + k$ for some $k \in \mathbb{N}$ (Proposition 1.6.3 (a)), so $m + m' = (n' + k) - n' = k$ in $\mathbb{Z}$, and

$$\begin{aligned} m \cdot 1_F + m' \cdot 1_F &= (n' + k) \cdot 1_F - n' \cdot 1_F \\ &= n' \cdot 1_F + k \cdot 1_F - n' \cdot 1_F = k \cdot 1_F, \end{aligned}$$

by Proposition 4.2.1 (a). If $n < n'$, then $n' = n + k$, $m + m' = -k$, and the same computation gives $n \cdot 1_F - (n+k) \cdot 1_F = -(k \cdot 1_F) = (-k) \cdot 1_F$. The case $m = -n$, $m' = n'$ follows by (F1). If $m = -n$ and $m' = -n'$, then $m + m' = -(n + n')$, and $-((n+n') \cdot 1_F) = -(n \cdot 1_F) - (n' \cdot 1_F)$, by the sum law and uniqueness of negatives.

Products: $(mm') \cdot 1_F = (m \cdot 1_F)(m' \cdot 1_F)$. If $m, m' \in \mathbb{N}$, this is Proposition 4.2.1 (a). We

write out $m = -n$, $m' = n'$: then $mm' = -(nn')$ by Proposition 2.3.2 (c), so

$$\begin{aligned}(mm') \cdot 1_F &= -((nn') \cdot 1_F) = -((n \cdot 1_F)(n' \cdot 1_F)) \\ &= (-(n \cdot 1_F))(n' \cdot 1_F) = (m \cdot 1_F)(m' \cdot 1_F),\end{aligned}$$

by negation, Proposition 4.2.1 (a), and Proposition 2.3.2 (c) in F. The two remaining sign patterns are the same with signs: $(-n)(-n') = nn'$ in $\mathbb{Z}$ and $(-(n \cdot 1_F))(-(n' \cdot 1_F)) = (n \cdot 1_F)(n' \cdot 1_F)$ in F, both by Proposition 2.3.2 (c), and $m = n$, $m' = -n'$ follows from the case written out by (F6).

Step 2: the definition on $\mathbb{Q}$. By Proposition 2.4.4 (b), every $x \in \mathbb{Q}$ is p/q for some $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$. Since $q \in \mathbb{Z}_{>0}$, that is, $q \in \mathbb{N}$ and $q \neq 0$, we have $q \geq 1$ (Theorem 1.6.1 (d)), and Proposition 4.2.1 (b) gives $q \cdot 1_F \geq 1_F > 0$, so $q \cdot 1_F \neq 0$. Define

$$\iota_F\Big(\frac{p}{q}\Big) \coloneqq (p \cdot 1_F)(q \cdot 1_F)^{-1} = \frac{p \cdot 1_F}{q \cdot 1_F}.$$

The value does not depend on the fraction chosen: if $p/q = r/s$ with $r \in \mathbb{Z}$ and $s \in \mathbb{Z}_{>0}$, then $ps = rq$ (Proposition 2.4.4 (b)), so $(p \cdot 1_F)(s \cdot 1_F) = (r \cdot 1_F)(q \cdot 1_F)$ by the product rule of Step 1. Multiplying both sides by $(q \cdot 1_F)^{-1}(s \cdot 1_F)^{-1}$ and using (F5), (F6), (F8) gives $(p \cdot 1_F)(q \cdot 1_F)^{-1} = (r \cdot 1_F)(s \cdot 1_F)^{-1}$. So ι_F is a function $\mathbb{Q} \to F$. For $m \in \mathbb{Z}$, $m = m/1$, so $\iota_F(m) = (m \cdot 1_F)(1_F)^{-1} = m \cdot 1_F$; in particular $\iota_F(n) = n \cdot 1_F$ for $n \in \mathbb{N}$, and $\iota_F(1) = 1_F$.

Step 3: sums and products. Let $x = p/q$ and $y = r/s$ as in Step 2. In the field $\mathbb{Q}$, Lemma 4.1.1 (b) gives

$$x + y = \frac{ps + rq}{qs}, \qquad xy = \frac{pr}{qs},$$

with $qs \in \mathbb{Z}_{>0}$ by (O4) in $\mathbb{Z}$. By the definition, Step 1, and Lemma 4.1.1 (b) in F,

$$\begin{aligned}\iota_F(x + y) &= \frac{(p \cdot 1_F)(s \cdot 1_F) + (r \cdot 1_F)(q \cdot 1_F)}{(q \cdot 1_F)(s \cdot 1_F)} \\ &= \frac{p \cdot 1_F}{q \cdot 1_F} + \frac{r \cdot 1_F}{s \cdot 1_F} = \iota_F(x) + \iota_F(y),\end{aligned}$$

and in the same way $\iota_F(xy) = \dfrac{(p \cdot 1_F)(r \cdot 1_F)}{(q \cdot 1_F)(s \cdot 1_F)} = \iota_F(x)\,\iota_F(y)$.

Step 4: order. Let $x < y$ in $\mathbb{Q}$. Then $y - x > 0$, and writing $y - x = p/q$ with $q \in \mathbb{Z}_{>0}$, the comparison formula (Theorem 2.5.1 (a)) gives $p > 0$, so $p \in \mathbb{Z}_{>0}$ (Theorem 2.3.4 (b)). By Proposition 4.2.1 (b), $p \cdot 1_F > 0$ and $q \cdot 1_F > 0$; by Proposition 2.5.2 (a) and (O4), $\iota_F(y - x) > 0$. By Step 3, $\iota_F(y) = \iota_F(x) + \iota_F(y - x) > \iota_F(x)$, using (O3). So ι_F is strictly increasing. If $x \neq y$, then $x < y$ or $y < x$, and the images differ by (O1): ι_F is injective. If $\iota_F(x) < \iota_F(y)$, then neither $x = y$ nor $y < x$ can hold, since they would give $\iota_F(x) = \iota_F(y)$ or $\iota_F(y) < \iota_F(x)$, against (O1); so $x < y$ by trichotomy.

Step 5: uniqueness. Let $\psi\colon \mathbb{Q} \to F$ satisfy the three equations of the statement. Then $\psi(0) = \psi(0 + 0) = \psi(0) + \psi(0)$, and adding $-\psi(0)$ gives $\psi(0) = 0$. By Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)), $\psi(n) = n \cdot 1_F$ for every $n \in \mathbb{N}$: the base is $\psi(0) = 0 = 0 \cdot 1_F$, and $\psi(n + 1) = \psi(n) + \psi(1) = n \cdot 1_F + 1_F = (n + 1) \cdot 1_F$. For $x \in \mathbb{Q}$, $\psi(x) + \psi(-x) = \psi(0) = 0$, so $\psi(-x) = -\psi(x)$ by Proposition 2.3.2 (a); hence $\psi(m) = m \cdot 1_F$ for every $m \in \mathbb{Z}$. For $q \in \mathbb{Z}_{>0}$, $\psi(q)\,\psi(q^{-1}) = \psi(1) = 1_F$, so $\psi(q^{-1}) = (q \cdot 1_F)^{-1}$ by

Lemma 4.1.1 (a). Finally $p/q = pq^{-1}$ (Proposition 2.4.4 (a)), so

$$\psi\Big(\frac{p}{q}\Big) = \psi(p)\,\psi(q^{-1}) = (p \cdot 1_F)(q \cdot 1_F)^{-1} = \iota_F\Big(\frac{p}{q}\Big).$$

So $\psi = \iota_F$. This proves the theorem.

■

For $F = \mathbb{Q}$, the identity map of $\mathbb{Q}$ satisfies the three equations, so by uniqueness $\iota_{\mathbb{Q}}$ is the identity. In general ι_F is the map of Step 2, and it carries $\mathbb{Q}$ onto a copy inside F that adds, multiplies and compares as $\mathbb{Q}$ does. So "$\mathbb{Q} \subseteq F$" is an honest abuse of notation, as $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$ was in Chapter 2, and where only one field is in play we write q for $\iota_F(q)$. The statements below keep ι_F visible.

Remark. Uniqueness has a consequence that later chapters of this volume use when comparing two ordered fields. Let L be another ordered field and $\chi\colon F \to L$ a map with $\chi(1_F) = 1_L$ that respects $+$ and $\cdot$. Then $\chi \circ \iota_F\colon \mathbb{Q} \to L$ satisfies the three equations of the theorem, so $\chi \circ \iota_F = \iota_L$: such a map fixes the rationals.

4.2.3 The Archimedean property

Section 2.5 proved that every rational lies below a natural number (Theorem 2.5.3). By Proposition 4.2.1 (d), that theorem says: for every $x \in \mathbb{Q}$ there is $n \in \mathbb{N}$ with $x < n \cdot 1_{\mathbb{Q}}$. This statement makes sense in any ordered field, and it deserves a name, because it can fail.

No element of the field lies above all the natural numbers.

Definition 4.2.2. Archimedean Ordered Field

An ordered field F is **Archimedean** if **for every** $x \in F$ **there is** $n \in \mathbb{N}$ with $x < n \cdot 1_F$. In symbols,

$$\forall x \in F \quad \exists n \in \mathbb{N} \quad x < n \cdot 1_F.$$

In words: whatever element of F we pick, some natural number, read in F, lies strictly above it. The n may depend on x; it is chosen after x.

Examples. $\mathbb{Q}$ is Archimedean: this is Theorem 2.5.3, by Proposition 4.2.1 (d). The constructed real numbers of the next chapter are Archimedean too (proved there). A non-example, $\mathbb{Q}(t)$, comes at the end of this section: there the element t lies above every $n \cdot 1$.

Why the name. Archimedes used, as an axiom about lengths, that a smaller length laid off often enough exceeds any larger one. That is Exercise B1: in an Archimedean field, for $y > 0$ and any x, some multiple $(n \cdot 1_F)y$ exceeds x. The definition is the case $y = 1_F$.

The form in which the next section uses the property is about reciprocals, as in Corollary 2.5.4, now in an arbitrary Archimedean ordered field.

Proposition 4.2.3. Small Reciprocals in an Archimedean Field

In an Archimedean ordered field F, **for every** $\varepsilon \in F_{>0}$ **there is** $N \in \mathbb{Z}_{>0}$ such that

$1/(n \cdot 1_F) < \varepsilon$ **for every** $n \in \mathbb{N}$ with $n \geq N$. In symbols,

$$\forall \varepsilon \in F_{>0} \;\; \exists N \in \mathbb{Z}_{>0} \;\; \forall n \in \mathbb{N} \;\; \Big(n \geq N \Rightarrow \frac{1}{n \cdot 1_F} < \varepsilon \Big).$$

Proof. Let $\varepsilon \in F_{>0}$. Then $1/\varepsilon > 0$ by Proposition 2.5.2 (a). By Definition 4.2.2 there is $N \in \mathbb{N}$ with $1/\varepsilon < N \cdot 1_F$; it depends on ε. Then $N \neq 0$, since $0 \cdot 1_F = 0 < 1/\varepsilon$, so $N \in \mathbb{Z}_{>0}$. Let $n \geq N$. Then $n \cdot 1_F \geq N \cdot 1_F$ (Proposition 4.2.1 (b), or equality), so $0 < 1/\varepsilon < n \cdot 1_F$, and Proposition 2.5.2 (b) gives

$$\frac{1}{n \cdot 1_F} < \frac{1}{1/\varepsilon} = \varepsilon,$$

where $1/(1/\varepsilon) = (\varepsilon^{-1})^{-1} = \varepsilon$ by Lemma 4.1.1 (a).

■

The order of the quantifiers is as in Section 2.5: ε first, then N, depending on ε, and then **every** $n \geq N$ at once. That last clause is what Section 4.3 will call "eventually".

Quick check. Find $n \in \mathbb{N}$ with $n > 355/113$ in $\mathbb{Q}$. Then find the least $n \in \mathbb{Z}_{>0}$ with $1/n < 7/500$.

Solution. By the comparison formula (Theorem 2.5.1 (a)), $355/113 < n/1$ if and only if $355 < 113n$. Since $113 \cdot 3 = 339 < 355$ and $113 \cdot 4 = 452 > 355$, $n = 4$ works (and is the least such n).

For $n \in \mathbb{Z}_{>0}$, $1/n < 7/500$ if and only if $500 < 7n$. Since $7 \cdot 71 = 497 < 500$ and $7 \cdot 72 = 504 > 500$, the least such n is 72; every $n \geq 72$ works too, as Proposition 4.2.3 promises.

4.2.4 Integer part and density

In $\mathbb{Q}$, the Archimedean property gave every rational an integer part (Proposition 2.5.5). About $\mathbb{Q}$ the proof used only Archimedes, applied to x and to $-x$; the rest (least elements and discreteness of $\mathbb{Z}$, totality of the order) holds in any ordered field. So it runs in any Archimedean field, with ι_F carrying the integers into F.

Proposition 4.2.4. Integer Part in an Archimedean Field

Suppose that the ordered field F is Archimedean. Then for every $x \in F$ there is **exactly one** $m \in \mathbb{Z}$ with

$$\iota_F(m) \leq x < \iota_F(m+1).$$

Idea. The integer just above x should be the least integer k with $x < \iota_F(k)$. Archimedes applied to x shows that such k exist; applied to $-x$, it bounds them below, so Proposition 2.3.6 supplies a least one. The comparisons happen in F, and ι_F reflecting order (Theorem 4.2.2) carries them back to $\mathbb{Z}$.

Proof. Let $x \in F$ and $S := \{k \in \mathbb{Z} : x < \iota_F(k)\}$. By Definition 4.2.2 there is $n \in \mathbb{N}$ with $x < n \cdot 1_F = \iota_F(n)$ (Theorem 4.2.2), so $n \in S$ and $S \neq \emptyset$. Applied to $-x$, it gives $n' \in \mathbb{N}$ with $-x < \iota_F(n')$, that is, $\iota_F(-n') = -\iota_F(n') < x$ (Proposition 2.3.5 (a); ι_F respects negatives, as in Step 5 of the proof of Theorem 4.2.2). For $k \in S$, $\iota_F(-n') < x < \iota_F(k)$, so $-n' < k$ because ι_F reflects order. Thus $-n'$ is a lower bound of S, and by Proposition 2.3.6 S has a least element l. Let $m := l - 1$, so $l = m + 1$.

Existence. $x < \iota_F(m+1)$, since $l \in S$. And $m \notin S$, since $m < l$ and l is least. So $x < \iota_F(m)$ fails, and $\iota_F(m) \leq x$ by totality (Proposition 4.1.2 (a)).

Uniqueness. Let $m, m' \in \mathbb{Z}$ both satisfy the condition, and suppose $m < m'$. By discreteness (Theorem 2.3.4 (c)), $m + 1 \leq m'$, so $\iota_F(m+1) \leq \iota_F(m')$ (Theorem 4.2.2). Then

$$x < \iota_F(m+1) \leq \iota_F(m') \leq x,$$

so $x < x$, against (O1). By symmetry $m' < m$ is impossible, so $m = m'$.

■

For $F = \mathbb{Q}$, $\iota_{\mathbb{Q}}$ is the identity and this is Proposition 2.5.5. The constructed real numbers of the next chapter will use it with F the new field, where m is an integer approximant of a number not yet known to be anything else.

Between two rationals there is always a third, their midpoint (Section 2.5). In an Archimedean field more is true: between any two elements, however close, lies a **rational**.

Theorem 4.2.5. The Rationals Are Dense in an Archimedean Field

In an Archimedean ordered field F, for all $x, y \in F$ with $x < y$ there is $q \in \mathbb{Q}$ with $x < \iota_F(q) < y$.

Idea. Look for q with denominator n: we need an integer k with $nx < k < ny$, after scaling by n. Integers are spaced 1 apart, so such a k exists once the scaled gap $n(y - x)$ exceeds 1; that is where Archimedes enters, through Proposition 4.2.3. The integer to take is the one just above nx, which the integer part supplies.

Proof. Since $y - x > 0$, Proposition 4.2.3 gives $n \in \mathbb{Z}_{>0}$ with $1/(n \cdot 1_F) < y - x$. Let $c := n \cdot 1_F$, which is positive (Proposition 4.2.1 (b)). Multiplying by c (Proposition 2.3.5 (c)) gives $1_F < c(y - x) = cy - cx$, by (F9) and Proposition 2.3.2 (c); adding cx by (O3), $cx + 1_F < cy$. Let m be the integer part of cx (Proposition 4.2.4), so that $\iota_F(m) \leq cx < \iota_F(m+1)$. Then

$$cx < \iota_F(m+1) = \iota_F(m) + 1_F \leq cx + 1_F < cy,$$

by Theorem 4.2.2 and Proposition 2.3.5 (b). Multiplying by $c^{-1} > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)) gives $x < \iota_F(m+1)c^{-1} < y$. Let $q := (m+1)/n \in \mathbb{Q}$. Since $c = \iota_F(n)$ and ι_F respects products, $\iota_F(q)\, c = \iota_F(m+1)$, so $\iota_F(q) = \iota_F(m+1)c^{-1}$, and $x < \iota_F(q) < y$.

■

Applied again, the theorem shows that an interval of an Archimedean field contains infinitely many rationals (Exercise C1). Chapter 6 uses the theorem with F a field of real numbers; Chapter 5 uses it only in an exercise.

4.2.5 A field that is not Archimedean

Section 2.5 proved the Archimedean property of $\mathbb{Q}$ from the construction of $\mathbb{Q}$: every rational is a quotient of integers. Nothing in (F1)–(F9) and (O1)–(O4) forces it. The standard witness uses polynomials, which this volume has not built; *Algebraic Preliminaries*, Chapter 4 builds them as formal polynomials. We use the school algebra of polynomials in this one example, and say so. No proof of a numbered result in this chapter cites it; it serves only as a counterexample, in prose, warnings and exercises.

Example. A Non-Archimedean Ordered Field

Let $\mathbb{Q}(t)$ be the set of quotients p/q of polynomials p, q in an indeterminate t with rational coefficients, $q \neq 0$, where $p/q = r/s$ means $ps = rq$, with the usual sum and product of fractions. Call p/q with $p \neq 0$ **positive** if the leading coefficients of p and q have the same sign. With this positive set, $\mathbb{Q}(t)$ is an ordered field; we state this and do not verify it. Show that $t > n \cdot 1$ for every $n \in \mathbb{N}$, so that $\mathbb{Q}(t)$ is **not** Archimedean, and that $0 < 1/t < 1/(n \cdot 1)$ for every $n \geq 1$.

Solution. In $\mathbb{Q}(t)$ the element $n \cdot 1$ is the constant polynomial n, by the recursion of Definition 4.2.1 (a) and the addition of constants. The difference $t - n \cdot 1 = (t - n)/1$ has numerator $t - n$, with leading coefficient 1, and denominator 1, with leading coefficient 1. The signs agree, so $t - n \cdot 1$ is positive, that is, $n \cdot 1 < t$ (Proposition 2.3.3). This holds for every $n \in \mathbb{N}$, so $x = t$ has no n with $x < n \cdot 1$: $\mathbb{Q}(t)$ is not Archimedean.

Next, $1/t$ has numerator 1 and denominator t, both with leading coefficient 1, so $1/t > 0$. For $n \geq 1$, Lemma 4.1.1 (c) gives

$$\frac{1}{n \cdot 1} - \frac{1}{t} = \frac{t - n}{nt},$$

whose numerator has leading coefficient 1 and whose denominator nt has leading coefficient $n > 0$. So the difference is positive, and $1/t < 1/(n \cdot 1)$.

Warning. The Archimedean property does not follow from the ordered-field axioms: $\mathbb{Q}(t)$ satisfies (F1)–(F9) and (O1)–(O4) and fails it. So in a general ordered field the move "take $\varepsilon = 1/n$" does not reach every ε: in $\mathbb{Q}(t)$, the tolerance $\varepsilon = 1/t$ is below every $1/n$. Each later use of "$1/n$ eventually drops below ε" must name the Archimedean property.

The density theorem fails in $\mathbb{Q}(t)$ too, so its hypothesis is needed. There $0 < 1/t$, and no rational lies strictly between them. Suppose $q \in \mathbb{Q}$ had $0 < \iota_{\mathbb{Q}(t)}(q) < 1/t$, with $\iota_{\mathbb{Q}(t)}$ the embedding of Theorem 4.2.2. Then $q > 0$, because $\iota_{\mathbb{Q}(t)}$ reflects order. By Proposition 2.5.2 (b) and Lemma 4.1.1 (a), $t = 1/(1/t) < 1/\iota_{\mathbb{Q}(t)}(q) = \iota_{\mathbb{Q}(t)}(1/q)$. By Theorem 2.5.3, $1/q < n$ for some $n \in \mathbb{N}$, so $\iota_{\mathbb{Q}(t)}(1/q) < \iota_{\mathbb{Q}(t)}(n) = n \cdot 1$, and $t < n \cdot 1$, against the example.

4.2.6 Exercises

4.2.6.1 A. Check your understanding

Exercise. A1

(a) Write "F is **not** Archimedean" in symbols, moving the negation inside by Logic and Proof, Proposition 6.2.1.

(b) Give two different elements of $\mathbb{Q}(t)$ (Example (A Non-Archimedean Ordered Field)) that lie above every $n \cdot 1$.

(c) In an ordered field, can $n \cdot 1_F = 0$ for some $n \geq 1$? Give a reason.

(d) What is $\iota_{\mathbb{Q}}$? Which part of Theorem 4.2.2 tells you?

Solution.

(a) The definition is $\forall x \in F\ \exists n \in \mathbb{N}\ \ x < n \cdot 1_F$. By Logic and Proof, Proposition 6.2.1, in its bounded form Logic and Proof, Proposition 6.2.3, applied twice, its negation is $\exists x \in F\ \forall n \in \mathbb{N}\ \ \neg(x < n \cdot 1_F)$, and by totality (Proposition 4.1.2 (a)) $\neg(x < n \cdot 1_F)$ is $n \cdot 1_F \leq x$:

$$\exists x \in F\ \ \forall n \in \mathbb{N}\ \ n \cdot 1_F \leq x.$$

(b) t, as in Example (A Non-Archimedean Ordered Field), and $t+1$, since $t+1 > t > n \cdot 1$. Also t^2: $t^2 - n \cdot 1 = (t^2 - n)/1$ has leading coefficients 1 and 1.

(c) No. By Proposition 4.2.1 (b), $n \cdot 1_F \geq 1_F > 0$ for $n \geq 1$. (The two-element field of Example (A field with two elements that cannot be ordered) has $2 \cdot 1 = 0$, and it cannot be ordered.)

(d) The identity map of $\mathbb{Q}$: it satisfies $\mathrm{id}(1) = 1$ and respects $+$ and $\cdot$, and the theorem says **exactly one** map does.

4.2.6.2 B. Practice

Exercise. B1: Archimedes' form

Suppose the ordered field F is Archimedean, and take $x \in F$ and $y \in F_{>0}$. Prove that there is $n \in \mathbb{N}$ with $(n \cdot 1_F)\, y > x$.

Solution. Since $y > 0$, $y^{-1} > 0$ (Proposition 2.5.2 (a)). By Definition 4.2.2 applied to x/y, there is $n \in \mathbb{N}$ with $x/y < n \cdot 1_F$. Multiplying by $y > 0$ (Proposition 2.3.5 (c)) gives $(xy^{-1})y < (n \cdot 1_F)y$, and $(xy^{-1})y = x(y^{-1}y) = x$ by (F5), (F6), (F8). Hence $(n \cdot 1_F)y > x$.

Exercise. B2: Smaller than every reciprocal

(a) Let F be Archimedean, and let $x \in F$ satisfy $0 \leq x < 1/(n \cdot 1_F)$ for every $n \in \mathbb{N}$ with $n \geq 1$. Prove that $x = 0$.

(b) Give an element $x \neq 0$ of $\mathbb{Q}(t)$ (Example (A Non-Archimedean Ordered Field)) with this property.

Solution.

(a) Suppose, for RAA (Logic and Proof, Definition 2.5.1), that $x \neq 0$. Then $x > 0$, since $x \geq 0$. By Proposition 4.2.3 with $\varepsilon := x$, there is $N \in \mathbb{Z}_{>0}$ with $1/(N \cdot 1_F) < x$. The hypothesis with $n = N \geq 1$ gives $x < 1/(N \cdot 1_F)$. By (O2), $x < x$, against (O1). So $x = 0$.

(b) $x = 1/t$: by Example (A Non-Archimedean Ordered Field), $0 < 1/t < 1/(n \cdot 1)$ for every $n \geq 1$, and $1/t \neq 0$. So part (a) fails without the Archimedean hypothesis.

Exercise. B3: Rationals with a power of two as denominator

(a) Prove that $2^k \geq k + 1$ in $\mathbb{Q}$ for every $k \in \mathbb{N}$.

(b) Let $a, b \in \mathbb{Q}$ with $a < b$. Prove that there are $m \in \mathbb{Z}$ and $k \in \mathbb{N}$ with $a < (m+1)/2^k < b$.

(c) Carry out the construction of (b) for $a = 1/3$, $b = 1/2$, taking k least with $1/(k+1) < b - a$.

Hint: follow the proof of Theorem 4.2.5, with 2^k in place of n.

Solution.

(a) Read 2^k as $(2^k) \cdot 1_{\mathbb{Q}} = (2 \cdot 1_{\mathbb{Q}})^k$ (Proposition 4.2.1 (c), (d)). We use Logic and Proof, Axiom 4.2.1 (through Corollary 1.7.3 (b)) on k for $P(k)$: "$(2^k) \cdot 1_{\mathbb{Q}} \geq (k+1) \cdot 1_{\mathbb{Q}}$". *Base:* $2^0 = 1 = 0 + 1$, so both sides are $1 \cdot 1_{\mathbb{Q}}$. *Step:* assume $P(k)$. Then

$$\begin{aligned}(2^{k+1}) \cdot 1_{\mathbb{Q}} &= ((2^k) \cdot 1_{\mathbb{Q}})(2 \cdot 1_{\mathbb{Q}}) \geq ((k+1) \cdot 1_{\mathbb{Q}})(2 \cdot 1_{\mathbb{Q}}) \\ &= (2k+2) \cdot 1_{\mathbb{Q}} \geq (k+2) \cdot 1_{\mathbb{Q}},\end{aligned}$$

by Proposition 4.2.1 (a) and $2^{k+1} = 2^k \cdot 2$; then Proposition 4.1.2 (d), all factors being non-negative by Proposition 4.2.1 (b); then (a) again with $(k+1)2 = 2k+2$ (Theorem 1.5.3); and finally Proposition 4.2.1 (b), since $k + 2 \leq 2k + 2$. This is $P(k+1)$.

(b) By Proposition 4.2.3 with $F = \mathbb{Q}$ and $\varepsilon := b - a > 0$, there is $k \in \mathbb{N}$ with $1/(k+1) < b - a$ (take $k + 1 \geq N$). By (a) and Proposition 4.1.2 (c), $1/2^k \leq 1/(k+1) < b - a$. Let $m := \lfloor 2^k a \rfloor$ (Proposition 2.5.5), so that $m \leq 2^k a < m + 1$. Then $2^k a < m + 1 \leq 2^k a + 1$, so

$$a < \frac{m+1}{2^k} \leq a + \frac{1}{2^k} < b,$$

dividing by $2^k > 0$ (Proposition 2.3.5 (c), Proposition 2.5.2 (a)) and using $1/2^k < b - a$ with (O3).

(c) $b - a = 1/6$. For $k \in \mathbb{N}$, $1/(k+1) < 1/6$ if and only if $6 < k + 1$, so the least k is 6, with $2^6 = 64$. Then $2^6 a = 64/3$, and $\lfloor 64/3 \rfloor = 21$, since $21 \cdot 3 = 63 \leq 64 < 66 = 22 \cdot 3$. So $(m+1)/2^k = 22/64 = 11/32$. Check: $1/3 < 11/32$ since $32 < 33$, and $11/32 < 1/2$ since $22 < 32$. (The bound $1/2^k \leq 1/(k+1)$ is far from sharp: $k = 3$ already gives $3/8$.)

4.2.6.3 C. Going deeper

Exercise. C1: Many rationals, and rational approximation

Throughout, F is an Archimedean ordered field.

(a) Let $x < y$ in F. Prove that the set $S := \{q \in \mathbb{Q} : x < \iota_F(q) < y\}$ is infinite (Sets, Relations, and Cardinality, Definition 3.1.3).

(b) Let $x \in F$ and $n \in \mathbb{Z}_{>0}$. Prove that there is $q \in \mathbb{Q}$ with $\iota_F(q) \leq x < \iota_F(q + 1/n)$.

Solution (Hint 1). For (a), apply Theorem 4.2.5 twice; for (b), apply Proposition 4.2.4 to a multiple of x.

Solution (Hint 2). For (a): let $r, s \in \mathbb{Q}$ with $x < \iota_F(r) < \iota_F(s) < y$; it is enough to find infinitely many rationals strictly between r and s. For (b): let m be the integer part of $(n \cdot 1_F)x$.

Solution.

(a) By Theorem 4.2.5 there is $s \in \mathbb{Q}$ with $x < \iota_F(s) < y$, and, applied to $x < \iota_F(s)$, there is $r \in \mathbb{Q}$ with $x < \iota_F(r) < \iota_F(s)$. Then $r < s$, since ι_F reflects order (Theorem 4.2.2). For $k \in \mathbb{Z}_{>0}$, let

$$s_k := r + \frac{s - r}{k + 1} \in \mathbb{Q}.$$

Since $k+1 > 1 > 0$, $0 < 1/(k+1) < 1$ (Proposition 2.5.2), so $0 < (s-r)/(k+1) < s-r$ (Proposition 2.3.5 (c)), and $r < s_k < s$ by (O3). As ι_F is strictly increasing, $x < \iota_F(r) < \iota_F(s_k) < \iota_F(s) < y$, so $s_k \in S$. If $k < l$, then $1/(l+1) < 1/(k+1)$ (Proposition 2.5.2 (b), Proposition 4.2.1 (b)), so $s_l < s_k$: the map $k \mapsto s_k$ is injective. Suppose S were finite, with m elements for some $m \in \mathbb{N}$ (Sets, Relations, and Cardinality, Definition 3.1.2). The map $\{1, \ldots, m+1\} \to S$, $k \mapsto s_k$, would be injective, against Sets, Relations, and Cardinality, Corollary 3.1.3 (a). So S is infinite.

(b) Let $c := n \cdot 1_F > 0$ (Proposition 4.2.1 (b)) and let m be the integer part of cx (Proposition 4.2.4): $\iota_F(m) \leq cx < \iota_F(m+1)$. Multiplying by $c^{-1} > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c), with equality allowed on the left) gives $\iota_F(m)c^{-1} \leq x < \iota_F(m+1)c^{-1}$. Let $q := m/n$. Since $c = \iota_F(n)$ and ι_F respects products, $\iota_F(m)c^{-1} = \iota_F(m/n) = \iota_F(q)$ and $\iota_F(m+1)c^{-1} = \iota_F((m+1)/n) = \iota_F(q + 1/n)$, using $(m+1)/n = m/n + 1/n$ (Lemma 4.1.1 (b) in $\mathbb{Q}$). So $\iota_F(q) \leq x < \iota_F(q + 1/n)$.

Marking scheme. (10 points) - 2: (a) obtains two rationals $r < s$ inside the interval, by density twice, with $r < s$ justified (order reflected). - 2: (a) exhibits infinitely many distinct rationals between r and s (any explicit injective family, or an induction producing one more each time), each checked to lie strictly between. - 2: (a) concludes that S is infinite from an injection, citing a finiteness result (pigeonhole or subsets of finite sets). - Second route for (a), same 6 points: 2 for a largest element of a finite non-empty S (Sets, Relations, and Cardinality, Proposition 2.3.8), 2 for density above it, 2 for the contradiction. - 2: (b) applies the integer part to $(n \cdot 1_F)x$. - 2: (b) divides by $n \cdot 1_F > 0$ and identifies the bounds as $\iota_F(q)$, $\iota_F(q + 1/n)$. Fatal errors (0 for the step): using density in an arbitrary ordered field (the Archimedean hypothesis must be named); "infinitely many because the construction never stops" with no finiteness argument; dividing by $n \cdot 1_F$ without its sign.

Exercise. C2: Where the proof for the rationals breaks

In $\mathbb{Q}(t)$ (Example (A Non-Archimedean Ordered Field)):

(a) Show that the set $\{n \cdot 1 : n \in \mathbb{N}\}$ is bounded above.

(b) The proof of Theorem 2.5.3 shows that every rational lies below a natural number. Which step of that proof has no analogue for the element t of $\mathbb{Q}(t)$? Prove that it has none.

Solution.

(a) By Example (A Non-Archimedean Ordered Field), $n \cdot 1 < t$ for every $n \in \mathbb{N}$, so t is an upper bound.

(b) The first step: "write $x = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}_{>0}$". The rest of the proof (the case split on the sign of p, discreteness of $\mathbb{Z}$, the comparison formula) works for any element of the form $\iota_F(p)/\iota_F(q)$ of an ordered field F, in particular of $\mathbb{Q}(t)$. But t is not of this form. Indeed, $\iota_{\mathbb{Q}(t)}(p)/\iota_{\mathbb{Q}(t)}(q) = \iota_{\mathbb{Q}(t)}(p/q)$ by Theorem 4.2.2; if $t = \iota_{\mathbb{Q}(t)}(r)$ with $r \in \mathbb{Q}$, then Theorem 2.5.3 gives $n \in \mathbb{N}$ with $r < n$, so $t = \iota_{\mathbb{Q}(t)}(r) < \iota_{\mathbb{Q}(t)}(n) = n \cdot 1$, against (a). So t is not the image of any rational, and the proof has nothing to start from.

Marking scheme. (10 points) - 3: (a) exhibits an upper bound and justifies it from the positive set of $\mathbb{Q}(t)$. - 3: (b) names the representation $x = p/q$ with integers p, q as the step that fails. - 4: (b) proves that t is not $\iota_{\mathbb{Q}(t)}(r)$ for any rational r, using the Archimedean property of $\mathbb{Q}$ and that $\iota_{\mathbb{Q}(t)}$ is strictly increasing. Fatal errors (0 for the step): naming the comparison formula or discreteness as the failing step (both hold for images of rationals); asserting that t "is not a fraction" without proof.

Exercise. C3: The order of the rationals is the only one

Prove that the usual order of $\mathbb{Q}$ (Definition 2.5.1) is the only relation $<$ on $\mathbb{Q}$ that makes $\mathbb{Q}$, with its usual operations, an ordered field.

Solution (Hint 1). Which elements must be positive in any such order? Start with 1, then $n \cdot 1$, then p/q.

Solution (Hint 2). By Proposition 2.3.3, an order on a commutative ring is determined by its set of positive elements. Let P be the positive set of an order $\prec$ making $\mathbb{Q}$ an ordered field; show $P = \mathbb{Q}_{>0}$.

Solution. Let $\prec$ be a relation making $\mathbb{Q}$ an ordered field, and $P \coloneqq \{x \in \mathbb{Q} : 0 \prec x\}$. By Proposition 2.3.3, P satisfies (P1) and (P2), and $x \prec y$ holds if and only if $y - x \in P$. The usual order has positive set $\mathbb{Q}_{>0}$ (Theorem 2.5.1) and is described by $\mathbb{Q}_{>0}$ in the same way. So it suffices to prove $P = \mathbb{Q}_{>0}$.

The multiples $n \cdot 1_{\mathbb{Q}}$ are defined by the recursion of Definition 4.2.1 (a), which uses only addition, so they do not depend on the order; by Proposition 4.2.1 (d), $n \cdot 1_{\mathbb{Q}} = n$. Applying Proposition 2.3.5 (h) to the ordered field $(\mathbb{Q}, \prec)$, in which $1 \neq 0$, gives $0 \prec n$ for every $n \in \mathbb{Z}_{>0}$: so $\mathbb{Z}_{>0} \subseteq P$.

Let $x \in \mathbb{Q}_{>0}$. Then $x = p/q$ with $p, q \in \mathbb{Z}_{>0}$ (Proposition 2.4.4 (b) and the comparison formula, Theorem 2.5.1 (a)), so $x = p \cdot q^{-1}$. Here $p, q \in P$, so $q^{-1} \in P$ by Proposition 2.5.2 (a) applied in $(\mathbb{Q}, \prec)$, and $x \in P$ by (P1). Hence $\mathbb{Q}_{>0} \subseteq P$.

Conversely, let $x \in P$. Then $x \neq 0$ by (P2). If $x \notin \mathbb{Q}_{>0}$, then $x < 0$ in the usual order, so $-x \in \mathbb{Q}_{>0} \subseteq P$; but then both x and $-x$ lie in P, against (P2). So $x \in \mathbb{Q}_{>0}$, and $P = \mathbb{Q}_{>0}$. Hence $\prec$ is the usual order.

Marking scheme. (10 points) - 2: reduces the claim to the positive sets, citing Proposition 2.3.3. - 2: shows every positive integer is positive for the new order, noting that $n \cdot 1$ does not depend on the order. - 3: shows every usual-positive p/q is in P, via inverses and products. - 3: shows $P \subseteq \mathbb{Q}_{>0}$ by (P2). - Route through Theorem 4.2.2 for $F = (\mathbb{Q}, \prec)$, same 10 points: 4 for $\iota_F = \mathrm{id}$ by uniqueness (the equations ignore the order), 6 for reading off the order. Fatal errors (0 for the step): assuming $1 \succ 0$ without $1 \neq 0$ or a rule; using the usual order's comparison formula to compare elements in the new order.

4.3 Sequences and Limits in an Ordered Field

In $\mathbb{Q}$ the fractions $1/2, 2/3, 3/4, \ldots$ "tend to 1". Sections 4.1 and 4.2 supplied the tools to say what that means without leaving $\mathbb{Q}$: an absolute value to measure distance, and the Archimedean property to make $1/(n+1)$ small. This section defines the limit of a sequence in an arbitrary ordered field, with tolerances taken from that field. It then proves that a sequence has at most one limit, and that a convergent sequence is bounded. Chapter 5 applies these results with $F = \mathbb{Q}$, at a stage when no other number system is available.

Throughout, F is an ordered field (Definition 2.3.2), and $F_{>0} = \{x \in F : x > 0\}$ is its set of positive elements, as in Section 4.1; for $F = \mathbb{Q}$ this is $\mathbb{Q}_{>0}$. As agreed after Proposition 4.2.1, we write n for $n \cdot 1_F$ where no confusion arises. So $1/(n+1)$ in F means $1/((n+1) \cdot 1_F)$, which exists and is positive, since $(n+1) \cdot 1_F \geq 1 > 0$ by Proposition 4.2.1 (b) and reciprocals of positive elements are positive (Proposition 2.5.2 (a)).

4.3.1 Sequences

A list $1, 1/2, 1/3, \ldots$ is a rule that assigns a term to each position, and a rule of that kind already has a name: a function on $\mathbb{N}$.

Definition 4.3.1. Sequence; Eventually

A **sequence** in F is a function $a \colon \mathbb{N} \to F$ (Sets, Relations, and Cardinality, Definition 1.4.1). Its value at n is written a_n and called the n-th **term**; the sequence is written $(a_n)_{n \in \mathbb{N}}$, or (a_n), and $\{a_n : n \in \mathbb{N}\}$ is its set of values.

A property $P(n)$ of natural numbers holds **eventually** if **there is** $N_0 \in \mathbb{N}$ such that $P(n)$ holds **for every** $n \in \mathbb{N}$ with $n \geq N_0$.

Examples.

- In $\mathbb{Q}$: $a_n := 1/(n+1)$, with terms $1, 1/2, 1/3, \ldots$.
- In any F: $a_n := n$, that is, $n \cdot 1_F$, with terms $0, 1, 2, \ldots$; and $a_n := (-1)^n$, with the powers of Definition 4.2.1, with terms $1, -1, 1, \ldots$. The second sequence has infinitely many terms but only two values: its set of values is $\{1, -1\}$. A sequence is ordered and may repeat; its set of values forgets both.
- The degenerate case: a **constant sequence** $a_n := c$ for a fixed $c \in F$. It matters because every definition below must give the expected answer on it.
- A sequence may be defined by recursion (Sets, Relations, and Cardinality, Theorem 2.3.1, read on $\mathbb{N} = \omega$ as Theorem 1.5.1 through Corollary 1.7.3 (b)). The Newton sequence of Example (A Rational Sequence Aimed at the Gap), $x_0 = 2$ and $x_{n+1} = (x_n^2 + 2)/(2x_n)$, is a sequence in $\mathbb{Q}$; Section 4.6 studies it.

Non-example by minimal change. The formula $n \mapsto 1/n$ defines no sequence: at $n = 0$ it asks for $1/0$, and 0 has no inverse. Everything else about the first example survives the change, and only "a function on all of $\mathbb{N}$" fails. This is why the book's sequences start at $n = 0$ and why we write $1/(n+1)$.

"Eventually" absorbs any finite beginning. In $\mathbb{Q}$, "$1/(n+1) < 1/10$" holds eventually, from $N_0 = 10$ on, although it fails for $n \leq 9$; "$(-1)^n = 1$" does not hold eventually, since it fails at every odd n. Two eventual properties hold together eventually: if $P(n)$ holds for $n \geq N_0$ and $Q(n)$ for $n \geq N_1$, then both hold for $n \geq \max\{N_0, N_1\}$, the larger of the two numbers, which exists by trichotomy in $\mathbb{N}$ (Theorem 1.6.1 (a)). The same goes for any finite number of properties, taking the larger number one comparison at a time. For infinitely many properties it fails: "$1/(n+1) < 1/(k+1)$" holds eventually for each $k \in \mathbb{N}$, from $N_0 = k + 1$ on, but no single N_0 serves every k, since $k = N_0$ fails at $n = N_0$.

4.3.2 The limit of a sequence

We want a precise sense in which $1/2, 2/3, 3/4, \ldots$ "tends to 1" in $\mathbb{Q}$. The first idea is "the terms get closer and closer to 1". Here is why it fails. With $a_n := (n+1)/(n+2)$,

$$1 - a_n = \frac{1}{n+2}, \qquad 2 - a_n = \frac{n+3}{n+2} = 1 + \frac{1}{n+2}.$$

Both distances shrink at every step, so the terms get "closer and closer" to 2 as well. What singles out 1 is that the distance to 1 falls below **any** tolerance we name, and stays below it, while the distance to 2 never falls below 1. The tolerances must come from F itself, since in $\mathbb{Q}$ there is nothing else to measure with.

Every tolerance in F is eventually met.

Definition 4.3.2. Limit of a Sequence in an Ordered Field

Let (a_n) be a sequence in F and $a \in F$. The sequence **converges to** a in F, written $a_n \to a$, if **for every** $\varepsilon \in F_{>0}$ **there exists** $N \in \mathbb{N}$ such that $|a_n - a| < \varepsilon$ **for every** $n \in \mathbb{N}$ with $n \geq N$. In symbols,

$$\forall \varepsilon \in F_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall n \in \mathbb{N} \;\; \big(n \geq N \Rightarrow |a_n - a| < \varepsilon\big).$$

Then a is **a limit** of (a_n). The sequence is **convergent** if it converges to some $a \in F$, and **divergent** otherwise.

In words: the first clause hands us an arbitrary tolerance ε, a positive element of F. The second says that some index N can be found, and it may depend on ε. The third says that from N on, **every** term lies within ε of a: the inequality holds **eventually**. The absolute value is that of Definition 4.1.1, and by Proposition 4.1.3 (d) the condition on each term reads

$$|a_n - a| < \varepsilon \iff a - \varepsilon < a_n < a + \varepsilon.$$

The domain of ε is part of the definition. For $F = \mathbb{Q}$, ε is a **rational** number, and "$a_n \to a$ in $\mathbb{Q}$" mentions nothing outside $\mathbb{Q}$. The formula is the one translated in Logic and Proof, Example (Five definitions, translated) (a), where it was read in the school real numbers. Here it is read in F, and every fact used about it will be proved from the ordered-field axioms.

Well-definedness. Nothing needs checking for "converges to a". We say "**a** limit" for now, because a sequence might have two; Theorem 4.3.4 below rules that out.

Negating the definition, quantifier by quantifier (Logic and Proof, Proposition 6.2.3 for the bounded quantifiers, Logic and Proof, Proposition 6.2.2 for the implication), and using that $\neg(x < \varepsilon)$ means $x \geq \varepsilon$ in the totally ordered F (Proposition 4.1.2 (a)): $a_n \not\to a$ means

$$\exists \varepsilon \in F_{>0} \;\; \forall N \in \mathbb{N} \;\; \exists n \in \mathbb{N} \;\; (n \geq N \wedge |a_n - a| \geq \varepsilon).$$

One bad tolerance, and beyond every threshold a term that misses it. A sequence is divergent when this holds for **every** $a \in F$.

Example. First Limits

Decide, from Definition 4.3.2, which of these converge, and to what.

(a) A constant sequence $a_n = c$ in any ordered field F.
(b) $a_n = 1/(n+1)$ in $\mathbb{Q}$.
(c) $a_n = (n+1)/(n+2)$ in $\mathbb{Q}$.
(d) $a_n = (-1)^n$ in any ordered field F.

Solution.

(a) $a_n \to c$. Let $\varepsilon \in F_{>0}$ and take $N := 0$. For every n, $|a_n - c| = |0| = 0 < \varepsilon$. Here N does not depend on ε at all: the degenerate case.

(b) $a_n \to 0$. The field $\mathbb{Q}$ is Archimedean (Definition 4.2.2, by Theorem 2.5.3). Let $\varepsilon \in \mathbb{Q}_{>0}$. By Proposition 4.2.3 there is $N \in \mathbb{Z}_{>0}$ with $1/m < \varepsilon$ for every $m \geq N$. For $n \geq N$ also $n + 1 \geq N$, so $|a_n - 0| = 1/(n+1) < \varepsilon$, where the absolute value drops because $1/(n+1) > 0$. The N depends on ε, and Archimedes is what supplies it.

(c) $a_n \to 1$. By the computation above the definition, $|a_n - 1| = 1/(n+2)$. Given $\varepsilon \in \mathbb{Q}_{>0}$, take the N of (b). For $n \geq N$, $n + 2 > n + 1 \geq N$, so $1/(n+2) < \varepsilon$.

(d) Divergent, in **every** ordered field. By the recursion for powers, $(-1)^{n+1} = (-1)^n(-1) = -(-1)^n$ (Proposition 2.3.2 (c)), and $|(-1)^n| = 1$ for every n, by induction with Proposition 4.1.3 (b). Hence

$$|a_{n+1} - a_n| = |-2a_n| = 2|a_n| = 2$$

for every n. Suppose $a_n \to a$ for some $a \in F$. Take $\varepsilon := 1$ and its N, and let $n := N$, so that n and $n + 1$ are both $\geq N$. The three-point triangle inequality (Proposition 4.1.3

(e)) gives

$$2 = |a_{n+1} - a_n| \leq |a_{n+1} - a| + |a - a_n| < 1 + 1 = 2,$$

so $2 < 2$, against (O1). So no $a \in F$ is a limit. The bad tolerance is $\varepsilon = 1$: two consecutive terms are 2 apart and cannot both lie within 1 of one point.

Quick check. For $a_n = 1/(n+1)$ in $\mathbb{Q}$ and $\varepsilon = 1/100$, which $N \in \mathbb{N}$ satisfy "$|a_n| < \varepsilon$ for every $n \geq N$"?

Solution. Exactly the $N \geq 100$. By the comparison formula (Theorem 2.5.1 (a)), $1/(n+1) < 1/100$ if and only if $100 < n+1$, that is, $n \geq 100$. So every $N \geq 100$ works, and an $N \leq 99$ fails at $n = N$ itself, where $1/(N+1) \geq 1/100$. The definition asks only for **some** working N; any larger one works too.

Non-example by minimal change. Keep the sequence $a_n = (n+1)/(n+2)$ and change the candidate limit from 1 to 2. The distance $|a_n - 2| = 1 + 1/(n+2)$ still shrinks at every step, and still the definition fails, at its first clause: for $\varepsilon := 1$, every term has $|a_n - 2| > 1$, so no N exists. This is the negation displayed above, with $\varepsilon = 1$ and $n = N$ for every N.

Why this definition.

- *Strict or not.* Replacing $|a_n - a| < \varepsilon$ by $|a_n - a| \leq \varepsilon$ gives the same notion: given ε, apply the weak form to $\varepsilon/2$, and $\varepsilon/2 < \varepsilon$ (Proposition 4.1.2 (b)).
- *Every $n \geq N$, not some.* With "some $n \geq N$" the sequence $(-1)^n$ would "converge" both to 1 and to -1: beyond any N there are terms equal to each.
- *The order of the quantifiers.* With $\exists N$ before $\forall \varepsilon$, the condition says that $a_n = a$ for every $n \geq N$: for each such n, $|a_n - a|$ lies below every element of $F_{>0}$, so it is not itself positive, and it is 0 by Proposition 4.1.3 (a). (Exercise A1 (b) asks for the details.) Then $1/(n+1)$ would have no limit in $\mathbb{Q}$.
- *Tolerances from F.* In a field with more elements than $\mathbb{Q}$, "every rational tolerance is eventually met" can be strictly weaker than "every tolerance in F is eventually met". The remark after Proposition 4.3.1 shows this happen.

Warning. $a_n \to a$ does **not** say that some term equals a: $1/(n+1) \to 0$ in $\mathbb{Q}$ (Example (First Limits) (b)), and no term is 0. Nor does it say that $|a_n - a|$ decreases: $a_n := (1+(-1)^n)/(n+1)$ in $\mathbb{Q}$ has $0 \leq a_n \leq 2/(n+1)$, so $|a_n| < \varepsilon$ as soon as $1/(n+1) < \varepsilon/2$, and it tends to 0; yet its terms run $2, 0, 2/3, 0, 2/5, \ldots$, up and down, hitting 0 at $n = 1, 3, 5, \ldots$.

Two kinds of sequence get their own names, because they recur throughout the rest of the chapter.

Definition 4.3.3. Null Sequence

A sequence (a_n) in F is **null** if $a_n \to 0$.

Definition 4.3.4. Bounded Sequence

A sequence (a_n) in F is **bounded** if **there exists** $M \in F_{>0}$ with $|a_n| \leq M$ **for every** $n \in \mathbb{N}$. Such an M is a **bound** for (a_n).

Examples. In $\mathbb{Q}$, $(1/(n+1))$ is null and bounded by 1; in every F, $((-1)^n)$ is bounded by 1 and not null (Example (First Limits) (d)); a constant sequence c is bounded by $|c|+1$, which is positive even when $c = 0$. **Non-example.** In $\mathbb{Q}$ the sequence (n) is not bounded: for any $M \in \mathbb{Q}_{>0}$, Archimedes (Theorem 2.5.3) gives $n \in \mathbb{N}$ with $M < n = |n|$. The bound is required to be positive only for convenience, so that we may divide by it; any $M \geq 0$ that works can be replaced by $M + 1$.

4.3.3 Archimedes and the sequence 1/(n + 1)

In $\mathbb{Q}$, $1/(n+1) \to 0$ came straight from Archimedes. In a general ordered field, the two statements turn out to be the same.

Proposition 4.3.1. Archimedes as a Null Sequence

The sequence $\big(1/((n+1)\cdot 1_F)\big)_{n\in\mathbb{N}}$ is null in F if and only if F is Archimedean.

Idea. One direction is Proposition 4.2.3 read as a limit. For the other, a field that is not Archimedean has an element x above every $n \cdot 1_F$; then every $1/(n+1)$ is at least $1/x$, and $1/x$ is a tolerance that is never met.

Proof. ($\Leftarrow$) Let F be Archimedean and $\varepsilon \in F_{>0}$. By Proposition 4.2.3 there is $N \in \mathbb{Z}_{>0}$ (depending on ε) with $1/(m \cdot 1_F) < \varepsilon$ for every $m \geq N$. For $n \geq N$ also $n+1 \geq N$, so

$$\left|\frac{1}{(n+1)\cdot 1_F} - 0\right| = \frac{1}{(n+1)\cdot 1_F} < \varepsilon,$$

where the absolute value drops because the term is positive. So the sequence is null.

($\Rightarrow$) We prove the contrapositive. Suppose F is not Archimedean. Negating Definition 4.2.2 (Logic and Proof, Proposition 6.2.3, and Proposition 4.1.2 (a) for $\neg(x < n \cdot 1_F) \iff x \geq n \cdot 1_F$), there is $x \in F$ with $x \geq n \cdot 1_F$ for every $n \in \mathbb{N}$. In particular $x \geq 1 \cdot 1_F = 1 > 0$, so $\varepsilon := 1/x \in F_{>0}$ (Proposition 2.5.2 (a)). For every $n \in \mathbb{N}$, $0 < (n+1)\cdot 1_F \leq x$ (Proposition 4.2.1 (b)), so Proposition 4.1.2 (c) gives

$$\left|\frac{1}{(n+1)\cdot 1_F} - 0\right| = \frac{1}{(n+1)\cdot 1_F} \geq \frac{1}{x} = \varepsilon.$$

So for this ε no N works, even with $n = N$, and the sequence is not null. ∎

So in $\mathbb{Q}(t)$, the field of Example (A Non-Archimedean Ordered Field), which is not Archimedean, $1/(n+1)$ does not tend to 0: the tolerance $1/t$ is never met. In $\mathbb{Q}$, $1/(n+1) \to 0$ is Example (First Limits) (b), which rests on the Archimedean property; every result proved without the

Archimedean property holds in $\mathbb{Q}(t)$ as well.

Remark. The same example shows why tolerances must come from F. In $\mathbb{Q}(t)$, every **rational** tolerance is eventually met by $1/(n+1)$. Write ι for the embedding ι_F of Theorem 4.2.2. Since ι respects products, $\iota(1) = 1$ and $\iota(n+1) = (n+1)\cdot 1$, we have $((n+1)\cdot 1)\cdot\iota(1/(n+1)) = \iota(1) = 1$, so the term $1/((n+1)\cdot 1)$ is $\iota(1/(n+1))$. Given $q \in \mathbb{Q}_{>0}$, Example (First Limits) (b) gives N with $1/(n+1) < q$ in $\mathbb{Q}$ for every $n \geq N$, and ι, being strictly increasing, carries this to $\iota(1/(n+1)) < \iota(q)$. Yet the sequence is not null in $\mathbb{Q}(t)$. For $F = \mathbb{Q}$ the two readings coincide, because there every tolerance is rational.

4.3.4 Changing finitely many terms

Whether a sequence converges, and to what, depends only on its terms from some point on. Two forms of this are used constantly: altering finitely many terms, and dropping the first k.

Proposition 4.3.2. Only the Tail Matters

Let (a_n) and (b_n) be sequences in F, and $a \in F$.

(a) If there is $N_0 \in \mathbb{N}$ with $a_n = b_n$ for every $n \geq N_0$, then $a_n \to a$ if and only if $b_n \to a$; in particular (a_n) is null if and only if (b_n) is; and (a_n) is bounded if and only if (b_n) is.

(b) For $k \in \mathbb{N}$, the shifted sequence $(a_{n+k})_{n\in\mathbb{N}}$ converges to a if and only if (a_n) does.

Idea. For limits, a threshold for one sequence becomes a threshold for the other once it is pushed past N_0, or moved by k. For bounds, the finitely many terms before N_0 are absorbed by a finite maximum.

Proof.

(a) *Limits.* Suppose $a_n \to a$, and let $\varepsilon \in F_{>0}$. Choose N with $|a_n - a| < \varepsilon$ for every $n \geq N$, and let $N' \coloneqq \max\{N, N_0\}$. For $n \geq N'$, $b_n = a_n$ and $|b_n - a| = |a_n - a| < \varepsilon$. So $b_n \to a$. The hypothesis is symmetric in the two sequences, so the converse follows by exchanging them. The case $a = 0$ is the statement about null sequences.
Bounds. Suppose $|a_n| \leq M$ for every n, with $M \in F_{>0}$. If $N_0 = 0$, the two sequences are equal. Otherwise let

$$M' \coloneqq \max\{|b_0|, \ldots, |b_{N_0-1}|, M\},$$

the maximum of finitely many elements of F, which exists by Sets, Relations, and Cardinality, Proposition 2.3.8 (its proof works in any ordered field, as *Sets, Relations, and Cardinality*, Section 2.6 remarks). Then $M' \geq M > 0$. For $n < N_0$, $|b_n| \leq M'$ because $|b_n|$ is one of the listed elements; for $n \geq N_0$, $|b_n| = |a_n| \leq M \leq M'$. So M' bounds (b_n). The converse is the same with the sequences exchanged.

(b) Suppose $a_n \to a$, and let $\varepsilon \in F_{>0}$, with N as in the definition. For $n \geq N$, also $n + k \geq N$, so $|a_{n+k} - a| < \varepsilon$. Hence $(a_{n+k})_n$ converges to a, with the same N.

Conversely, suppose $a_{n+k} \to a$, and let $\varepsilon \in F_{>0}$, with N such that $|a_{m+k} - a| < \varepsilon$ for every $m \geq N$. Let $n \geq N + k$. Then $n = m + k$ for the natural number $m \coloneqq n - k$, and $m \geq N$

(Proposition 1.6.3 (a), (c)). So $|a_n - a| = |a_{m+k} - a| < \varepsilon$, and $N + k$ is a threshold for (a_n). This proves the proposition.

■

The shift clause is about convergence only; it is used to start a sequence later, for instance to read a formula such as $1/n$, which needs $n \geq 1$, as the shift by 1 of a sequence indexed from 0. Section 4.5 adds the Cauchy property to clause (a).

One more reformulation moves every limit to 0.

Proposition 4.3.3. Limits Through Null Sequences

For a sequence (a_n) in F and $a \in F$: $a_n \to a$ if and only if the sequence $(a_n - a)$ is null.

Proof. For every n, $|(a_n - a) - 0| = |a_n - a|$. So the two conditions of Definition 4.3.2, for (a_n) with limit a and for $(a_n - a)$ with limit 0, say word for word the same thing, with the same N for each ε.

■

4.3.5 A sequence has at most one limit

If "the terms are eventually within every tolerance of a" is to name **the** number the sequence tends to, it must single out one a. It does.

Theorem 4.3.4. Uniqueness of Limits

Let (a_n) be a sequence in F and $a, b \in F$. If $a_n \to a$ and $a_n \to b$, then $a = b$.

Try it first (ten minutes): suppose $a_n \to a$ and $a_n \to b$ with $a \neq b$. Which ε makes the two tails collide?

Idea. The natural first choice is $\varepsilon := |a - b|$, positive because $a \neq b$; it is the choice that worked in Logic and Proof, Proposition 6.2.5. Beyond both thresholds a term a_n is within $|a - b|$ of a and of b, and the triangle inequality through a_n yields only $|a - b| < 2|a - b|$: true, and no contradiction. The two tolerances **add**. So give each half: with $\varepsilon := |a - b|/2$, the same chain ends in $|a - b| < |a - b|$.

Proof. Let $a_n \to a$ and $a_n \to b$, and suppose, for a contradiction, that $a \neq b$. Then $a - b \neq 0$, so $|a - b| > 0$ (Proposition 4.1.3 (a)), and

$$\varepsilon := \frac{|a - b|}{2} \in F_{>0}$$

by Proposition 4.1.2 (b); it depends only on a and b. Since $a_n \to a$, there is N_1 with $|a_n - a| < \varepsilon$ for every $n \geq N_1$; since $a_n \to b$, there is N_2 with $|a_n - b| < \varepsilon$ for every $n \geq N_2$. Let $n := \max\{N_1, N_2\}$. Then both inequalities hold at this n, and the three-point triangle

inequality (Proposition 4.1.3 (e)), with $|a-a_n| = |a_n-a|$ by part (b) of the same proposition, gives

$$|a-b| \le |a-a_n| + |a_n-b| < \varepsilon + \varepsilon = |a-b|,$$

where the strict inequality is Proposition 2.3.5 (b) and the last equality is Proposition 4.1.2 (b). So $|a-b| < |a-b|$, against (O1). Hence the assumption $a \neq b$ is refuted, and $a = b$ (Logic and Proof, Definition 2.5.1).

■

Remark. This is the argument of Logic and Proof, Proposition 6.2.5, rerun in F. There a hypothesis valid for **every** ε was fed the one ε that makes the numbers collide. Here the two hypotheses give $|a-b| < 2\varepsilon$ for every $\varepsilon \in F_{>0}$, and the colliding choice is $|a-b|/2$. The facts about the absolute value that the old argument took from school are now Proposition 4.1.3, valid in every ordered field.

Since a convergent sequence has exactly one limit, we may speak of **the** limit and write it $\lim a_n$, or $\lim_{n\to\infty} a_n$. So $\lim a_n = a$ means the same as $a_n \to a$. In $\mathbb{Q}$, Example (First Limits) gives $\lim 1/(n+1) = 0$ and $\lim (n+1)/(n+2) = 1$, while $\lim (-1)^n$ names nothing.

4.3.6 Convergent sequences are bounded

A convergent sequence cannot run off: from some index on, its terms are trapped near the limit, and only finitely many terms come before that index.

Proposition 4.3.5. Convergent Sequences Are Bounded

Every convergent sequence in F is bounded.

Idea. Use the definition once, with the tolerance 1: from some N on, $|a_n| \le |a_n - a| + |a| < 1 + |a|$. The terms before N are finitely many, so one of them, or $|a| + 1$, is the largest absolute value in sight.

Proof. Let $a_n \to a$. Applying Definition 4.3.2 with $\varepsilon := 1 \in F_{>0}$ gives $N \in \mathbb{N}$ with $|a_n - a| < 1$ for every $n \ge N$. For such n, the triangle inequality (Proposition 4.1.3 (e), with $x = a_n - a$ and $y = a$) gives

$$|a_n| = |(a_n - a) + a| \le |a_n - a| + |a| < 1 + |a|.$$

If $N = 0$, let $M := |a| + 1$. Otherwise let

$$M := \max\{|a_0|, \dots, |a_{N-1}|, |a| + 1\},$$

which exists by Sets, Relations, and Cardinality, Proposition 2.3.8 in the totally ordered F (as in the proof of Proposition 4.3.2). In both cases $M \ge |a|+1 \ge 1 > 0$, by Proposition 4.1.3 (a), so $M \in F_{>0}$. For $n < N$, $|a_n| \le M$ because $|a_n|$ is one of the listed elements; for $n \ge N$, $|a_n| < |a| + 1 \le M$. So M bounds (a_n).

■

The bound M depends on the whole sequence, through a, N and the first N terms, but on no tolerance: it is one element of F, fixed once. Section 4.4 uses exactly that, in the product of two limits. The converse fails: $((-1)^n)$ is bounded by 1 and divergent (Example (First Limits) (d)).

4.3.7 Exercises

4.3.7.1 A. Check your understanding

Exercise. A1

(a) Write "$a_n \not\to a$" in symbols, and name the law used at each step.

(b) What does $\exists N \in \mathbb{N}\ \forall \varepsilon \in F_{>0}\ \forall n \in \mathbb{N}\ (n \geq N \Rightarrow |a_n - a| < \varepsilon)$ say about (a_n)? Does $(1/(n+1))$ satisfy it in $\mathbb{Q}$, with $a = 0$?

(c) For $a_n = (2n+1)/(n+1)$ in $\mathbb{Q}$, find the least N that works in Definition 4.3.2 for the limit 2 and $\varepsilon = 1/50$.

Solution.

(a) Start from $\neg\forall \varepsilon \in F_{>0}\ \exists N \in \mathbb{N}\ \forall n \in \mathbb{N}\ (n \geq N \Rightarrow |a_n - a| < \varepsilon)$. The bounded negation laws (Logic and Proof, Proposition 6.2.3 (a), then (b), then (a)) move $\neg$ past the three quantifiers, turning each $\forall$ into $\exists$ and the $\exists$ into $\forall$. That leaves $\neg(n \geq N \Rightarrow |a_n - a| < \varepsilon)$, which is $n \geq N \wedge \neg(|a_n - a| < \varepsilon)$ by Logic and Proof, Proposition 6.2.2. Finally $\neg(|a_n - a| < \varepsilon)$ is $|a_n - a| \geq \varepsilon$, by trichotomy in F (Proposition 4.1.2 (a)). The result:

$$\exists \varepsilon \in F_{>0}\ \ \forall N \in \mathbb{N}\ \ \exists n \in \mathbb{N}\ \ (n \geq N \wedge |a_n - a| \geq \varepsilon).$$

(b) It says that $a_n = a$ for every $n \geq N$: the sequence is eventually constant, equal to a. Indeed, fix $n \geq N$. Then $|a_n - a| < \varepsilon$ for every $\varepsilon \in F_{>0}$. If $a_n \neq a$, then $\varepsilon := |a_n - a|$ is in $F_{>0}$ (Proposition 4.1.3 (a)) and gives $|a_n - a| < |a_n - a|$, against (O1); so $a_n = a$. This is the argument of Logic and Proof, Proposition 6.2.5, rerun in F for one n at a time (no halving is needed: only one tolerance is in play). Conversely, an eventually constant sequence satisfies the formula, since $|0| = 0 < \varepsilon$. The sequence $(1/(n+1))$ does not satisfy it: no term equals 0. It converges to 0, but its N must depend on ε. The quantifier order is the whole difference.

(c) $a_n - 2 = (2n + 1 - 2(n+1))/(n+1) = -1/(n+1)$, so $|a_n - 2| = 1/(n+1)$. By the comparison formula, $1/(n+1) < 1/50$ if and only if $n + 1 > 50$, that is, $n \geq 50$. So $N = 50$ works, and every $N \leq 49$ fails at $n = 49$ (which is $\geq N$), where $|a_{49} - 2| = 1/50$. The least N is 50.

4.3.7.2 B. Practice

Exercise. B1

Prove from Definition 4.3.2 that, in $\mathbb{Q}$,

$$\frac{2n+1}{n+1} \to 2 \qquad \text{and} \qquad \frac{n^2+1}{2n^2+3} \to \frac{1}{2}.$$

Say exactly where the Archimedean property is used.

Solution. First limit. As in A1 (c), $|(2n+1)/(n+1) - 2| = 1/(n+1)$. Let $\varepsilon \in \mathbb{Q}_{>0}$. Since $\mathbb{Q}$ is Archimedean (Theorem 2.5.3), Proposition 4.2.3 gives $N \in \mathbb{Z}_{>0}$, depending on ε, with $1/m < \varepsilon$ for every $m \geq N$. For $n \geq N$, $n+1 \geq N$, so the distance $1/(n+1)$ is $< \varepsilon$.

Second limit. Bring the difference to one fraction:

$$\frac{n^2+1}{2n^2+3} - \frac{1}{2} = \frac{2(n^2+1) - (2n^2+3)}{2(2n^2+3)} = \frac{-1}{2(2n^2+3)},$$

so the distance is $1/(2(2n^2+3))$. We bound it by $1/(n+1)$. For $n \geq 1$, $n^2 = n \cdot n \geq n \cdot 1 = n$ (Proposition 4.1.2 (d), applied to $0 \leq 1 \leq n$ and $0 \leq n \leq n$), so

$$2(2n^2+3) - (n+1) = 4n^2 - n + 5 \geq 3n + 5 > 0;$$

for $n = 0$ the left side is $5 > 0$. So $0 < n+1 < 2(2n^2+3)$, and Proposition 4.1.2 (c) gives $1/(2(2n^2+3)) \leq 1/(n+1)$. With the N of the first part, every $n \geq N$ has distance $\leq 1/(n+1) < \varepsilon$.

Where Archimedes is used. Only in obtaining N from ε, through Proposition 4.2.3. The algebra of the distances holds in any ordered field. In $\mathbb{Q}(t)$ the same algebra gives the same distances, and the proofs fail exactly there: both take N from $1/(n+1) < \varepsilon$, and by the proof of Proposition 4.3.1 with $x = t$ (Example (A Non-Archimedean Ordered Field)) no N makes $1/(n+1) < 1/t$ for all $n \geq N$.

Exercise. B2

Prove that if $a_n \to a$ in F, then $|a_n| \to |a|$. Show by an example that the converse fails, and prove that it holds when $a = 0$.

Solution. Let $\varepsilon \in F_{>0}$, and choose N with $|a_n - a| < \varepsilon$ for $n \geq N$. By the reverse triangle inequality (Proposition 4.1.3 (f)), for $n \geq N$,

$$\big||a_n| - |a|\big| \leq |a_n - a| < \varepsilon.$$

So $|a_n| \to |a|$, with the same N.

The converse fails for $a_n = (-1)^n$ and $a = 1$: $|a_n| = 1$ for every n, a constant sequence, so $|a_n| \to 1 = |1|$ (Example (First Limits) (a)); but (a_n) converges to nothing (part (d) of that example).

For $a = 0$ the converse holds: $\big||a_n| - |0|\big| = \big||a_n|\big| = |a_n| = |a_n - 0|$, by Proposition 4.1.3 (a), since $|a_n| \geq 0$. So the two conditions coincide: a sequence is null if and only if its absolute values form a null sequence.

Exercise. B3: Bounded but not convergent, in $\mathbb{Q}(t)$

In the field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field), show that the sequence $(n \cdot 1)_{n \in \mathbb{N}}$ is bounded and does not converge. Is (n) bounded in $\mathbb{Q}$?

Solution. Bounded. By Example (A Non-Archimedean Ordered Field), $t > n \cdot 1$ for every $n \in \mathbb{N}$. Also $n \cdot 1 \geq 0$ (Proposition 4.2.1 (b), and $0 \cdot 1 = 0$), so $|n \cdot 1| = n \cdot 1 < t$. And $t > 1 \cdot 1 = 1 > 0$, so $M := t \in \mathbb{Q}(t)_{>0}$ is a bound.

Not convergent. Suppose $n \cdot 1 \to a$ for some $a \in \mathbb{Q}(t)$. Take $\varepsilon := 1/2$, which is positive (Proposition 4.1.2 (b)), and its N. With $n := N$, both n and $n+1$ are $\geq N$, and $(n+1) \cdot 1 - n \cdot 1 = 1$ by Definition 4.2.1 (a). The three-point triangle inequality gives

$$1 = |(n+1) \cdot 1 - n \cdot 1| \leq |(n+1) \cdot 1 - a| + |a - n \cdot 1| < \tfrac{1}{2} + \tfrac{1}{2} = 1,$$

against (O1). So the sequence diverges.

In $\mathbb{Q}$. No: by the non-example after Definition 4.3.4, Archimedes puts some n above any proposed bound. The same sequence is bounded in one field and unbounded in another. Boundedness is a property of the sequence **in** F, like convergence.

4.3.7.3 C. Going deeper

Exercise. C1: A positive limit keeps the terms positive

(a) Prove that if $a_n \to a$ in F and $a > 0$, then $a_n > a/2$ eventually; in particular $a_n > 0$ eventually.

(b) Give a sequence in $\mathbb{Q}$ with $a_n > 0$ for every n whose limit is not positive. What does this say about the converse of (a)?

Solution.

(a) Let $\varepsilon := a/2$, which lies in $F_{>0}$ by Proposition 4.1.2 (b), since $a > 0$. Choose N with $|a_n - a| < a/2$ for every $n \geq N$. By Proposition 4.1.3 (d), for such n,

$$a_n > a - \frac{a}{2} = \frac{a}{2} > 0,$$

where $a - a/2 = a/2$ because $a/2 + a/2 = a$ (Proposition 4.1.2 (b)). So $a_n > a/2$ for every $n \geq N$.

(b) $a_n = 1/(n+1)$: every term is positive, and the limit is 0 (Example (First Limits) (b)). So "every term positive" does not give "limit positive": the strict inequality is lost in the limit. What survives is the weak inequality $a \geq 0$, proved in Section 4.4.

Exercise. C2: Find the error

Here is an argument for Proposition 4.3.5. Is the conclusion true? Is the argument correct? If not, name the line that fails, say what makes it illegal, and give a counterexample to the step. Then repair the argument.

> (1) Let $a_n \to a$. (2) Take $\varepsilon := 1$ in the definition of the limit, and let N be the index it provides. (3) Then for every $n \in \mathbb{N}$, $|a_n| \leq |a_n - a| + |a| < 1 + |a|$. (4) So $M := 1 + |a|$ is a bound for (a_n).

Solution. The conclusion is **true**: it is Proposition 4.3.5. The argument is **not** correct.

The failing line is (3). The first inequality is the triangle inequality and is fine. The second uses $|a_n - a| < 1$ "for every $n \in \mathbb{N}$". Line (2) supplies it only for $n \geq N$: it is the bounded universal $\forall n \in \mathbb{N}\,(n \geq N \Rightarrow |a_n - a| < 1)$. Instantiating it at n needs the line $n \geq N$ (bounded $\forall$E, Logic and Proof, Proposition 6.1.1), and for $n < N$ no such line is available. Line (3) cites a line that is not there: entry (N9) of Logic and Proof, Example (What is not allowed).

Counterexample to the step. In $\mathbb{Q}$, let $a_0 := 5$ and $a_n := 0$ for $n \geq 1$, with $a = 0$. Then $a_n \to 0$, and $N = 1$ is a legitimate choice in line (2) for $\varepsilon = 1$, since $|a_n - 0| = 0 < 1$ for $n \geq 1$. At $n = 0$, line (3) claims $|a_0| < 1 + |0| = 1$, but $|a_0| = 5$; and the $M = 1$ of line (4) is not a bound. No counterexample to the conclusion exists: 5 is a bound.

Repair. Keep line (3) for $n \geq N$ only, and take $M := \max\{|a_0|, \ldots, |a_{N-1}|, |a|+1\}$, the finite maximum of the proof of Proposition 4.3.5.

Marking scheme. (10 points) - 2: says that the conclusion is true and the argument is not correct. - 3: names line (3) and the use of $|a_n - a| < 1$ for $n < N$, where the hypothesis does not apply (bounded $\forall$ used outside its range; entry (N9)). - 3: gives a convergent sequence and a legitimate N for which the step's inequality fails at some $n < N$, and checks it. - 2: repairs the argument with a finite maximum over the terms before N. Fatal errors (0 for the step): declaring the conclusion false; offering a counterexample to the conclusion; blaming the triangle inequality.

4.4 The Algebra and Order of Limits

So far every limit has been found by hand: we computed the distance from a_n to the candidate limit and bounded it, as in Example (First Limits) and Exercise B1 of Section 4.3. This section proves the rules that make most such work unnecessary. Sums, multiples, products and quotients of convergent sequences converge to the sum, multiple, product and quotient of the limits; limits respect $\leq$; and a sequence trapped between two sequences with a common limit has that limit too. Everything holds in an arbitrary ordered field F. No result of this section uses the Archimedean property; only the examples and exercises set in $\mathbb{Q}$ do, each time through Example (First Limits) (b) or the non-example after Definition 4.3.4.

Throughout, F is an ordered field and $F_{>0}$ its set of positive elements, as in Section 4.3.

4.4.1 Two tools

The first tool says that a small factor wins against a factor that merely stays bounded.

Proposition 4.4.1. Null Times Bounded Is Null

Let (a_n) be a null sequence in F, and let (b_n) be a sequence with $|b_n| \leq M$ for every n, where $M \in F_{>0}$. Then $(a_n b_n)$ is null.

Proof. Let $\varepsilon \in F_{>0}$. Then $\varepsilon/M \in F_{>0}$ (Proposition 2.5.2 (a) and (O4)). Since (a_n) is null, there is N (depending on ε and M) with $|a_n| < \varepsilon/M$ for every $n \geq N$. For such n, by

Proposition 4.1.3 (b),

$$|a_n b_n - 0| = |a_n||b_n| \leq |a_n| M < \frac{\varepsilon}{M} M = \varepsilon,$$

where the first inequality multiplies $|b_n| \leq M$ by $|a_n| \geq 0$ (Proposition 4.1.2 (d)), and the second multiplies $|a_n| < \varepsilon/M$ by $M > 0$ (Proposition 2.3.5 (c)).

■

For instance, $((-1)^n/(n+1))$ is null in $\mathbb{Q}$: it is $1/(n+1)$, null by Example (First Limits) (b), times $(-1)^n$, bounded by 1. The bound M is fixed before ε and enters the tolerance ε/M. That is the pattern of the product law below.

The second tool is the fact behind every division by b_n: if $b_n \to b \neq 0$, the terms eventually stay away from 0, by a margin we can name.

Lemma 4.4.2. Terms Near a Non-zero Limit

Let $b_n \to b$ in F with $b \neq 0$. Then there is $N_0 \in \mathbb{N}$ with $|b_n| > |b|/2$ for every $n \geq N_0$. If $b > 0$, then moreover $b_n > b/2$ for every $n \geq N_0$.

Proof. Since $b \neq 0$, $|b| > 0$ (Proposition 4.1.3 (a)), so $\varepsilon := |b|/2 \in F_{>0}$ (Proposition 4.1.2 (b)). Choose N_0 (depending on (b_n) and b) with $|b_n - b| < |b|/2$ for every $n \geq N_0$. For such n, the reverse triangle inequality (Proposition 4.1.3 (f)), together with $x \leq |x|$ and $|-x| = |x|$ (parts (c), (b)), gives

$$|b| - |b_n| \leq \big||b| - |b_n|\big| \leq |b - b_n| = |b_n - b| < \frac{|b|}{2},$$

so $|b_n| > |b| - |b|/2 = |b|/2$, adding $|b_n| - |b|/2$ to both sides (Proposition 2.3.5 (b)). If $b > 0$, then $|b| = b$, and Proposition 4.1.3 (d) turns $|b_n - b| < b/2$ into $b - b/2 < b_n$, that is, $b_n > b/2$.

■

In particular $b_n \neq 0$ for every $n \geq N_0$, since $|b_n| > |b|/2 > 0$. The lemma needs $b \neq 0$. In $\mathbb{Q}$, $b_n := (1 + (-1)^n)/(n+1)$ tends to 0 (the warning after Example (First Limits) in Section 4.3: $0 \leq b_n \leq 2/(n+1)$, and $1/(n+1) \to 0$ by Example (First Limits) (b), which uses Archimedes), yet $b_n = 0$ at every odd n, so for $b = 0$ no N_0 has $|b_n| > |b|/2 = 0$ for every $n \geq N_0$. Even $1/(n+1)$, which tends to 0 (Example (First Limits) (b)) and is never 0, is bounded away from 0 by no positive margin.

4.4.2 The limit laws

Theorem 4.4.3. Limit Laws

Let $a_n \to a$ and $b_n \to b$ in F, and let $c \in F$. Then:

(a) $a_n + b_n \to a + b$;

(b) $ca_n \to ca$; in particular $-a_n \to -a$ and $a_n - b_n \to a - b$;

(c) $a_n b_n \to ab$;
(d) if $b \neq 0$ and $b_n \neq 0$ for every n, then $a_n/b_n \to a/b$.

Idea. Work backwards from each target distance. For (a), $|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b|$: a sum of two terms, each of which we can make small, so we give each half the tolerance. For (c), write

$$a_n b_n - ab = a_n(b_n - b) + b(a_n - a).$$

The factors $b_n - b$ and $a_n - a$ can be made small, but a_n cannot. It is, however, bounded (Proposition 4.3.5), so we replace $|a_n|$ by a bound M and ask $|b_n - b|$ to be smaller than $\varepsilon/(2M)$. For (d), it is enough to treat $1/b_n$ and then use (c). *Before reading on: why must M be fixed before ε?*

Proof.

(a) Let $\varepsilon \in F_{>0}$. Then $\varepsilon/2 \in F_{>0}$ (Proposition 4.1.2 (b)). Since $a_n \to a$, choose N_1 (depending on ε) with $|a_n - a| < \varepsilon/2$ for every $n \geq N_1$; since $b_n \to b$, choose N_2 (depending on ε) with $|b_n - b| < \varepsilon/2$ for every $n \geq N_2$. Let $N \coloneqq \max\{N_1, N_2\}$. For $n \geq N$, both estimates hold, and the triangle inequality (Proposition 4.1.3 (e)) gives

$$\begin{aligned} |(a_n + b_n) - (a + b)| &= |(a_n - a) + (b_n - b)| \\ &\leq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where the strict inequality is Proposition 2.3.5 (b) and the last equality Proposition 4.1.2 (b).

(b) If $c = 0$, then $ca_n = 0 = ca$ for every n (Proposition 2.3.2 (b)), a constant sequence, which converges to 0 (Example (First Limits) (a)). Let $c \neq 0$, so $|c| > 0$ (Proposition 4.1.3 (a)), and let $\varepsilon \in F_{>0}$. Then $\varepsilon/|c| \in F_{>0}$ (Proposition 2.5.2 (a) and (O4)). Choose N (depending on ε and c) with $|a_n - a| < \varepsilon/|c|$ for every $n \geq N$. For such n, by distributivity and Proposition 4.1.3 (b),

$$|ca_n - ca| = |c(a_n - a)| = |c||a_n - a| < |c| \frac{\varepsilon}{|c|} = \varepsilon,$$

multiplying by $|c| > 0$ (Proposition 2.3.5 (c)). With $c = -1$, and $(-1)x = -x$ (Proposition 2.3.2 (c)), this gives $-a_n \to -a$. Then (a), applied to (a_n) and $(-b_n)$, gives $a_n - b_n = a_n + (-b_n) \to a + (-b) = a - b$.

(c) By Proposition 4.3.5 there is $M \in F_{>0}$ with $|a_n| \leq M$ for every $n \in \mathbb{N}$. We fix this M now; it depends on the sequence (a_n) only. Now let $\varepsilon \in F_{>0}$. Since $|b| \geq 0$, the element $|b| + 1$ is $\geq 1 > 0$ (Proposition 2.3.5 (b), (e)). Since $2 > 0$ (Proposition 4.1.2 (b)), $2M > 0$ and $2(|b| + 1) > 0$ by (O4), so Proposition 2.5.2 (a) and (O4) give

$$\frac{\varepsilon}{2M} \in F_{>0} \qquad \text{and} \qquad \frac{\varepsilon}{2(|b| + 1)} \in F_{>0}.$$

Since $b_n \to b$, choose N_1 with $|b_n - b| < \varepsilon/(2M)$ for every $n \geq N_1$. Since $a_n \to a$, choose N_2 with $|a_n - a| < \varepsilon/(2(|b| + 1))$ for every $n \geq N_2$. Both depend only on ε, M and b. Let $N \coloneqq \max\{N_1, N_2\}$, and let $n \geq N$. Using the identity of the Idea, the

triangle inequality and the product rule (Proposition 4.1.3 (e), (b)),

$$\begin{aligned} |a_n b_n - ab| &\le |a_n||b_n - b| + |b||a_n - a| \\ &\le M|b_n - b| + (|b| + 1)|a_n - a| \\ &< M\,\frac{\varepsilon}{2M} + (|b| + 1)\,\frac{\varepsilon}{2(|b| + 1)} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

The second line uses Proposition 4.1.2 (d) twice, with $0 \le |a_n| \le M$ and $0 \le |b| \le |b|+1$, and adds the two estimates (Proposition 2.3.5 (b)). The third multiplies the two estimates by the positive elements M and $|b|+1$ (Proposition 2.3.5 (c)) and adds them (Proposition 2.3.5 (b)); the last equality is Proposition 4.1.2 (b). So $a_n b_n \to ab$.

(d) First we show $1/b_n \to 1/b$. By Lemma 4.4.2 there is N_0 (depending on (b_n) and b) with $|b_n| > |b|/2$ for every $n \ge N_0$. For such n, Lemma 4.1.1 (c) and Proposition 4.1.3 (b) give

$$\Big|\frac{1}{b_n} - \frac{1}{b}\Big| = \Big|\frac{b - b_n}{b_n b}\Big| = \frac{|b_n - b|}{|b_n||b|}.$$

Multiplying $|b_n| > |b|/2$ by $|b| > 0$ gives $|b_n||b| > |b|^2/2 > 0$, so $1/(|b_n||b|) < 2/|b|^2$ (Proposition 2.5.2 (b)). Multiplying by $|b_n - b| \ge 0$,

$$\Big|\frac{1}{b_n} - \frac{1}{b}\Big| \le \frac{2}{|b|^2}\,|b_n - b| \qquad (n \ge N_0).$$

Now let $\varepsilon \in F_{>0}$. Then $\varepsilon|b|^2/2 \in F_{>0}$ ((O4), Proposition 4.1.2 (b)); choose N_1 (depending on ε and b) with $|b_n - b| < \varepsilon|b|^2/2$ for every $n \ge N_1$, and let $N := \max\{N_0, N_1\}$. For $n \ge N$, the last display and multiplication of $|b_n - b| < \varepsilon|b|^2/2$ by $2/|b|^2 > 0$ (Proposition 2.5.2 (a), (O4); Proposition 2.3.5 (c)) give $|1/b_n - 1/b| < (2/|b|^2)(\varepsilon|b|^2/2) = \varepsilon$. So $1/b_n \to 1/b$. Finally $a_n/b_n = a_n \cdot (1/b_n)$, and (c) gives $a_n/b_n \to a \cdot (1/b) = a/b$. This proves the theorem.

■

The answer to the prompt: the tolerance $\varepsilon/(2M)$ contains M, so M must exist before that tolerance is written down. And M must not depend on n: it has to bound $|a_n|$ for **every** $n \ge N$ at once, including indices not yet chosen. That is why it comes from Proposition 4.3.5, a single bound for the whole sequence, and not from "$|a_n|$ is some element of F". The $+1$ in $|b| + 1$ is there because b may be 0, and $\varepsilon/(2|b|)$ would then mean nothing.

Part (a) is the move we call **split** ε: to make a sum of k small terms smaller than ε, give each term the tolerance ε/k, and combine the finitely many thresholds by a maximum. Part (c) is **bound, then shrink**: replace the factor you cannot control by a bound fixed in advance, then make the other factor small.

Alternatively, (c) follows from the earlier tools without new estimates. By Proposition 4.3.3, $(b_n - b)$ and $(a_n - a)$ are null; $a_n(b_n - b)$ is null by Proposition 4.4.1, with the bound M; $b(a_n - a)$ is null by (b); so their sum $a_n b_n - ab$ is null by (a), and Proposition 4.3.3 again gives $a_n b_n \to ab$. The bound M is still fixed first: it is the hypothesis of Proposition 4.4.1.

In (d), the hypothesis "$b_n \ne 0$ for every n" is there only so that a_n/b_n is defined for every n. By Lemma 4.4.2, $b_n \ne 0$ holds eventually anyway once $b \ne 0$. If some early b_n vanish, the quotient is a sequence from some N_0 on, and Proposition 4.3.2 (b) applies the law to the shifted sequences (a_{n+N_0}) and (b_{n+N_0}).

Warning. Part (d) needs $b \neq 0$, even when every $b_n \neq 0$. In $\mathbb{Q}$, $a_n = 1/(n+1)$ and $b_n = 1/(n+1)^2$ are both null (the first by Example (First Limits) (b), the second by (c), since $b_n = a_n a_n$), with every $b_n \neq 0$, and $a_n/b_n = n+1$. Since $n+1 > n$, the sequence $(n+1)$ is not bounded (the non-example after Definition 4.3.4), and so has no limit (Proposition 4.3.5). The "law" $\lim(a_n/b_n) = 0/0$ names nothing.

Quick check. Find $\lim(3n^2 - n)/(n^2 + 5)$ in $\mathbb{Q}$, citing a law, or another result, for each step.

Solution. The limit is 3. Let $u_0 := 0$ and $u_n := 1/n$ for $n \geq 1$. The shifted sequence $(u_{n+1})_n = (1/(n+1))_n$ is null (Example (First Limits) (b); this is where Archimedes enters), so (u_n) is null by Proposition 4.3.2 (b). By Theorem 4.4.3: $3 - u_n \to 3 - 0 = 3$ (the constant 3 converges to 3 by Example (First Limits) (a); then (b) and (a)); $u_n^2 \to 0$ ((c)); $1 + 5u_n^2 \to 1$ ((b), (a)). Every $1 + 5u_n^2 \geq 1 > 0$: $u_n^2 \geq 0$ (Proposition 2.3.5 (e) if $u_n \neq 0$, and $u_0^2 = 0$), so $5u_n^2 \geq 0$ (Proposition 4.1.2 (d)), and $1 + 5u_n^2 \geq 1$ by Proposition 2.3.5 (b). Also the limit $1 \neq 0$, so (d) gives

$$c_n := \frac{3 - u_n}{1 + 5u_n^2} \to \frac{3}{1} = 3.$$

For $n \geq 1$, multiplying numerator and denominator by n^2 shows $c_n = (3n^2 - n)/(n^2 + 5)$. The two sequences agree from $N_0 = 1$ on (at $n = 0$ they differ: 3 versus 0), so Proposition 4.3.2 (a) gives the limit 3.

4.4.3 Limits and order

Inequalities between terms pass to the limits, provided we accept that a strict inequality may become an equality.

Proposition 4.4.4. Limits Preserve Weak Inequalities

Let $a_n \to a$ and $b_n \to b$ in F, and let $N_0 \in \mathbb{N}$.

(a) If $a_n \leq b_n$ for every $n \geq N_0$, then $a \leq b$.

(b) In particular, for $c \in F$: if $a_n \leq c$ for every $n \geq N_0$, then $a \leq c$; and if $a_n \geq c$ for every $n \geq N_0$, then $a \geq c$.

(c) *(Squeeze.)* Let (c_n) be a sequence in F. If $a = b$ and $a_n \leq c_n \leq b_n$ for every $n \geq N_0$, then $c_n \to a$.

Idea. For (a), suppose instead that $b < a$. The midpoint $(a + b)/2$ separates them, and from some index on a_n lies above it and b_n below it, contradicting $a_n \leq b_n$. The tolerance that achieves this is half the gap, $(a - b)/2$. For (c), from some index on a_n and b_n both lie in the window $(a - \varepsilon, a + \varepsilon)$, and c_n is trapped between them.

Proof.

(a) Suppose, for a contradiction, that $b < a$. Let

$$\varepsilon := \frac{a-b}{2},$$

which lies in $F_{>0}$ by Proposition 2.3.5 (a) and Proposition 4.1.2 (b). Choose N_1 (depending on ε) with $|a_n - a| < \varepsilon$ for every $n \geq N_1$, and N_2 (depending on ε) with $|b_n - b| < \varepsilon$ for every $n \geq N_2$. Let $n := \max\{N_0, N_1, N_2\}$. By Proposition 4.1.3 (d), at this n,

$$a_n > a - \varepsilon = \frac{a+b}{2} = b + \varepsilon > b_n.$$

This contradicts $a_n \leq b_n$, which holds because $n \geq N_0$. So $b < a$ is false, and $a \leq b$ by trichotomy (O1).

(b) A constant sequence c converges to c (Example (First Limits) (a)). Apply (a) to (a_n) and the constant sequence c for the first claim, and to the constant sequence c and (a_n) for the second.

(c) Let $\varepsilon \in F_{>0}$. Choose N_1 (depending on ε) with $|a_n - a| < \varepsilon$ for $n \geq N_1$ and N_2 (depending on ε) with $|b_n - a| < \varepsilon$ for $n \geq N_2$ (recall $b = a$), and let $N := \max\{N_0, N_1, N_2\}$. For $n \geq N$, Proposition 4.1.3 (d) gives

$$a - \varepsilon < a_n \leq c_n \leq b_n < a + \varepsilon,$$

so $|c_n - a| < \varepsilon$ by the same part (d). Hence $c_n \to a$.

■

The two midpoint equalities in (a) are field algebra: $a - (a-b)/2 = (2a - a + b)/2 = (a+b)/2$, and likewise $b + (a-b)/2 = (a+b)/2$. Chapter 5 uses the squeeze (c) repeatedly, mostly with $F = \mathbb{Q}$ to show that a sequence is null. In the squeeze (c), the middle sequence is not assumed to converge: its convergence is the conclusion.

Warning. Strict inequalities do not pass to the limit. In $\mathbb{Q}$, $1/(n+1) > 0$ for every n, and the limit is 0, not positive (Example (First Limits) (b)). From $a_n < b_n$ for every n, (a) gives only $a \leq b$.

4.4.4 Exercises

4.4.4.1 A. Check your understanding

Exercise. A1

These questions are about the proof of the product law, Theorem 4.4.3 (c).

(a) Why is M fixed before ε, and what does it depend on?

(b) Which hypothesis supplies N_1, which supplies N_2, and which supplies M?

(c) Why does the proof use $|b| + 1$ rather than $|b|$?

(d) State the idea of the proof in two sentences.

(e) Run the proof in $\mathbb{Q}$ on $a_n = 1 + 1/(n+1)$, $b_n = 2$ and $\varepsilon = 1/10$: give a valid M, the least N_1, the least N_2, and N.

Solution.

(a) The tolerance $\varepsilon/(2M)$ is built from M, so M must be available first. It depends only on the sequence (a_n): it bounds $|a_n|$ for every n, independently of ε and of n. A "bound" depending on n would be useless, since the tolerance for $|b_n - b|$ must be named before n is.

(b) N_1 comes from $b_n \to b$, applied with the tolerance $\varepsilon/(2M)$; N_2 from $a_n \to a$, with the tolerance $\varepsilon/(2(|b|+1))$; M also from $a_n \to a$, through Proposition 4.3.5.

(c) b may be 0, and then $\varepsilon/(2|b|)$ is undefined. Since $|b| + 1 \geq 1 > 0$, the tolerance $\varepsilon/(2(|b|+1))$ always exists, and $|b| \leq |b| + 1$ keeps the estimate valid.

(d) Write $a_n b_n - ab = a_n(b_n - b) + b(a_n - a)$, so that each term has a factor tending to 0. The other factor of the first term is not small, but it is bounded by a constant M fixed in advance, so making $|b_n - b| < \varepsilon/(2M)$ and $|a_n - a| < \varepsilon/(2(|b|+1))$ makes each term smaller than $\varepsilon/2$.

(e) Here $a = 1$ and $b = 2$. Since $0 < 1/(n+1) \leq 1$, $|a_n| \leq 2$ for every n, so $M = 2$ is valid. Then $\varepsilon/(2M) = 1/40$, and $|b_n - b| = 0 < 1/40$ for every n: the least N_1 is 0. Next $\varepsilon/(2(|b|+1)) = (1/10)/6 = 1/60$, and $|a_n - a| = 1/(n+1) < 1/60$ if and only if $n + 1 > 60$, that is, $n \geq 60$: the least N_2 is 60. So $N = \max\{0, 60\} = 60$. Check at $n = 60$: $a_{60}b_{60} - ab = 2(1 + 1/61) - 2 = 2/61 < 1/10$.

4.4.4.2 B. Practice

Exercise. B1

Find the following limits in $\mathbb{Q}$ by the limit laws, naming the result used at each step.

(a) $\dfrac{n+2}{3n+1}$.

(b) $\dfrac{n^2+1}{(n+1)^2}$.

(c) $\Big(\dfrac{1}{n+1} - 2\Big)\Big(3 + \dfrac{(-1)^n}{n+1}\Big)$.

Hint: write each term through $u_n := 1/(n+1)$.

Solution. Let $u_n := 1/(n+1)$, so $u_n \to 0$ (Example (First Limits) (b)), and constant sequences converge to their value (part (a) of that example). All laws cited are parts of Theorem 4.4.3.

(a) With $m := n + 1 \geq 1$: $(n+2)/(3n+1) = (m+1)/(3m-2) = (1+u_n)/(3-2u_n)$, dividing numerator and denominator by m. Now $1 + u_n \to 1$ by (a), and $3 - 2u_n \to 3$ by (b) and (a). Since $0 < u_n \leq 1$, $3 - 2u_n \geq 1 > 0$ for every n, and the limit $3 \neq 0$, so (d) gives the limit $1/3$.

(b) With $m = n + 1$: $(n^2+1)/(n+1)^2 = ((m-1)^2+1)/m^2 = 1 - 2u_n + 2u_n^2$. By (c), $u_n^2 \to 0$; by (b) and (a), $1 - 2u_n + 2u_n^2 \to 1 - 0 + 0 = 1$. No quotient law is needed.

(c) $u_n - 2 \to -2$ by (b) and (a). The sequence $(-1)^n u_n$ is null by Proposition 4.4.1, since u_n is null and $|(-1)^n| = 1$ (Example (First Limits) (d)); so $3 + (-1)^n u_n \to 3$ by (a). By (c), the product tends to $(-2) \cdot 3 = -6$.

In each part Archimedes enters once, through $u_n \to 0$; every other step is a law valid in any ordered field.

Exercise. B2: Put the proof in order

Here are lines of the proof of Proposition 4.4.4 (a), shuffled, together with two lines that do not belong. Recall the hypotheses: $a_n \to a$, $b_n \to b$, and $a_n \le b_n$ for every $n \ge N_0$. Write the seven lines of a correct proof in a correct order, say which other orders are also admissible, and say why each of the other two lines fails.

(i) Let $n := \max\{N_0, N_1, N_2\}$.
(ii) Let $\varepsilon := a - b$, which lies in $F_{>0}$.
(iii) Choose N_2 with $|b_n - b| < \varepsilon$ for every $n \ge N_2$.
(iv) Suppose, for a contradiction, that $b < a$.
(v) This contradicts $a_n \le b_n$; so $b < a$ is false, and $a \le b$ by trichotomy.
(vi) Let $\varepsilon := (a - b)/2$, which lies in $F_{>0}$ by Proposition 4.1.2.
(vii) Let $n := N_1$.
(viii) Then $a_n > a - \varepsilon = (a + b)/2 = b + \varepsilon > b_n$.
(ix) Choose N_1 with $|a_n - a| < \varepsilon$ for every $n \ge N_1$.

Solution. An admissible order is

$$\text{(iv)},\ \text{(vi)},\ \text{(ix)},\ \text{(iii)},\ \text{(i)},\ \text{(viii)},\ \text{(v)}.$$

The only other admissible order exchanges (ix) and (iii), which use only ε and one hypothesis each. Nothing else moves: (vi) uses the assumption $b < a$ of (iv) to make ε positive; (ix) and (iii) use ε; (i) uses N_1 and N_2; (viii) uses the two estimates at this n, and needs $n \ge N_1$ and $n \ge N_2$; (v) uses (viii) and $n \ge N_0$, and closes the assumption of (iv).

Line (ii) is an ε of the wrong size. It is positive, but with it (viii) becomes $a_n > a - (a - b) = b$ and $b_n < b + (a - b) = a$, and "$a_n > b$ and $b_n < a$" is no contradiction with $a_n \le b_n$: for instance $a_n = b_n = (a + b)/2$ satisfies all three at once. The two windows of half-width $a - b$ around a and b overlap; half-width $(a - b)/2$ makes them touch only at the midpoint, which neither open window contains.

Line (vii) chooses an index that depends on the wrong thing. With $n := N_1$, the estimate for a_n holds, but nothing makes $n \ge N_2$ or $n \ge N_0$: the estimate for b_n and the hypothesis $a_n \le b_n$ may both fail at this n. The index must lie beyond **all three** thresholds, which the maximum in (i) ensures.

Exercise. B3: Close to a convergent sequence

Let $a_n \to a$ in F, and let (b_n) be a sequence in F such that $(b_n - a_n)$ is null. Prove that $b_n \to a$. Here is the start of a proof.

> Let $\varepsilon \in F_{>0}$. Then $\varepsilon/2 \in F_{>0}$. Since $a_n \to a$, choose N_1 with $|a_n - a| < \varepsilon/2$ for every $n \ge N_1$. Since $(b_n - a_n)$ is null, choose N_2 with $|b_n - a_n| < \varepsilon/2$ for every

$n \geq N_2$. Let $N \coloneqq \max\{N_1, N_2\}$.

...

Name the move, then complete it.

Solution. The move is **split** ε, with two terms: each gets $\varepsilon/2$, and the two thresholds are combined by a maximum.

Completion. Let $n \geq N$. Then $n \geq N_1$ and $n \geq N_2$, and the three-point triangle inequality (Proposition 4.1.3 (e)) gives

$$|b_n - a| \leq |b_n - a_n| + |a_n - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon \in F_{>0}$ was arbitrary, $b_n \to a$. (The same result also follows from the laws: $b_n = (b_n - a_n) + a_n \to 0 + a$ by Theorem 4.4.3 (a). The exercise is the law's proof, run once more.)

4.4.4.3 C. Going deeper

Exercise. C1: Exit check: the product law

Prove, from Definition 4.3.2, that if $a_n \to a$ and $b_n \to b$ in an ordered field F, then $a_n b_n \to ab$. You may use that convergent sequences are bounded (Proposition 4.3.5), the rules for the absolute value (Proposition 4.1.3), Proposition 4.3.3 and Proposition 4.4.1, but not Theorem 4.4.3.

Solution. By Proposition 4.3.5 there is $M \in F_{>0}$ with $|a_n| \leq M$ for every $n \in \mathbb{N}$. It depends only on (a_n), and we fix it before any tolerance is named.

Let $\varepsilon \in F_{>0}$. Since $|b| + 1 \geq 1 > 0$, both $\varepsilon/(2M)$ and $\varepsilon/(2(|b| + 1))$ lie in $F_{>0}$. By $b_n \to b$ there is N_1 with $|b_n - b| < \varepsilon/(2M)$ for every $n \geq N_1$; by $a_n \to a$ there is N_2 with $|a_n - a| < \varepsilon/(2(|b| + 1))$ for every $n \geq N_2$. These depend only on ε, M and b. Let $N \coloneqq \max\{N_1, N_2\}$.

Let $n \geq N$. Since

$$a_n b_n - ab = a_n(b_n - b) + b(a_n - a),$$

the triangle inequality and $|xy| = |x||y|$ give

$$\begin{aligned} |a_n b_n - ab| &\leq |a_n||b_n - b| + |b||a_n - a| \\ &\leq M|b_n - b| + (|b| + 1)|a_n - a| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where the second line uses $0 \leq |a_n| \leq M$ and $0 \leq |b| \leq |b| + 1$ (Proposition 4.1.2 (d)), and the third multiplies the two estimates by $M > 0$ and $|b| + 1 > 0$. So for every $\varepsilon \in F_{>0}$ there is $N \in \mathbb{N}$ with $|a_n b_n - ab| < \varepsilon$ for every $n \geq N$: $a_n b_n \to ab$.

Marking scheme. (10 points) - 2: fixes a bound $M \in F_{>0}$ for $(|a_n|)$, justified (convergent sequences are bounded), before ε. - 2: writes $a_n b_n - ab$ as a sum of two terms, each with a factor tending to 0. - 3: chooses N_1, N_2 depending only on ε, M and b, with the case $b = 0$ handled, and N as their maximum. - 2: carries out the estimate with the triangle

inequality and the product rule for $|\cdot|$. - 1: concludes with the quantifiers in the order $\forall\varepsilon\,\exists N\,\forall n \geq N$. Fatal errors (0 for the step): a threshold or tolerance that depends on n; the bound "$|a_n| \leq |a| + 1$ for every n" without the finite part for $n < N$; dividing by $|b|$ without $b \neq 0$. A route through null sequences ($(b_n - b)$ and $(a_n - a)$ are null by Proposition 4.3.3; $a_n(b_n - b)$ and $b(a_n - a)$ are null by Proposition 4.4.1, with bounds M and $|b| + 1$; the sum of two null sequences is null, proved from Definition 4.3.2 by splitting ε) earns the same points: decomposition 2, the bound 2, the two null products and the proved sum 5, conclusion 1. Citing Theorem 4.4.3 for the sum or for $b(a_n - a)$ gets 0 for that step. A three-term route ($a_nb_n - ab = (a_n - a)(b_n - b) + a(b_n - b) + b(a_n - a)$, with a tolerance such as $\delta = \min\{1, \varepsilon/(1 + |a| + |b|)\}$ chosen from ε, a and b only) earns the same points: decomposition 2; the choice of δ, independent of n and with $\delta \leq 1$ controlling the product term (in place of the bound M), 2; a threshold N for both sequences at tolerance δ (the case $a = 0$ or $b = 0$ covered by the 1 in the denominator) 3; the estimate 2; conclusion 1.

Exercise. C2: Non-negative terms and reciprocals

Let $a_n \to a$ in F, with $a_n \geq 0$ for every n.

(a) Prove that $a \geq 0$.

(b) Suppose moreover $a > 0$. Prove that there is $N_0 \in \mathbb{N}$ with $a_n > 0$ for every $n \geq N_0$, and that the sequence $(1/a_{n+N_0})_{n\in\mathbb{N}}$ converges to $1/a$.

(c) Give an example in $\mathbb{Q}$ where $a_n > 0$ for every n and $(1/a_n)$ does not converge.

Solution (Hint 1). For (a), compare with a constant sequence. For (b), which result keeps terms away from a non-zero limit?

Solution (Hint 2). For (b), start: "By Lemma 4.4.2, since $a > 0$, there is N_0 with $a_n > a/2$ for every $n \geq N_0$."

Solution.

(a) Since $a_n \geq 0$ for every $n \geq 0$, Proposition 4.4.4 (b) with $c = 0$ gives $a \geq 0$.

(b) Since $a > 0$, Lemma 4.4.2 gives N_0 with $a_n > a/2 > 0$ for every $n \geq N_0$. The shifted sequence $(a_{n+N_0})_n$ converges to a by Proposition 4.3.2 (b), and each of its terms is non-zero. By Theorem 4.4.3 (d), applied to the constant sequence 1 and to $(a_{n+N_0})_n$, whose limit a is non-zero, $1/a_{n+N_0} \to 1/a$.

(c) $a_n = 1/(n+1)$: every term is positive, and $1/a_n = n + 1$ is not bounded in $\mathbb{Q}$: for any $M \in \mathbb{Q}_{>0}$, the non-example after Definition 4.3.4 gives n with $M < n < n+1 = |1/a_n|$. So it does not converge (Proposition 4.3.5). The hypothesis $a > 0$ of (b) fails here: $a = 0$.

Marking scheme. (10 points) - 3: (a), by comparison with the constant sequence 0 (or by contradiction with $\varepsilon = -a/2$). - 3: (b), an index N_0 beyond which a_n is bounded away from 0, from $a > 0$. - 2: (b), $1/a_{n+N_0} \to 1/a$, either by the quotient law (Theorem 4.4.3 (d)) applied to the shift, with its convergence to a and $a \neq 0$ checked, or directly from

$|1/a_{n+N_0} - 1/a| \le (2/a^2)|a_{n+N_0} - a|$, which uses $a_{n+N_0} > a/2$. - 2: (c), an example with the failure of convergence justified. Fatal errors (0 for the step): concluding $a > 0$ in (a) from $a_n > 0$; applying the quotient law without checking that the limit is non-zero.

4.5 Cauchy Sequences

Every convergence proof so far began with the limit in hand: to show $a_n \to a$ we estimated $|a_n - a|$, and the limit laws of Section 4.4 compute a new limit only from limits already known. That is no help for the sequence of Chapter 2 that closes in on the gap at $\sqrt{2}$ (Example (A Rational Sequence Aimed at the Gap)), whose terms $2, 3/2, 17/12, 577/408, \ldots$ visibly settle down while no rational is available to be their limit. This section names a property that can be checked from the terms alone: eventually they are all close **to each other**. We prove the facts about it that Chapter 5 needs: convergent sequences have it, sequences with it are bounded and closed under sums and products, and, when such a sequence does not tend to 0, its terms eventually keep one sign and stay away from 0.

Throughout, F is an ordered field (Definition 2.3.2), sequences are indexed by $\mathbb{N}$ (Definition 4.3.1), and ε ranges over $F_{>0}$. We use the absolute-value rules of Proposition 4.1.3 and the halving rules of Proposition 4.1.2 without restating them, and write 2 for $1 + 1$ in F.

4.5.1 Terms that agree with one another

A limit relates a sequence to a point of F; boundedness (Definition 4.3.4) is a property of the sequence alone. We want a second property of the sequence alone, saying "the terms settle down" without saying where.

In $\mathbb{Q}$ the Newton terms satisfy $17/12 - 577/408 = 1/408$, and the later terms are closer still to one another (Section 4.6 proves it). The way to say "settle down" without a target is to compare terms with terms: however small a tolerance is named, from some index on, any two terms differ by less than it.

The terms eventually agree with one another to within any tolerance.

Definition 4.5.1. Cauchy Sequence

A sequence $(a_n)_{n\in\mathbb{N}}$ in F is a **Cauchy sequence in** F (or **Cauchy**) if **for every** $\varepsilon \in F_{>0}$ **there exists** $N \in \mathbb{N}$ such that $|a_m - a_n| < \varepsilon$ **for all** $m, n \in \mathbb{N}$ with $m \ge N$ and $n \ge N$. In symbols:

$$\forall \varepsilon \in F_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall m \in \mathbb{N} \;\; \forall n \in \mathbb{N}$$
$$\big((m \ge N \wedge n \ge N) \Rightarrow |a_m - a_n| < \varepsilon\big).$$

In words: the first quantifier says the tolerance is arbitrary and is named first; the second says a threshold exists, and since it comes after ε, it may depend on ε and on nothing else; the last two say that **every** pair of indices beyond the threshold, near or far apart, gives terms within ε of each other. Compared with Definition 4.3.2, the point a has been replaced by a second index m.

The tolerances are the elements of $F_{>0}$, so the field is part of the definition. For $F = \mathbb{Q}$, the tolerances are rational, and "Cauchy in $\mathbb{Q}$" is a statement about rationals only, as Chapter 5 needs. Since $|a_m - a_n| = |a_n - a_m|$ (Proposition 4.1.3 (b)), the roles of m and n are symmetric,

and there is nothing else to check for the definition to make sense.

Negation. Moving ¬ inward quantifier by quantifier (Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2) and reading $\neg(x < \varepsilon)$ as $x \geq \varepsilon$ by (O1), "(a_n) is not Cauchy" reads

$$\exists \varepsilon \in F_{>0} \;\; \forall N \in \mathbb{N} \;\; \exists m, n \in \mathbb{N} \;\; (m \geq N \wedge n \geq N \wedge |a_m - a_n| \geq \varepsilon).$$

One tolerance is beaten beyond every threshold.

Examples.

A constant sequence. Let $a_n = c$ for every n. Then $|a_m - a_n| = |0| = 0 < \varepsilon$ for all m, n, so $N = 0$ works for every ε. This degenerate case matters: Chapter 5 represents each rational c by its constant sequence.

The reciprocals in $\mathbb{Q}$. Let $a_n = 1/(n+1)$ in $\mathbb{Q}$. Given $\varepsilon \in \mathbb{Q}_{>0}$, Corollary 2.5.4 gives $N \in \mathbb{Z}_{>0}$ with $1/N < \varepsilon$. Let $m, n \geq N$. Then $m + 1 > N$ and $n + 1 > N$, so $0 < 1/(m+1) < 1/N$ and $0 < 1/(n+1) < 1/N$ (Proposition 2.5.2). Two elements x, y of $(0, c)$ satisfy $x - y < x < c$ and $y - x < y < c$, so $|x - y| < c$ by Proposition 4.1.3 (d). Hence $|a_m - a_n| < 1/N < \varepsilon$. The threshold depends on ε only: for $\varepsilon = 1/100$, $N = 101$ works.

Sequences with a limit. By the first result of the next subsection, the convergent sequences of Section 4.3 (Example (First Limits)) are all examples. The definition earns its keep on sequences with no limit available, which Section 4.6 supplies.

Non-example by a minimal change. Change the constant sequence 1 at every odd index: let $a_n = 1$ for even n and $a_n = 0$ for odd n. The sequence is still bounded, by 1, and every two terms still differ by at most 1. What fails is "for every ε": take $\varepsilon = 1$. Whatever N is offered, the indices N and $N + 1$ are both $\geq N$, one even and one odd, and their terms differ by $1 \not< 1$. So the negation above holds with $\varepsilon = 1$, $m = N$, $n = N + 1$.

Why this definition. Each clause is there for a reason.

- *Every pair, not only neighbors.* Asking only that $|a_{n+1} - a_n| < \varepsilon$ for $n \geq N$ is too weak: the example at the end of this section has consecutive differences tending to 0 and is not Cauchy.
- *The threshold after the tolerance.* With the quantifiers swapped, "$\exists N \; \forall \varepsilon \; \forall m, n \geq N$", the condition forces $|a_m - a_n| < \varepsilon$ for every ε, hence $a_m = a_n$ for all $m, n \geq N$ (the argument of Theorem 4.3.4): only sequences that are eventually constant would qualify.
- *Tolerances from the field.* Changing the set of tolerances changes the notion: in the non-Archimedean field $\mathbb{Q}(t)$ of Section 4.2 there are more tolerances than in $\mathbb{Q}$, among them $1/t$, and Exercise C2 of Section 4.6 shows a sequence of rationals that passes every rational test and fails the test $1/t$.

Warning. The definition asks for **one** N that serves **every** pair $m, n \geq N$. A proof that checks only some pairs, say each n against $n + 2$, proves nothing: in the non-example above, $a_{n+2} = a_n$ for every n, so those pairs pass every test, and the sequence still fails.

Payoff. The tail-change principle extends at once. If $a_n = b_n$ for every $n \geq N_0$ and (a_n) is Cauchy, then for $\varepsilon \in F_{>0}$ and its threshold N for (a_n), every $m, n \geq \max\{N, N_0\}$ give $|b_m - b_n| = |a_m - a_n| < \varepsilon$. So (b_n) is Cauchy, with thresholds $\max\{N, N_0\}$, as in Proposition 4.3.2.

4.5.2 Convergent sequences are Cauchy, and Cauchy sequences are bounded

The first result says the new property is no stronger than convergence: it is a necessary condition, one we can test without knowing the limit.

Proposition 4.5.1. Convergent Sequences Are Cauchy

If $a_n \to a$ in F, then (a_n) is a Cauchy sequence in F.

Idea. Two terms that are each within $\varepsilon/2$ of a are within ε of each other: insert the point a between them with the triangle inequality, and spend half the tolerance on each side.

Proof. Let $\varepsilon \in F_{>0}$. Then $\varepsilon/2 \in F_{>0}$ by Proposition 4.1.2 (b), so, since $a_n \to a$, there is $N \in \mathbb{N}$ with $|a_n - a| < \varepsilon/2$ for every $n \geq N$; N depends only on ε. Let $m, n \geq N$. By the three-point triangle inequality (Proposition 4.1.3 (e)) with a in the middle, and by $|a - a_n| = |a_n - a|$ (Proposition 4.1.3 (b)),

$$|a_m - a_n| \leq |a_m - a| + |a - a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

where the strict inequality adds two strict inequalities (Proposition 2.3.5 (b)) and the equality is Proposition 4.1.2 (b). This proves that (a_n) is Cauchy.

■

The converse is the question of the chapter. In $\mathbb{Q}$ it fails (Section 4.6); Chapter 5 builds a field in which it holds.

Quick check. Is the sequence $\left((-1)^n/(n+1)\right)_{n\in\mathbb{N}}$ Cauchy in $\mathbb{Q}$? Answer without estimating $|a_m - a_n|$.

Solution. Yes. The sequence $(1/(n+1))$ is null in $\mathbb{Q}$ (Example (First Limits)), and $((-1)^n)$ is bounded by 1, since $|(-1)^n| = 1$ (by induction, with $|xy| = |x||y|$, Proposition 4.1.3 (b)). So the product is null by Proposition 4.4.1. A null sequence converges, to 0, so it is Cauchy by Proposition 4.5.1.

The second result is the Cauchy version of Proposition 4.3.5. There the tail was trapped near the limit a; here no limit is available, so we trap the tail near one of its own terms.

Proposition 4.5.2. Cauchy Sequences Are Bounded

Every Cauchy sequence in F is bounded.

Idea. Apply the definition once, with the tolerance 1. It gives an index N beyond which every term is within 1 of the term a_N, a base point we can name. The finitely many terms before N are handled by a finite maximum, as for convergent sequences.

Proof. Let (a_n) be Cauchy in F. Since $1 \in F_{>0}$ (Proposition 2.3.5 (e)), there is $N \in \mathbb{N}$ with $|a_m - a_n| < 1$ for all $m, n \geq N$. For $n \geq N$, taking $m = N$ and using the triangle inequality (Proposition 4.1.3 (e), (b)),

$$|a_n| = |(a_n - a_N) + a_N| \leq |a_n - a_N| + |a_N| < 1 + |a_N|.$$

Let

$$M \coloneqq \max\{|a_0|, \ldots, |a_{N-1}|, |a_N| + 1\},$$

a finite maximum in the totally ordered set F (Sets, Relations, and Cardinality, Proposition 2.3.8, whose proof works in any totally ordered set; for $N = 0$ the set is $\{|a_0| + 1\}$ and $M = |a_0| + 1$). Then $M \geq |a_N| + 1 \geq 1 > 0$, so $M \in F_{>0}$. For $n < N$, $|a_n| \leq M$ because $|a_n|$ is in the set; for $n \geq N$, $|a_n| < |a_N| + 1 \leq M$. So $|a_n| \leq M$ for every $n \in \mathbb{N}$, and (a_n) is bounded (Definition 4.3.4).

■

The bound M depends on the whole sequence, not on any tolerance. That is what makes it usable in the next proof: it can be fixed before ε is named.

4.5.3 Sums and products of Cauchy sequences

Chapter 5 will add and multiply real numbers by adding and multiplying Cauchy sequences of rationals term by term. For that, sums and products of Cauchy sequences must again be Cauchy. The limit laws (Theorem 4.4.3) do not apply, since they need limits; the proofs below copy their estimates with a second index in place of the limit.

Theorem 4.5.3. Sums and Products of Cauchy Sequences

Let (a_n) and (b_n) be Cauchy sequences in F, and let $c \in F$.

(a) Every constant sequence and every null sequence in F is Cauchy.

(b) The sequences $(a_n + b_n)$, $(c\,a_n)$, $(-a_n)$ and $(a_n - b_n)$ are Cauchy.

(c) The sequence $(a_n b_n)$ is Cauchy.

Idea.

(a) is Proposition 4.5.1. For (b), split ε: give each of (a_n) and (b_n) the tolerance $\varepsilon/2$ and take the larger threshold. For (c), the difference $a_n b_n - a_m b_m$ splits as $a_n(b_n - b_m) + b_m(a_n - a_m)$. The factors a_n and b_m do not get small, but they are bounded (Proposition 4.5.2). So **bound, then shrink**: replace each factor we cannot control by a bound fixed in advance, then make the other factor small enough to absorb the bound.

Proof.

(a) A constant sequence (c) converges to c, and a null sequence converges to 0 (Example (First Limits), Definition 4.3.3). Both are Cauchy by Proposition 4.5.1.

(b) *Sums.* Let $\varepsilon \in F_{>0}$, so $\varepsilon/2 \in F_{>0}$ (Proposition 4.1.2 (b)). Since (a_n) is Cauchy, there is N_1 with $|a_m - a_n| < \varepsilon/2$ for all $m, n \geq N_1$; since (b_n) is Cauchy, there is N_2 with $|b_m - b_n| < \varepsilon/2$ for all $m, n \geq N_2$. Let $N \coloneqq \max\{N_1, N_2\}$, which depends only on ε. For $m, n \geq N$, we have $(a_m + b_m) - (a_n + b_n) = (a_m - a_n) + (b_m - b_n)$ (Proposition 2.3.2), so the triangle inequality (Proposition 4.1.3 (e)) gives

$$\begin{aligned} |(a_m + b_m) - (a_n + b_n)| &\leq |a_m - a_n| + |b_m - b_n| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where the strict step adds two strict inequalities (Proposition 2.3.5 (b)) and the equality is Proposition 4.1.2 (b).
Multiples. If $c = 0$, then $(c\,a_n)$ is the constant sequence 0 (Proposition 2.3.2 (b)), Cauchy by (a). If $c \neq 0$, then $|c| > 0$ (Proposition 4.1.3 (a)), so $\varepsilon/|c| \in F_{>0}$ (Proposition 2.5.2, (O4)); choose N (depending on ε and c) with $|a_m - a_n| < \varepsilon/|c|$ for all $m, n \geq N$. For such m, n,

$$|c\,a_m - c\,a_n| = |c(a_m - a_n)| = |c|\,|a_m - a_n| < |c| \cdot \frac{\varepsilon}{|c|} = \varepsilon,$$

by distributivity (F9) and Proposition 2.3.2 (c), then Proposition 4.1.3 (b) and Proposition 2.3.5 (c); the last equality $|c| \cdot \varepsilon |c|^{-1} = \varepsilon$ is (F5)–(F8).
Negatives and differences. Since $-a_n = (-1)a_n$ (Proposition 2.3.2 (c)), $(-a_n)$ is the multiple with $c = -1$. Then $a_n - b_n = a_n + (-b_n)$ is the sum of two Cauchy sequences.

(c) By Proposition 4.5.2, fix $A, B \in F_{>0}$ with $|a_n| \leq A$ and $|b_n| \leq B$ for every $n \in \mathbb{N}$; they depend on the two sequences only. Now let $\varepsilon \in F_{>0}$. Then $\varepsilon/(2A)$ and $\varepsilon/(2B)$ lie in $F_{>0}$ (Proposition 4.1.2 (b), (O4), Proposition 2.5.2 (a)), so there are N_1 with $|b_n - b_m| < \varepsilon/(2A)$ for all $m, n \geq N_1$, and N_2 with $|a_n - a_m| < \varepsilon/(2B)$ for all $m, n \geq N_2$. Let $N \coloneqq \max\{N_1, N_2\}$, depending on ε, A, B, and let $m, n \geq N$. Expanding the right side shows

$$a_n b_n - a_m b_m = a_n(b_n - b_m) + b_m(a_n - a_m).$$

By the triangle inequality and $|xy| = |x||y|$ (Proposition 4.1.3 (e), (b)), then Proposition 4.1.2 (d) applied to $0 \leq |a_n| \leq A$ and to $0 \leq |b_m| \leq B$,

$$\begin{aligned} |a_n b_n - a_m b_m| &\leq |a_n|\,|b_n - b_m| + |b_m|\,|a_n - a_m| \\ &\leq A|b_n - b_m| + B|a_n - a_m| \\ &< A \cdot \frac{\varepsilon}{2A} + B \cdot \frac{\varepsilon}{2B} = \varepsilon, \end{aligned}$$

where the strict step multiplies by $A > 0$ and $B > 0$ (Proposition 2.3.5 (c)) and adds (Proposition 2.3.5 (b)); the equality uses $A \cdot \varepsilon/(2A) = \varepsilon/2$ (Lemma 4.1.1 (a), (F5)–(F8)), likewise for B, and Proposition 4.1.2 (b). So $(a_n b_n)$ is Cauchy.

■

The order of choices in (c) is the whole proof: A and B are chosen before ε, from the sequences alone, so that $\varepsilon/(2A)$ is one fixed tolerance. Exercise C3 shows what goes wrong when the bound is skipped.

Remark. In Chapter 5, $\mathbb{R}$ is a set of classes of Cauchy sequences in $\mathbb{Q}$, and the sum and product of two classes are the classes of the termwise sum and product. Clauses (b) and (c) with $F = \mathbb{Q}$ are exactly what makes these termwise operations stay inside the set of Cauchy sequences.

4.5.4 The sign of a Cauchy sequence that is not null

A null sequence is the Cauchy analogue of 0. Chapter 5 must decide, for each Cauchy sequence of rationals that is not null, whether it represents a positive or a negative number, and it needs an inverse for it. Both rest on one fact: such a sequence eventually keeps one sign, at a definite distance from 0.

Theorem 4.5.4. The Sign of a Non-Null Cauchy Sequence

Let (a_n) be a Cauchy sequence in F that is not null. Then there are $c \in F_{>0}$ and $N \in \mathbb{N}$ such that either

$$a_n \geq c \text{ for every } n \geq N, \qquad \text{or} \qquad a_n \leq -c \text{ for every } n \geq N.$$

The two alternatives do not both hold.

Try it first (ten minutes): write "(a_n) is not null" in symbols. Which ε does it hand you, and what does the Cauchy condition do with it?

Idea. "Not null" is the negation of Definition 4.3.3:

$$\exists \varepsilon \in F_{>0} \;\; \forall N \in \mathbb{N} \;\; \exists n \in \mathbb{N} \;\; (n \geq N \wedge |a_n| \geq \varepsilon).$$

The obvious attempt reads this as "$|a_n| \geq \varepsilon$ for all large n", and that is not what it says: it promises, beyond each threshold, **one** term at least ε from 0, and between such terms others might come near 0 or change sign. The Cauchy condition is what forbids that wandering. So **fix the bad** ε and keep it for the rest of the proof. Ask the Cauchy condition for a threshold N_0 with tolerance $\varepsilon/2$, and only then ask the negation for one term a_{n_0}, beyond N_0, with $|a_{n_0}| \geq \varepsilon$. Every later term is within $\varepsilon/2$ of a_{n_0}, so it has the sign of a_{n_0} and size at least $\varepsilon/2$. *Before reading on: where is $n_0 \geq N_0$ used?*

Proof. Since (a_n) is not null, the negation of Definition 4.3.3 holds. Moving $\neg$ inward by Logic and Proof, Proposition 6.2.3 (a), (b), Logic and Proof, Proposition 6.2.1 (a) and Logic and Proof, Proposition 6.2.2, and reading $\neg(|a_n - 0| < \varepsilon)$ as $|a_n| \geq \varepsilon$ by (O1), it says: there is $\varepsilon \in F_{>0}$ such that for every $N \in \mathbb{N}$ some $n \geq N$ has $|a_n| \geq \varepsilon$. Fix such an ε.

By Proposition 4.1.2 (b), $\varepsilon/2 \in F_{>0}$. Since (a_n) is Cauchy, there is $N_0 \in \mathbb{N}$ (depending only on the sequence, through ε) with

$$|a_m - a_n| < \frac{\varepsilon}{2} \quad \text{for all } m, n \geq N_0.$$

Applying the negation with $N := N_0$ gives $n_0 \in \mathbb{N}$ with $n_0 \geq N_0$ and $|a_{n_0}| \geq \varepsilon$. By Definition 4.1.1, either $a_{n_0} \geq 0$ and $a_{n_0} = |a_{n_0}| \geq \varepsilon$, or $a_{n_0} < 0$ and $-a_{n_0} \geq \varepsilon$, that is, $a_{n_0} \leq -\varepsilon$ (Proposition 2.3.5 (a)).

Let $n \geq N_0$. Since both n and n_0 are $\geq N_0$, $|a_n - a_{n_0}| < \varepsilon/2$, that is, $-\varepsilon/2 < a_n - a_{n_0} < \varepsilon/2$ (Proposition 4.1.3 (d)).

Case 1: $a_{n_0} \geq \varepsilon$. Adding a_{n_0} to the left inequality (Proposition 2.3.5 (b)), $a_n > a_{n_0} - \varepsilon/2 \geq \varepsilon - \varepsilon/2 = \varepsilon/2$, where the last equality is Proposition 4.1.2 (b).

Case 2: $a_{n_0} \leq -\varepsilon$. Adding a_{n_0} to the right inequality, $a_n < a_{n_0} + \varepsilon/2 \leq -\varepsilon + \varepsilon/2 = -\varepsilon/2$.

Let $c := \varepsilon/2$ and $N := N_0$; both depend only on (a_n). In Case 1, $a_n \geq c$ for every $n \geq N$; in Case 2, $a_n \leq -c$ for every $n \geq N$.

Finally, both alternatives cannot hold, even with different constants and thresholds: if $a_n \geq c$ for $n \geq N$ and $a_n \leq -c'$ for $n \geq N'$, with $c, c' \in F_{>0}$, then $n := \max\{N, N'\}$ gives $0 < c \leq a_n \leq -c' < 0$, against (O1). This proves the theorem.

■

The answer to the prompt: $n_0 \geq N_0$ is used once, to apply the Cauchy estimate to the pair n, n_0. Had we taken n_0 from the negation first and N_0 afterwards, n_0 might lie below N_0, and the term a_{n_0} would say nothing about the tail.

Both hypotheses are needed; the witnesses are in $\mathbb{Q}$.

- *Not null.* The sequence $(1/(n+1))$ is Cauchy (the examples after Definition 4.5.1) and null. For any $c \in \mathbb{Q}_{>0}$ and any N, Corollary 2.5.4 gives an index $n \geq N$ with $0 < 1/(n+1) < c$, so neither alternative holds.
- *Cauchy.* The sequence $((-1)^n)$ is not null, since $|(-1)^n| = 1 \geq 1$ for every n. Its terms alternate between 1 and -1, so neither alternative holds for any $c > 0$ and N: a_N and a_{N+1} have opposite signs.

4.5.5 Sign stability and reciprocals

Two smaller results complete the toolkit: the sign found by Theorem 4.5.4 survives a null change of the sequence, and a sequence that stays away from 0 can be inverted.

Proposition 4.5.5. Sign Stability under Null Differences

Let (a_n) and (b_n) be sequences in F such that $(a_n - b_n)$ is null. If $c \in F_{>0}$, $N \in \mathbb{N}$, and $a_n \geq c$ for every $n \geq N$, then there is $N' \in \mathbb{N}$ with $b_n \geq c/2$ for every $n \geq N'$.

Proof. Since $c/2 \in F_{>0}$ (Proposition 4.1.2 (b)) and $(a_n - b_n)$ is null, there is N_1 (depending on c) with $|a_n - b_n| < c/2$ for every $n \geq N_1$. Let $N' := \max\{N, N_1\}$, which depends on c and N, and $n \geq N'$. Then $a_n - b_n < c/2$ (Proposition 4.1.3 (c)), so

$$b_n > a_n - \frac{c}{2} \geq c - \frac{c}{2} = \frac{c}{2}$$

by Proposition 2.3.5 (a), (b) and Proposition 4.1.2 (b). In particular $b_n \geq c/2$.

■

Neither sequence needs to be Cauchy here. In Chapter 5, two Cauchy sequences with null difference represent the same real number, and this proposition is what shows that "represented by a sequence eventually $\geq c$ for some $c > 0$" does not depend on the representative chosen.

Proposition 4.5.6. Reciprocals of Cauchy Sequences

Let (a_n) be a Cauchy sequence in F, and let $c \in F_{>0}$ and $N \in \mathbb{N}$ be such that $|a_n| \geq c$ for every $n \geq N$. Define

$$b_n := \begin{cases} 1/a_n & \text{if } n \geq N, \\ 1 & \text{if } n < N. \end{cases}$$

Then (b_n) is a Cauchy sequence in F, and $a_n b_n = 1$ for every $n \geq N$.

Idea. For $n \geq N$, $a_n \neq 0$, so b_n is defined. The difference of reciprocals is $(a_m - a_n)/(a_n a_m)$; its numerator can be made small, and the denominator is at least c^2 in absolute value. So $|b_n - b_m| \leq |a_m - a_n|/c^2$, and the tolerance $c^2\varepsilon$ for (a_n) gives the tolerance ε for (b_n).

Proof. For $n \geq N$, $|a_n| \geq c > 0$, so $a_n \neq 0$ (Proposition 4.1.3 (a)), and $b_n = a_n^{-1}$ is defined, with $a_n b_n = 1$. For $m, n \geq N$, Lemma 4.1.1 (c) gives

$$b_n - b_m = \frac{1}{a_n} - \frac{1}{a_m} = \frac{a_m - a_n}{a_n a_m}.$$

By Proposition 4.1.3 (b), $|b_n - b_m| = |a_m - a_n| \cdot (1/(|a_n|\,|a_m|))$. From $0 < c \leq |a_n|$ and $0 < c \leq |a_m|$, Proposition 4.1.2 (d) gives $c^2 \leq |a_n|\,|a_m|$, and then Proposition 4.1.2 (c) gives $1/(|a_n|\,|a_m|) \leq 1/c^2$. As $1/(|a_n|\,|a_m|) > 0$ (Proposition 2.5.2 (a)), Proposition 4.1.2 (d) with $0 \leq |a_m - a_n| \leq |a_m - a_n|$ gives

$$|b_n - b_m| \leq \frac{|a_m - a_n|}{c^2} \quad \text{for all } m, n \geq N.$$

Now let $\varepsilon \in F_{>0}$. Then $c^2\varepsilon \in F_{>0}$ by (O4), so there is N_1 (depending on ε and c) with $|a_m - a_n| < c^2\varepsilon$ for all $m, n \geq N_1$. Let $N' := \max\{N, N_1\}$. For $m, n \geq N'$, multiplying by $1/c^2 > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)),

$$|b_n - b_m| \leq \frac{|a_m - a_n|}{c^2} < \frac{c^2\varepsilon}{c^2} = \varepsilon.$$

So (b_n) is Cauchy. The values $b_n = 1$ for $n < N$ play no part; any values would do, by the tail-change remark after Definition 4.5.1.

■

The hypothesis "$|a_n| \geq c$" cannot be weakened to "$a_n \neq 0$". In $\mathbb{Q}$, $a_n = 1/(n+1)$ is Cauchy and never 0, but $1/a_n = n + 1$ is not bounded (for each $M \in \mathbb{Q}_{>0}$, Theorem 2.5.3 gives an n with $n + 1 > M$), so it is not Cauchy, by Proposition 4.5.2. Combined with Theorem 4.5.4, which

supplies c and N for every Cauchy sequence that is not null, this proposition is how Chapter 5 inverts a non-zero real number.

4.5.6 Small steps do not make a Cauchy sequence

The commonest wrong proof that a sequence is Cauchy looks only at neighboring terms.

Warning. "If $a_{n+1} - a_n \to 0$, then (a_n) is Cauchy" is **false**. The next example gives a sequence of rationals whose steps shrink to 0 and whose terms do not settle down at all.

Example. Shrinking Steps, Terms That Do Not Settle

Here is an argument for the false claim in the warning.

"Let $\varepsilon \in F_{>0}$. Since $a_{n+1} - a_n \to 0$, choose N with $|a_{k+1} - a_k| < \varepsilon$ for every $k \geq N$. Let $m > n \geq N$. Then $a_m - a_n$ is the sum of the consecutive differences $a_{k+1} - a_k$ for $k = n, \ldots, m-1$, and each of them is smaller than ε in absolute value. So $|a_m - a_n| < \varepsilon$, and (a_n) is Cauchy."

In $\mathbb{Q}$, let $H_0 := 0$ and $H_{n+1} := H_n + 1/(n+1)$ for $n \in \mathbb{N}$. Show that $H_{n+1} - H_n \to 0$, that $H_{2n} - H_n \geq 1/2$ for every $n \geq 1$, and that (H_n) is not Cauchy. Find the line of the argument that fails.

Solution. The sequence. By Sets, Relations, and Cardinality, Theorem 2.3.1 (read on $\mathbb{N} = \omega$ as Theorem 1.5.1, by Corollary 1.7.3 (b)), with $X = \mathbb{Q}$, starting value 0 and $g(n, s) := s + 1/(n+1)$, there is exactly one sequence (H_n) in $\mathbb{Q}$ with $H_0 = 0$ and $H_{n+1} = H_n + 1/(n+1)$. Its first terms are $0, 1, 3/2, 11/6, 25/12$.

Shrinking steps. $H_{n+1} - H_n = 1/(n+1)$, and $(1/(n+1))$ is null in $\mathbb{Q}$ (Example (First Limits)).

Long stretches. Fix $n \geq 1$. We show, by Logic and Proof, Axiom 4.2.1 on k (through Corollary 1.7.3 (b)), that

$$H_{n+k} - H_n \geq \frac{k}{n+k} \quad \text{for every } k \in \mathbb{N}.$$

Base $P(0)$: both sides are 0. Step: assume it for k. Since $n + k \geq 1$, we have $0 < n + k < n + k + 1$, so $0 < 1/(n+k+1) < 1/(n+k)$ (Proposition 2.5.2 (b)). With $0 \leq k \leq k$, Proposition 4.1.2 (d) gives $k/(n+k+1) \leq k/(n+k)$. Hence, by the induction hypothesis, transitivity (Proposition 2.3.5 (i)) and adding $1/(n+k+1)$ to both sides (Proposition 2.3.5 (b)),

$$\begin{aligned} H_{n+k+1} - H_n &= (H_{n+k} - H_n) + \frac{1}{n+k+1} \\ &\geq \frac{k}{n+k+1} + \frac{1}{n+k+1} = \frac{k+1}{n+k+1}. \end{aligned}$$

Taking $k = n$ gives $H_{2n} - H_n \geq n/(2n) = 1/2$. (Directly: each of the n terms $1/(n+1), \ldots, 1/(2n)$ is at least $1/(2n)$.)

Not Cauchy. We prove the negation displayed after Definition 4.5.1 with $\varepsilon = 1/2$. Let $N \in \mathbb{N}$, and put $n := N + 1$ and $m := 2n$. Then $m, n \geq N$, $n \geq 1$, and $|H_m - H_n| = H_{2n} - H_n \geq 1/2$.

So no threshold works for the tolerance $1/2$.

The failing line. "Each of them is smaller than ε ... so $|a_m - a_n| < \varepsilon$." A sum of $m-n$ terms, each of absolute value below ε, is only below $(m-n)\varepsilon$ (triangle inequality, Proposition 4.1.3 (e), applied $m-n-1$ times). The number of summands $m-n$ is not fixed: m is chosen **after** N, and the definition allows it to be as far beyond n as we like. To repair the line one would need each step below $\varepsilon/(m-n)$, a tolerance that depends on m and n; no single N is promised for it. For (H_n), with $n = N+1$ and $m = 2n$, the n steps are each below $1/n \leq 1$, and their sum is at least $1/2$.

Section 4.6 goes further: its Exercise B2 shows $H_{2^k} \geq 1 + k/2$ for every $k \in \mathbb{N}$, so (H_n) even grows without bound.

4.5.7 Exercises

4.5.7.1 A. Check your understanding

Exercise. A1

These questions are about Theorem 4.5.4 and its proof.

(a) Write "(a_n) is not null" in symbols, with every quantifier and its domain.

(b) The proof asks the Cauchy condition for the tolerance $\varepsilon/2$. Why not ε itself?

(c) Where is the inequality $n_0 \geq N_0$ used?

(d) Run the proof in $\mathbb{Q}$ on $a_n = 1/(n+1) - 1$ with $\varepsilon = 1/2$: check that this ε witnesses "not null", give the least N_0 that the Cauchy condition with tolerance $1/4$ allows, choose n_0, and state the c and N the proof produces.

(e) State the idea of the proof in two sentences.

Solution.

(a) $\exists \varepsilon \in F_{>0}\ \forall N \in \mathbb{N}\ \exists n \in \mathbb{N}\ (n \geq N \wedge |a_n| \geq \varepsilon)$: one tolerance and, beyond every threshold, some term at least that far from 0.

(b) With tolerance ε, Case 1 would give only $a_n > a_{n_0} - \varepsilon \geq \varepsilon - \varepsilon = 0$: every later term is positive, but no positive c with $a_n \geq c$ comes out, and the terms might still approach 0. Any tolerance strictly smaller than ε leaves a positive margin; $\varepsilon/2$ is the simplest.

(c) In the line "since both n and n_0 are $\geq N_0$, $|a_n - a_{n_0}| < \varepsilon/2$": the Cauchy estimate holds only for pairs of indices beyond N_0. It is the reason N_0 is chosen first and the negation is applied **at** N_0.

(d) Here $|a_n| = 1 - 1/(n+1) = n/(n+1)$, which is 0 at $n = 0$ and at least $1/2$ for every $n \geq 1$ (as $2n \geq n+1$). So beyond any N the index $n = N+1$ has $|a_n| \geq 1/2$: $\varepsilon = 1/2$ witnesses "not null". Next, $|a_m - a_n| = |1/(m+1) - 1/(n+1)|$. For $m, n \geq 3$, both reciprocals lie in $(0, 1/4]$, so their difference is below $1/4$, as in the examples after Definition 4.5.1. For $N_0 = 2$ the pair $m = 2$, $n = 11$ gives $1/3 - 1/12 = 1/4 \not< 1/4$. So the least N_0 is 3. Take $n_0 = 3 \geq N_0$: $a_3 = -3/4$, $|a_3| = 3/4 \geq 1/2$, Case 2. The proof gives $a_n < a_3 + 1/4 = -1/2 \leq -1/4$ for $n \geq 3$, and it states the result as $c = \varepsilon/2 = 1/4$, $N = 3$: $a_n \leq -1/4$ for every $n \geq 3$. ($n_0 = 0$ is not allowed: it is below N_0, and $|a_0| = 0 < 1/2$.)

(e) "Not null" provides one tolerance ε and, beyond any threshold, a term at least ε from 0. Taking the threshold from the Cauchy condition with $\varepsilon/2$ first, every later term stays within $\varepsilon/2$ of that one term, so it keeps its sign and stays at least $\varepsilon/2$ from 0.

4.5.7.2 B. Practice

Exercise. B1

Determine which of the following sequences in $\mathbb{Q}$ are Cauchy. Justify your answers, and for the first one work directly from Definition 4.5.1.

(a) $a_n = (n+1)/(n+2)$.
(b) $a_n = (-1)^n$.
(c) $a_n = n$.

Solution.

(a) Cauchy. For every $k \in \mathbb{N}$, $(k+1)/(k+2) = 1 - 1/(k+2)$, so $|a_m - a_n| = |1/(n+2) - 1/(m+2)|$. Let $\varepsilon \in \mathbb{Q}_{>0}$, and use Corollary 2.5.4 to choose $N \in \mathbb{Z}_{>0}$ with $1/N < \varepsilon$. For $m, n \geq N$, both $1/(m+2)$ and $1/(n+2)$ lie in $(0, 1/N)$, since $m+2, n+2 > N$ (Proposition 2.5.2 (b)). Two elements of $(0, 1/N)$ differ by less than $1/N$ (as in the examples after the definition), so $|a_m - a_n| < 1/N < \varepsilon$.

(b) Not Cauchy. Take $\varepsilon = 1$. For any N, the indices N and $N+1$ have $a_{N+1} = -a_N$ and $|a_N| = 1$, so $|a_N - a_{N+1}| = |2a_N| = 2 \geq 1$.

(c) Not Cauchy. If it were, it would be bounded (Proposition 4.5.2): $|n| \leq M$ for some $M \in \mathbb{Q}_{>0}$ and every n. By Theorem 2.5.3 there is $n \in \mathbb{N}$ with $M < n = |n|$, a contradiction. (Archimedes is used here; in the non-Archimedean field $\mathbb{Q}(t)$ of Section 4.2 the sequence $(n \cdot 1)$ is bounded, by t, and the boundedness argument says nothing.)

Exercise. B2

Let (a_n) and (b_n) be Cauchy sequences in F. Prove directly from Definition 4.5.1 that $(a_n - b_n)$ is Cauchy, naming the rule behind each step.

Solution. Let $\varepsilon \in F_{>0}$. By Proposition 4.1.2 (b), $\varepsilon/2 \in F_{>0}$. Since (a_n) is Cauchy, choose N_1 with $|a_m - a_n| < \varepsilon/2$ for all $m, n \geq N_1$; since (b_n) is Cauchy, choose N_2 with $|b_m - b_n| < \varepsilon/2$ for all $m, n \geq N_2$. Let $N := \max\{N_1, N_2\}$ (a maximum of two natural numbers; it depends only on ε). Let $m, n \geq N$; then $m, n \geq N_1$ and $m, n \geq N_2$. By the ring rules (Proposition 2.3.2),

$$(a_m - b_m) - (a_n - b_n) = (a_m - a_n) + (-(b_m - b_n)).$$

By the triangle inequality (Proposition 4.1.3 (e)) and $|-x| = |x|$ (Proposition 4.1.3 (b)),

$$|(a_m - b_m) - (a_n - b_n)| \leq |a_m - a_n| + |b_m - b_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

adding strict inequalities by Proposition 2.3.5 (b) and using Proposition 4.1.2 (b) for the equality. Since ε was arbitrary, $(a_n - b_n)$ is Cauchy.

Exercise. B3

Let $a_n = 2 - 1/(n+1)$ in $\mathbb{Q}$. Show that (a_n) is Cauchy and not null, and find $c \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ as in Theorem 4.5.4. Then run the proof of the theorem with $\varepsilon = 1$ and the least admissible N_0: which c and N does it produce?

Solution. Cauchy. (a_n) is the difference of the constant sequence 2 and the Cauchy sequence $(1/(n+1))$ (examples after Definition 4.5.1), so it is Cauchy by Theorem 4.5.3 (a), (b).

Not null, and c, N directly. For every n, $0 < 1/(n+1) \leq 1$, so $a_n \geq 1$. Then $|a_n| \geq 1$ for every n, so $\varepsilon = 1$ witnesses "not null", and $c = 1$, $N = 0$ satisfy the first alternative. (Any smaller positive c, for instance $c = 1/2$, works with $N = 0$ as well.)

The proof run. With $\varepsilon = 1$, the Cauchy condition is asked for tolerance $1/2$. Here $|a_m - a_n| = |1/(n+1) - 1/(m+1)|$; for $m, n \geq 1$ both reciprocals lie in $(0, 1/2]$, so the difference is below $1/2$, while $N_0 = 0$ fails ($m = 0$, $n = 1$: $1 - 1/2 = 1/2 \not< 1/2$). So $N_0 = 1$. Take $n_0 = 1$: $a_1 = 3/2 \geq 1$, Case 1. The proof gives $c = 1/2$, $N = 1$: $a_n > 3/2 - 1/2 = 1 \geq 1/2$ for every $n \geq 1$.

4.5.7.3 C. Going deeper

Exercise. C1

Let (a_n) be a Cauchy sequence in F and (b_n) a sequence in F such that $(a_n - b_n)$ is null. Prove that (b_n) is Cauchy.

Here is the beginning of a proof. *Name the move, then complete it.*

"Let $\varepsilon \in F_{>0}$. For all $m, n \in \mathbb{N}$, the triangle inequality (Proposition 4.1.3 (e)), applied twice, gives

$$|b_n - b_m| \leq |b_n - a_n| + |a_n - a_m| + |a_m - b_m|.$$

..."

Solution. The move is **split** ε, now into three parts, one for each term on the right: two come from the null difference, one from the Cauchy condition.

Completion. Let $3 \coloneqq 1 + 1 + 1 = 3 \cdot 1_F$; then $3 > 0$ (Proposition 4.2.1 (b)), so $\varepsilon/3 \in F_{>0}$ (Proposition 2.5.2 (a), (O4)), and $\varepsilon/3 + \varepsilon/3 + \varepsilon/3 = 3 \cdot (\varepsilon/3) = \varepsilon$: by (F5)–(F9), with $3 \neq 0$. Since $(a_n - b_n)$ is null, choose N_1 with $|a_n - b_n| < \varepsilon/3$ for every $n \geq N_1$. Since (a_n) is Cauchy, choose N_2 with $|a_n - a_m| < \varepsilon/3$ for all $m, n \geq N_2$. Let $N \coloneqq \max\{N_1, N_2\}$, which depends only on ε. For $m, n \geq N$, using $|b_n - a_n| = |a_n - b_n|$ (Proposition 4.1.3 (b)),

$$|b_n - b_m| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

adding three strict inequalities (Proposition 2.3.5 (b)). So (b_n) is Cauchy.

The split $\varepsilon/4$, $\varepsilon/2$, $\varepsilon/4$, which needs only halving, works as well.

Exercise. C2: Exit check: the sign theorem

Let (a_n) be a Cauchy sequence in an ordered field F that is not null. Prove that there are $c \in F_{>0}$ and $N \in \mathbb{N}$ such that either $a_n \geq c$ for every $n \geq N$, or $a_n \leq -c$ for every $n \geq N$.

Solution. "Null" means $\forall \varepsilon \in F_{>0}\ \exists N \in \mathbb{N}\ \forall n \in \mathbb{N}\ (n \geq N \Rightarrow |a_n| < \varepsilon)$. Its negation (Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.1 (a), Logic and Proof, Proposition 6.2.2, then (O1)) is

$$\exists \varepsilon \in F_{>0}\ \ \forall N \in \mathbb{N}\ \ \exists n \in \mathbb{N}\ \ (n \geq N \wedge |a_n| \geq \varepsilon).$$

Fix such an ε; it stays fixed from here on.

Apply the Cauchy condition with the tolerance $\varepsilon/2 \in F_{>0}$ (Proposition 4.1.2 (b)): there is N_0 with $|a_m - a_n| < \varepsilon/2$ for all $m, n \geq N_0$. Apply the negation at N_0: there is $n_0 \geq N_0$ with $|a_{n_0}| \geq \varepsilon$. By the definition of $|\cdot|$, $a_{n_0} \geq \varepsilon$ or $a_{n_0} \leq -\varepsilon$.

Let $n \geq N_0$. As $n, n_0 \geq N_0$, $-\varepsilon/2 < a_n - a_{n_0} < \varepsilon/2$ (Proposition 4.1.3 (d)). If $a_{n_0} \geq \varepsilon$, then $a_n > a_{n_0} - \varepsilon/2 \geq \varepsilon/2$. If $a_{n_0} \leq -\varepsilon$, then $a_n < a_{n_0} + \varepsilon/2 \leq -\varepsilon/2$. With $c := \varepsilon/2$ and $N := N_0$, the first case gives $a_n \geq c$ and the second $a_n \leq -c$, for every $n \geq N$.

Marking scheme. (10 points) - 2: negates "null" correctly, with the quantifiers in order, and fixes one ε. - 2: applies the Cauchy condition with a fixed tolerance δ such that $0 < \delta < \varepsilon$ (for example $\delta = \varepsilon/2$) to get N_0. - 2: obtains $n_0 \geq N_0$ with $|a_{n_0}| \geq \varepsilon$ from the negation applied **at** N_0. - 3: derives the one-signed bound (with $c = \varepsilon - \delta$) for every $n \geq N_0$, in both sign cases. - 1: names c and N and concludes. Fatal errors (0 for the step): n_0 chosen before N_0, or from the negation at a threshold unrelated to N_0; "$|a_n| \geq \varepsilon$ for all large n" asserted from "not null" alone. A complete proof by contradiction (assume neither alternative holds for any c, N, and refute it at $c = \varepsilon - \delta$, $N = N_0$) maps onto the same $2 + 2 + 2 + 3 + 1$.

Exercise. C3: Find the error

Here is an argument that the product of two Cauchy sequences (a_n), (b_n) in F is Cauchy.

"For all m, n, $|a_n b_n - a_m b_m| \leq |a_n|\,|b_n - b_m| + |b_m|\,|a_n - a_m|$. Given $\varepsilon \in F_{>0}$, choose N such that $|b_n - b_m| < \varepsilon/(2|a_n|)$ and $|a_n - a_m| < \varepsilon/(2|b_m|)$ for all $m, n \geq N$. Then for $m, n \geq N$, $|a_n b_n - a_m b_m| < \varepsilon/2 + \varepsilon/2 = \varepsilon$."

Is the conclusion true? Does the argument prove it? If not, name the line that fails, say why, and repair it.

Solution. The conclusion is true: it is Theorem 4.5.3 (c).

The argument does not prove it. The failing line is "choose N such that $|b_n - b_m| < \varepsilon/(2|a_n|)$ ... for all $m, n \geq N$". The tolerances $\varepsilon/(2|a_n|)$ and $\varepsilon/(2|b_m|)$ depend on n and m, which are introduced **after** N. The definition of Cauchy supplies a threshold for one fixed tolerance $\varepsilon' \in F_{>0}$, named before the threshold; here the threshold is asked to serve a tolerance that changes with the very indices it controls. This is a witness depending on variables introduced later, entry (N2) of Logic and Proof, Example (What is not allowed).

A counterexample to the step. The step gets one N for all the tolerances $\varepsilon/(2|a_n|)$ from

their being positive. In $\mathbb{Q}$ take the Cauchy sequence $b_n = 1/(n+1)$, the never-zero (but unbounded) $a_n = n+1$, and $\varepsilon = 1$. For any N, $n = N$ and $m = 2N+1$ give

$$|b_n - b_m| = \frac{1}{N+1} - \frac{1}{2N+2} = \frac{1}{2(N+1)} = \frac{\varepsilon}{2|a_n|},$$

which is not smaller: no N works. Inside the exercise (a_n) is Cauchy, hence bounded, but the argument never uses that; the repair does.

Repair. Fix bounds first: by Proposition 4.5.2, choose $A, B \in F_{>0}$ with $|a_n| \le A$ and $|b_n| \le B$ for every n, before ε is named. Then the fixed tolerances $\varepsilon/(2A) \le \varepsilon/(2|a_n|)$ (when $a_n \neq 0$) and $\varepsilon/(2B)$ give thresholds N_1, N_2, and $N := \max\{N_1, N_2\}$ works, as in the proof of Theorem 4.5.3 (c).

4.6 A Cauchy Sequence with No Rational Limit

Section 4.5 proved that every convergent sequence is Cauchy (Proposition 4.5.1). In $\mathbb{Q}$ the converse fails, and the failure sits exactly at the hole Chapter 2 found at $\sqrt{2}$: no rational squares to 2 (Theorem 2.6.1). This section proves that the rational sequence of Chapter 2 that closes in on that hole (Example (A Rational Sequence Aimed at the Gap)) is Cauchy in $\mathbb{Q}$ and converges to no rational. Chapter 5 starts from this sequence and builds the number it is aiming at. The section ends by listing what the chapter has not proved, and why, in $\mathbb{Q}$, it could not have.

Throughout this section the field is $F = \mathbb{Q}$: every tolerance ε is a positive rational, and "converges" means "converges in $\mathbb{Q}$" (Definition 4.3.2).

4.6.1 The sequence again

Recall the sequence of Example (A Rational Sequence Aimed at the Gap). For $s \in \mathbb{Q}_{>0}$ let $t(s) := (s^2+2)/(2s)$, the average of s and $2/s$; let $x_0 := 2$ and $x_{n+1} := t(x_n)$, and let $e_n := x_n^2 - 2$, the amount by which x_n^2 overshoots 2. The first terms are

$$x_0 = 2, \quad x_1 = \frac{3}{2}, \quad x_2 = \frac{17}{12}, \quad x_3 = \frac{577}{408},$$

with

$$e_0 = 2, \quad e_1 = \frac{1}{4}, \quad e_2 = \frac{1}{144}, \quad e_3 = \frac{1}{166464}.$$

Chapter 2 proved, for every $n \in \mathbb{N}$, the following facts, and we cite them from Example (A Rational Sequence Aimed at the Gap) without proving them again:

$$x_n \in \mathbb{Q}_{>0}, \qquad e_n > 0 \ \ (\text{so } x_n^2 > 2),$$
$$e_{n+1} = \frac{e_n^2}{4x_n^2} \le \frac{e_n^2}{8}, \qquad x_n - x_{n+1} = \frac{e_n}{2x_n} > 0.$$

The second line says two things. The overshoot is roughly squared at each step, so it shrinks very fast; and the terms decrease. Section 4.5 asks for more than decrease: any two late terms must be close. The overshoot e_n measures how close, because every later term lies between $2/x_n$ and x_n, an interval of length e_n/x_n.

For the proof we do not need the true speed of e_n. A crude bound is enough, as long as it can be pushed below every rational tolerance:

$$e_1 = \frac{1}{4}, \quad e_2 = \frac{1}{144} \le \frac{1}{8}, \quad e_3 = \frac{1}{166464} \le \frac{1}{12};$$

that is, $e_n \le 1/(4n)$ for $n = 1, 2, 3$, and the theorem proves it for every $n \ge 1$. Archimedes then turns $1/(4n)$ into anything below a given ε.

4.6.2 The theorem

Theorem 4.6.1. The Newton Sequence Is Cauchy with No Rational Limit

Let $(x_n)_{n\in\mathbb{N}}$ and $e_n = x_n^2 - 2$ be as in Example (A Rational Sequence Aimed at the Gap). Then $x_n > 1$ for every $n \in \mathbb{N}$, $x_m \le x_n$ whenever $m \ge n$, and $e_n \le 1/(4n)$ for every $n \ge 1$. The sequence (x_n) is a Cauchy sequence in $\mathbb{Q}$, and there is **no** $L \in \mathbb{Q}$ with $x_n \to L$.

The first three clauses are part of the statement, rather than steps hidden in the proof, so that Chapter 5 can cite them.

Idea. Cauchy. For $m \ge n$, the later term x_m lies below x_n, since the terms decrease, and above $2/x_n$, since $x_m > 2/x_m \ge 2/x_n$. So $0 \le x_n - x_m \le x_n - 2/x_n = e_n/x_n \le e_n$. With $e_n \le 1/(4n)$, every two terms beyond N differ by at most $1/(4N)$, and Archimedes makes $1/N$ smaller than any rational ε.

No rational limit. **Locate the missing limit**: suppose $x_n \to L$ in $\mathbb{Q}$ and find out what L would have to satisfy. The limit laws turn $x_n \to L$ into $x_n^2 \to L^2$. But $x_n^2 = 2 + e_n \to 2$, because e_n is squeezed between 0 and $1/(4n)$. Limits are unique, so $L^2 = 2$, and no rational satisfies that. *Before reading on: which law turns $x_n \to L$ into $x_n^2 \to L^2$, and applied to which two sequences?*

Proof. We prove six facts, then the two claims. Induction is Logic and Proof, Axiom 4.2.1, read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b).

(i), (ii), (iv) in part: cited. By Example (A Rational Sequence Aimed at the Gap), for every $n \in \mathbb{N}$: (i) $x_n \in \mathbb{Q}_{>0}$, $e_n > 0$ and $x_n^2 > 2$; (ii) $e_{n+1} = e_n^2/(4x_n^2) \le e_n^2/8$; and $x_{n+1} < x_n$.

(iii) $x_n > 1$. Suppose $x_n \le 1$. Since $x_n > 0$ by (i), either $x_n < 1$, and then $x_n^2 < 1^2 = 1$ by Proposition 2.3.5 (g), or $x_n = 1$ and $x_n^2 = 1$. Either way $x_n^2 \le 1 < 2$, against $x_n^2 > 2$ and (O1). So $x_n > 1$, by totality (Proposition 2.3.5 (i)).

(iv) $x_m \le x_n$ *for* $m \ge n$. Fix n. By induction on k, $x_{n+k} \le x_n$ for every $k \in \mathbb{N}$: the base $P(0)$ is $x_n \le x_n$, and if $x_{n+k} \le x_n$, then $x_{n+k+1} < x_{n+k} \le x_n$. Every $m \ge n$ in $\mathbb{N}$ is $n + k$ for some $k \in \mathbb{N}$ (Proposition 1.6.3 (a)), so $x_m \le x_n$.

(v) $e_n \le 1/(4n)$ *for* $n \ge 1$. By induction on k, for $P(k)$: $e_{k+1} \le 1/(4(k+1))$. Base $P(0)$: $e_1 = 1/4$. Step: assume $P(k)$, and let $n := k + 1 \ge 1$, so $0 < e_n \le 1/(4n) \le 1/4$ (by (i), $P(k)$, and Proposition 4.1.2 (c) with $4 \le 4n$). Then, by (ii) and Proposition 4.1.2 (d) (with the factors $0 \le e_n/8 \le e_n/8$ and $0 \le e_n \le 1/4$, then $0 \le 1/32 \le 1/32$ and $0 \le e_n \le 1/(4n)$),

$$e_{n+1} \le \frac{e_n \cdot e_n}{8} \le \frac{e_n}{8} \cdot \frac{1}{4} = \frac{e_n}{32} \le \frac{1}{128n} \le \frac{1}{4(n+1)},$$

where the last step is Proposition 4.1.2 (c), since $0 < 4(n+1) = 4n+4 \le 128n$ for $n \ge 1$ (as $4 \le 124n$). This is $P(k+1)$.

(vi) For $m \ge n \ge 1$: $0 \le x_n - x_m \le 1/(4n)$. By (iv), $x_m \le x_n$, so $x_n - x_m \ge 0$. By (i), $x_m^2 > 2$; multiplying by $1/x_m > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)) gives $x_m > 2/x_m$. From $0 < x_m \le x_n$, Proposition 4.1.2 (c) gives $1/x_n \le 1/x_m$, so $2/x_n \le 2/x_m < x_m$ (Proposition 4.1.2 (d), with $0 \le 2 \le 2$). Hence

$$x_n - x_m \le x_n - \frac{2}{x_n} = \frac{x_n^2 - 2}{x_n} = \frac{e_n}{x_n} \le e_n \le \frac{1}{4n}.$$

Here $e_n/x_n \le e_n$ because $0 < 1 < x_n$ by (iii), so $1/x_n < 1$ (Proposition 2.5.2 (b)), and $e_n > 0$; the last step is (v).

Cauchy. Let $\varepsilon \in \mathbb{Q}_{>0}$. By Corollary 2.5.4 there is $N \in \mathbb{Z}_{>0}$ with $1/N < \varepsilon$; it depends only on ε. Let $m, n \ge N$, and let $k := \min\{m,n\} \ge N \ge 1$ and $l := \max\{m,n\}$. By (vi) with the pair $l \ge k$,

$$|x_m - x_n| = x_k - x_l \le \frac{1}{4k} \le \frac{1}{4N} \le \frac{1}{N} < \varepsilon,$$

where $1/(4k) \le 1/(4N) \le 1/N$ is Proposition 4.1.2 (c) applied to $N \le 4N \le 4k$. So (x_n) is Cauchy in $\mathbb{Q}$.

No rational limit. Suppose, for a contradiction, that $x_n \to L$ for some $L \in \mathbb{Q}$. By the product law, Theorem 4.4.3 (c), applied with both sequences equal to (x_n), $x_n^2 \to L^2$.

Next, (e_n) is null. We use the squeeze, Proposition 4.4.4 (c), with $N_0 = 1$, the lower sequence constantly 0 (which tends to 0) and the upper sequence $(1/(n+1))_{n\in\mathbb{N}}$. The upper sequence is defined for every $n \in \mathbb{N}$, and it is null in $\mathbb{Q}$ by Proposition 4.2.3, as in Example (First Limits). For $n \ge 1$ we have $n+1 \le 4n$ (as $1 \le 3n$), so by (i), (v) and Proposition 4.1.2 (c),

$$0 < e_n \le \frac{1}{4n} \le \frac{1}{n+1} \qquad (n \ge 1).$$

(At $n = 0$ the middle term is undefined and $e_0 = 2$; the squeeze needs the inequalities only from $N_0 = 1$ on.) So $e_n \to 0$.

Then $x_n^2 = 2 + e_n$ for every n, and the sum law, Theorem 4.4.3 (a), with the constant sequence 2 gives $x_n^2 \to 2 + 0 = 2$. The sequence (x_n^2) therefore tends to L^2 and to 2, and by Theorem 4.3.4, $L^2 = 2$. This contradicts Theorem 2.6.1. Hence no $L \in \mathbb{Q}$ has $x_n \to L$.

■

The answer to the prompt: the product law, Theorem 4.4.3 (c), applied to $a_n = b_n = x_n$. It is the only step that uses the hypothesis $x_n \to L$; the rest of the second half is about e_n, which we control without any limit, and then uniqueness compares the two limits of (x_n^2).

Archimedes is used twice, both times in $\mathbb{Q}$, where Theorem 2.5.3 provides it: for the Cauchy threshold, through Corollary 2.5.4, and for the upper sequence $(1/(n+1))$ of the squeeze, through Proposition 4.2.3.

Remark. The sequence is Cauchy and not null, since $|x_n| = x_n > 1$ for every n, so Theorem 4.5.4 applies to it. Its proof, run with $\varepsilon = 1$, gives $c = 1/2$; directly, $x_n \ge 1$ for every n. Chapter 5 turns the class of (x_n) into a positive real number whose square is 2.

Warning. "Cauchy" does not mean "convergent to a limit we have not named yet". In $\mathbb{Q}$ there is nothing to name: by Theorem 4.6.1, the statement "every Cauchy sequence of rationals converges in $\mathbb{Q}$" is **false**, not merely unproved. A proof that says "(a_n) is Cauchy, so let a be its limit" is wrong in $\mathbb{Q}$, whatever it goes on to do.

Quick check. Give an upper bound and a positive lower bound for the terms x_n in $\mathbb{Q}$, with reasons.

Solution. Upper bound 2: $x_n \leq x_0 = 2$ for every n, by clause (iv) of the proof with $n = 0$. Lower bound 1: $x_n > 1$ for every n, by (iii). The upper bound is attained, at x_0; the lower bound 1 is crude (every term also exceeds $2/x_1 = 4/3$, by the argument of (vi), and $x_0 = 2 > 4/3$), but any positive lower bound answers the question.

4.6.3 What this chapter did not prove

Looking back over the chapter, three familiar tools of real analysis are absent, and their absence is checkable.

- **No result asserts convergence without naming the limit.** Each result of the chapter that concludes $a_n \to a$ has a in its hypotheses or in its statement: the examples of Section 4.3 (Example (First Limits)), the tail-change rules (Proposition 4.3.2), the limit laws (Theorem 4.4.3) and the squeeze (Proposition 4.4.4 (c)). The results of Section 4.5, about sequences whose limit is unknown, conclude "Cauchy", "bounded", or a sign, never "convergent".
- **No supremum.** No result uses the supremum of a subset of F (Sets, Relations, and Cardinality, Definition 2.6.5), so nothing here rests on a least upper bound property.
- **No monotone convergence theorem, and no Bolzano–Weierstrass theorem.** Neither is stated or used.

In $\mathbb{Q}$ each of the three would fail. A least upper bound property fails by Section 2.6: the set of rationals with square below 2 is non-empty and bounded above, with no rational supremum. A monotone convergence theorem fails by the sequence of this section: (x_n) is decreasing (clause (iv) of the proof), bounded (the Quick check above) and Cauchy, and it has no limit in $\mathbb{Q}$. And so does Bolzano–Weierstrass, which would give the bounded sequence (x_n) a subsequence converging in $\mathbb{Q}$. Indeed, let $n_0 < n_1 < n_2 < \cdots$ be indices and suppose $x_{n_k} \to L \in \mathbb{Q}$ as $k \to \infty$. By induction, $n_k \geq k$ for every k (if $n_k \geq k$, then $n_{k+1} > n_k \geq k$, so $n_{k+1} \geq k+1$). So for $k \geq 1$, $0 < e_{n_k} \leq 1/(4n_k) \leq 1/(4k) \leq 1/(k+1)$, and the argument of the proof, run on the sequence $(x_{n_k})_{k\in\mathbb{N}}$, gives $L^2 = 2$ again.

The reader who knows real analysis will have met all three as theorems about $\mathbb{R}$. They are true there because of an axiom about $\mathbb{R}$, completeness, that no ordered field satisfies automatically, and this chapter never assumed it. Chapter 5 builds, from Cauchy sequences of rationals, a field in which every Cauchy sequence converges, and proves that in it every non-empty subset that is bounded above has a supremum: a complete ordered field. Chapter 6 proves that any two complete ordered fields are copies of each other, by exactly one order-preserving field isomorphism. *Analysis on the Real Line*, Chapter 1 takes the axioms of a complete ordered field as its starting point, and its Chapter 2 revisits this chapter's sequences, now in $\mathbb{R}$.

4.6.4 Exercises

4.6.4.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 4.6.1.

(a) Which result gives $x_n^2 \to L^2$ from $x_n \to L$, applied to which two sequences?

(b) Why does the proof need the bound $e_n \le 1/(4n)$, and not only "$e_n > 0$ and (e_n) is decreasing"?

(c) Where is the Archimedean property used?

(d) Check fact (vi) at $n = 2$, $m = 3$: compute $x_2 - x_3$ and compare it with e_2/x_2, e_2 and $1/(4 \cdot 2)$.

(e) State the idea of the proof that (x_n) has no rational limit, in two sentences.

Solution.

(a) The product law, Theorem 4.4.3 (c), with $a_n = b_n = x_n$ and $a = b = L$.

(b) The proof must show that (e_n) is null, that is, below every rational ε from some index on. A sequence can be positive and decreasing without being null: $1 + 1/(n+1)$ decreases and stays above 1. The bound $1/(4n) \le 1/(n+1)$ ties e_n to a sequence known to be null, and the same bound gives the Cauchy threshold.

(c) Twice: for the Cauchy threshold, by Corollary 2.5.4 (an N with $1/N < \varepsilon$); and for the upper sequence $(1/(n+1))$ of the squeeze, null by Proposition 4.2.3.

(d) $x_2 - x_3 = 17/12 - 577/408 = (17 \cdot 34 - 577)/408 = (578 - 577)/408 = 1/408$. Next, $e_2/x_2 = (1/144) \cdot (12/17) = 1/204$. So

$$\frac{1}{408} \le \frac{1}{204} \le \frac{1}{144} \le \frac{1}{8},$$

which is the chain $x_2 - x_3 \le e_2/x_2 \le e_2 \le 1/(4 \cdot 2)$ of (vi).

(e) A rational limit L of (x_n) would make x_n^2 tend to L^2, by the product law. But $x_n^2 = 2 + e_n$ tends to 2, since e_n is squeezed to 0, so uniqueness of limits gives $L^2 = 2$, which no rational satisfies.

4.6.4.2 B. Practice

Exercise. B1

With (x_n) as in Theorem 4.6.1, prove that the sequence $(x_n - 2/x_n)$ is null in $\mathbb{Q}$, that $(2/x_n)$ is Cauchy in $\mathbb{Q}$, and that $(2/x_n)$ has no limit in $\mathbb{Q}$.

Solution. Null. For every n, $x_n - 2/x_n = (x_n^2 - 2)/x_n = e_n/x_n$, and as in step (vi) of the proof, $0 < e_n/x_n \le e_n \le 1/(4n) \le 1/(n+1)$ for $n \ge 1$ (using $x_n > 1$ and $e_n \le 1/(4n)$ from the theorem, and $e_n > 0$ from Example (A Rational Sequence Aimed at the Gap) (i)). By the squeeze (Proposition 4.4.4 (c)) with $N_0 = 1$, lower sequence 0 and upper sequence $(1/(n+1))$, null in $\mathbb{Q}$ (Example (First Limits)), $x_n - 2/x_n \to 0$.

Cauchy. $2/x_n = x_n - (x_n - 2/x_n)$ is the difference of a Cauchy sequence (Theorem 4.6.1) and a null one, which is Cauchy by Theorem 4.5.3 (a); so it is Cauchy by Theorem 4.5.3 (b).

No rational limit. Suppose $2/x_n \to L \in \mathbb{Q}$. Then $x_n = 2/x_n + (x_n - 2/x_n) \to L + 0 = L$ by the sum law (Theorem 4.4.3 (a)), against Theorem 4.6.1.

The terms $2/x_n$ approach the gap from below ($2/x_n < x_m$ for all m, and $(2/x_n)^2 = 4/x_n^2 < 2$), while the x_n approach it from above; their differences are null, so in Chapter 5 the two sequences will represent the same real number.

Exercise. B2

Let (H_n) be the sequence of Example (Shrinking Steps, Terms That Do Not Settle): $H_0 = 0$ and $H_{n+1} = H_n + 1/(n+1)$ in $\mathbb{Q}$. Prove that $H_{2^k} \geq 1 + k/2$ for every $k \in \mathbb{N}$, and hence that (H_n) is not bounded.

Solution. The powers 2^k of the natural number 2 are those of Section 1.6 (defined just before Proposition 1.6.4 by Theorem 1.5.1), with $2^0 = 1$ and $2^{k+1} = 2^k \cdot 2 = 2^k + 2^k$.

We use Logic and Proof, Axiom 4.2.1 on k. Base $P(0)$: $H_{2^0} = H_1 = H_0 + 1/1 = 1 \geq 1 + 0/2$. Step: assume $H_{2^k} \geq 1 + k/2$. The example shows $H_{2n} - H_n \geq 1/2$ for every $n \geq 1$; with $n = 2^k \geq 1$ (by induction, $2^k \geq 1$) it gives $H_{2^{k+1}} \geq H_{2^k} + 1/2 \geq 1 + (k+1)/2$.

Not bounded. Let $M \in \mathbb{Q}_{>0}$. By Theorem 2.5.3 there is $k \in \mathbb{N}$ with $2M < k$. Then $H_{2^k} \geq 1 + k/2 > M$, and $H_{2^k} > 0$, so $|H_{2^k}| > M$. No M bounds the sequence.

So (H_n) is not Cauchy for a second reason: a Cauchy sequence is bounded (Proposition 4.5.2).

4.6.4.3 C. Going deeper

Exercise. C1: Exit check: a Cauchy sequence with no rational limit

Let $y_0 := 2$ and $y_{n+1} := (y_n^2 + 3)/(2y_n)$, a sequence in $\mathbb{Q}_{>0}$, and let $d_n := y_n^2 - 3$.

(a) Prove that no rational number squares to 3.

(b) Prove that (y_n) is a Cauchy sequence in $\mathbb{Q}$ and that it converges to no element of $\mathbb{Q}$.

Solution (Hint 1). For (b), copy facts (i)–(vi) of the proof of Theorem 4.6.1 with 3 in place of 2. Since only the case 2 was proved in Chapter 2, prove the analogues of (i), (ii) and the decrease yourself, as Example (A Rational Sequence Aimed at the Gap) does for 2. For (a), run the descent of Theorem 2.6.1.

Solution (Hint 2). First verify the identities $d_{n+1} = d_n^2/(4y_n^2)$ and $y_n - y_{n+1} = d_n/(2y_n)$ by clearing denominators. In the descent for (a), from $p^2 = 3q^2$ show $q < p < 2q$, and try $p' := 3q - p$, $q' := p - q$.

Solution.

(a) Suppose $x \in \mathbb{Q}$ with $x^2 = 3$. Then $x \neq 0$, and replacing x by $-x$ if needed, $x > 0$; write $x = a/b$ with $a, b \in \mathbb{Z}_{>0}$ (Proposition 2.4.4 (b), Theorem 2.5.1 (a)), so $a^2 = 3b^2$. The set of $q \in \mathbb{Z}_{>0}$ for which some $p \in \mathbb{Z}_{>0}$ has $p^2 = 3q^2$ contains b and is bounded below by 0, so it has a least element q (Proposition 2.3.6); fix such a p.
$q < p$: if $p \leq q$, then $p^2 \leq q^2$ (Proposition 2.3.5 (g), or equality), so $3q^2 \leq q^2$ and $2q^2 \leq 0$, against $2q^2 > 0$ (Proposition 2.3.5 (e), (O4)). *$p < 2q$:* if $2q \leq p$, then $4q^2 \leq p^2 = 3q^2$, so $q^2 \leq 0$, against $q^2 > 0$. By totality, $q < p < 2q$.
Let $p' := 3q - p$ and $q' := p - q$. Then $q' > 0$ since $q < p$; $q' < q$ since $p < 2q$; and $p' > 0$ since $p < 2q < 3q$. Expanding,

$$\begin{aligned} p'^2 - 3q'^2 &= (9q^2 - 6pq + p^2) - (3p^2 - 6pq + 3q^2) \\ &= 6q^2 - 2p^2 = 2(3q^2 - p^2) = 0. \end{aligned}$$

So q' is an admissible denominator smaller than q, a contradiction. No rational squares to 3.

(b) *Positivity and the two identities.* For $s \in \mathbb{Q}_{>0}$, $(s^2 + 3)/(2s) \in \mathbb{Q}_{>0}$ ($s^2 + 3 > 0$, $2s > 0$, Proposition 2.5.2 (a), (O4)), so the recursion (Sets, Relations, and Cardinality, Theorem 2.3.1, read on ω through Corollary 1.7.3 (b)) defines (y_n) in $\mathbb{Q}_{>0}$. With $u(s) := (s^2 + 3)/(2s)$, clearing the non-zero denominators $4s^2$ and $2s$,

$$u(s)^2 - 3 = \frac{(s^2+3)^2 - 12s^2}{4s^2} = \frac{(s^2-3)^2}{4s^2}, \qquad s - u(s) = \frac{s^2 - 3}{2s}.$$

So $d_{n+1} = d_n^2/(4y_n^2)$ and $y_n - y_{n+1} = d_n/(2y_n)$. By induction, $d_n > 0$: $d_0 = 1$, and $d_n > 0$ gives $d_n^2 > 0$, so $d_{n+1} > 0$. Hence $y_n^2 > 3$, $4y_n^2 > 12$, and $d_{n+1} \leq d_n^2/12$ (Proposition 2.5.2 (b)). Also $y_n - y_{n+1} > 0$.

Values. $y_1 = 7/4$, $y_2 = 97/56$; $d_0 = 1$, $d_1 = 49/16 - 3 = 1/16$, $d_2 = 9409/3136 - 3 = 1/3136$, and indeed $d_2 = (1/16)^2/(4 \cdot 49/16) = 1/3136$.

Monotone, and $y_n > 1$. As in (iii) and (iv): $y_n > 1$, since $0 < y_n \leq 1$ would give $y_n^2 \leq 1 < 3$; and $y_m \leq y_n$ for $m \geq n$, by induction from $y_{n+1} < y_n$.

A bound on d_n. $d_n \leq 1/(16n)$ for $n \geq 1$, by induction: $d_1 = 1/16$; if $0 < d_n \leq 1/(16n) \leq 1/16$, then $d_{n+1} \leq d_n \cdot d_n/12 \leq d_n/192 \leq 1/(3072n) \leq 1/(16(n+1))$, since $16n + 16 \leq 3072n$ for $n \geq 1$.

Close terms. For $m \geq n \geq 1$: $y_m \leq y_n$, and $y_m > 3/y_m \geq 3/y_n$, so

$$0 \leq y_n - y_m \leq y_n - \frac{3}{y_n} = \frac{d_n}{y_n} \leq d_n \leq \frac{1}{16n},$$

using $y_n > 1$.

Cauchy. Given $\varepsilon \in \mathbb{Q}_{>0}$, Corollary 2.5.4 gives $N \in \mathbb{Z}_{>0}$ with $1/N < \varepsilon$. For $m, n \geq N$, with $k := \min\{m, n\} \geq N$, $|y_m - y_n| \leq 1/(16k) \leq 1/N < \varepsilon$.

No rational limit. If $y_n \to L \in \mathbb{Q}$, then $y_n^2 \to L^2$ (Theorem 4.4.3 (c)). For $n \geq 1$, $0 < d_n \leq 1/(16n) \leq 1/(n+1)$, so $d_n \to 0$ by the squeeze (Proposition 4.4.4 (c), with $N_0 = 1$ and the null upper sequence $(1/(n+1))$), and $y_n^2 = 3 + d_n \to 3$ (Theorem 4.4.3 (a)). By Theorem 4.3.4, $L^2 = 3$, against (a).

Marking scheme. (10 points) - 2: (a), a descent with a least denominator (Proposition 2.3.6), with the inequalities $q < p < 2q$ and the positivity of p', q' justified (another correct route, such as divisibility by 3 with a least denominator, earns the same). - 1: positivity of y_n and d_n, and the recursion identity for d_{n+1}. - 2: a comparison of y_n and y_m for $m \geq n$: monotonicity together with $y_m \geq 3/y_n$, or (factorization route, as in R4) $|y_n - y_m| = |d_n - d_m|/(y_n + y_m)$ with $y_n + y_m > 2$, since $y_n > 1$ follows from $y_n^2 > 3$. - 2: an explicit bound on d_n (or on $y_n - y_m$) tending to 0, with Archimedes giving N for each rational ε. - 1: Cauchy concluded with the quantifiers in order. - 2: no rational limit, by the product law, uniqueness of limits and (a). Fatal errors (0 for the step): assuming that $\sqrt{3}$ exists in $\mathbb{Q}$ or in a larger field and comparing with it; invoking a monotone convergence theorem. A route by factorization as in R4 earns the next 2 for showing $d_n \to 0$ (by an explicit bound with Archimedes, or by the squeeze), and the 1 for concluding that (d_n) is Cauchy by Proposition 4.5.1 and transferring this to (y_n) with the quantifiers in order. A route through the steps $y_n - y_{n+1}$ and a geometric bound earns the middle $2 + 1$ only if the sum of the steps from n to m is bounded explicitly, independently of m (the example at the end of Section 4.5 shows why).

Exercise. C2: Cauchy depends on the field

In the ordered field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field), consider the sequence $a_n := 1/((n+1)\cdot 1)$, whose terms are the rationals $1/(n+1)$. Explain why (a_n) is **not** a Cauchy sequence in $\mathbb{Q}(t)$, although it is Cauchy in $\mathbb{Q}$.

Solution. In $\mathbb{Q}$, $(1/(n+1))$ is Cauchy (Section 4.5, the examples after Definition 4.5.1).

In $\mathbb{Q}(t)$, take the tolerance $\varepsilon := 1/t$, which is positive (the example shows $0 < 1/t$). Let $N \in \mathbb{N}$, and take $m := N$, $n := N + 1$. Then

$$|a_m - a_n| = \frac{1}{N+1} - \frac{1}{N+2} = \frac{1}{(N+1)(N+2)},$$

by Lemma 4.1.1 (c), a positive rational. The natural number $k := (N+1)(N+2)$ satisfies $k \cdot 1 < t$, because $t - k \cdot 1$ has leading coefficient $1 > 0$ (Example (A Non-Archimedean Ordered Field)). Since $0 < k \cdot 1 < t$, Proposition 2.5.2 (b) gives $1/t < 1/(k \cdot 1) = |a_m - a_n|$. So for the tolerance $1/t$ no threshold N works, and (a_n) is not Cauchy in $\mathbb{Q}(t)$.

The terms are the same in both fields; what changed is the set of tolerances. $\mathbb{Q}(t)$ has the positive element $1/t$, below every positive rational, and the rational differences $1/((N+1)(N+2))$ never get below it. Being Cauchy is a property of a sequence **in a field**, not of its terms alone. (The example $\mathbb{Q}(t)$ uses school polynomial algebra; nothing else in the book depends on this exercise.)

4.6.4.4 R. Review

Exercise. R1

Let $\mathcal{C}$ be the set of Cauchy sequences in $\mathbb{Q}$, and for $(a_n), (b_n) \in \mathcal{C}$ write $(a_n) \sim (b_n)$ if $(a_n - b_n)$

is null. Prove that $\sim$ is an equivalence relation on $\mathcal{C}$, and that

$$[(a_n)] + [(b_n)] \coloneqq [(a_n + b_n)]$$

is a well-defined operation on the quotient set $\mathcal{C}/\!\sim$.

Solution. The tool is Chapter 3's quotient of a commutative ring by an ideal (Theorem 3.4.1). Let $\mathcal{N}$ be the set of null sequences in $\mathbb{Q}$. We show that $\mathcal{C}$ is a commutative ring, that $\mathcal{N}$ is an ideal of it, and that $\sim$ is the relation $\sim_{\mathcal{N}}$.

$\mathcal{C}$ is a commutative ring (Definition 2.3.1), with termwise $+$ and $\cdot$. By Theorem 4.5.3, $\mathcal{C}$ contains the constant sequences 0 and 1 by (a), and it is closed under termwise sums and negatives by (b) and under termwise products by (c); so termwise $+$ and $\cdot$ are functions $\mathcal{C} \times \mathcal{C} \to \mathcal{C}$. Each of (F1)–(F7) and (F9), with the constant 0 as zero, the constant 1 as one and $(-a_n)$ as the negative of (a_n), is an equation between sequences, and it holds at every index n because it holds in $\mathbb{Q}$ (Theorem 2.5.1). Two sequences that agree at every index are equal, so each clause holds in $\mathcal{C}$.

$\mathcal{N}$ is an ideal of $\mathcal{C}$ (Definition 3.4.1). First, $\mathcal{N} \subseteq \mathcal{C}$ by Theorem 4.5.3 (a). (I1): the constant sequence 0 tends to 0 (Example (First Limits) (a)). (I2): if $(y_n), (z_n) \in \mathcal{N}$, then $y_n + z_n \to 0 + 0 = 0$ by Theorem 4.4.3 (a). (I3): let $(z_n) \in \mathcal{N}$ and $(r_n) \in \mathcal{C}$. By Proposition 4.5.2 there is $M \in \mathbb{Q}_{>0}$ with $|r_n| \leq M$ for every n, so $(z_n r_n)$ is null by Proposition 4.4.1, and $r_n z_n = z_n r_n$ by (F6) in $\mathbb{Q}$. So $(r_n z_n) \in \mathcal{N}$.

The relation is $\sim_{\mathcal{N}}$. The difference $(a_n) - (b_n)$ in $\mathcal{C}$ is the sequence $(a_n - b_n)$, so $(a_n) \sim (b_n)$ holds exactly when $(a_n) - (b_n) \in \mathcal{N}$. Thus $\sim$ is $\sim_{\mathcal{N}}$ and $\mathcal{C}/\!\sim$ is $\mathcal{C}/\mathcal{N}$. By Theorem 3.4.1 (a), $\sim$ is an equivalence relation, and by (b), $[(a_n)] + [(b_n)] \coloneqq [(a_n + b_n)]$ is well defined; so is the product $[(a_n)] \cdot [(b_n)] \coloneqq [(a_n b_n)]$, and by (c) $\mathcal{C}/\mathcal{N}$ is a commutative ring.

Chapter 3 proved (b) by Chapter 2's lemma on operations on classes (Theorem 2.2.1), the method by which $\mathbb{Z}$ and $\mathbb{Q}$ were built; here it runs on sequences. Chapter 5 builds $\mathbb{R}$ as this quotient, and nothing in this exercise is used there.

Exercise. R2

Let F be an ordered field. Define $r_0 \coloneqq 1$ and $r_{n+1} \coloneqq r_n/2 + 1$ for $n \in \mathbb{N}$. Prove that $1 \leq r_n < 2$ and $r_n < r_{n+1}$ for every $n \in \mathbb{N}$.

Solution. The tools are definition by recursion and proof by induction (Sets, Relations, and Cardinality, Theorem 2.3.1, Logic and Proof, Axiom 4.2.1), both read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b). Here $2 \coloneqq 1+1 > 0$ (Proposition 4.1.2 (b)), so $2 \neq 0$ and $x/2$ is defined for every $x \in F$. The recursion theorem with $X = F$, starting value 1 and rule $g(n, x) \coloneqq x/2+1$ gives exactly one sequence (r_n) in F satisfying the two equations. Its first terms are $r_1 = 3/2$ and $r_2 = 7/4$.

The bounds. We prove $P(n)$: "$1 \leq r_n < 2$" by induction on n. *Base* $P(0)$*:* $r_0 = 1$, and $1 = 2/2 < 2$ by Proposition 4.1.2 (b) with $\varepsilon = 2$. *Step:* assume $1 \leq r_n < 2$. Then $r_n > 0$, so $r_n/2 > 0$ (Proposition 2.5.2 (a), (O4)). Multiplying $r_n < 2$ by $1/2 > 0$ gives $r_n/2 < 2/2 = 1$

(Proposition 2.3.5 (c)). Adding 1 (Proposition 2.3.5 (b)),

$$1 = 0 + 1 < \frac{r_n}{2} + 1 = r_{n+1} < 1 + 1 = 2.$$

So $1 \le r_{n+1} < 2$, which is $P(n+1)$.

The increase. Let $n \in \mathbb{N}$. By $P(n)$, $r_n > 0$, so $r_n = r_n/2 + r_n/2$ (Proposition 4.1.2 (b) with $\varepsilon = r_n$), and $r_n/2 < 1$ as in the step. Hence

$$r_{n+1} - r_n = \left(\frac{r_n}{2} + 1\right) - \left(\frac{r_n}{2} + \frac{r_n}{2}\right) = 1 - \frac{r_n}{2} > 0,$$

and $r_n < r_{n+1}$ by Proposition 2.3.5 (a).

Exercise. R3

Say that a sequence (a_n) in an ordered field F **eventually keeps one sign** if there are $c \in F_{>0}$ and $N \in \mathbb{N}$ such that either $a_n \ge c$ for every $n \ge N$, or $a_n \le -c$ for every $n \ge N$. Write this in symbols, and form its negation by moving $\neg$ inward one law at a time, naming each law. Then use the negation to show that $a_n := (-1)^n/(n+1)$ in $\mathbb{Q}$ does not eventually keep one sign, and explain why this does not contradict the theorem that a Cauchy sequence that is not null eventually keeps one sign.

Solution. "(a_n) eventually keeps one sign" is

$$\exists c \in F_{>0}\ \exists N \in \mathbb{N}\ \Big(\forall n \in \mathbb{N}\ (n \ge N \to a_n \ge c)\ \vee\ \forall n \in \mathbb{N}\ (n \ge N \to a_n \le -c)\Big).$$

Negating, with the laws for bounded quantifiers (Logic and Proof, Proposition 6.2.3), De Morgan's laws (Logic and Proof, Proposition 2.5.5) and negated implication (Logic and Proof, Proposition 6.2.2):

1. $\forall c \in F_{>0}\ \neg\exists N \in \mathbb{N}\ (\dots)$, by (b) of the bounded law;
2. $\forall c \in F_{>0}\ \forall N \in \mathbb{N}\ \neg(\forall n\,(\dots) \vee \forall n\,(\dots))$, by (b) again;
3. $\forall c \in F_{>0}\ \forall N \in \mathbb{N}\ (\neg\forall n\,(\dots) \wedge \neg\forall n\,(\dots))$, by De Morgan (a) (and (b) for the converse);
4. $\forall c \in F_{>0}\ \forall N \in \mathbb{N}\ (\exists n \in \mathbb{N}\ \neg(n \ge N \to a_n \ge c) \wedge \exists n \in \mathbb{N}\ \neg(n \ge N \to a_n \le -c))$, by (a) of the bounded law, twice;
5. $\forall c \in F_{>0}\ \forall N \in \mathbb{N}\ (\exists n \in \mathbb{N}\ (n \ge N \wedge \neg(a_n \ge c)) \wedge \exists n \in \mathbb{N}\ (n \ge N \wedge \neg(a_n \le -c)))$, by Logic and Proof, Proposition 6.2.2, twice;
6. and $\neg(a_n \ge c)$ is $a_n < c$, $\neg(a_n \le -c)$ is $a_n > -c$, by totality (Proposition 4.1.2 (a)).

(The $\vdash$ directions of steps 4 and 5 use RAA, as Logic and Proof, Proposition 6.2.3 and Logic and Proof, Proposition 6.2.2 record.) So the negation says: **for every** $c \in F_{>0}$ and **every** $N \in \mathbb{N}$, some $n \ge N$ has $a_n < c$, and some $n \ge N$ has $a_n > -c$.

The example. Here $(-1)^n$ is the power in $\mathbb{Q}$ (Definition 4.2.1), so $(-1)^{n+1} = -(-1)^n$ (Proposition 2.3.2 (c)), and by induction $(-1)^n$ is 1 or -1 for every n. Let $c \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$. Of $(-1)^N$ and $(-1)^{N+1} = -(-1)^N$, one is 1 and the other -1; call p the index among $N, N+1$ with $(-1)^p = 1$, and q the other. Both are $\ge N$, and $1/(p+1) > 0$, $1/(q+1) > 0$ (Proposition 2.5.2 (a)). So $a_q = -1/(q+1) < 0 < c$ and $a_p = 1/(p+1) > 0 > -c$ (Proposition 2.3.5 (a)). This is the negation, so (a_n) does not eventually keep one sign.

No contradiction. $|a_n| = |(-1)^n| \cdot |1/(n+1)| = 1/(n+1)$ (Proposition 4.1.3 (b)), and $(1/(n+1))$ is null in $\mathbb{Q}$ (Example (First Limits) (b)). So for each $\varepsilon \in \mathbb{Q}_{>0}$ the threshold that serves $(1/(n+1))$ gives $|a_n - 0| < \varepsilon$: the sequence (a_n) is null. It is Cauchy (Theorem 4.5.3 (a)), but the theorem (Theorem 4.5.4) assumes "not null", so it says nothing about (a_n). The example shows that this hypothesis cannot be dropped.

Exercise. R4

Let (a_n) be a sequence in $\mathbb{Q}$ with $a_n > 0$ for every n and $a_n^2 \to 2$ in $\mathbb{Q}$. Is (a_n) a Cauchy sequence in $\mathbb{Q}$? Prove your answer.

Solution. **Yes.** The tool is the factorization $a_n^2 - a_m^2 = (a_n - a_m)(a_n + a_m)$, which turns closeness of the squares into closeness of the terms once the factor $a_n + a_m$ is bounded below by a positive constant.

A lower bound. By Lemma 4.4.2 applied to $a_n^2 \to 2 > 0$, there is N_0 with $a_n^2 > 2/2 = 1$ for every $n \geq N_0$. For such n, $a_n > 1$: otherwise $0 < a_n \leq 1$ would give $a_n^2 \leq 1$ (Proposition 2.3.5 (g)). So $a_n + a_m > 2$ for $m, n \geq N_0$, and $1/(a_n + a_m) < 1/2$ (Proposition 2.5.2 (b)).

The estimate. For $m, n \geq N_0$, dividing the factorization by $a_n + a_m \neq 0$,

$$|a_n - a_m| = \frac{|a_n^2 - a_m^2|}{a_n + a_m} \leq \frac{|a_n^2 - a_m^2|}{2}.$$

The sequence (a_n^2) converges, so it is Cauchy (Proposition 4.5.1). Given $\varepsilon \in \mathbb{Q}_{>0}$, choose N_1 with $|a_n^2 - a_m^2| < 2\varepsilon$ for all $m, n \geq N_1$, and let $N := \max\{N_0, N_1\}$. For $m, n \geq N$, $|a_n - a_m| < 2\varepsilon/2 = \varepsilon$. So (a_n) is Cauchy.

Why the look-alike tools fail. The limit laws would need a limit of (a_n), and there is none in $\mathbb{Q}$: $a_n \to L$ would give $a_n^2 \to L^2$, so $L^2 = 2$, against Theorem 2.6.1. Theorem 4.5.4 and Proposition 4.5.6 presuppose that (a_n) is Cauchy, which is what is asked. The Newton sequence of Theorem 4.6.1 is one such (a_n), so the question is not empty.

Chapter 5

Constructing the Real Numbers from Cauchy Sequences

Chapter 4 ended at a hole. The Newton sequence $2, 3/2, 17/12, 577/408, \ldots$ of Chapter 2 is Cauchy in $\mathbb{Q}$, and no rational is its limit (Theorem 4.6.1). Its terms crowd around a place where $\mathbb{Q}$ has no element. This chapter does not search for a number in the hole. It **takes the sequence itself as the number**. Many sequences crowd around the same hole: the decimal truncations $1, 7/5, 141/100, \ldots$ approach it from below and never meet the Newton sequence. So two Cauchy sequences of rationals must count as one number when their difference dies out. That is an equivalence relation, and *Sets, Relations, and Cardinality*, Section 2.5 knows what to do with one: take its classes. *A real number is a Cauchy sequence of rationals, up to a null difference.* The chapter defines the sum, product and order of these classes on representatives and proves each well defined. The sum and product come from the quotient of a ring by an ideal of Chapter 3. The order needs a new idea, because comparing terms does not compare classes. The chapter then places $\mathbb{Q}$ inside the new field, proves it Archimedean with $\mathbb{Q}$ dense in it, and finds there a positive number whose square is 2. Until the field exists, every tolerance, margin and threshold is rational, so nothing about real numbers is assumed on the way to building them.

Then come the two forms of completeness. Every Cauchy sequence of real numbers converges. The proof builds the limit from rationals by a rule, so it uses no axiom of choice, and Section 5.5 shows where choice would have entered. Every non-empty set of real numbers that is bounded above has a least upper bound. This is proved by bisection in any ordered field that is Archimedean and Cauchy-complete, and one line of that proof needs Archimedes. So the constructed $\mathbb{R}$ is a complete ordered field, and this ends the construction of the number systems that began with the axioms of set theory in Chapter 1. Chapter 6 shows that any two complete ordered fields are isomorphic, and the optional Chapter 7 builds one from Dedekind cuts. *Analysis on the Real Line*, Chapter 1 takes the axioms of a complete ordered field as its starting point, and this chapter gives them a model. *Algebraic Preliminaries*, Chapter 1 builds the complex numbers on this $\mathbb{R}$. *Metric and Topological Spaces*, Chapter 2 repeats the construction for an arbitrary metric space.

What you need. Chapters 2 and 4 above all; Section 3.4 for one step of Section 5.2; a few facts about $\mathbb{N}$ from Chapter 1; *Sets, Relations, and Cardinality*, Chapters 1–3, chiefly Section 2.5; and *Logic and Proof*, Chapters 2–5, chiefly Chapter 4. From Chapter 4, nearly all of it, applied with $F = \mathbb{Q}$ and, once Section 5.3 has made $\mathbb{R}$ an ordered field, with $F = \mathbb{R}$. This covers the rules of an ordered field and the absolute value (Proposition 4.1.2, Lemma 4.1.1, Definition 4.1.1, Proposition 4.1.3), and powers and the copy of $\mathbb{Q}$ in an ordered field (Definition 4.2.1,

Proposition 4.2.1, Theorem 4.2.2). It covers the Archimedean property, its reciprocal form, the integer part and density (Definition 4.2.2, Proposition 4.2.3, Proposition 4.2.4, Theorem 4.2.5). It covers sequences, limits and null sequences (Definition 4.3.1, Definition 4.3.2, Definition 4.3.3, Example (First Limits), Proposition 4.3.2, Theorem 4.3.4, Proposition 4.3.5, Proposition 4.4.1, Theorem 4.4.3, Proposition 4.4.4, Proposition 4.3.1, Proposition 4.3.3, Lemma 4.4.2). It covers Cauchy sequences, with the sign theorem and its companions (Definition 4.5.1, Proposition 4.5.1, Proposition 4.5.2, Theorem 4.5.3, Theorem 4.5.4, Proposition 4.5.5, Proposition 4.5.6). And it covers the Newton sequence with no rational limit (Theorem 4.6.1). The non-Archimedean field $\mathbb{Q}(t)$ (Example (A Non-Archimedean Ordered Field)) appears only in prose, a warning and an exercise; no proof cites it. From Chapter 2: commutative rings, fields and ordered fields (Definition 2.3.1, Definition 2.3.2, Proposition 2.3.2, Proposition 2.3.5, Proposition 2.5.2); the order from a positive set (Proposition 2.3.3), used here a third time; the lemma on rules on classes (Theorem 2.2.1); least elements and discreteness in $\mathbb{Z}$ (Proposition 2.3.6, Theorem 2.3.4); the copies of $\mathbb{N}$ in $\mathbb{Z}$ and of $\mathbb{Z}$ in $\mathbb{Q}$ (Proposition 2.2.3, Proposition 2.4.4); $\mathbb{Q}$ as an Archimedean ordered field (Theorem 2.5.1, Theorem 2.5.3, Corollary 2.5.4, Proposition 2.5.5); and the Newton sequence and the gaps (Example (A Rational Sequence Aimed at the Gap), Theorem 2.6.1, Theorem 2.6.2). Review exercise R2 returns to the fractions of Sections 2.4 and 2.5 (Definition 2.4.1, Definition 2.5.1). From Section 3.4: ideals and the quotient ring (Definition 3.4.1, Theorem 3.4.1), which give Section 5.2 the ring laws of $\mathbb{R}$; exercise C2 of Section 5.2 mentions $\mathbb{Z}/4\mathbb{Z}$ (Definition 3.4.2). From Chapter 1: the citing rule (Corollary 1.7.3), by which induction and recursion apply to $\mathbb{N}$, and the arithmetic and order of $\mathbb{N}$ (Definition 1.5.1, Theorem 1.5.3, Theorem 1.6.1, Proposition 1.6.3), in Sections 5.1 and 5.6. From *Sets, Relations, and Cardinality*, Section 2.5: equivalence relations, classes and quotient sets (Sets, Relations, and Cardinality, Definition 2.5.1, Sets, Relations, and Cardinality, Definition 2.5.2, Sets, Relations, and Cardinality, Definition 2.5.4, Sets, Relations, and Cardinality, Theorem 2.5.1), with the two examples that showed what "well defined" means (Sets, Relations, and Cardinality, Proposition 2.5.5, Sets, Relations, and Cardinality, Example (A rule on classes that is not a function)). Further from that volume: functions and their equality (Sets, Relations, and Cardinality, Definition 1.4.1, Sets, Relations, and Cardinality, Proposition 1.4.1), from its Chapter 1; the recursion theorem (Sets, Relations, and Cardinality, Theorem 2.3.1), from its Section 2.3; and upper bounds and suprema (Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5), from its Section 2.6. From its Chapter 3, the axiom of choice (Sets, Relations, and Cardinality, Axiom 3.3.1) is named only to mark the step that Section 5.5 avoids. Review exercise R3 uses countability (Sets, Relations, and Cardinality, Theorem 3.2.2, Sets, Relations, and Cardinality, Theorem 3.4.1, Sets, Relations, and Cardinality, Theorem 1.4.4), and names Sets, Relations, and Cardinality, Proposition 3.5.2, from its Section 3.5, only as the route not taken, since its ($\Leftarrow$) direction uses choice. From *Logic and Proof*: the negation laws of its Chapter 4 (Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2) and its rule $\exists$E (Logic and Proof, Definition 5.6.1), which fixes one object at a time and is contrasted with choice. *Logic and Proof*, Chapter 2 gives the rules $\neg$I and RAA (Logic and Proof, Definition 2.2.1, Logic and Proof, Definition 2.5.1), used in Section 5.3, and its Chapter 3 gives induction (Logic and Proof, Axiom 4.2.1). *Logic and Proof*, Chapter 5's catalogue of what a proof may not do (Logic and Proof, Example (What is not allowed)) serves exercise C2 of Section 5.3. Nothing from *Analysis on the Real Line* is used, and no fact about $\mathbb{R}$ is assumed.

Time. About 25 hours on the core track (all of A, one B and one C per section, the exit checks); about 35 hours for everything.

5.0.1 Roadmap

- **Sequences That Should Converge.** The Newton sequence and the decimal truncations, defined inside $\mathbb{Q}$ by powers of ten and a largest-integer argument in $\mathbb{Z}$, as two sequences aimed at one hole; the set $\mathcal{C}$ of Cauchy sequences of rationals; the relation $a \sim b$ when $a-b$ is null, proved an equivalence relation; the truncations equivalent to the Newton sequence; and finite changes and shifts, which never change the class.
- **The Real Numbers and Their Arithmetic.** $\mathbb{R}$ defined as the quotient set $\mathcal{C}/\!\sim$; sum, product and negative defined on representatives and proved well defined, the product through "null times bounded is null"; the null sequences as an ideal of $\mathcal{C}$, so that Chapter 3's quotient theorem makes $\mathbb{R}$ a commutative ring; and inverses through the sign theorem of Chapter 4, after the failed termwise reciprocal.
- **Positive Classes and the Order.** Why termwise comparison defines no order on classes; positive sequences, eventually above a fixed rational margin, and why the margin is named before the index; positivity as a property of the class; the positive classes satisfying Chapter 2's recipe, which makes $\mathbb{R}$ an ordered field; termwise $\leq$ passing to classes, which strict $<$ does not; and the absolute value computed on representatives.
- **The Rationals Inside the Reals.** The map $\kappa(q) = [(q)_n]$, which preserves the operations and the order and is Chapter 4's embedding $\iota_{\mathbb{R}}$; $\mathbb{R}$ Archimedean, with a unique integer part; $\mathbb{Q}$ dense in $\mathbb{R}$, proved from the Cauchy property alone; rational tolerances sufficing for limits in $\mathbb{R}$; every Cauchy sequence of rationals converging in $\mathbb{R}$ to its own class; and the class of the Newton sequence, a positive real number whose square is 2.
- **Every Cauchy Sequence of Reals Converges.** Cauchy completeness of $\mathbb{R}$, proved through the rational approximants $m_k/(k+1)$, named by unique integer parts; where the axiom of choice would have entered (picking a representative of each term) and why the diagonal of arbitrary representatives does not work.
- **The Real Numbers Form a Complete Ordered Field.** Complete and Cauchy-complete ordered fields; the bound $2^k \geq k+1$ in any ordered field; the theorem that an Archimedean, Cauchy-complete ordered field is complete, proved by bisection with the Archimedean line marked; the constructed $\mathbb{R}$ a complete ordered field; a table of what $\mathbb{R}$ has and $\mathbb{Q}$ lacks; and review exercises that return to Chapter 4's sequences, Chapter 2's fractions, uncountability and recursion.

5.0.2 Named moves

- **Representatives, then independence.** To define an operation or a property on classes, state it on representatives, then show that the class of the answer survives replacing each input by an equivalent one. It proves the sum, product and negative well defined in Section 5.2 and positivity a property of the class in Section 5.3. The termwise reciprocal of Section 5.2 and the termwise order of Section 5.3 fail it. Exercise B2 of Section 5.2 and Exercise C1 of Section 5.3 leave the last step, and then the last half, to you.
- **A term approximates its class.** For $a \in \mathcal{C}$, the rationals a_k converge in $\mathbb{R}$ to $[a]$. Estimate $|[a] - a_k|$ through the representative $(a_m - a_k)_m$, and aim the termwise bound below the target, since only $\leq$ passes to classes. It is proved in Section 5.4 and used in Section 5.5 to find the limit's candidate.
- **$\varepsilon/3$ through an approximant.** To compare two objects you cannot compare directly, here two rationals q_j, q_k known only through the reals x_j, x_k they approximate, pass through the approximated objects and give each of the three distances a third of the tolerance. It proves in Section 5.5 that the approximants form a Cauchy sequence of rationals. The triangle inequality is used in $\mathbb{R}$, and the result is carried back to $\mathbb{Q}$ through κ.

- **Bisect and keep the invariant.** Halve an interval by a rule that keeps its right end an upper bound and its left end not one; the ends converge to the supremum once the lengths tend to 0. Making the lengths tend to 0 is the one step that needs Archimedes. It proves in Section 5.6 that an Archimedean, Cauchy-complete ordered field is complete. The Quick check there and Exercise C2 of Section 5.6 run it by hand, and the latter runs it in $\mathbb{Q}$, where it finds no supremum.

5.1 Sequences That Should Converge

Chapter 4 ended with a rational sequence that is Cauchy in $\mathbb{Q}$ and converges to no rational (Theorem 4.6.1). Its terms crowd around a place where $\mathbb{Q}$ has a hole. This chapter builds the number that belongs in the hole, and the first step is to notice that many different sequences crowd around the same hole. This section sets up the raw material, the set $\mathcal{C}$ of Cauchy sequences of rationals, and the relation that will decide when two of them aim at the same number: their difference dies out. Nothing about real numbers is assumed. Every tolerance, bound and threshold in this section is a rational number.

5.1.1 Two sequences aimed at one hole

The first sequence is the Newton sequence of Chapter 2 (Example (A Rational Sequence Aimed at the Gap)): $x_0 := 2$ and

$$x_{n+1} := \frac{x_n^2 + 2}{2x_n} = \frac{x_n}{2} + \frac{1}{x_n} \qquad (n \in \mathbb{N}),$$

with error $e_n := x_n^2 - 2$. Its first terms are $2, 3/2, 17/12, 577/408$. We shall use the following facts, each proved earlier, for every $n \in \mathbb{N}$:

(NS1) $x_n \in \mathbb{Q}_{>0}$, and $e_n > 0$, that is, $x_n^2 > 2$ (Example (A Rational Sequence Aimed at the Gap));

(NS2) $x_n > 1$, and $e_n \le 1/(4n)$ when $n \ge 1$ (the first clauses of the statement of Theorem 4.6.1);

(NS3) (x_n) is a Cauchy sequence in $\mathbb{Q}$, and there is no $L \in \mathbb{Q}$ with $x_n \to L$ (Theorem 4.6.1).

From (NS1) and (NS2), $0 < x_n^2 - 2 \le 1/(4n)$ for every $n \ge 1$. At $n = 0$ the expression $1/(4n)$ is not defined, and indeed $x_0^2 - 2 = 2$ is not small.

The second sequence approaches the hole from below, by decimal truncation. To define it inside $\mathbb{Q}$, without the school decimals, we need powers of ten.

Powers of ten. For $k \in \mathbb{N}$, 10^k is the power in the ordered field $\mathbb{Q}$ of Definition 4.2.1 (b): $10^0 = 1$ and $10^{k+1} = 10^k \cdot 10$. (Here $10 = 10 \cdot 1_{\mathbb{Q}}$, by the identification $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$ of Chapter 2 and Proposition 4.2.1 (d).) Two facts about it are used below, both by induction (Logic and Proof, Axiom 4.2.1, read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b)) with base $P(0)$.

- $10^k \in \mathbb{Z}$. Base: $10^0 = 1 \in \mathbb{Z}$. Step: if $10^k \in \mathbb{Z}$, then $10^{k+1} = 10^k \cdot 10$ is a product of two integers, and products of integers computed in $\mathbb{Q}$ are integers (Proposition 2.4.4: $j(m)j(n) = j(mn)$).
- $10^k \ge k + 1$. Base: $10^0 = 1 = 0 + 1$. Step: suppose $10^k \ge y$, where $y := k + 1$. Then $y \ge 1$ (Proposition 4.2.1 (b)), and multiplying by $10 > 0$ gives $10^{k+1} = 10^k \cdot 10 \ge 10y$ (by Proposition 2.3.5 (c) if $10^k > y$; trivially if $10^k = y$). Moreover $10y - (y + 1) = 9y - 1 \ge 9 - 1 > 0$. So $10^{k+1} > y + 1 = k + 2$, where $(k + 1) + 1 = k + 2$ is arithmetic in $\mathbb{N}$ (Theorem 1.5.3), carried into $\mathbb{Q}$ by the sum law of Proposition 4.2.1 (a).

In particular $10^k \geq 1 > 0$, so $1/10^k > 0$ (Proposition 2.5.2 (a)), and Proposition 4.1.2 (c), applied to $0 < k+1 \leq 10^k$, gives $1/10^k \leq 1/(k+1)$. Together,

$$0 < \frac{1}{10^k} \leq \frac{1}{k+1} \qquad (k \in \mathbb{N}). \tag{5.1.1}$$

The inequality is not strict at $k = 0$, where both sides are 1. We write $1/10^k$ throughout; negative powers have not been defined.

The truncations. For $k \in \mathbb{N}$, let m_k be the **largest** natural number m with

$$m^2 \leq 2 \cdot (10^k)^2,$$

and let $d_k := m_k/10^k$. In school terms, d_k is $\sqrt{2}$ cut off after k decimal places; but the definition mentions only integers and the rational 10^k.

That such a largest m exists, and is unique, is a fact about sets of integers. Let

$$S_k := \{m \in \mathbb{Z} : m \geq 0 \text{ and } m^2 > 2 \cdot (10^k)^2\}.$$

It is bounded below by 0, and it is non-empty: $2 \cdot 10^k$ is a non-negative integer, and $(2 \cdot 10^k)^2 = 4(10^k)^2 > 2(10^k)^2$ because $(10^k)^2 > 0$. By Proposition 2.3.6, S_k has a least element m^*. Since $0^2 = 0 \leq 2(10^k)^2$, $0 \notin S_k$, so $m^* \geq 1$ by discreteness (Theorem 2.3.4 (c)), hence $m_k := m^* - 1 \geq 0$ is a natural number (Theorem 2.3.4 (b)). It is not in S_k, being smaller than the least element, so $m_k^2 \leq 2(10^k)^2$. Every natural number $m > m_k$ satisfies $m \geq m^*$ (discreteness again), so $m^2 \geq (m^*)^2 > 2(10^k)^2$ (by Proposition 2.3.5 (g) if $m > m^*$; trivially if $m = m^*$). So m_k is the largest natural number with $m^2 \leq 2(10^k)^2$, and there is exactly one. Because the description singles out one number for each k, the rule $k \mapsto d_k$ is a function $\mathbb{N} \to \mathbb{Q}$ (Sets, Relations, and Cardinality, Definition 1.4.1), with no choices made.

The first values are:

k	0	1	2	3	4	5	6
m_k	1	14	141	1414	14142	141421	1414213

For instance $m_1 = 14$, because $14^2 = 196 \leq 200 = 2 \cdot 10^2$ and $15^2 = 225 > 200$. So $d_0 = 1$, $d_1 = 7/5$, $d_2 = 141/100$, $d_3 = 707/500$.

Different sequences, one intended number. The two sequences never meet: $x_k^2 > 2 \geq d_k^2$ for every k, so $x_k \neq d_k$. They approach from opposite sides, and their differences $x_1 - d_1 = 1/10$, $x_2 - d_2 = 1/150$, $x_3 - d_3 = 11/51000$ shrink fast. Both "should" converge to the same number, and $\mathbb{Q}$ has no such number. The plan of the chapter is to make the sequences themselves into the numbers, and for that we must first say when two sequences count as one.

5.1.2 Cauchy sequences of rationals

The raw material is every sequence that behaves as if it had a limit.

Definition 5.1.1. Cauchy Sequences of Rationals

$\mathcal{C}$ is the set of all sequences $a\colon \mathbb{N} \to \mathbb{Q}$ (Definition 4.3.1, $F = \mathbb{Q}$) that are Cauchy sequences in $\mathbb{Q}$ (Definition 4.5.1): **for every** $\varepsilon \in \mathbb{Q}_{>0}$ **there is** $N \in \mathbb{N}$ with $|a_m - a_n| < \varepsilon$ **for all**

$m, n \geq N$. In symbols, $a \in \mathcal{C}$ means

$$\forall \varepsilon \in \mathbb{Q}_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall m \in \mathbb{N} \;\; \forall n \in \mathbb{N} \\ ((m \geq N \wedge n \geq N) \Rightarrow |a_m - a_n| < \varepsilon).$$

A sequence a of rationals is **null** if $a_n \to 0$ in $\mathbb{Q}$ (Definition 4.3.3, $F = \mathbb{Q}$).

We write a sequence as a single letter, $a = (a_n)_{n \in \mathbb{N}}$, and do arithmetic on sequences **termwise**: $a + b := (a_n + b_n)_n$, $ab := (a_n b_n)_n$, $-a := (-a_n)_n$ and $a - b := (a_n - b_n)_n$. For $q \in \mathbb{Q}$, $(q)_n$ is the constant sequence with every term q.

Nothing is new here except the name: this is Chapter 4's definition with $F = \mathbb{Q}$, so every tolerance ε is **rational**. That is the point. The definition can be read by someone who knows only $\mathbb{Q}$, and it will be used to build a larger field without assuming anything about it.

Examples. Every constant sequence $(q)_n$ is in $\mathcal{C}$, with $N = 0$ for every ε. The sequence $(1/(n+1))_n$ is null (Example (First Limits) (b)), so it is in $\mathcal{C}$ (Theorem 4.5.3 (a)). The Newton sequence (x_n) is in $\mathcal{C}$, by (NS3), and is not null, since $x_n > 1$ for every n. The sequence $(n)_n$ is not in $\mathcal{C}$: a Cauchy sequence is bounded (Proposition 4.5.2), and for every $M \in \mathbb{Q}_{>0}$ some $n \in \mathbb{N}$ has $n > M$ (Theorem 2.5.3).

The toolkit from Chapter 4, with $F = \mathbb{Q}$. This chapter uses the following facts so often that we list them once. Each was proved in Chapter 4 for an arbitrary ordered field; here $F = \mathbb{Q}$.

- Constant sequences and null sequences are in $\mathcal{C}$, and $\mathcal{C}$ is closed under $+$, $-$, $\cdot$ and negatives (Theorem 4.5.3).
- Sums, negatives and constant multiples of null sequences are null (Theorem 4.4.3 (a), (b)).
- A null sequence times a bounded sequence is null (Proposition 4.4.1).
- Every member of $\mathcal{C}$ is bounded (Proposition 4.5.2).
- Changing finitely many terms changes no limit, and so no nullity (Proposition 4.3.2 (a)).
- A sequence squeezed, from some index on, between two sequences with the same limit has that limit too (Proposition 4.4.4 (c)); in particular one squeezed between $(0)_n$, or $(-u_n)_n$, and a null $(u_n)_n$ is null.

5.1.3 When two sequences name the same number

The Newton sequence and the truncations should name one number, yet they share no term. "Eventually equal" is therefore too strict a test. What the two sequences do share is visible in the differences above: $x_k - d_k$ shrinks to nothing. In $\mathbb{Q}$ we cannot compare the two sequences through a common limit, since there is none; but their difference is again a sequence of rationals, and whether it tends to 0 is a question $\mathbb{Q}$ can answer.

Two Cauchy sequences are equivalent when their difference dies out.

Definition 5.1.2. Equivalent Cauchy Sequences

For $a, b \in \mathcal{C}$, we write $a \sim b$, and say that a is **equivalent** to b, if the sequence $a - b = (a_n - b_n)_n$ is null. In symbols, $a \sim b$ means

$$\forall \varepsilon \in \mathbb{Q}_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall n \in \mathbb{N} \;\; (n \geq N \Rightarrow |a_n - b_n| < \varepsilon).$$

In words: whatever rational tolerance is named, **from some index on**, the two sequences agree to within it **at the same index**. The threshold N comes after ε, so it may depend on ε. The

comparison is always between a_n and b_n with the same n; nothing is said about a_m against b_n. The relation is defined only on $\mathcal{C}$: both sequences must be Cauchy before we ask whether they are equivalent.

Examples.

- *Constant sequences (the degenerate case).* For $p, q \in \mathbb{Q}$, $(p)_n \sim (q)_n$ if and only if $p = q$. The difference is the constant $(p-q)_n$, which converges to $p-q$ (Example (First Limits) (a)); by uniqueness of limits (Theorem 4.3.4) it is null exactly when $p-q=0$. So on constant sequences $\sim$ is plain equality, as it must be if $(q)_n$ is to stand for q.
- *A null sequence and zero.* $(1/(n+1))_n \sim (0)_n$: the difference is $(1/(n+1))_n$, null by Example (First Limits) (b). No term of the first sequence is 0.
- *The truncations and the Newton sequence.* $(d_k) \sim (x_k)$; this is Example (The Truncations and the Newton Sequence Are Equivalent) below.

Non-example by a minimal change. Shift the null sequence of the second example up by one: $b_n := 1 + 1/(n+1)$. It is still in $\mathcal{C}$ (a constant plus a null sequence, Theorem 4.5.3 (a), (b)), and $b-(1)_n = (1/(n+1))_n$ is still null, so $b \sim (1)_n$. But $b \not\sim (0)_n$: the clause that fails is "for every ε". With $\varepsilon = 1$, every n has $|b_n - 0| = 1 + 1/(n+1) \geq 1$, so no threshold exists.

Why this definition.

- *Why "null", and not "same limit".* Two sequences with a common limit in $\mathbb{Q}$ have null difference (Theorem 4.4.3 (b)). Conversely, if $a \to L$ and $b \to L'$ with $a-b$ null, then $a-b$ tends to both $L-L'$ and 0, so $L = L'$ (Theorem 4.3.4). So for convergent sequences the new relation says exactly "same limit". But it is stated without mentioning any limit, so it applies equally to (x_k) and (d_k), which have none in $\mathbb{Q}$.
- *Why not "eventually equal".* Then (x_k) and (d_k) would name different numbers, and so would $(1/(n+1))_n$ and $(0)_n$. Sequences that agree ever more closely, without ever agreeing exactly, would never be identified.
- *Why only on $\mathcal{C}$.* On all sequences of rationals the same relation would not respect products, and Section 5.2 needs it to. For instance $(1/(n+1))_n$ and $(0)_n$ have null difference, but multiplying both termwise by $(n+1)_n$, which is not Cauchy, gives $(1)_n$ and $(0)_n$, whose difference is not null. Inside $\mathcal{C}$ every multiplier is bounded, and that is what saves the product.

It must be checked that $\sim$ really sorts sequences into groups, that is, that it is an equivalence relation in the sense of *Sets, Relations, and Cardinality*, Section 2.5.

Proposition 5.1.1. Equivalence of Cauchy Sequences Is an Equivalence Relation

The relation $\sim$ of Definition 5.1.2 is an equivalence relation on $\mathcal{C}$ (Sets, Relations, and Cardinality, Definition 2.5.1).

Idea. Each of the three clauses becomes a statement about one difference sequence. Reflexivity looks at $a-a$, the zero sequence. Symmetry looks at $b-a$, the negative of $a-b$. Transitivity writes $a-c$ as the sum $(a-b)+(b-c)$ of two null sequences. *Before reading on: which fact from Chapter 4 does transitivity use, and where is the triangle inequality hiding?*

Proof. Let $a, b, c \in \mathcal{C}$.

(Eq1) $a - a = (0)_n$, which is null (Example (First Limits) (a)). So $a \sim a$.

(Eq2) Suppose $a \sim b$. For every n, $b_n - a_n = -(a_n - b_n)$ (since $(a_n - b_n) + (b_n - a_n) = 0$ by (F1)–(F4), and negatives are unique, Proposition 2.3.2 (a)), so $b - a = -(a - b)$, which is null by Theorem 4.4.3 (b) ($-0 = 0$). So $b \sim a$.

(Eq3) Suppose $a \sim b$ and $b \sim c$. For every n,

$$a_n - c_n = (a_n - b_n) + (b_n - c_n),$$

so $a - c = (a - b) + (b - c)$, a sum of two null sequences. By the sum rule Theorem 4.4.3 (a), it is null ($0 + 0 = 0$). So $a \sim c$.

This proves that $\sim$ is an equivalence relation on $\mathcal{C}$.

■

The answer to the prompt: transitivity uses the sum rule of the limit laws, Theorem 4.4.3 (a), and the triangle inequality is inside its proof, where $|(a_n - b_n) + (b_n - c_n)| \le |a_n - b_n| + |b_n - c_n|$ splits the tolerance into two halves. Written out directly: given ε, take thresholds for $a - b$ and $b - c$ with tolerance $\varepsilon/2$, and their maximum serves $a - c$.

By Sets, Relations, and Cardinality, Theorem 2.5.1, $\mathcal{C}$ now falls apart into disjoint classes, and two Cauchy sequences lie in the same class exactly when their difference is null. Section 5.2 takes these classes as the real numbers.

5.1.4 The truncations and the Newton sequence

The two sequences of the hook are the first test of the definition.

Example. The Truncations and the Newton Sequence Are Equivalent

Show that (d_k) is non-decreasing, that $(d_k) \in \mathcal{C}$, and that $(d_k) \sim (x_k)$.

Solution. Step 1: (d_k) is non-decreasing. Let $k \in \mathbb{N}$. By the choice of m_k and the defining equation $10^{k+1} = 10^k \cdot 10$,

$$(10m_k)^2 = 100\, m_k^2 \le 200\, (10^k)^2 = 2\, (10^k \cdot 10)^2 = 2\, (10^{k+1})^2.$$

So $10m_k$ is a natural number whose square is at most $2(10^{k+1})^2$, and m_{k+1} is the largest such: $10m_k \le m_{k+1}$. Dividing by $10^{k+1} > 0$ (Proposition 2.5.2 (a), and Proposition 2.3.5 (c) if $10m_k < m_{k+1}$; trivially if they are equal),

$$d_{k+1} = \frac{m_{k+1}}{10^{k+1}} \ge \frac{10m_k}{10^k \cdot 10} = \frac{m_k}{10^k} = d_k.$$

By Logic and Proof, Axiom 4.2.1 on i, with base $P(0)$ ($d_k \le d_k$) and the step just proved, $d_k \le d_{k+i}$ for every $i \in \mathbb{N}$. Every $j \ge k$ is $k + i$ for some $i \in \mathbb{N}$ (Proposition 1.6.3 (a)), so $d_k \le d_j$ whenever $j \ge k$. In particular $d_k \ge d_0 = 1$ for every k ($m_0 = 1$, since $1^2 \le 2 < 2^2$). No law of exponents was used, only the defining equation of 10^{k+1}.

Step 2: $d_k \le d_j < d_k + 1/10^k$ for $j \ge k$, and $(d_k) \in \mathcal{C}$. The lower bound is Step 1. For the upper bound, $m_j^2 \le 2(10^j)^2$ gives $d_j^2 \le 2$. Since $m_k + 1 > m_k$, the maximality of m_k gives

$(m_k+1)^2 > 2(10^k)^2$, that is,

$$\left(d_k + \frac{1}{10^k}\right)^2 = \frac{(m_k+1)^2}{(10^k)^2} > 2.$$

If we had $d_j \geq d_k + 1/10^k$, then, both sides being non-negative, $d_j^2 \geq (d_k + 1/10^k)^2 > 2$ (Proposition 2.3.5 (g)), against $d_j^2 \leq 2$. By trichotomy, $d_j < d_k + 1/10^k$. With Eq. 5.1.1,

$$0 \leq d_j - d_k < \frac{1}{10^k} \leq \frac{1}{k+1} \qquad (j \geq k). \tag{5.1.2}$$

Now let $\varepsilon \in \mathbb{Q}_{>0}$. By Corollary 2.5.4 there is $n_\varepsilon \in \mathbb{Z}_{>0}$ with $1/m < \varepsilon$ for every integer $m \geq n_\varepsilon$. Let $N := n_\varepsilon$, which depends on ε only. For $j, k \geq N$, say with $j \geq k$ (the other case is symmetric, as $|d_j - d_k| = |d_k - d_j|$), Eq. 5.1.2 gives $|d_j - d_k| < 1/(k+1) < \varepsilon$, since $k+1 > N$. So $(d_k) \in \mathcal{C}$.

Step 3: $0 \leq 2 - d_k^2 < 5/10^k$. The first inequality is $d_k^2 \leq 2$. For the second, expand the display of Step 2 (Proposition 2.3.2 (d)):

$$2 < d_k^2 + \frac{2d_k}{10^k} + \frac{1}{10^k} \cdot \frac{1}{10^k}.$$

Here $d_k < 2$, since $0 \leq d_k$ and $d_k^2 \leq 2 < 4 = 2^2$ (Proposition 2.3.5 (g)); and $(1/10^k)(1/10^k) \leq 1/10^k$, since $0 < 1/10^k \leq 1$ by Eq. 5.1.1. Multiplying $d_k < 2$ by $2/10^k > 0$ (Proposition 2.3.5 (c)) and adding the two bounds (Proposition 2.3.5 (b)), $2 - d_k^2 < 4/10^k + 1/10^k = 5/10^k$.

Step 4: $(d_k) \sim (x_k)$. Let $k \geq 1$. By (NS2), $x_k > 1$, and $d_k \geq 1$ by Step 1, so $x_k + d_k \geq 1$. Since $|x_k - d_k| \geq 0$, Proposition 4.1.2 (d) gives

$$\begin{aligned} |x_k - d_k| &\leq |x_k - d_k|(x_k + d_k) = |x_k^2 - d_k^2| \\ &\leq |x_k^2 - 2| + |2 - d_k^2| \leq \frac{1}{4k} + \frac{5}{10^k}, \end{aligned}$$

by Proposition 4.1.3 (b) and (e), the bound $0 < x_k^2 - 2 \leq 1/(4k)$ from (NS1)–(NS2), and Step 3. Since $4k - (k+1) = 3k - 1 > 0$, $1/(4k) \leq 1/(k+1)$ (Proposition 4.1.2 (c)). Define, for **every** $k \in \mathbb{N}$,

$$u_k := \frac{1}{k+1} + \frac{5}{10^k}.$$

Then $-u_k \leq x_k - d_k \leq u_k$ for every $k \geq 1$ (Proposition 4.1.3 (d)). The sequence (u_k) is null: $(1/(k+1))$ is null (Example (First Limits) (b)); $(1/10^k)$ is null by the squeeze Proposition 4.4.4 (c) between $(0)_k$ and $(1/(k+1))$, using Eq. 5.1.1; $(5/10^k)$ is null by Theorem 4.4.3 (b) with $c = 5$; and the sum of two null sequences is null (Theorem 4.4.3 (a)). So $(-u_k)$ is null too, by (b) with $c = -1$. The squeeze Proposition 4.4.4 (c), with $N_0 = 1$, lower sequence $(-u_k)$ and upper sequence (u_k), shows that $(x_k - d_k)$ is null. Both sequences are in $\mathcal{C}$ (by (NS3) and Step 2), so $(x_k) \sim (d_k)$, and $(d_k) \sim (x_k)$ by symmetry.

The bound of Step 4 is crude but honest: at $k = 1$ it gives $|x_1 - d_1| \leq 1/4 + 5/10 = 3/4$, while in fact $x_1 - d_1 = 3/2 - 7/5 = 1/10$.

The upper sequence (u_k) is defined for every k, including $k = 0$, where $1/(4k)$ is not; the squeeze only needs the inequality from $N_0 = 1$ on. One consequence is worth recording at once: (d_k) **has no limit in $\mathbb{Q}$ either.** If $d_k \to L \in \mathbb{Q}$, then $x_k = d_k + (x_k - d_k) \to L + 0 = L$ by the sum rule, against (NS3). So the truncations, like the Newton sequence, are Cauchy and aim at

nothing in $\mathbb{Q}$, and Example (The Truncations and the Newton Sequence Are Equivalent) says they aim at the same nothing.

5.1.5 Finite changes, and shifts

A class will contain far more than the two sequences of the hook. The next example shows how loose the relation is at the start of a sequence.

Example. Finite Changes and Shifts

(a) Let $a \in \mathcal{C}$, let $b\colon \mathbb{N} \to \mathbb{Q}$, and suppose there is $N_0 \in \mathbb{N}$ with $a_n = b_n$ for every $n \geq N_0$. Show that $b \in \mathcal{C}$ and $a \sim b$.

(b) Let $a \in \mathcal{C}$, and let $s \coloneqq (a_{n+1})_{n\in\mathbb{N}}$ be the shifted sequence. Show that $s \in \mathcal{C}$ and $s \sim a$.

Solution.

(a) $b \in \mathcal{C}$ by the tail-change remark after Definition 4.5.1 (Section 4.5). For $n \geq N_0$, $a_n - b_n = 0$. So $a - b$ agrees with the null sequence $(0)_n$ from N_0 on, and is null by Proposition 4.3.2 (a). Hence $a \sim b$.

(b) Let $\varepsilon \in \mathbb{Q}_{>0}$, and let N be a threshold for a with tolerance ε (Definition 5.1.1). For $m, n \geq N$, also $m+1, n+1 \geq N$, so $|s_m - s_n| = |a_{m+1} - a_{n+1}| < \varepsilon$: $s \in \mathcal{C}$, with the same thresholds. For $n \geq N$, apply the Cauchy condition to the pair $(n+1, n)$, both indices $\geq N$:

$$|s_n - a_n| = |a_{n+1} - a_n| < \varepsilon.$$

So $s - a$ is null, and $s \sim a$.

By (a), a representative of a class may begin with **any** finite list of rationals: $(5, -3, 7, 1/4, 1/5, 1/6, \dots)$, which is $1/(n+1)$ from $n = 3$ on, is equivalent to $(1/(n+1))_n$, and so to $(0)_n$. Part (b) uses the Cauchy condition with the two indices $n+1$ and n; a sequence of rationals that is not Cauchy can fail it (for $((-1)^n)_n$, every $|a_{n+1} - a_n|$ is 2).

Warning. **Equivalent is not "eventually equal", and not "close at the start".** $(1/(n+1))_n \sim (0)_n$, though no term of the first sequence is 0. And $(0)_n$ and $(1-1/(n+1))_n$ have the same first term, 0, but are **not** equivalent: their difference $(0)_n - (1 - 1/(n+1))_n$ converges to -1 (Theorem 4.4.3 (a), (b) with Example (First Limits) (a), (b)), so it is not null, by uniqueness of limits (Theorem 4.3.4). Only the behavior far out counts, and it counts at every tolerance.

Quick check. Let $a \coloneqq (1, 0, 1, 0, \dots)$, with $a_n = 1$ for even n and $a_n = 0$ for odd n; and let $b \coloneqq (1, 0, 1/2, 0, 1/3, 0, \dots)$, with $b_{2k} = 1/(k+1)$ and $b_{2k+1} = 0$. Which of them lies in $\mathcal{C}$? For the one that does, find a constant sequence equivalent to it.

Solution. $a \notin \mathcal{C}$. The negation of the Cauchy condition (moving $\neg$ inward as in Logic and

Proof, Proposition 6.2.3) reads

$$\exists \varepsilon \in \mathbb{Q}_{>0} \quad \forall N \in \mathbb{N} \quad \exists m, n \geq N \quad |a_m - a_n| \geq \varepsilon,$$

and it holds with $\varepsilon = 1/2$: for any N, the indices $m = N$ and $n = N + 1$ have one even and one odd (Proposition 1.6.4 (d)), so $|a_m - a_n| = 1 \geq 1/2$.

$b \in \mathcal{C}$, and $b \sim (0)_n$. Indeed b is null: given $\varepsilon \in \mathbb{Q}_{>0}$, Corollary 2.5.4 gives $K \in \mathbb{Z}_{>0}$ with $1/m < \varepsilon$ for every integer $m \geq K$. For $n \geq 2K$, n is odd or of the form $2k$ (Proposition 1.6.4 (d)), and in the second case $k \geq K$. So either n is odd and $b_n = 0 < \varepsilon$, or $n = 2k$ with $k \geq K$ and $b_n = 1/(k+1) < \varepsilon$. So b is null, hence in $\mathcal{C}$ (Theorem 4.5.3 (a)), and $b - (0)_n = b$ is null. The pattern "a non-zero term every other index" is the same in a and b; what differs is whether those terms die out.

The relation is in place, and it is an equivalence relation. The next section takes its classes as the real numbers and gives them an arithmetic.

5.1.6 Exercises

5.1.6.1 A. Check your understanding

Exercise. A1

(a) For each sequence of rationals, decide whether it lies in $\mathcal{C}$ and whether it is null, with a one-line reason: $((-1)^n/(n+1))_n$; $(n)_n$; $(1 + 1/(n+1))_n$; $(d_k)_k$; $((-1)^n)_n$.

(b) Write "$a \not\sim b$", for $a, b \in \mathcal{C}$, in symbols, with no $\neg$ in front of a quantifier.

(c) True or false: if $a \sim b$, then $a_n = b_n$ for at least one n. Give a reason.

Solution.

(a)
- $((-1)^n/(n+1))$: in $\mathcal{C}$ and null. Its absolute value is $1/(n+1)$ (Proposition 4.1.3 (b), $|(-1)^n| = 1$), which is below any $\varepsilon \in \mathbb{Q}_{>0}$ from some index on (Example (First Limits) (b)); null sequences are in $\mathcal{C}$ (Theorem 4.5.3 (a)).
- (n): neither. It is not bounded (Theorem 2.5.3), so it is not Cauchy (Proposition 4.5.2), and not null (a null sequence is convergent, hence bounded, Proposition 4.3.5).
- $(1 + 1/(n+1))$: in $\mathcal{C}$, not null. It converges to 1 (Theorem 4.4.3 (a)), so it is Cauchy (Proposition 4.5.1), and its limit is not 0 (Theorem 4.3.4).
- (d_k): in $\mathcal{C}$ (Example (The Truncations and the Newton Sequence Are Equivalent)), not null: $d_k \geq 1$ for every k, so the tolerance $\varepsilon = 1$ has no threshold.
- $((-1)^n)$: neither. Consecutive terms differ by 2, so $\varepsilon = 1$ has no Cauchy threshold; and it is not null, since every term has absolute value 1.

(b) By Logic and Proof, Proposition 6.2.3 and Logic and Proof, Proposition 6.2.2, reading $\neg(x < \varepsilon)$ as $x \geq \varepsilon$:

$$\exists \varepsilon \in \mathbb{Q}_{>0} \quad \forall N \in \mathbb{N} \quad \exists n \in \mathbb{N} \quad (n \geq N \wedge |a_n - b_n| \geq \varepsilon).$$

One rational tolerance is beaten at indices beyond every threshold.

(c) False. $(1/(n+1))_n \sim (0)_n$, and $1/(n+1) \neq 0$ for every n.

5.1.6.2 B. Practice

Exercise. B1: Equivalent or not

(a) Show that $(1/(n+1))_n \sim (1/(n+2))_n$ and that both are equivalent to $(0)_n$.

(b) Determine which constant sequences $(c)_n$, $c \in \mathbb{Q}$, are equivalent to $(3 + 1/(n+1))_n$. Justify your answer.

Solution.

(a) Both sequences are null: the first by Example (First Limits) (b); the second is the first shifted by one, $(a_{n+1})_n$ with $a_n = 1/(n+1)$, so it is null by Proposition 4.3.2 (b). Null sequences lie in $\mathcal{C}$ (Theorem 4.5.3 (a)), and a null sequence a satisfies $a \sim (0)_n$, since $a - (0)_n = a$. By symmetry and transitivity (Proposition 5.1.1), $(1/(n+1)) \sim (0)_n \sim (1/(n+2))$ gives $(1/(n+1)) \sim (1/(n+2))$.

(b) Only $c = 3$. The difference $(3+1/(n+1)) - (c)_n = (3-c) + 1/(n+1)$ converges to $3-c$, by the sum rule (Theorem 4.4.3 (a)) with the constant $3 - c$ and the null $(1/(n+1))$. By uniqueness of limits (Theorem 4.3.4) it is null if and only if $3 - c = 0$. Hence $(3+1/(n+1)) \sim (c)_n$ exactly when $c = 3$; this is consistent with the degenerate example after Definition 5.1.2, since two constants equivalent to one sequence are equivalent to each other, hence equal.

Exercise. B2: Null differences preserve Cauchy

(a) Let $a \in \mathcal{C}$, and let $b\colon \mathbb{N} \to \mathbb{Q}$ be a sequence such that $a - b$ is null. Prove that $b \in \mathcal{C}$ and $b \sim a$.

(b) Hence show that $(2/x_n)_n \in \mathcal{C}$ and $(2/x_n)_n \sim (x_n)_n$, where (x_n) is the Newton sequence.

Solution.

(a) For every n, $b_n = a_n - (a_n - b_n)$, so $b = a - (a - b)$. Here $a \in \mathcal{C}$, and $a - b$ is null, hence in $\mathcal{C}$ (Theorem 4.5.3 (a)); the difference of two members of $\mathcal{C}$ is in $\mathcal{C}$ (Theorem 4.5.3 (b)). So $b \in \mathcal{C}$. Then $a \sim b$ by definition, and $b \sim a$ by symmetry.

(b) By (NS1), $x_n > 0$, so $2/x_n$ is defined. For every n,

$$x_n - \frac{2}{x_n} = \frac{x_n^2 - 2}{x_n} = \frac{e_n}{x_n}.$$

By (NS1) and (NS2), $e_n > 0$ and $x_n > 1$, so $0 < 1/x_n < 1$ (Proposition 2.5.2) and $0 < e_n/x_n < e_n$ (Proposition 2.3.5 (c)). For $n \geq 1$, $e_n \leq 1/(4n) \leq 1/(n+1)$. So $0 \leq x_n - 2/x_n \leq 1/(n+1)$ for every $n \geq 1$, and the squeeze Proposition 4.4.4 (c), with $N_0 = 1$ and the null sequence $(1/(n+1))$ (Example (First Limits) (b)), shows that $(x_n - 2/x_n)$ is null. Part (a), with $a = (x_n) \in \mathcal{C}$ (NS3) and $b = (2/x_n)$, gives the claim. The terms $2/x_n$ satisfy $(2/x_n)^2 = 4/x_n^2 < 2$, so this sequence approaches the hole from below, as the truncations do.

Exercise. B3: Every other Newton term

Show that $(x_{2n})_n \sim (x_n)_n$ for the Newton sequence. *Hint: imitate Example (Finite Changes and Shifts) (b).*

Solution. First, $(x_{2n})_n \in \mathcal{C}$. Let $\varepsilon \in \mathbb{Q}_{>0}$, and let N be a threshold for (x_n) with tolerance ε (NS3). For $m, n \geq N$, also $2m \geq m \geq N$ and $2n \geq N$, so $|x_{2m} - x_{2n}| < \varepsilon$.

Second, the difference is null. For $n \geq N$, the two indices $2n$ and n are both $\geq N$, so $|x_{2n} - x_n| < \varepsilon$. So $(x_{2n} - x_n)_n$ is null, with the same thresholds as (x_n), and $(x_{2n}) \sim (x_n)$.

The Cauchy condition was used with two indices that are far apart ($2n$ and n). This is exactly where "every pair of indices beyond N" is needed, and not only neighbors.

5.1.6.3 C. Going deeper

Exercise. C1: Subsequences

Let $a \in \mathcal{C}$, and let $k \mapsto n_k$ be a function $\mathbb{N} \to \mathbb{N}$ with $n_k < n_{k+1}$ for every $k \in \mathbb{N}$ (for instance $n_k = 2k$, or any sequence of indices defined by Sets, Relations, and Cardinality, Theorem 2.3.1 with this property).

(a) Prove that $n_k \geq k$ for every k, that $(a_{n_k})_{k \in \mathbb{N}} \in \mathcal{C}$, and that $(a_{n_k})_k \sim a$.

(b) Deduce: if $a, b \in \mathcal{C}$ and $a_{n_k} = b_{n_k}$ for every k, then $a \sim b$. Compare with the warning after Example (Finite Changes and Shifts): why does this not say that sequences which are "close at the start" are equivalent?

Solution (Hint 1). For (a), the indices n_k run ahead of k. Once you know $n_k \geq k$, the Cauchy condition applies to the pair (n_k, k).

Solution (Hint 2). For (a), prove $n_k \geq k$ by Logic and Proof, Axiom 4.2.1 on k; the step uses $n_{k+1} > n_k \geq k$.

Solution.

(a) $n_k \geq k$. By Logic and Proof, Axiom 4.2.1 on k, base $P(0)$: $n_0 \geq 0$ (Theorem 1.6.1 (d)). Step: if $n_k \geq k$, then $n_{k+1} > n_k \geq k$, so $k < n_{k+1}$; by Theorem 1.6.1 (b), $n_{k+1} < k+1$ would give $n_{k+1} \leq k$, so by trichotomy $n_{k+1} \geq k+1$.
In $\mathcal{C}$. Let $\varepsilon \in \mathbb{Q}_{>0}$, and let N be a threshold for a with tolerance ε. For $j, k \geq N$, $n_j \geq j \geq N$ and $n_k \geq k \geq N$, so $|a_{n_j} - a_{n_k}| < \varepsilon$.
Equivalent to a. For $k \geq N$, both $n_k \geq N$ and $k \geq N$, so $|a_{n_k} - a_k| < \varepsilon$. So $(a_{n_k} - a_k)_k$ is null, with the same thresholds, and $(a_{n_k})_k \sim a$.

(b) By (a), $(a_{n_k})_k \sim a$ and $(b_{n_k})_k \sim b$. The two subsequences are the same sequence, since $a_{n_k} = b_{n_k}$ for every k. So $a \sim (a_{n_k})_k = (b_{n_k})_k \sim b$, using symmetry for the first step, and transitivity (Proposition 5.1.1) gives $a \sim b$.

The hypothesis of (b) is that the sequences agree at **infinitely many** indices, $n_0 < n_1 < \cdots$, arbitrarily far out. In the warning, $(0)_n$ and $(1-1/(n+1))_n$ agree only at $n = 0$. Agreement far out, along a whole subsequence, forces equivalence for Cauchy sequences; agreement at finitely many places says nothing. (For sequences that are not Cauchy, (b) fails: $(1, 0, 1, 0, \dots)$ and $(1)_n$ agree at every even index, and their difference is not null.)

Marking scheme. (10 points)

- 2: proves $n_k \geq k$ for every k, justifying the passage from $n_{k+1} > n_k \geq k$ (or, by a least counterexample, from $n_k > n_{k-1} \geq k-1$) to the next whole number.
- 3: shows $(a_{n_k}) \in \mathcal{C}$, with one threshold that serves all $j, k \geq N$ and with $n_j, n_k \geq N$ justified from $n_k \geq k$.
- 3: shows $(a_{n_k} - a_k)_k$ is null by applying the Cauchy condition to the pair (n_k, k).
- 2: deduces (b) by symmetry and transitivity of $\sim$.

Fatal errors (0 for the step): using a limit of a in $\mathbb{Q}$; a threshold that depends on k.

5.2 The Real Numbers and Their Arithmetic

Section 5.1 sorted the Cauchy sequences of rationals by a relation, $a \sim b$ when $a - b$ is null, and proved it an equivalence relation (Proposition 5.1.1). This section takes the classes as the new numbers. It defines their sum, product and negative on representatives, proves that the definitions do not depend on the representatives, and shows that the result is a field. The construction follows a pattern seen three times already: the integers modulo n in *Sets, Relations, and Cardinality*, Section 2.5, the integers and the rationals in Chapter 2, and the quotient of a ring by an ideal in Chapter 3. The order comes in Section 5.3.

5.2.1 The real numbers as classes

The Newton sequence and the decimal truncations are different sequences that aim at one number (Example (The Truncations and the Newton Sequence Are Equivalent)). Rather than search $\mathbb{Q}$ for that number, we let the whole group of sequences aiming at it **be** the number.

A real number is a Cauchy sequence of rationals, up to a null difference.

Definition 5.2.1. The Real Numbers

The set of **real numbers** is the quotient set (Sets, Relations, and Cardinality, Definition 2.5.4)

$$\mathbb{R} := \mathcal{C}/\sim$$

of the set $\mathcal{C}$ of Cauchy sequences of rationals (Definition 5.1.1) by the relation $\sim$ of Definition 5.1.2. Its elements are the classes $[a] = \{b \in \mathcal{C} : b \sim a\}$ for $a \in \mathcal{C}$ (Sets, Relations, and Cardinality, Definition 2.5.2), and any $b \in [a]$ is a **representative** of $[a]$.

In words: a real number is a set of Cauchy sequences of rationals, namely all those equivalent to one of them. By Sets, Relations, and Cardinality, Theorem 2.5.1 (b), for all $a, b \in \mathcal{C}$,

$$[a] = [b] \iff a \sim b \iff a - b \text{ is null.}$$

This is the only tool we have, so far, for deciding when two real numbers are equal: compare representatives, and ask whether their difference is null.

Examples.

- *The Newton class.* By Example (The Truncations and the Newton Sequence Are Equivalent), $[(x_n)_n] = [(d_k)_k]$: one real number, with two quite different representatives. By Example (Finite Changes and Shifts) (b), $(x_{n+1})_n$ is a third.
- *The class of a constant.* For $q \in \mathbb{Q}$, the class $[(q)_n]$ contains $(q)_n$, and also $(q+1/(n+1))_n$, which lies in $\mathcal{C}$ by Theorem 4.5.3 (a), (b) and is equivalent to $(q)_n$, since the difference $(1/(n+1))_n$ is null (Example (First Limits) (b)). Different constants give different classes, because constant sequences are equivalent only when they are equal (the degenerate example after Definition 5.1.2). Section 5.4 uses these classes to place a copy of $\mathbb{Q}$ inside $\mathbb{R}$.
- *Zero.* $[(0)_n]$ is the set of all null sequences in $\mathcal{C}$: $b \sim (0)_n$ means that $b - (0)_n = b$ is null. Every null sequence lies in $\mathcal{C}$ (Theorem 4.5.3 (a)). So $(1/(n+1))_n$ (Example (First Limits) (b)), $((-1)^n/(n+1))_n$ (null times bounded, as $|(-1)^n| = 1$; Proposition 4.4.1) and every sequence that is eventually 0 (null by Proposition 4.3.2 (a)) are representatives of it.

What has been done, and what has not. The symbol $\mathbb{R}$ now names one definite set, built from $\mathbb{Q}$ by the set-building tools of *Sets, Relations, and Cardinality*. In earlier volumes, the "familiar real numbers" were used informally; from here on in this volume $\mathbb{R}$ means this set and nothing else, and every fact about it must be proved from the definition. At this moment $\mathbb{R}$ has no addition, no multiplication, no order, and it does not contain $\mathbb{Q}$: the rational number 2 is not a class of sequences. All of these must be built. This section builds the arithmetic.

5.2.2 Arithmetic on representatives

To add two real numbers, add representatives term by term; likewise for products and negatives. The terms are rationals, where the arithmetic is known.

Definition 5.2.2. Sum, Product and Negative of Real Numbers

For $a, b \in \mathcal{C}$, let

$$[a] + [b] := [a + b], \qquad [a]\,[b] := [ab], \qquad -[a] := [-a],$$

where $a + b$, ab and $-a$ are the termwise sum, product and negative. Let $0_{\mathbb{R}} := [(0)_n]$ and $1_{\mathbb{R}} := [(1)_n]$.

The right-hand sides are classes of members of $\mathcal{C}$, because $\mathcal{C}$ is closed under termwise sums, products and negatives and contains the constants (Theorem 4.5.3). That makes the right-hand sides **meaningful**. It does not yet make the definition **legitimate**. Each formula takes a class and reads off a representative; if two representatives of the same class gave different answers, the "operation" would assign two values to one input, and so would not be a function (Sets, Relations, and Cardinality, Definition 1.4.1).

This danger is real, and it has been met before. In *Sets, Relations, and Cardinality*, Section 2.5, the arithmetic of the integers modulo n was defined by the same kind of formula and survived the check (Sets, Relations, and Cardinality, Proposition 2.5.5), while the rule $[a]_6 \mapsto [a]_4$ failed it (Sets, Relations, and Cardinality, Example (A rule on classes that is not a function)). Chapter 2 made the same check for the integers and for the rationals, through the lemma Theorem 2.2.1, which says that a rule on classes is a function as soon as equivalent inputs give equivalent

outputs. Here the check is the same, and only the computation behind it is new.

Representatives, then independence. To define an operation on classes, define it on representatives; then show that the output class survives replacing each input by an equivalent one.

For sums this is a one-line computation. For products it is not, and the next theorem leans on Chapter 4's result that Cauchy sequences are bounded (Proposition 4.5.2).

Theorem 5.2.1. The Operations on Real Numbers Are Well Defined

Let $a, a', b, b' \in \mathcal{C}$ with $a \sim a'$ and $b \sim b'$. Then

$$a + b \sim a' + b', \qquad ab \sim a'b', \qquad -a \sim -a'.$$

Hence the rules of Definition 5.2.2 define functions $+\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\cdot\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $-\colon \mathbb{R} \to \mathbb{R}$.

Idea. For the product we must show that $a_n b_n - a'_n b'_n$ dies out, knowing only that $a_n - a'_n$ and $b_n - b'_n$ do. Replace one factor at a time: subtract and add the mixed term $a_n b'_n$, so that

$$a_n b_n - a'_n b'_n = a_n(b_n - b'_n) + (a_n - a'_n)\, b'_n.$$

Each summand is a null sequence times another sequence. A null sequence times an **unbounded** one need not be null, so we need the multipliers (a_n) and (b'_n) bounded; they are, because they are Cauchy. Then each summand is null, and so is the sum.

Proof. Let $a, a', b, b' \in \mathcal{C}$ with $a \sim a'$ and $b \sim b'$; so $a - a'$ and $b - b'$ are null (Definition 5.1.2).

Sum. For every n, $(a_n + b_n) - (a'_n + b'_n) = (a_n - a'_n) + (b_n - b'_n)$, by the field axioms of $\mathbb{Q}$. So $(a + b) - (a' + b')$ is the sum of two null sequences, which is null by Theorem 4.4.3 (a). Hence $a + b \sim a' + b'$.

Negative. For every n, $(-a_n) - (-a'_n) = -(a_n - a'_n)$ (Proposition 2.3.2 (c)). So $(-a) - (-a') = -(a - a')$, which is null by Theorem 4.4.3 (b). Hence $-a \sim -a'$.

Product. For every n, by the field axioms of $\mathbb{Q}$,

$$a_n b_n - a'_n b'_n = a_n(b_n - b'_n) + (a_n - a'_n)\, b'_n. \tag{5.2.1}$$

The sequence a is in $\mathcal{C}$, so it is bounded (Proposition 4.5.2): there is $M \in \mathbb{Q}_{>0}$ with $|a_n| \le M$ for every n. Since $b - b'$ is null, the sequence $\big((b_n - b'_n)\, a_n\big)_n$ is null by Proposition 4.4.1, and it is the first summand of Eq. 5.2.1, by commutativity in $\mathbb{Q}$. Likewise $b' \in \mathcal{C}$ is bounded, by some $M' \in \mathbb{Q}_{>0}$ (Proposition 4.5.2), and $a - a'$ is null, so the second summand $\big((a_n - a'_n)\, b'_n\big)_n$ is null (Proposition 4.4.1). By the sum rule (Theorem 4.4.3 (a)), $ab - a'b'$ is null. Hence $ab \sim a'b'$.

Functions. The maps $(a, b) \mapsto a + b$ and $(a, b) \mapsto ab$ are functions $\mathcal{C} \times \mathcal{C} \to \mathcal{C}$, and $a \mapsto -a$ is a function $\mathcal{C} \to \mathcal{C}$ (Theorem 4.5.3). Taking $b' = b$ above ($b \sim b$ by reflexivity) shows that each binary map respects $\sim$ in its first argument; taking $a' = a$ shows the same for the second. These are the hypotheses of Theorem 2.2.1 (b), and $-a \sim -a'$ is the hypothesis of

Theorem 2.2.1 (a). So the three rules define functions $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $\mathbb{R} \to \mathbb{R}$, as claimed.

■

The proof used two different bounds: one for a, which multiplies $b_n - b'_n$, and one for b', which multiplies $a_n - a'_n$. A bound for b would be of no use, since b appears nowhere in Eq. 5.2.1 as a multiplier. Every Cauchy sequence is bounded, so both bounds exist; this is where the restriction of $\sim$ to $\mathcal{C}$ pays (compare the "Why only on $\mathcal{C}$" paragraph of Section 5.1).

So $\mathbb{R}$ has a sum, a product and a negative. For example, $[(x_n)] + [(1)_n] = [(x_n + 1)_n]$, and the answer does not change if (x_n) is replaced by (d_k): the theorem says $(x_n + 1) \sim (d_n + 1)$.

5.2.3 The null sequences form an ideal

Chapter 3 made the check of well-definedness once and for all, for the quotient of a commutative ring by an ideal (Definition 3.4.1, Theorem 3.4.1). The construction of $\mathbb{R}$ is an instance of it, and seeing it that way gives the laws of arithmetic in $\mathbb{R}$ without checking them one by one.

Let $\mathcal{N} \subseteq \mathcal{C}$ be the set of null sequences of rationals; it lies in $\mathcal{C}$ by Theorem 4.5.3 (a). Then

$$a \sim b \iff a - b \in \mathcal{N} \qquad (a, b \in \mathcal{C}),$$

which is the relation $\sim_{\mathcal{N}}$ of Theorem 3.4.1 (a). So $\mathbb{R} = \mathcal{C}/\sim$ is $\mathcal{C}/\mathcal{N}$ in the notation of Chapter 3, and the rules of Definition 5.2.2 are the rules of Theorem 3.4.1 (b). What must be checked is that $\mathcal{C}$ is a commutative ring and $\mathcal{N}$ an ideal of it. Once this is checked, Theorem 3.4.1 (b) gives the well-definedness in Theorem 5.2.1 a second time. Its proof changes one argument at a time, so the multipliers are b and a' rather than a and b', and (I3) is supplied by "null times bounded is null".

Theorem 5.2.2. The Real Numbers Form a Commutative Ring

$(\mathbb{R}, +, \cdot)$, with the operations of Definition 5.2.2, is a commutative ring (Definition 2.3.1) with zero $0_{\mathbb{R}}$ and one $1_{\mathbb{R}}$: it satisfies (F1)–(F7) and (F9), the additive inverse of $[a]$ is $-[a]$, and $0_{\mathbb{R}} \neq 1_{\mathbb{R}}$.

Idea. Two claims, then Chapter 3's theorem. Termwise operations inherit every law of $\mathbb{Q}$, because a law about sequences is an equation between functions on $\mathbb{N}$, checked one index at a time. And the null sequences satisfy the three clauses of an ideal: $(0)_n$ is null, sums of null sequences are null, and a null sequence times a Cauchy sequence is null. The quotient theorem then hands over all the laws at once. For $0_{\mathbb{R}} \neq 1_{\mathbb{R}}$, the constant 1 is not null.

Proof.

Claim. $\mathcal{C}$, with the termwise operations, is a commutative ring with zero $(0)_n$, one $(1)_n$, and $-a$ the additive inverse of a.

Proof. By Theorem 4.5.3, the termwise sum and product of two members of $\mathcal{C}$, the termwise negative of a member, and the constants $(0)_n$, $(1)_n$ lie in $\mathcal{C}$; so $+$ and $\cdot$ are functions $\mathcal{C} \times \mathcal{C} \to \mathcal{C}$. Each of (F1)–(F7) and (F9) asserts an equality of sequences, that is, of functions $\mathbb{N} \to \mathbb{Q}$, so it holds if and only if it holds at every index n (Sets, Relations, and Cardinality,

Proposition 1.4.1). At the index n it is the same clause in $\mathbb{Q}$, with a_n, b_n, c_n in place of a, b, c, and $\mathbb{Q}$ is a field (Theorem 2.5.1). For example, (F9) at n reads $a_n(b_n + c_n) = a_n b_n + a_n c_n$, and (F4) reads $a_n + (-a_n) = 0$.

■

Claim. $\mathcal{N}$ is an ideal of $\mathcal{C}$ (Definition 3.4.1).

Proof. (I1) The zero sequence $(0)_n$ is null (Example (First Limits) (a)). (I2) If $a, b \in \mathcal{N}$, then $a + b \in \mathcal{N}$, by the sum rule Theorem 4.4.3 (a). (I3) Let $a \in \mathcal{N}$ and $r \in \mathcal{C}$. Then r is bounded (Proposition 4.5.2), so $(a_n r_n)_n$ is null (Proposition 4.4.1), and $ra = (r_n a_n)_n = (a_n r_n)_n$ by commutativity in $\mathbb{Q}$. So $ra \in \mathcal{N}$.

■

As noted before the theorem, $\sim$ is the relation $\sim_{\mathcal{N}}$, so $\mathbb{R} = \mathcal{C}/\mathcal{N}$, and the operations of Definition 5.2.2 are those of Theorem 3.4.1 (b). By Theorem 3.4.1 (c), $\mathbb{R}$ is a commutative ring with zero $[(0)_n] = 0_{\mathbb{R}}$ and one $[(1)_n] = 1_{\mathbb{R}}$. Its additive inverse of $[a]$ is $[-a] = -[a]$, since $[a] + [-a] = [a + (-a)] = [(0)_n]$ by the first claim.

Finally, by Theorem 3.4.1 (d), $1_{\mathbb{R}} = 0_{\mathbb{R}}$ if and only if $\mathcal{N} = \mathcal{C}$. But $(1)_n \in \mathcal{C}$ is not null: for $\varepsilon = 1/2$, every n has $|1 - 0| = 1 \geq 1/2$. So $\mathcal{N} \neq \mathcal{C}$, and $0_{\mathbb{R}} \neq 1_{\mathbb{R}}$. This proves the theorem.

■

To see what the quotient theorem did, here is distributivity unfolded. For $a, b, c \in \mathcal{C}$, using Definition 5.2.2 at every step but the middle one, and (F9) in $\mathcal{C}$ there,

$$\begin{aligned}[a]([b] + [c]) &= [a]\,[b + c] = [a(b + c)] \\ &= [ab + ac] = [ab] + [ac] = [a][b] + [a][c].\end{aligned}$$

Every law goes the same way: move to representatives, use the law in $\mathcal{C}$, which is the law in $\mathbb{Q}$ at each index, and move back. Only the legitimacy of "move to representatives" needed an argument, and Theorem 5.2.1 supplied it.

5.2.4 Inverses

A commutative ring is not yet a field. The missing clause is (F8): every $x \neq 0_{\mathbb{R}}$ needs an inverse. The obvious idea is to invert term by term, and it fails; the example after the theorem shows how. What saves it is a fact from Chapter 4: a Cauchy sequence that is not null stays **a fixed distance** away from 0 from some index on (Theorem 4.5.4).

Theorem 5.2.3. The Real Numbers Form a Field

Every $x \in \mathbb{R}$ with $x \neq 0_{\mathbb{R}}$ has a multiplicative inverse: there is $y \in \mathbb{R}$ with $xy = 1_{\mathbb{R}}$. Hence $\mathbb{R}$, with the operations of Definition 5.2.2, is a field (Definition 2.3.1).

Idea. Write $x = [a]$. Since $x \neq 0_{\mathbb{R}}$, a is not null, so from some index N on its terms are at least $c > 0$ in absolute value. Invert a_n for $n \geq N$, and put anything harmless, say 1, before N; Chapter 4 showed the result is again Cauchy (Proposition 4.5.6). The product with a is 1 from N on, and the first N terms do not matter. *Before reading on: the inverse is built*

from one representative a. Why does the class $[b]$ not depend on which representative of x we started from?

Proof. Let $x \in \mathbb{R}$ with $x \neq 0_{\mathbb{R}}$. By Sets, Relations, and Cardinality, Definition 2.5.4, $x = [a]$ for some $a \in \mathcal{C}$; fix one such a ($\exists$E, Logic and Proof, Definition 5.6.1). The sequence a is not null: otherwise $a \sim (0)_n$, and $x = [(0)_n] = 0_{\mathbb{R}}$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

By Theorem 4.5.4 (with $F = \mathbb{Q}$) there are $c \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ such that $a_n \geq c$ for every $n \geq N$, or $a_n \leq -c$ for every $n \geq N$. Fix one such pair (c, N) ($\exists$E again). This fixes two objects, once; it is not a selection from infinitely many sets. In either case $|a_n| \geq c$ for every $n \geq N$: if $a_n \geq c > 0$, then $|a_n| = a_n$; if $a_n \leq -c$, then $|a_n| = -a_n \geq c$ (Definition 4.1.1, Proposition 2.3.5 (a)).

By Proposition 4.5.6, the sequence b with $b_n := 1/a_n$ for $n \geq N$ and $b_n := 1$ for $n < N$ is in $\mathcal{C}$, and $a_n b_n = 1$ for every $n \geq N$. So the sequence $ab - (1)_n$ agrees with $(0)_n$ from N on, and it is null by Proposition 4.3.2 (a). Hence $ab \sim (1)_n$, and

$$x\,[b] = [a]\,[b] = [ab] = [(1)_n] = 1_{\mathbb{R}}.$$

So $y := [b]$ is an inverse of x, and (F8) holds. Together with Theorem 5.2.2, which gives (F1)–(F7), (F9) and $0_{\mathbb{R}} \neq 1_{\mathbb{R}}$, this shows that $\mathbb{R}$ is a field.

■

The answer to the prompt: **inverses are unique.** By the theorem, $\mathbb{R}$ is a field, so x has only one inverse (Lemma 4.1.1 (a) with $F = \mathbb{R}$); the reason is the one-line computation $y = y(xy') = (yx)y' = y'$ for two inverses y, y'. So starting from another representative a' of x, with its own c', N' and b', gives $[b'] = [b]$; we write x^{-1} for it.

Example. Termwise Reciprocals Are Not a Definition

A shorter route to inverses would be: for $x = [a] \neq 0_{\mathbb{R}}$, let $x^{-1} := [(1/a_n)_n]$. Explain why this is not a definition, and why $0_{\mathbb{R}}$ must be excluded in any case.

Solution. It need not make sense. The sequences $a := (0, 1, 1, 1, \dots)$ and $(1)_n$ agree from $n = 1$ on; a is in $\mathcal{C}$, since $|a_m - a_n| = 0$ for all $m, n \geq 1$ (threshold $N = 1$ for every tolerance), so the two sequences are equivalent (Example (Finite Changes and Shifts) (a)) and represent the same real number $1_{\mathbb{R}} \neq 0_{\mathbb{R}}$. With the representative a, the rule asks for $1/a_0 = 1/0$, which does not exist. A rule that cannot be applied to some representatives does not define a function on classes; it would need a statement of which representatives are allowed, and a proof that the answer does not depend on the allowed one. The proof of Theorem 5.2.3 supplies both at once: it inverts only from an index N on, and the uniqueness of inverses makes the class independent of all the choices.

Why $0_{\mathbb{R}}$ is excluded. The null sequence $(1/(n+1))_n$ represents $0_{\mathbb{R}}$ and has no zero term, so the termwise reciprocal $(n+1)_n$ exists. But it is not bounded (Theorem 2.5.3), so it is not in $\mathcal{C}$ (Proposition 4.5.2), and it represents no real number. No inverse of $0_{\mathbb{R}}$ is possible in any case, since $0_{\mathbb{R}} \cdot y = y \cdot 0_{\mathbb{R}} = 0_{\mathbb{R}} \neq 1_{\mathbb{R}}$ for every y ((F6), Proposition 2.3.2 (b), Theorem 5.2.2).

Warning. **A termwise formula is not an operation until it is checked on every representative.** "$[a]^{-1} := [(1/a_n)]$" fails on $(0, 1, 1, 1, \dots)$, a representative of $1_{\mathbb{R}}$, whose first term cannot be inverted. Every definition on $\mathbb{R}$ must say what it does to **each** representative of a class, and then prove that the class of the answer is the same for all of them.

Quick check. Find a representative of $[(1 + 1/(n+1))_n] \cdot [(2)_n] - [(2)_n]$, and name its class.

Solution. By Definition 5.2.2, with $x - y := x + (-y)$ in the ring $\mathbb{R}$, the real number is $[(1+1/(n+1))\cdot 2]+[(-2)_n] = [(2+2/(n+1)-2)_n]$, so $(2/(n+1))_n$ represents it. This sequence is null (Theorem 4.4.3 (b) with $c = 2$, applied to $(1/(n+1))$, which is null by Example (First Limits) (b)), so the class is $0_{\mathbb{R}}$. In other words, $[(1 + 1/(n+1))_n] \cdot [(2)_n] = [(2)_n]$. The first factor is $1_{\mathbb{R}}$, since $(1 + 1/(n+1))_n - (1)_n = (1/(n+1))_n$ is null, so $(1 + 1/(n+1))_n \sim (1)_n$; it is $1_{\mathbb{R}}$ represented by a sequence none of whose terms is 1.

$\mathbb{R}$ is now a field. It has no order yet: "positive" for a class of sequences needs care, and Section 5.3 defines it.

5.2.5 Exercises

5.2.5.1 A. Check your understanding

Exercise. A1

(a) Prove, directly from Definition 5.2.2 and the field axioms of $\mathbb{Q}$, that $([a][b])[c] = [a]([b][c])$ and $1_{\mathbb{R}}\,[a] = [a]$ for all $a, b, c \in \mathcal{C}$.

(b) In the product part of the proof of Theorem 5.2.1, which two of the four sequences a, a', b, b' must be bounded, and why is "b is bounded" not what Eq. 5.2.1 needs?

(c) Run that proof on $a = (2 + 1/(n+1))_n$, $a' = (2)_n$, $b = (x_n)_n$ (the Newton sequence) and $b' = (d_n)_n$ (the truncations): check that $a \sim a'$ and $b \sim b'$, write $a_n b_n - a'_n b'_n$ as in Eq. 5.2.1, and give a bound for each multiplier.

(d) State the idea of the proof of Theorem 5.2.1 in two sentences.

Solution.

(a) By Definition 5.2.2 twice, $([a][b])[c] = [ab][c] = [(ab)c]$ and $[a]([b][c]) = [a][bc] = [a(bc)]$. For every n, $(a_n b_n)c_n = a_n(b_n c_n)$ by (F5) in $\mathbb{Q}$, so $(ab)c = a(bc)$ as sequences (Sets, Relations, and Cardinality, Proposition 1.4.1), and the two classes are equal. Likewise $1_{\mathbb{R}}[a] = [(1)_n][a] = [(1 \cdot a_n)_n] = [a]$, by (F6) and (F7) in $\mathbb{Q}$.

(b) a and b'. In Eq. 5.2.1, a_n multiplies the null $b_n - b'_n$, and b'_n multiplies the null $a_n - a'_n$; Proposition 4.4.1 needs the multiplier bounded. The sequence b appears in Eq. 5.2.1 only inside the null factor $b_n - b'_n$, so its bound is never used.

(c) $a - a' = (1/(n+1))_n$ is null (Example (First Limits) (b)), so $a \sim a'$ (both are in $\mathcal{C}$ by Theorem 4.5.3 (a), (b)); $b \sim b'$ is Example (The Truncations and the Newton Sequence

Are Equivalent). The split is

$$\Big(2+\frac{1}{n+1}\Big)x_n - 2d_n = \Big(2+\frac{1}{n+1}\Big)(x_n - d_n) + \frac{1}{n+1}\,d_n.$$

Multipliers: $|2+1/(n+1)| \le 3$, since $0 < 1/(n+1) \le 1$; and $|d_n| \le 2$, since $1 \le d_n < 2$ (Steps 1 and 3 of the solution of Example (The Truncations and the Newton Sequence Are Equivalent)). So the first summand is null (null times bounded by 3) and the second is null (null times bounded by 2).

(d) Changing both factors of a product at once is split into two changes, one factor at a time, by adding and subtracting a mixed term. Each change is a null sequence times a bounded one, because Cauchy sequences are bounded, so each is null, and so is their sum.

5.2.5.2 B. Practice

Exercise. B1: The negative, directly

Without using Theorem 3.4.1, prove that the rule $-[a] := [-a]$ defines a function $\mathbb{R} \to \mathbb{R}$, and that $[a] + (-[a]) = 0_{\mathbb{R}}$ for every $a \in \mathcal{C}$. Deduce that $[a] - [b] = [a - b]$ for all $a, b \in \mathcal{C}$.

Solution. Well defined. $-a \in \mathcal{C}$ for $a \in \mathcal{C}$ (Theorem 4.5.3 (b)). Let $a \sim a'$. For every n, $(-a_n) - (-a'_n) = -(a_n - a'_n)$ (Proposition 2.3.2 (c)), and $(-(a_n - a'_n))_n$ is null because $a - a'$ is (Theorem 4.4.3 (b)). So $-a \sim -a'$, and Theorem 2.2.1 (a) gives a function $\mathbb{R} \to \mathbb{R}$.

Inverse. $[a] + (-[a]) = [a] + [-a] = [a + (-a)] = [(a_n + (-a_n))_n] = [(0)_n] = 0_{\mathbb{R}}$, by Definition 5.2.2 and (F4) in $\mathbb{Q}$ at each index.

Differences. $[a] - [b] = [a] + (-[b]) = [a] + [-b] = [a + (-b)] = [a - b]$, by the definition $x - y := x + (-y)$ in a ring and Definition 5.2.2.

Exercise. B2: The larger of two reals, faded

For $x, y \in \mathbb{Q}$ let $\max(x, y) := x$ if $x \ge y$, and $\max(x, y) := y$ otherwise. For $a, b \in \mathcal{C}$ consider the rule

$$g([a], [b]) := [(\max(a_n, b_n))_n].$$

(a) Prove that $|\max(x, y) - \max(x', y')| \le |x - x'| + |y - y'|$ for all $x, y, x', y' \in \mathbb{Q}$. *Hint:* $\max(x, y)$ *is one of* x, y, *and is* $\ge$ *both.*

(b) Deduce from (a) that $(\max(a_n, b_n))_n \in \mathcal{C}$ whenever $a, b \in \mathcal{C}$.

(c) Here is the start of a proof that the rule g is well defined. *Name the move, then complete it.*
"Let $a, a', b, b' \in \mathcal{C}$ with $a \sim a'$ and $b \sim b'$. By (b), the sequences $(\max(a_n, b_n))_n$ and $(\max(a'_n, b'_n))_n$ are in $\mathcal{C}$. Let $e_n := \max(a_n, b_n) - \max(a'_n, b'_n)$. By (a), for every n,

$$0 \le |e_n| \le |a_n - a'_n| + |b_n - b'_n|.$$

⟨ blank ⟩"

Solution.

(a) Write $M := \max(x, y)$ and $M' := \max(x', y')$. Note $M \geq x$, $M \geq y$, and M is one of x, y; likewise for M'. Suppose first $M \geq M'$. If $M = x$, then $0 \leq M - M' \leq x - x' \leq |x - x'|$, since $M' \geq x'$ (Proposition 2.3.5 (b), Proposition 4.1.3 (c)). If $M = y$, then likewise $0 \leq M - M' \leq y - y' \leq |y - y'|$. Either way $|M - M'| = M - M' \leq |x - x'| + |y - y'|$. If $M' > M$, exchange the roles of (x, y) and (x', y'); the right-hand side is unchanged, as $|x' - x| = |x - x'|$.

(b) Let $\varepsilon \in \mathbb{Q}_{>0}$, and let N be a threshold with tolerance $\varepsilon/2$ for both a and b (the larger of the two thresholds). For $m, n \geq N$, (a) with $(x, y, x', y') = (a_m, b_m, a_n, b_n)$ gives $|\max(a_m, b_m) - \max(a_n, b_n)| \leq |a_m - a_n| + |b_m - b_n| < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

(c) *The move:* representatives, then independence; here its last step, "a sequence squeezed between 0 and a null sequence is null, so the outputs are equivalent and the classes agree". *Completion:* The sequences $(|a_n - a'_n|)_n$ and $(|b_n - b'_n|)_n$ are null, since $||t| - 0| = |t - 0|$ for every $t \in \mathbb{Q}$ (Definition 4.3.3). So their sum is null (Theorem 4.4.3 (a)). By the squeeze Proposition 4.4.4 (c), with lower sequence $(0)_n$ and this upper sequence, $(|e_n|)_n$ is null; hence $(e_n)_n$ is null, by the same remark on $|e_n - 0|$. So $(\max(a_n, b_n))_n \sim (\max(a'_n, b'_n))_n$. Taking $b' = b$, and then $a' = a$, gives the two hypotheses of Theorem 2.2.1 (b), so g is a function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Exercise. B3: A square root of two

Let (x_n) be the Newton sequence. Show that $[(x_n)] \cdot [(x_n)] = [(2)_n]$.

Solution. By Definition 5.2.2, $[(x_n)] \cdot [(x_n)] = [(x_n^2)_n]$, so we must show that $(x_n^2 - 2)_n$ is null. By the bound $0 < x_n^2 - 2 \leq 1/(4n)$ for $n \geq 1$ (Section 5.1, (NS1)–(NS2)), and $1/(4n) \leq 1/(n+1)$ for $n \geq 1$ (since $4n - (n+1) = 3n - 1 > 0$; Proposition 4.1.2 (c)),

$$0 \leq x_n^2 - 2 \leq \frac{1}{n+1} \qquad (n \geq 1).$$

The squeeze Proposition 4.4.4 (c), with $N_0 = 1$, lower sequence $(0)_n$ and the null upper sequence $(1/(n+1))_n$ (Example (First Limits) (b)), defined for every $n \in \mathbb{N}$, shows that $(x_n^2 - 2)_n$ is null. So $(x_n^2) \sim (2)_n$ and the classes are equal. (The inequality fails at $n = 0$, where $x_0^2 - 2 = 2$; the squeeze does not need it there.) Section 5.4 returns to this class.

5.2.5.3 C. Going deeper

Exercise. C1: Exit check: multiplication is well defined

State precisely what must be proved for the rule $[a][b] := [ab]$ to define a function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and prove it, including the fact that $ab \in \mathcal{C}$ for $a, b \in \mathcal{C}$.

Solution. What must be shown. (i) For $a, b \in \mathcal{C}$, the termwise product ab is in $\mathcal{C}$, so that $[ab]$ is a real number. (ii) If $a \sim a'$ and $b \sim b'$, then $ab \sim a'b'$. By Theorem 2.2.1 (b), (i) and (ii) give exactly one function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $([a], [b]) \mapsto [ab]$; without (ii), one pair of classes could receive two values, and (Fn2) of Sets, Relations, and Cardinality, Definition 1.4.1 would

fail.

(i) Bound, then shrink. Let $a, b \in \mathcal{C}$. Both are bounded (Proposition 4.5.2); let $M \in \mathbb{Q}_{>0}$ be a bound for both (the larger of the two bounds). For all m, n,

$$\begin{aligned} |a_m b_m - a_n b_n| &= |a_m(b_m - b_n) + (a_m - a_n)b_n| \\ &\leq M|b_m - b_n| + M|a_m - a_n|, \end{aligned}$$

by Proposition 4.1.3 (b), (e). Let $\varepsilon \in \mathbb{Q}_{>0}$, and let N be a threshold with tolerance $\varepsilon/(2M)$ for both a and b. For $m, n \geq N$, the right side is less than $M \cdot \varepsilon/(2M) + M \cdot \varepsilon/(2M) = \varepsilon$. So $ab \in \mathcal{C}$. (This is Theorem 4.5.3 (c), which may be cited instead.)

(ii) The add-and-subtract decomposition. For every n,

$$a_n b_n - a'_n b'_n = a_n(b_n - b'_n) + (a_n - a'_n)b'_n.$$

The multipliers a and b' are in $\mathcal{C}$, hence bounded (Proposition 4.5.2). The sequences $b - b'$ and $a - a'$ are null. By Proposition 4.4.1, both summands are null, and by the sum rule (Theorem 4.4.3 (a)) so is $ab - a'b'$. Hence $ab \sim a'b'$.

Marking scheme. (10 points)

- 2: states what must be shown (the product lies in $\mathcal{C}$; equivalent inputs give equivalent outputs) and why this makes the rule a function.
- 2: proves $ab \in \mathcal{C}$, by "bound, then shrink" with the threshold's dependence on ε and M clear, or by citing Theorem 4.5.3 (c).
- 2: the add-and-subtract decomposition, or any identity writing $a_n b_n - a'_n b'_n$ as a sum of (null) $\times$ (sequence) terms.
- 2: boundedness of the two multipliers, justified by "Cauchy implies bounded".
- 2: concludes with null $\times$ bounded is null and null + null is null.

A proof that changes one factor at a time (first $ab \sim a'b$, then $a'b \sim a'b'$) earns the same points; the multipliers are then b and a'. A proof through the quotient ring is also valid: show that $\mathcal{N}$ is an ideal of the commutative ring $\mathcal{C}$, and cite Theorem 3.4.1 (b). It earns the first two items as above; the decomposition item for identifying $\sim$ with $\sim_{\mathcal{N}}$ and proving (I1) and (I2); the multiplier item for "a Cauchy multiplier is bounded" (Proposition 4.5.2); and the last item for (I3), null $\times$ Cauchy is null, together with a correct citation of the quotient theorem. The ideal clauses must be proved, not cited from Theorem 5.2.2. Fatal errors (0 for the step): "both differences are small, so the product difference is small" with no bound on the multipliers; a bound for b used where b' is the multiplier.

Exercise. C2: No square is zero except zero

(a) Let a be a sequence of rationals such that $(a_n^2)_n$ is null. Prove that a is null.

(b) Deduce, using only Definition 5.2.2 and (a), and not Theorem 5.2.3, that for $x \in \mathbb{R}$, $x^2 = 0_{\mathbb{R}}$ implies $x = 0_{\mathbb{R}}$.

Solution.

(a) Let $\varepsilon \in \mathbb{Q}_{>0}$. Then $\varepsilon^2 \in \mathbb{Q}_{>0}$, so there is N with $|a_n^2| < \varepsilon^2$ for every $n \geq N$. For such n, $|a_n|^2 = a_n^2 = |a_n^2| < \varepsilon^2$ (Proposition 4.1.3 (b), and $a_n^2 \geq 0$), and $|a_n| \geq 0$, $\varepsilon > 0$; so

$|a_n| < \varepsilon$ by Proposition 2.3.5 (g). The threshold for ε is the threshold of (a_n^2) for ε^2. So a is null.

(b) Let $x = [a]$ with $a \in \mathcal{C}$. Then $x^2 = [a][a] = [(a_n^2)_n]$ by Definition 5.2.2. If $x^2 = 0_{\mathbb{R}} = [(0)_n]$, then $(a_n^2)_n \sim (0)_n$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)), that is, $(a_n^2)_n$ is null. By (a), a is null, so $a \sim (0)_n$ and $x = 0_{\mathbb{R}}$.

In a field this also follows from inverses: if $x \neq 0_{\mathbb{R}}$, then $x = x^{-1}x^2 = x^{-1}0_{\mathbb{R}} = 0_{\mathbb{R}}$, a contradiction. The direct proof shows the fact is already visible in the representatives, before any inverse exists. It fails in some commutative rings: in $\mathbb{Z}/4\mathbb{Z}$ of Chapter 3 (Definition 3.4.2), $[2]^2 = [4] = [0]$ while $[2] \neq [0]$.

5.3 Positive Classes and the Order

Section 5.2 made $\mathbb{R}$ a field: its elements are classes of Cauchy sequences of rationals, added and multiplied through representatives. A field is not yet a number line. To compare two real numbers we need an order, and it must cooperate with the arithmetic: $\mathbb{R}$ should be an ordered field in the sense of Definition 2.3.2. Chapter 2 ordered $\mathbb{Q}$ in two steps. It first said which rationals are positive, and then let $a < b$ mean that $b - a$ is positive (Proposition 2.3.3). This section takes the same two steps for $\mathbb{R}$. All the work is in the first: deciding which classes are positive, in a way that does not depend on the representative.

Throughout, $\mathcal{C}$ is the set of Cauchy sequences of rationals (Definition 5.1.1), $a \sim b$ means that $a - b$ is null in $\mathbb{Q}$ (Definition 5.1.2), and $[a]$ is the class of a, an element of $\mathbb{R}$ (Definition 5.2.1). Until $\mathbb{R}$ has an order, every tolerance, margin and threshold is a rational or a natural number: we are still building $\mathbb{R}$, and may not measure anything with it. The operations are those of Definition 5.2.2: $[a] + [b] = [a + b]$, $[a][b] = [ab]$, $-[a] = [-a]$, with $0_{\mathbb{R}} = [(0)_n]$ and $1_{\mathbb{R}} = [(1)_n]$, where $(q)_n$ is the constant sequence with value q. In particular

$$[a] - [b] = [a] + [-b] = [a - b],$$

since $a + (-b) = a - b$ termwise. We use this without comment.

5.3.1 Comparing terms does not compare classes

The first idea is to compare two classes by comparing their terms. It fails, and the way it fails shows what the definition must contain.

Example. Termwise Order Is Not an Order on Classes

Consider the rule

> $[a] < [b]$ means that $a_n < b_n$ for every n from some index on.

Let $a_n := 0$ and $b_n := 1/(n+1)$ for $n \in \mathbb{N}$. Show that $[a] = [b]$. Then apply the rule to the classes $[a]$ and $[a]$ twice: once through the representatives a and a, once through a and b. Conclude that the rule does not define a relation on $\mathbb{R}$.

Solution. Both sequences are in $\mathcal{C}$: a constant sequence and a null sequence are Cauchy (Theorem 4.5.3 (a)), and $(1/(n+1))$ is null in $\mathbb{Q}$ (Example (First Limits) (b)). Termwise, $b - a = (1/(n+1))$, which is null, so $b \sim a$, and $[a] = [b]$ by Sets, Relations, and Cardinality, Theorem 2.5.1 (b).

Through the representatives a and a: the rule asks whether $0 < 0$ from some index on, and $0 < 0$ never holds. So the rule says that $[a] < [a]$ is false.

Through the representatives a and b, which name the same two classes, since $b \in [b] = [a]$: for **every** n, $a_n = 0 < 1/(n+1) = b_n$. So the rule says that $[a] < [a]$ is true.

A relation on $\mathbb{R}$ must give one answer for each pair of classes. The rule gives two answers for the pair $(0_{\mathbb{R}}, 0_{\mathbb{R}})$, depending on the representatives, so it defines no relation on $\mathbb{R}$. Worse, the second answer, $0_{\mathbb{R}} < 0_{\mathbb{R}}$, is excluded by trichotomy (O1) in any ordered field.

The trouble is that the gap between a_n and b_n is allowed to shrink to nothing, and a gap that shrinks to nothing is exactly what $\sim$ ignores. A class that deserves to be called positive must have terms that stay above 0 by a **fixed** margin.

5.3.2 Positive sequences

We look for a property of a single sequence $a \in \mathcal{C}$ that says "the number a names is greater than 0". The example shows that "eventually $a_n > 0$" is too weak, because a null sequence can have it. The repair is to name the margin first.

Positive means: eventually above one fixed positive rational.

Definition 5.3.1. Positive Cauchy Sequence

A sequence $a \in \mathcal{C}$ is **positive** if **there exist** $\delta \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ such that $a_n \geq \delta$ **for every** $n \in \mathbb{N}$ with $n \geq N$. In symbols,

$$\exists \delta \in \mathbb{Q}_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall n \in \mathbb{N} \;\; (n \geq N \Rightarrow a_n \geq \delta).$$

In words: the first quantifier names a margin δ, a positive rational, before anything else is named; the second names a threshold; the third says that **every** term from the threshold on is at least the margin. Because δ comes first, one margin serves all the late terms at once. The word "positive" describes the sequence as a name for a number, not its terms one by one: the terms before N may be negative, and, as the warning below shows, terms that are all > 0 are not enough.

Compare the definition with the weaker condition "the terms are eventually positive":

$$\exists N \in \mathbb{N} \;\; \forall n \in \mathbb{N} \;\; (n \geq N \Rightarrow a_n > 0).$$

Here each late term a_n is above 0 by its own amount, namely a_n itself, and these amounts may shrink as n grows. In the definition the amount is chosen before n, so it cannot depend on n. The two conditions differ only in the order of "there is a margin" and "for every late n", and that order is the whole point.

Examples.

A positive constant. For $q \in \mathbb{Q}_{>0}$, the constant sequence $(q)_n$ is positive, with $\delta = q$ and $N = 0$. This degenerate case is how Section 5.4 will see that the rational q, placed inside $\mathbb{R}$, is greater than 0.

The Newton sequence. Let (x_n) be the sequence of Example (A Rational Sequence Aimed at the Gap), which is in $\mathcal{C}$ by Theorem 4.6.1. The same theorem states that $x_n > 1$ for every $n \in \mathbb{N}$. So (x_n) is positive, with $\delta = 1$ and $N = 0$.

Late positivity. Let $a_n := 1 - 2/(n+1)$. It is in $\mathcal{C}$, as the difference of a constant sequence and a constant multiple of the null sequence $(1/(n+1))$ (Theorem 4.5.3 (a), (b)). Its first terms are $-1, 0, 1/3, 1/2$, so it is not positive "from the start". For $n \geq 3$, $n+1 \geq 4$, so $1/(n+1) \leq 1/4$ (Proposition 4.1.2 (c)), hence $2/(n+1) \leq 2/4 = 1/2$ (Proposition 4.1.2 (d), with $0 \leq 2 \leq 2$), and $a_n \geq 1/2$. So a is positive, with $\delta = 1/2$ and $N = 3$. No smaller N works with this δ, since $a_2 = 1/3 < 1/2$; the definition allows any finite stretch of bad terms.

Non-example by a minimal change. Keep every term positive, but let the margin shrink: $a_n := 1/(n+1)$. Every term is > 0, so the weaker condition above holds with $N = 0$. What fails is the order of the quantifiers. The negation of the definition, by Logic and Proof, Proposition 6.2.3 and Logic and Proof, Proposition 6.2.2, reads

$$\forall \delta \in \mathbb{Q}_{>0} \;\; \forall N \in \mathbb{N} \;\; \exists n \in \mathbb{N} \;\; (n \geq N \wedge a_n < \delta),$$

and it holds: given δ and N, Corollary 2.5.4 gives $n_0 \in \mathbb{Z}_{>0}$ with $1/m < \delta$ for every $m \geq n_0$. With $n := \max\{N, n_0\}$, we have $n \geq N$, and $m := n+1 \geq n_0$, so $a_n = 1/(n+1) < \delta$. Every margin is eventually undercut.

Why this definition. Each clause has a job.

- *The margin δ.* Without it, the definition would be the weaker condition, and Example (Termwise Order Is Not an Order on Classes) shows the result: $(1/(n+1))$ would be positive, while $(0)_n$, which names the same number $0_{\mathbb{R}}$, is not.
- *A rational margin.* The only positive numbers available are rationals. A real margin would use the order of $\mathbb{R}$, which is what we are defining.
- *The threshold N.* Changing finitely many terms does not change the class (Example (Finite Changes and Shifts)), so a definition meant for classes must ignore every finite stretch of terms. The late-positivity example needs $N = 3$.
- *$\geq$ rather than $>$.* Either would do: $a_n \geq \delta$ gives $a_n > \delta/2$, and $a_n > \delta$ gives $a_n \geq \delta$. We use $\geq$, the form in which Theorem 4.5.4 delivers its conclusion.

Warning. "Every term of a is positive" does **not** make a positive. The sequence $(1/(n+1))$ has every term > 0, and it is not positive (non-example above); its class is $0_{\mathbb{R}}$.

Quick check. Let $a \in \mathcal{C}$, with δ ranging over $\mathbb{Q}_{>0}$. Which of the two conditions

$$\text{(A) } \exists \delta \; \exists N \; \forall n \geq N \;\; a_n \geq \delta, \qquad \text{(B) } \exists N \; \forall n \geq N \; \exists \delta \;\; a_n \geq \delta$$

implies the other? Which of them says the same as "eventually $a_n > 0$"?

Solution.

(A) implies (B): the margin found once serves as the δ for every late n. (B) does not imply (A): $(1/(n+1))$ satisfies (B), with $\delta := 1/(n+1)$ chosen after n, and fails (A). (B) says the same as "eventually $a_n > 0$": if $a_n > 0$, take $\delta := a_n$; if $a_n \geq \delta > 0$, then $a_n > 0$. So (A) is the definition of positive, and (B) is the weaker condition.

Payoff: positivity belongs to the class. A real number is a class, and a definition about sequences says something about the class only if equivalent sequences agree on it. Chapter 4

proved the fact we need while preparing for this chapter: a margin survives a null change of the sequence, at half its size (Proposition 4.5.5).

Proposition 5.3.1. Positivity Is a Property of the Class

Let $a, a' \in \mathcal{C}$ with $a' \sim a$. If a is positive, then a' is positive.

Proof. Since a is positive, there are $\delta \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ with $a_n \geq \delta$ for every $n \geq N$; fix them [∃E] (Logic and Proof, Definition 5.6.1). Since $\sim$ is symmetric (Proposition 5.1.1), $a \sim a'$, that is, $(a_n - a'_n)$ is null in $\mathbb{Q}$. Proposition 4.5.5, with $F = \mathbb{Q}$, the sequences a and a', and $c = \delta$, gives $N' \in \mathbb{N}$ with $a'_n \geq \delta/2$ for every $n \geq N'$. Since $\delta/2 \in \mathbb{Q}_{>0}$ (Proposition 4.1.2 (b)), a' is positive, with margin $\delta/2$ and threshold N'. ■

The margin halves, and that is harmless: the definition asks only for **some** margin. Exercise C2 shows a tempting proof of this proposition that keeps a margin of 0, and so does not prove it.

We can now name the set of positive real numbers. Let

$$P := \{[a] : a \in \mathcal{C} \text{ and } a \text{ is positive}\} \subseteq \mathbb{R}.$$

Chapter 2 built the positive sets $P_{\mathbb{Z}}$ and $P_{\mathbb{Q}}$; P is the same step for $\mathbb{R}$. For $x \in \mathbb{R}$:

$$x \in P \iff \text{some representative of } x \text{ is positive} \iff \text{every representative of } x \text{ is positive.}$$

The first equivalence is the definition of P. For the second, every class has a representative (Sets, Relations, and Cardinality, Theorem 2.5.1 (a)), which gives "every ⇒ some". Conversely, if $x = [a]$ with a positive and $b \in x$, then $b \sim a$ (Sets, Relations, and Cardinality, Definition 2.5.2), so b is positive by Proposition 5.3.1. Membership in P can therefore be tested on **any** representative we like. This is the well-definedness the termwise rule lacked.

5.3.3 The positive classes

Chapter 2's recipe needs two facts about P: it is closed under sum and product, and it splits $\mathbb{R}$ into three parts, P, $\{0_{\mathbb{R}}\}$ and the negatives of P. These are clauses (P1) and (P2) of Proposition 2.3.3.

Theorem 5.3.2. The Positive Real Numbers

(i) For every $x \in \mathbb{R}$, **exactly one** of $x \in P$, $x = 0_{\mathbb{R}}$, $-x \in P$ holds.

(ii) If $x \in P$ and $y \in P$, then $x + y \in P$ and $xy \in P$.

Idea. For (i), "at least one" is the sign theorem of Chapter 4: a Cauchy sequence that is not null is eventually above c or eventually below $-c$ (Theorem 4.5.4). A null representative gives $0_{\mathbb{R}}$. "At most one" holds because a sequence cannot be eventually above a positive margin and also eventually below a negative one, or null. For (ii), add or multiply the two margins and take the later of the two thresholds.

Proof. Let $x \in \mathbb{R}$, and write $x = [a]$ with $a \in \mathcal{C}$.

(i), *at least one.* Suppose first that a is null. Termwise $a - (0)_n = a$, so $a \sim (0)_n$ and $x = [(0)_n] = 0_{\mathbb{R}}$. Suppose next that a is not null. By Theorem 4.5.4 with $F = \mathbb{Q}$, there are $c \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ such that either $a_n \geq c$ for every $n \geq N$, or $a_n \leq -c$ for every $n \geq N$; fix them [∃E]. In the first case a is positive, so $x \in P$. In the second case $-a_n \geq c$ for every $n \geq N$ (Proposition 2.3.5 (a)), so $-a$ is positive and $-x = [-a] \in P$.

(i), *at most one.* We exclude the three pairs.

- $x \in P$ and $x = 0_{\mathbb{R}}$. Then $(0)_n$ is a representative of x, so it is positive, since every representative of an element of P is: some $\delta \in \mathbb{Q}_{>0}$ has $0 \geq \delta$. This contradicts $\delta > 0$ by (O1) in $\mathbb{Q}$.
- $-x \in P$ and $x = 0_{\mathbb{R}}$. Then $-x = [-(0)_n] = [(0)_n] = 0_{\mathbb{R}}$, and the previous case, applied to $-x$, excludes this.
- $x \in P$ and $-x \in P$. Then a and $-a$ are both positive: $a_n \geq \delta$ for $n \geq N$ and $-a_n \geq \delta'$ for $n \geq N'$, with $\delta, \delta' \in \mathbb{Q}_{>0}$. At $n := \max\{N, N'\}$, $\delta \leq a_n \leq -\delta'$ (Proposition 2.3.5 (a)), so $0 < \delta \leq -\delta' < 0$, against (O1) in $\mathbb{Q}$.

(ii) Let $x, y \in P$, and write $x = [a]$, $y = [b]$. Both a and b are positive: there are $\delta_a, \delta_b \in \mathbb{Q}_{>0}$ and $N_a, N_b \in \mathbb{N}$ with $a_n \geq \delta_a$ for $n \geq N_a$ and $b_n \geq \delta_b$ for $n \geq N_b$. Let $N := \max\{N_a, N_b\}$ and $n \geq N$. Adding the two inequalities (Proposition 2.3.5 (b)),

$$a_n + b_n \geq \delta_a + \delta_b,$$

and $\delta_a + \delta_b > 0 + 0 = 0$ by the same rule. Multiplying, by Proposition 4.1.2 (d) applied to $0 \leq \delta_a \leq a_n$ and $0 \leq \delta_b \leq b_n$,

$$a_n b_n \geq \delta_a \delta_b,$$

and $\delta_a \delta_b > 0$ by (O4). So $a + b$ and ab are positive, with margins $\delta_a + \delta_b$ and $\delta_a \delta_b$ and the threshold N. Hence $x + y = [a + b] \in P$ and $xy = [ab] \in P$ (Definition 5.2.2).

■

Both parts rest on a choice of representatives, and neither depends on it: membership in P can be tested on any representative, which is what Proposition 5.3.1 bought. In (ii), the margins and the threshold depend on a and b, and on nothing else.

5.3.4 The order of the reals

With P in hand the definition is the one Chapter 2 used for $\mathbb{Z}$ and $\mathbb{Q}$.

Definition 5.3.2. Order on the Reals

For $x, y \in \mathbb{R}$, $x < y$ means $y - x \in P$. As usual, $x \leq y$ means $x < y$ or $x = y$, and $y > x$ means $x < y$.

In terms of representatives: $[a] < [b]$ exactly when $b - a$ is positive, that is, when $b_n - a_n \geq \delta$ from some index on, for some margin $\delta \in \mathbb{Q}_{>0}$. The terms of b must eventually overtake those of a by a fixed rational amount. Since $0_{\mathbb{R}} - x = -x$ and $x - 0_{\mathbb{R}} = x$, we have $x > 0_{\mathbb{R}}$ if and only if $x \in P$, and $x < 0_{\mathbb{R}}$ if and only if $-x \in P$.

Theorem 5.3.3. The Reals Form an Ordered Field

With the operations of Definition 5.2.2 and the order of Definition 5.3.2, $\mathbb{R}$ is an ordered field (Definition 2.3.2), and its positive elements are exactly the elements of P.

Idea. Each order axiom is a statement about one difference, and Theorem 5.3.2 is exactly what Proposition 2.3.3 asks for. (O1) is (i) applied to $y-x$, since $-(y-x)=x-y$. (O2) is closure under sums, since $(z-y)+(y-x)=z-x$. (O3) holds because $(y+z)-(x+z)=y-x$. (O4) is closure under products.

Proof. By Theorem 5.2.3, $\mathbb{R}$ is a field, in particular a commutative ring (Definition 2.3.1). By Theorem 5.3.2, the subset $P \subseteq \mathbb{R}$ satisfies clause (P1) of Proposition 2.3.3 (part (ii)) and clause (P2) (part (i)). That proposition then says that the relation "$y-x \in P$", which is $x<y$ of Definition 5.3.2, makes $\mathbb{R}$ an ordered commutative ring with $P=\{x\in\mathbb{R} : x>0_{\mathbb{R}}\}$. An ordered commutative ring that is a field is an ordered field (Definition 2.3.2).

■

So $\mathbb{R}$ is an ordered field, and every result that Chapters 2 and 4 proved for an arbitrary ordered field holds in it. We write $\mathbb{R}_{>0} \coloneqq \{x\in\mathbb{R} : x>0_{\mathbb{R}}\}$, the set $F_{>0}$ of Chapter 4 for $F=\mathbb{R}$; by the theorem, $\mathbb{R}_{>0}=P$. The rules for inequalities (Proposition 2.3.5), the rules for halves (Proposition 4.1.2), the absolute value (Definition 4.1.1, Proposition 4.1.3), and limits, null sequences and Cauchy sequences with tolerances in $\mathbb{R}_{>0}$ (Definition 4.3.2, Definition 4.5.1) are now available in $\mathbb{R}$. From here on they may be applied with $F=\mathbb{R}$. It pays to name the field at each such use, so that a tolerance in $\mathbb{Q}_{>0}$ is never confused with one in $\mathbb{R}_{>0}$.

One consequence deserves a line. In any ordered field, $x\le y$ holds if and only if $y<x$ fails: by (O1), exactly one of $x<y$, $x=y$, $y<x$ holds.

5.3.5 Comparing classes through their terms

The order is defined through margins, but in practice we usually know a termwise inequality and want an inequality between classes. The next lemma is the bridge, and the example at the start of the section shows its limit.

Lemma 5.3.4. Termwise Inequalities Pass to Classes

Let $a,b\in\mathcal{C}$ and $N_0\in\mathbb{N}$. If $a_n\le b_n$ for every $n\ge N_0$, then $[a]\le[b]$.

Idea. Suppose instead that $[b]<[a]$. Then $a-b$ is positive, so from some index on a_n exceeds b_n by a fixed margin; one index beyond both thresholds then contradicts $a_n\le b_n$. By trichotomy, "not $[b]<[a]$" is "$[a]\le[b]$".

Proof. Suppose that $[b]<[a]$. By Definition 5.3.2, $[a]-[b]=[a-b]\in P$, so $a-b$ is positive, since every representative of an element of P is (Proposition 5.3.1). Fix $\delta\in\mathbb{Q}_{>0}$ and $N\in\mathbb{N}$

with $a_n - b_n \geq \delta$ for every $n \geq N$ [$\exists$E]. Let $n \coloneqq \max\{N, N_0\}$. Then $a_n - b_n \geq \delta > 0$. But $n \geq N_0$, so $a_n \leq b_n$, that is, $a_n - b_n \leq 0$ (Proposition 2.3.5 (a)). These contradict (O1) in $\mathbb{Q}$. Hence $[b] < [a]$ is false, by $\neg$I (Logic and Proof, Definition 2.2.1). By (O1) in the ordered field $\mathbb{R}$ (Theorem 5.3.3), $[a] < [b]$ or $[a] = [b]$; that is, $[a] \leq [b]$.

■

The proof opens its box with "$[b] < [a]$" and closes it with the negation of that statement, which is $\neg$I. It is not reductio ad absurdum (Logic and Proof, Definition 2.5.1), whose box opens with the negation of the goal. The step from "$[b] < [a]$ is false" to "$[a] \leq [b]$" is trichotomy in $\mathbb{R}$, a property of this particular order, not a rule of logic.

Warning. The strict version of the lemma is **false**. In Example (Termwise Order Is Not an Order on Classes), $a_n = 0 < 1/(n+1) = b_n$ for **every** n, and yet $[a] = [b]$, so $[a] < [b]$ fails. A strict termwise inequality passes to classes only as $\leq$. To get $<$ between classes, show that the difference is eventually above a fixed margin.

Most uses of the lemma compare a class with a constant: if $a_n \leq q$ for every $n \geq N_0$, then $[a] \leq [(q)_n]$; and if $a_n \geq q$ for every $n \geq N_0$, then $[a] \geq [(q)_n]$.

Quick check. Let $a_n \coloneqq 1/(n+1) - 1/(n+1)^2$. Every term satisfies $a_n \geq 0$. Is a positive?

Solution. No. Here $a_n = n/(n+1)^2$, so $0 \leq a_n \leq 1/(n+1)$ for every n (as $n \leq n+1$). The lower sequence $(0)_n$ and the upper sequence $(1/(n+1))$ are null, so a is null by the squeeze, Proposition 4.4.4 (c) with $F = \mathbb{Q}$; being null, $a \in \mathcal{C}$ (Theorem 4.5.3 (a)). Then $[a] = 0_{\mathbb{R}}$, as in the proof of Theorem 5.3.2 (i), and $0_{\mathbb{R}} \notin P$ by the same part. So a is not positive, although no term is negative. Lemma 5.3.4 gives only $[a] \geq 0_{\mathbb{R}}$, which is true.

5.3.6 The absolute value through representatives

Since $\mathbb{R}$ is an ordered field, every real number x has an absolute value $|x|$, by Definition 4.1.1 with $F = \mathbb{R}$: x if $x \geq 0_{\mathbb{R}}$, and $-x$ if $x < 0_{\mathbb{R}}$. The proofs of the next section estimate $|[a]-[b]|$, and for that we want to compute an absolute value on representatives, the way sums and products are computed.

Proposition 5.3.5. The Absolute Value Is Computed Termwise

Let $a \in \mathcal{C}$. Then $(|a_n|)_n \in \mathcal{C}$, and

$$|[a]| = [(|a_n|)_n],$$

where the left side is the absolute value in the ordered field $\mathbb{R}$ and $|a_n|$ is the absolute value in $\mathbb{Q}$.

Idea. The reverse triangle inequality makes $(|a_n|)$ Cauchy with the same thresholds as a. For the equality, sort a by Theorem 4.5.4: null, eventually positive by a margin, or eventually

negative by a margin. In each case $(|a_n|)$ agrees from some index on with a, with $-a$, or is null, and that settles which of $[a]$, $-[a]$, $0_{\mathbb{R}}$ it represents.

Proof. Membership. Let $\varepsilon \in \mathbb{Q}_{>0}$, and choose N with $|a_m - a_n| < \varepsilon$ for all $m, n \geq N$. By the reverse triangle inequality in $\mathbb{Q}$ (Proposition 4.1.3 (f)), $||a_m| - |a_n|| \leq |a_m - a_n| < \varepsilon$ for the same m, n. So $(|a_n|) \in \mathcal{C}$, with the same thresholds as a.

Equality. We use three facts. A class $[b]$ with b eventually equal to b' equals $[b']$ (Example (Finite Changes and Shifts)). A null sequence represents $0_{\mathbb{R}}$ (proof of Theorem 5.3.2 (i)). And, in the ordered field $\mathbb{R}$, $|x| = x$ for $x > 0_{\mathbb{R}}$, $|x| = -x$ for $x < 0_{\mathbb{R}}$, and $|0_{\mathbb{R}}| = 0_{\mathbb{R}}$ (Definition 4.1.1).

Case 1: a is null. Then $[a] = 0_{\mathbb{R}}$, so $|[a]| = 0_{\mathbb{R}}$. The sequence $(|a_n|)$ is null too, since $||a_n| - 0| = |a_n| = |a_n - 0|$ (Proposition 4.1.3 (a) and Definition 4.1.1), so the two sequences pass the same tests of Definition 4.3.3. Hence $[(|a_n|)_n] = 0_{\mathbb{R}} = |[a]|$.

Case 2: a is not null. By Theorem 4.5.4 with $F = \mathbb{Q}$, fix $c \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ with $a_n \geq c$ for every $n \geq N$, or with $a_n \leq -c$ for every $n \geq N$ [$\exists$E].

If $a_n \geq c$ for $n \geq N$, then a is positive, so $[a] \in P$ and $[a] > 0_{\mathbb{R}}$, whence $|[a]| = [a]$. For $n \geq N$, $a_n > 0$, so $|a_n| = a_n$. The sequences $(|a_n|)$ and a agree from N on, so $[(|a_n|)_n] = [a] = |[a]|$.

If $a_n \leq -c$ for $n \geq N$, then $-a$ is positive, so $-[a] = [-a] \in P$ and $[a] < 0_{\mathbb{R}}$, whence $|[a]| = -[a] = [-a]$. For $n \geq N$, $a_n < 0$, so $|a_n| = -a_n$. The sequences $(|a_n|)$ and $-a$ agree from N on, so $[(|a_n|)_n] = [-a] = |[a]|$.

■

The proposition does **not** say that $|a_n| = a_n$ or $-a_n$ for all n; a representative of a positive number may have negative terms early on, as the late-positivity example shows. It says that the class of $(|a_n|)$ is the right one, and classes do not see early terms. Section 5.4 uses the proposition to estimate the distance from a real number $[a]$ to the rational numbers a_k, a term at a time.

5.3.7 Exercises

5.3.7.1 A. Check your understanding

Exercise. A1

In parts (a)–(d), decide for each sequence whether it is in $\mathcal{C}$ and whether it is positive; when it is positive, give a margin δ and a threshold N.

(a) $a_n = 1 - 1/(n+1)$.

(b) $a_n = 1/(n+1)$.

(c) $a_k = d_k - 1$, where (d_k) is the sequence of decimal truncations of Example (The Truncations and the Newton Sequence Are Equivalent).

(d) $a_n = (-1)^n$.

(e) Show from Definition 5.3.2 and Definition 5.3.1 that $1_{\mathbb{R}} > 0_{\mathbb{R}}$. Which result of Chapter 2 gives the same conclusion in every ordered field?

Solution.

(a) In $\mathcal{C}$ (a constant minus a null sequence, Theorem 4.5.3 (a), (b)), and positive. For $n \geq 1$, $n+1 \geq 2$, so $1/(n+1) \leq 1/2$ (Proposition 4.1.2 (c)) and $a_n \geq 1/2$. So $\delta = 1/2$, $N = 1$ work. ($N = 0$ fails for this δ: $a_0 = 0$.)

(b) In $\mathcal{C}$, being null, and not positive: this is the non-example after Definition 5.3.1.

(c) In $\mathcal{C}$, as the difference of $(d_k) \in \mathcal{C}$ (Example (The Truncations and the Newton Sequence Are Equivalent)) and a constant, and positive. The first truncations are $d_0 = 1$ and $d_1 = 7/5$ (14 is the largest natural number whose square is at most $2 \cdot 10^2 = 200$), and (d_k) is non-decreasing (Example (The Truncations and the Newton Sequence Are Equivalent)). So $d_k \geq d_1 = 7/5$ for every $k \geq 1$, and $a_k \geq 2/5$. So $\delta = 2/5$, $N = 1$ work. ($N = 0$ fails for every $\delta > 0$: $a_0 = 0$.)

(d) Not in $\mathcal{C}$: for every N, the indices N and $N+1$ give terms differing by 2, so the tolerance 1 has no threshold. The question "positive?" does not arise; the definition applies only to members of $\mathcal{C}$.

(e) $1_{\mathbb{R}} - 0_{\mathbb{R}} = [(1)_n - (0)_n] = [(1)_n]$, and $(1)_n$ is positive with $\delta = 1$, $N = 0$. So $1_{\mathbb{R}} - 0_{\mathbb{R}} \in P$, that is, $0_{\mathbb{R}} < 1_{\mathbb{R}}$. In every ordered commutative ring with $1 \neq 0$, Proposition 2.3.5 (e) gives $1 > 0$; for $\mathbb{R}$, $1_{\mathbb{R}} \neq 0_{\mathbb{R}}$ holds by Theorem 5.2.2.

5.3.7.2 B. Practice

Exercise. B1

Prove part (ii) of Theorem 5.3.2 directly from Definition 5.3.1, naming the rule behind each step. Then take $a_n = 1 - 1/(n+1)$ and $b_n = x_n - 1$, with (x_n) the Newton sequence of Example (A Rational Sequence Aimed at the Gap), find margins and thresholds for a and b, and state the margins and threshold that your proof produces for $a + b$ and for ab.

Hint: for b, compare x_n with $7/5$ by squaring.

Solution. The proof. Let $x, y \in P$. Choose representatives $x = [a]$ and $y = [b]$; they are positive, since every representative of an element of P is (Proposition 5.3.1). By Definition 5.3.1, fix $\delta_a \in \mathbb{Q}_{>0}$, $N_a \in \mathbb{N}$ with $a_n \geq \delta_a$ for every $n \geq N_a$, and $\delta_b \in \mathbb{Q}_{>0}$, $N_b \in \mathbb{N}$ with $b_n \geq \delta_b$ for every $n \geq N_b$. Let $N := \max\{N_a, N_b\}$, and let $n \geq N$; then $n \geq N_a$ and $n \geq N_b$.

Sum. Adding $a_n \geq \delta_a$ and $b_n \geq \delta_b$ gives $a_n + b_n \geq \delta_a + \delta_b$ (Proposition 2.3.5 (b)), and adding $\delta_a > 0$ and $\delta_b > 0$ gives $\delta_a + \delta_b > 0$ (the same rule, strict form). So $a + b$ is positive with margin $\delta_a + \delta_b$ and threshold N.

Product. From $0 \leq \delta_a \leq a_n$ and $0 \leq \delta_b \leq b_n$, Proposition 4.1.2 (d) gives $\delta_a \delta_b \leq a_n b_n$, and $\delta_a \delta_b > 0$ by (O4). So ab is positive with margin $\delta_a \delta_b$ and threshold N.

By Definition 5.2.2, $x + y = [a + b]$ and $xy = [ab]$, so both lie in P.

The example. For a, Exercise A1 (a) gives $\delta_a = 1/2$, $N_a = 1$. For b: $x_n > 0$ and $x_n^2 > 2$ for every n (Example (A Rational Sequence Aimed at the Gap)), while $(7/5)^2 = 49/25 < 2$. Both $7/5$ and x_n are ≥ 0, so Proposition 2.3.5 (g) turns $(7/5)^2 < x_n^2$ into $7/5 < x_n$. Hence

$b_n = x_n - 1 > 2/5$ for every n, and $\delta_b = 2/5$, $N_b = 0$ work. The proof then gives the threshold $N = \max\{1, 0\} = 1$, the margin $1/2 + 2/5 = 9/10$ for $a + b$, and the margin $(1/2)(2/5) = 1/5$ for ab.

Exercise. B2

Let $a, b \in \mathcal{C}$. Prove that $[a] < [b]$ if and only if there are $\delta \in \mathbb{Q}_{>0}$ and $N \in \mathbb{N}$ with $b_n - a_n \geq \delta$ for every $n \geq N$.

Solution. By Definition 5.3.2, $[a] < [b]$ means $[b] - [a] \in P$. By Definition 5.2.2, $[b] - [a] = [b - a]$, where $b - a \in \mathcal{C}$ is the termwise difference (Theorem 4.5.3 (b)). An element of $\mathbb{R}$ lies in P if and only if one (equivalently, every) representative of it is positive, by the definition of P and Proposition 5.3.1. So $[a] < [b]$ holds if and only if $b - a$ is positive, which, by Definition 5.3.1, is the condition stated, since the n-th term of $b - a$ is $b_n - a_n$.

Exercise. B3

(a) Let $a \in \mathcal{C}$ with $0 \leq a_n \leq 1$ for every $n \in \mathbb{N}$. Prove that $0_{\mathbb{R}} \leq [a] \leq 1_{\mathbb{R}}$.

(b) Give an example of $a \in \mathcal{C}$ with $0 < a_n < 1$ for every n and $[a] = 1_{\mathbb{R}}$. What does your example show about part (a)?

Solution.

(a) The constant sequences $(0)_n$ and $(1)_n$ are in $\mathcal{C}$ (Theorem 4.5.3 (a)). By Lemma 5.3.4 with $N_0 = 0$, $(0)_n \leq a$ termwise gives $0_{\mathbb{R}} = [(0)_n] \leq [a]$, and $a \leq (1)_n$ termwise gives $[a] \leq [(1)_n] = 1_{\mathbb{R}}$.

(b) Let $a_n := 1 - 1/(n + 2)$. For every n, $n + 2 \geq 2 > 1$, so $0 < 1/(n + 2) < 1$ (Proposition 2.5.2 (b)), and $0 < a_n < 1$. The sequence $a - (1)_n = (-1/(n+2))$ is null: $(1/(n+2))$ is the shift by one of the null sequence $(1/(n+1))$ (Proposition 4.3.2 (b)), and the negative of a null sequence is null (Theorem 4.4.3 (b)). So $a = (1)_n + (-1/(n + 2))_n$ is in $\mathcal{C}$, as the sum of a constant and a null sequence (Theorem 4.5.3 (a), (b)). Since $a - (1)_n$ is null, $a \sim (1)_n$ and $[a] = 1_{\mathbb{R}}$. The example shows that part (a) cannot be sharpened to strict inequalities: $a_n < 1$ for every n, and still $[a] = 1_{\mathbb{R}}$. This is the warning after Lemma 5.3.4 once more. (Any a with $0 < a_n < 1$ for all n and $1 - a_n$ null is a correct answer.)

5.3.7.3 C. Going deeper

Exercise. C1

Show that the rule $[a] \mapsto [(|a_n|)_n]$ is a well-defined function $\mathbb{R} \to \mathbb{R}$, directly from the definitions and without using Proposition 5.3.5.

Here is the beginning of a proof. *Name the move, then complete it.*

"Let $a, a' \in \mathcal{C}$ with $a \sim a'$. The sequences $(|a_n|)$ and $(|a'_n|)$ are in $\mathcal{C}$, by the reverse triangle

inequality Proposition 4.1.3 (f), as in the membership part of Proposition 5.3.5. . . . ”

Solution. *The move* is **representatives, then independence**: the rule is defined on a representative, so we must show that replacing a by an equivalent a' gives the same output class. Then each class has exactly one output, and the rule is a function (Sets, Relations, and Cardinality, Definition 1.4.1).

Completion. Let $e_n \coloneqq |a_n| - |a'_n|$. By the reverse triangle inequality in $\mathbb{Q}$ (Proposition 4.1.3 (f)),

$$0 \le |e_n| \le |a_n - a'_n| \quad \text{for every } n \in \mathbb{N}.$$

The sequence $(a_n - a'_n)$ is null, since $a \sim a'$ (Definition 5.1.2); so is $(|a_n - a'_n|)$, because $||a_n - a'_n| - 0| = |a_n - a'_n - 0|$ and the two sequences pass the same tests of Definition 4.3.3. The squeeze Proposition 4.4.4 (c) in $\mathbb{Q}$, with lower sequence $(0)_n$ and upper sequence $(|a_n - a'_n|)$, shows that $(|e_n|)$ is null. Then (e_n) is null, for the same reason as before: $|e_n - 0| = ||e_n| - 0|$. So $(|a_n|) \sim (|a'_n|)$, and $[(|a_n|)_n] = [(|a'_n|)_n]$ (Sets, Relations, and Cardinality, Theorem 2.5.1 (b)).

So the output depends only on the class $[a]$, and the rule defines a function $\mathbb{R} \to \mathbb{R}$. (By Proposition 5.3.5, this function is $x \mapsto |x|$; the direct proof above does not need that.)

Marking scheme. (10 points) - 2: names the move (representatives, then independence) and states the goal $(|a_n|) \sim (|a'_n|)$. - 3: proves $||a_n| - |a'_n|| \le |a_n - a'_n|$ for every n, citing the reverse triangle inequality in $\mathbb{Q}$. - 3: from $a \sim a'$ and the inequality of the previous step, shows that $(|a_n| - |a'_n|)_n$ is null. - 2: concludes $[(|a_n|)_n] = [(|a'_n|)_n]$, so each class has one output and the rule is a function $\mathbb{R} \to \mathbb{R}$.

A direct ε–N proof of the third step, with no squeeze, earns its 3 points: given $\varepsilon \in \mathbb{Q}_{>0}$, take N for the null sequence $(a_n - a'_n)$; then $||a_n| - |a'_n|| \le |a_n - a'_n| < \varepsilon$ for every $n \ge N$. Fatal errors (0 for the step): using $|x|$ on $\mathbb{R}$, or citing Proposition 5.3.5; proving only that the two output sequences are Cauchy, without proving them equivalent.

Exercise. C2: Find the error

Here is an argument for Proposition 5.3.1.

“Let a be positive and $a' \sim a$. Fix $\delta \in \mathbb{Q}_{>0}$ and N with $a_n \ge \delta$ for every $n \ge N$. Since $a' - a$ is null, there is N' with $|a'_n - a_n| < \delta$ for every $n \ge N'$. So, for $n \ge \max\{N, N'\}$, $a'_n > a_n - \delta \ge 0$. Hence $a'_n > 0$ for all large n. Therefore a' is positive.”

Is the conclusion true? Does the argument prove it? If not, name the line that fails, say why, give a counterexample to that step, and repair the argument.

Solution (Hint 1). Write out in symbols what “a' is positive” requires, and compare it with what the argument has when it reaches the last line.

Solution (Hint 2). “Positive” asks for one margin $\delta' \in \mathbb{Q}_{>0}$, chosen before n, with $a'_n \ge \delta'$

for all large n. Find the line where the argument should produce such a δ', and ask what it produces instead.

Solution. The conclusion is true: it is Proposition 5.3.1.

The middle lines are correct. With the tolerance δ and the strict inequality $|a'_n - a_n| < \delta$, we get $a'_n > a_n - \delta \geq \delta - \delta = 0$ for every $n \geq \max\{N, N'\}$. So the line "$a'_n > 0$ for all large n" is proved.

The failing line is the last: "Therefore a' is positive." What has been proved is

$$\exists N \in \mathbb{N} \;\; \forall n \geq N \;\; a'_n > 0,$$

while "positive" (Definition 5.3.1) is

$$\exists \delta \in \mathbb{Q}_{>0} \;\; \exists N \in \mathbb{N} \;\; \forall n \geq N \;\; a'_n \geq \delta.$$

To pass from the first to the second, one would take the margin $\delta := a'_n$, which depends on n, and then claim one margin for all n. That swaps a "for every n, there is a margin" for a "there is a margin for every n". It is the analogue, for a margin depending on n, of entry (N2) of Logic and Proof, Example (What is not allowed), where a threshold depending on ε is treated as uniform.

Counterexample to the step. The sequence $(1/(n+1))$ has $a'_n > 0$ for all n, so it satisfies the premise of the last line, and it is not positive (the non-example after Definition 5.3.1). So "eventually > 0" does not imply "positive". The argument reached a true conclusion through a step that is not valid.

Repair. Use the tolerance $\delta/2 \in \mathbb{Q}_{>0}$ (Proposition 4.1.2 (b)) instead of δ. Since $a' - a$ is null, choose N'' with $|a'_n - a_n| < \delta/2$ for every $n \geq N''$. For $n \geq \max\{N, N''\}$,

$$a'_n > a_n - \frac{\delta}{2} \geq \delta - \frac{\delta}{2} = \frac{\delta}{2}.$$

So $a'_n \geq \delta/2$ for every $n \geq \max\{N, N''\}$: one margin, $\delta/2$, fixed before n, and a' is positive. This is the proof of Proposition 4.5.5.

Marking scheme. (10 points) - 2: says the conclusion is true, citing Proposition 5.3.1. - 3: identifies the last line as the failing one and names the quantifier swap (a margin depending on n treated as one margin for all n). - 2: a counterexample to the step: a sequence whose terms are eventually > 0 and which is not positive, such as $(1/(n+1))$. - 3: a repair that produces one margin in $\mathbb{Q}_{>0}$ fixed before n (for instance the tolerance $\delta/2$ above; any tolerance strictly between 0 and δ works equally well). Fatal errors (0 for the step): blaming one of the valid middle lines; a "counterexample" to the conclusion; a repair whose margin still depends on n.

5.4 The Rationals Inside the Reals

Section 5.3 made $\mathbb{R}$ an ordered field. We built it to enlarge $\mathbb{Q}$, but so far $\mathbb{Q}$ is not inside it: a rational number is a class of pairs of integers, and a real number is a class of Cauchy sequences. This section places a copy of $\mathbb{Q}$ in $\mathbb{R}$ and shows how the copy sits there. It is preserved by all

the operations and the order; it is spread through $\mathbb{R}$ without gaps, so that between any two real numbers there is a rational one; and every real number is the limit of rationals. Along the way we prove that $\mathbb{R}$ is Archimedean and that rational tolerances suffice for limits in $\mathbb{R}$. At the end we meet the number that Chapter 4 promised: a positive real whose square is 2.

Notation is as in Section 5.3: $\mathcal{C}$ is the set of Cauchy sequences of rationals, $[a]$ is the class of $a \in \mathcal{C}$, $(q)_n$ is the constant sequence with value q, $P = \mathbb{R}_{>0}$ is the set of positive reals, and $x < y$ means $y - x \in P$ (Definition 5.3.2). By Theorem 5.3.3, the results of Chapters 2 and 4 for an arbitrary ordered field apply with $F = \mathbb{R}$; we say "in $\mathbb{R}$" where the context does not make the field clear.

5.4.1 A copy of the rationals

Section 5.2 already represented each rational q by the class of its constant sequence, as Chapter 4 said it would. We make that a map.

Definition 5.4.1. The Rationals in the Reals

For $q \in \mathbb{Q}$, let

$$\kappa(q) := [(q)_n] \in \mathbb{R},$$

the class of the constant sequence with value q. This defines a function $\kappa\colon \mathbb{Q} \to \mathbb{R}$.

The right side is in $\mathbb{R}$ because constant sequences are Cauchy (Theorem 4.5.3 (a)), and κ is a function because each q determines one constant sequence and that sequence one class. In this notation $0_{\mathbb{R}} = \kappa(0)$ and $1_{\mathbb{R}} = \kappa(1)$ (Definition 5.2.2). The letter κ is new on purpose: Chapter 2 used ι and j for $\mathbb{N} \to \mathbb{Z} \to \mathbb{Q}$, and Chapter 4 wrote ι_F for the copy of $\mathbb{Q}$ in an ordered field F. Part (d) of the next theorem shows that κ is $\iota_{\mathbb{R}}$.

The theorem says that κ carries everything we know how to do with rationals into $\mathbb{R}$ without change: sums, products, negatives, reciprocals, the order and the absolute value.

Theorem 5.4.1. The Rationals Embed in the Reals

For all $q, s \in \mathbb{Q}$:

(a) $\kappa(q+s) = \kappa(q) + \kappa(s)$, $\kappa(qs) = \kappa(q)\kappa(s)$, $\kappa(0) = 0_{\mathbb{R}}$ and $\kappa(1) = 1_{\mathbb{R}}$;

(b) $q < s$ if and only if $\kappa(q) < \kappa(s)$; hence κ is injective;

(c) $\kappa(-q) = -\kappa(q)$, $\kappa(|q|) = |\kappa(q)|$, and, if $q \neq 0$, $\kappa(q) \neq 0_{\mathbb{R}}$ and $\kappa(1/q) = \kappa(q)^{-1}$;

(d) κ is the map $\iota_{\mathbb{R}}$ of Theorem 4.2.2 for the ordered field $F = \mathbb{R}$; in particular $\kappa(n) = n \cdot 1_{\mathbb{R}}$ for every $n \in \mathbb{N}$ (Definition 4.2.1).

Idea. Constant sequences add, multiply and negate termwise into constant sequences, so (a) and the first identity of (c) are the definitions of the operations. For (b), the difference of two constant sequences is constant, and a constant $s - q > 0$ is its own margin. The rest is algebra in a field: the absolute value by cases, reciprocals by uniqueness of inverses, and (d) by the uniqueness clause of Chapter 4's embedding theorem.

Proof.

(a) By Definition 5.2.2, $\kappa(q)+\kappa(s) = [(q)_n+(s)_n] = [(q+s)_n] = \kappa(q+s)$, since the termwise sum of two constant sequences is the constant sequence of the sum. The product is the same with $\cdot$ in place of $+$. The last two identities are the definitions of $0_{\mathbb{R}}$ and $1_{\mathbb{R}}$.
(c), first identity. By Definition 5.2.2, $-\kappa(q) = [-(q)_n] = [(-q)_n] = \kappa(-q)$.

(b) ($\Rightarrow$) Let $q < s$. By (a) and the first identity of (c), $\kappa(s) - \kappa(q) = \kappa(s) + \kappa(-q) = \kappa(s-q) = [(s-q)_n]$. The constant $s-q$ lies in $\mathbb{Q}_{>0}$ (Proposition 2.3.5 (a)), so the constant sequence $(s-q)_n$ is positive, with margin $\delta = s-q$ and threshold $N = 0$ (Definition 5.3.1). Hence $\kappa(s) - \kappa(q) \in P$, that is, $\kappa(q) < \kappa(s)$.

($\Leftarrow$) Let $\kappa(q) < \kappa(s)$. By trichotomy in $\mathbb{Q}$, $q < s$, $q = s$ or $s < q$. If $q = s$, then $\kappa(q) = \kappa(s)$; if $s < q$, then $\kappa(s) < \kappa(q)$ by ($\Rightarrow$). Either contradicts $\kappa(q) < \kappa(s)$ by (O1) in $\mathbb{R}$ (Theorem 5.3.3). So $q < s$.

Injectivity: if $q \neq s$, then $q < s$ or $s < q$, so $\kappa(q) < \kappa(s)$ or $\kappa(s) < \kappa(q)$, and in either case $\kappa(q) \neq \kappa(s)$ by (O1).

(c), the rest. If $q > 0$, then $\kappa(q) > \kappa(0) = 0_{\mathbb{R}}$ by (b), so $|\kappa(q)| = \kappa(q) = \kappa(|q|)$ (Definition 4.1.1, in $\mathbb{R}$ and in $\mathbb{Q}$). If $q = 0$, both sides are $0_{\mathbb{R}}$. If $q < 0$, then $\kappa(q) < 0_{\mathbb{R}}$ by (b), so $|\kappa(q)| = -\kappa(q) = \kappa(-q) = \kappa(|q|)$. Now let $q \neq 0$. Then $\kappa(q) \neq \kappa(0) = 0_{\mathbb{R}}$ by injectivity, and by (a),

$$\kappa(q)\,\kappa(1/q) = \kappa(q \cdot (1/q)) = \kappa(1) = 1_{\mathbb{R}}.$$

In the field $\mathbb{R}$ (Theorem 5.2.3), the inverse of $\kappa(q)$ is the only element with this property (Lemma 4.1.1 (a)), so $\kappa(1/q) = \kappa(q)^{-1}$.

(d) By Theorem 5.3.3, $\mathbb{R}$ is an ordered field, so Theorem 4.2.2 gives **exactly one** map $\iota_{\mathbb{R}} \colon \mathbb{Q} \to \mathbb{R}$ with $\iota_{\mathbb{R}}(1) = 1_{\mathbb{R}}$ that preserves sums and products. By (a), κ is such a map, so $\kappa = \iota_{\mathbb{R}}$. The same theorem states that $\iota_{\mathbb{R}}(n) = n \cdot 1_{\mathbb{R}}$ for every $n \in \mathbb{N}$.

■

By (d), part (b) and the injectivity of κ are also part of Theorem 4.2.2 with $F = \mathbb{R}$, and (d) needs only (a). The direct proof of (b) is kept because it shows how a constant sequence is positive, with its own value as margin.

Two more identities follow from (a) and (c):

$$\kappa(q-s) = \kappa(q) + \kappa(-s) = \kappa(q) - \kappa(s),$$
$$\kappa(q/s) = \kappa(q)\,\kappa(1/s) = \kappa(q)/\kappa(s) \quad \text{for } s \neq 0.$$

So any computation with rationals, done in $\mathbb{Q}$ and then moved into $\mathbb{R}$ by κ, gives the same answer as the computation done in $\mathbb{R}$ on the images. Part (b) adds that comparisons agree too.

5.4.2 Archimedes in the reals

Chapter 4 showed that an ordered field need not be Archimedean: in $\mathbb{Q}(t)$ the element t is larger than every $n \cdot 1$ (Example (A Non-Archimedean Ordered Field)). So the Archimedean property of $\mathbb{R}$ has to be proved. It comes from $\mathbb{Q}$, through the fact that a Cauchy sequence is bounded.

Theorem 5.4.2. The Reals Are Archimedean

$\mathbb{R}$ is an Archimedean ordered field (Definition 4.2.2): for every $x \in \mathbb{R}$ there is $n \in \mathbb{N}$ with $x < n \cdot 1_{\mathbb{R}}$.

Idea. A representative of x is a Cauchy sequence of rationals, so all its terms lie below one rational bound M. Archimedes in $\mathbb{Q}$ puts a natural number above M; one more unit gives a fixed margin above every term, and a fixed margin is what positivity asks for.

Proof. Let $x \in \mathbb{R}$, and write $x = [a]$ with $a \in \mathcal{C}$. By Proposition 4.5.2 with $F = \mathbb{Q}$, there is $M \in \mathbb{Q}_{>0}$ with $|a_k| \leq M$ for every $k \in \mathbb{N}$. By Theorem 2.5.3, there is $n \in \mathbb{N}$ with $M < n$. For every k,

$$(n+1) - a_k \geq (n+1) - M = (n - M) + 1 > 1,$$

since $a_k \leq |a_k| \leq M$ (Proposition 4.1.3 (c)), so $-M \leq -a_k$ and then $(n+1)-a_k \geq (n+1)-M$ (Proposition 2.3.5 (a), (b)), and $n - M > 0$. So the sequence $((n+1) - a_k)_k$ is positive, with margin 1 and threshold 0. It represents $\kappa(n+1)-x$, by Definition 5.2.2. Hence $x < \kappa(n+1)$, and $\kappa(n+1) = (n+1) \cdot 1_{\mathbb{R}}$ by Theorem 5.4.1 (d). Since $n+1 \in \mathbb{N}$, this proves the theorem. ■

Every result of Chapter 4 about Archimedean ordered fields now applies with $F = \mathbb{R}$. Two of them are used at once.

The integer part in $\mathbb{R}$. By Proposition 4.2.4 with $F = \mathbb{R}$, and $\iota_{\mathbb{R}} = \kappa$ (Theorem 5.4.1 (d)), for every $x \in \mathbb{R}$ there is **exactly one** $m \in \mathbb{Z}$ with

$$\kappa(m) \leq x < \kappa(m+1).$$

We call m the **integer part** of x. For a rational q, the integer part of $\kappa(q)$ is the integer part $\lfloor q \rfloor$ of Proposition 2.5.5, because κ preserves and reflects $\leq$ (Theorem 5.4.1 (b)). The word "exactly" matters: Section 5.5 builds a sequence by taking integer parts, and uniqueness is what makes that a rule rather than a choice.

The second consequence, Proposition 4.2.3 in $\mathbb{R}$, turns into the statement that rational tolerances suffice; it comes after the density theorem.

5.4.3 Between two reals there is a rational

Theorem 5.4.3. The Rationals Are Dense in the Reals

Let $x, y \in \mathbb{R}$ with $x < y$. Then there is $q \in \mathbb{Q}$ with $x < \kappa(q) < y$.

Try it first (ten minutes): write $x = [a]$ and $y = [b]$. Which rational would you put between them? Build it from the terms of a and b, and say how far out you must go.

Idea. The obvious attempt takes a rational between a_n and b_n, say their midpoint, and has to say **which** n. An early index is no good: a_0 may be larger than b_0, since early terms are irrelevant to the classes. Even where $a_n < b_n$, a later term a_m may climb past the midpoint,

and the midpoint must sit above **every** late a_m by a fixed margin.

So go far out, in two ways. Since $y - x$ is positive, from some index on $b_n - a_n \geq \delta$, a fixed rational gap. Since a and b are Cauchy, from some index N on all their terms stay within $\delta/4$ of a_N and of b_N, respectively. The midpoint q of a_N and b_N is at least $\delta/2$ above a_N; every later a_n is less than $\delta/4$ above a_N; so q beats every later a_n by more than $\delta/4$. The same holds below b_n.

Proof. Write $x = [a]$ and $y = [b]$ with $a, b \in \mathcal{C}$. By Definition 5.3.2, $y - x = [b - a] \in P$, so $b - a$ is positive (Proposition 5.3.1). Fix $\delta \in \mathbb{Q}_{>0}$ and $N_1 \in \mathbb{N}$ with

$$b_n - a_n \geq \delta \quad \text{for every } n \geq N_1$$

[∃E] (Logic and Proof, Definition 5.6.1). Then $\delta/4 = (\delta/2)/2 \in \mathbb{Q}_{>0}$ (Proposition 4.1.2 (b), twice). Since a and b are Cauchy in $\mathbb{Q}$, fix N_2 with $|a_m - a_n| < \delta/4$ for all $m, n \geq N_2$, and N_3 with $|b_m - b_n| < \delta/4$ for all $m, n \geq N_3$ [∃E]. Let

$$N \coloneqq \max\{N_1, N_2, N_3\}, \qquad q \coloneqq \frac{a_N + b_N}{2} \in \mathbb{Q}.$$

Let $n \geq N$. Expanding the right sides shows

$$q - a_n = (a_N - a_n) + \frac{b_N - a_N}{2},$$
$$b_n - q = (b_n - b_N) + \frac{b_N - a_N}{2}.$$

Since $n, N \geq N_2$, $a_N - a_n > -\delta/4$ (Proposition 4.1.3 (d)); since $N \geq N_1$, $b_N - a_N \geq \delta$, and multiplying by $1/2 > 0$ (Proposition 4.1.2 (b) with $\varepsilon = 1$) gives $(b_N - a_N)/2 \geq \delta/2$ (Proposition 2.3.5 (c), with equality allowed). Adding (Proposition 2.3.5 (b)),

$$q - a_n > -\frac{\delta}{4} + \frac{\delta}{2} = \frac{\delta}{4}.$$

In the same way, since $n, N \geq N_3$, $b_n - b_N > -\delta/4$, and $b_n - q > \delta/4$. So the sequences $(q - a_n)_n$ and $(b_n - q)_n$ are positive, each with margin $\delta/4$ and threshold N. By Definition 5.2.2 they represent $\kappa(q) - x$ and $y - \kappa(q)$. Hence $x < \kappa(q)$ and $\kappa(q) < y$.

■

The rational q depends on the representatives a, b and on the constants fixed in the proof; nothing makes it the "best" rational, and a different run of the proof may give a different q. The theorem only claims that one exists.

The proof cites neither Theorem 2.5.3 nor Theorem 5.4.2: density came from the Cauchy property of the representatives, not from Archimedes. Chapter 4 proved density in every Archimedean ordered field (Theorem 4.2.5), so, with Theorem 5.4.2, it gives the same conclusion for $\mathbb{R}$ by another route (Exercise B1). We keep the direct proof for what it shows: in this construction, the rationals are dense because each real number is made of rationals.

5.4.4 Rational tolerances suffice

A tolerance for a limit in $\mathbb{R}$ is an element of $\mathbb{R}_{>0}$, a class of sequences. That is awkward to work with. The next result lets us test convergence in $\mathbb{R}$ with the tolerances $\kappa(1/n)$ only.

Corollary 5.4.4. Rational Tolerances Suffice

(a) For every $\varepsilon \in \mathbb{R}_{>0}$ there is $N \in \mathbb{Z}_{>0}$ with $\kappa(1/n) < \varepsilon$ for every $n \in \mathbb{N}$ with $n \geq N$.

(b) Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and $x \in \mathbb{R}$. Then $x_k \to x$ in $\mathbb{R}$ (Definition 4.3.2 with $F = \mathbb{R}$) if and only if, **for every** $n \in \mathbb{Z}_{>0}$, **there is** $K \in \mathbb{N}$ with $|x_k - x| < \kappa(1/n)$ **for every** $k \geq K$. Likewise, (x_k) is Cauchy in $\mathbb{R}$ (Definition 4.5.1 with $F = \mathbb{R}$) if and only if, for every $n \in \mathbb{Z}_{>0}$, there is $K \in \mathbb{N}$ with $|x_j - x_k| < \kappa(1/n)$ for all $j, k \geq K$.

Proof.

(a) Let $\varepsilon \in \mathbb{R}_{>0}$. Since $\mathbb{R}$ is Archimedean (Theorem 5.4.2), Proposition 4.2.3 in $\mathbb{R}$ gives $N \in \mathbb{Z}_{>0}$ with $1/(n \cdot 1_{\mathbb{R}}) < \varepsilon$ for every $n \geq N$. For such n, $n \neq 0$, $n \cdot 1_{\mathbb{R}} = \kappa(n)$ (Theorem 5.4.1 (d)) and $1/\kappa(n) = \kappa(1/n)$ (Theorem 5.4.1 (c)). So $\kappa(1/n) < \varepsilon$.

(b) ($\Rightarrow$) For $n \in \mathbb{Z}_{>0}$, $1/n > 0$ in $\mathbb{Q}$ (Proposition 2.5.2 (a)), so $\kappa(1/n) > \kappa(0) = 0_{\mathbb{R}}$ (Theorem 5.4.1 (b)): it is one of the tolerances of the definition, which then supplies K.

($\Leftarrow$) Let $\varepsilon \in \mathbb{R}_{>0}$. By (a), there is $N \in \mathbb{Z}_{>0}$ with $\kappa(1/N) < \varepsilon$. The hypothesis with $n \coloneqq N$ gives K with $|x_k - x| < \kappa(1/N)$ for every $k \geq K$, and then $|x_k - x| < \varepsilon$ by (O2). So K serves for ε. The Cauchy version is the same argument with $|x_j - x_k|$ for $j, k \geq K$ in place of $|x_k - x|$.

■

In practice: to prove a limit in $\mathbb{R}$, fix $n \in \mathbb{Z}_{>0}$ and find a threshold for the rational tolerance $1/n$. The threshold may depend on n, and on nothing chosen later.

5.4.5 Each real number is the limit of its terms

A real number $[a]$ was introduced as a name for "what the rationals a_k are approaching". Inside $\mathbb{Q}$ there may be nothing to approach. Inside $\mathbb{R}$ there is, and it is $[a]$ itself.

Theorem 5.4.5. A Cauchy Sequence Converges to Its Class

Let $a \in \mathcal{C}$. Then the sequence $(\kappa(a_k))_{k\in\mathbb{N}}$ converges in $\mathbb{R}$ to $[a]$.

Idea. For a fixed k, the difference $[a] - \kappa(a_k)$ is represented by $(a_m - a_k)_m$: the sequence a with the single rational a_k subtracted from every term. For m and k past a Cauchy threshold, these terms are small, and Lemma 5.3.4 turns a termwise bound into a bound between classes. That lemma yields only $\leq$, so we aim the termwise bound at $1/(2n)$, strictly below the target $1/n$.

Proof. Let $n \in \mathbb{Z}_{>0}$. By Corollary 5.4.4 (b), it suffices to find $K \in \mathbb{N}$ with $|\kappa(a_k) - [a]| < \kappa(1/n)$ for every $k \geq K$. Since $\kappa(a_k) - [a] = -([a] - \kappa(a_k))$, Proposition 4.1.3 (b) in $\mathbb{R}$ gives $|\kappa(a_k) - [a]| = |[a] - \kappa(a_k)|$, so we bound the latter.

Since $a \in \mathcal{C}$ and $1/(2n) \in \mathbb{Q}_{>0}$, fix $N \in \mathbb{N}$ with $|a_m - a_k| < 1/(2n)$ for all $m, k \geq N$. Let $k \geq N$.

By Definition 5.2.2, $[a] - \kappa(a_k) = [a] + [(-a_k)_m] = [(a_m - a_k)_m]$, and $(a_m - a_k)_m \in \mathcal{C}$ as the difference of a and a constant sequence (Theorem 4.5.3 (a), (b)). By Proposition 5.3.5,

$$|[a] - \kappa(a_k)| = [(|a_m - a_k|)_m].$$

For every $m \geq N$, $|a_m - a_k| \leq 1/(2n)$. So Lemma 5.3.4, with the constant sequence $(1/(2n))_m$ and $N_0 = N$, gives

$$|[a] - \kappa(a_k)| \leq \kappa(1/(2n)).$$

Finally $0 < n < 2n$ in $\mathbb{Q}$, so $1/(2n) < 1/n$ (Proposition 2.5.2 (b)), and $\kappa(1/(2n)) < \kappa(1/n)$ by Theorem 5.4.1 (b). Combining $\leq$ with $<$ (Proposition 2.3.5 (i)), $|[a] - \kappa(a_k)| < \kappa(1/n)$ for every $k \geq N$. So $K := N$ works.

■

Why $1/(2n)$ and not $1/n$ directly? The Cauchy property gives $|a_m - a_k| < 1/n$, but Lemma 5.3.4 passes only the non-strict inequality to classes, so we would reach $|[a] - \kappa(a_k)| \leq \kappa(1/n)$, one symbol short of the definition. Halving the tolerance first leaves room for the strict inequality.

The sequence does converge to the number it names, once that number exists. The theorem has a consequence worth stating in words: **every real number is a limit of rationals**, namely of the terms of any of its representatives. It also explains the slogan of the chapter. Equivalent sequences have the same class, so their terms converge in $\mathbb{R}$ to the same number; by uniqueness of limits (Theorem 4.3.4 in $\mathbb{R}$), that number is the only limit they have. A real number is what all its representatives converge to.

Identification. From here on we write q for $\kappa(q)$ when $q \in \mathbb{Q}$, and $\mathbb{Q} \subseteq \mathbb{R}$. This is a convention about notation, of the kind Chapter 2 adopted twice, writing n for $\iota(n)$ and m for $j(m)$ (Proposition 2.2.3, Proposition 2.4.4). It is safe because of Theorem 5.4.1: κ is injective and preserves sums, products, negatives, reciprocals, the order and the absolute value, so a statement about rationals means the same whether it is read in $\mathbb{Q}$ or, through κ, in $\mathbb{R}$. It is not a set inclusion: the rational q is a class of pairs of integers, and $\kappa(q)$ is a class of sequences of rationals, a different set. In the new notation the theorem says $a_k \to [a]$ in $\mathbb{R}$, and Corollary 5.4.4 says that the tolerances $1/n$, $n \in \mathbb{Z}_{>0}$, suffice in $\mathbb{R}$.

5.4.6 The square root of two

Chapter 4 ended with the Newton sequence (x_n) of Example (A Rational Sequence Aimed at the Gap), Cauchy in $\mathbb{Q}$ and with no limit in $\mathbb{Q}$ (Theorem 4.6.1). Its class is the number it was aiming at.

Example. The Class of the Newton Sequence

Let (x_n) be the Newton sequence of Example (A Rational Sequence Aimed at the Gap), and let $r := [(x_n)] \in \mathbb{R}$. Show that $r > 0$, that $r^2 = 2$, that $r \neq q$ for every $q \in \mathbb{Q}$, and that

$r = [(d_k)]$, where (d_k) is the sequence of decimal truncations of Example (The Truncations and the Newton Sequence Are Equivalent).

Solution. The sequence (x_n) is in $\mathcal{C}$ by Theorem 4.6.1, so $r \in \mathbb{R}$. We use the facts $x_n > 1$ for every $n \in \mathbb{N}$ and $e_n \le 1/(4n)$ for every $n \ge 1$, where $e_n := x_n^2 - 2$, stated in Theorem 4.6.1, and $e_n > 0$ for every $n \in \mathbb{N}$, proved in Example (A Rational Sequence Aimed at the Gap). The first values are

$$e_0 = 2, \quad e_1 = \frac{1}{4}, \quad e_2 = \frac{1}{144}, \quad e_3 = \frac{1}{166464}, \quad e_4 = \frac{1}{221682772224}.$$

$r > 0$. Since $x_n > 1$ for every n, the sequence (x_n) is positive, with margin 1 and threshold 0. So $r \in P$, that is, $r > 0$. By Lemma 5.3.4, also $r \ge 1$.

$r^2 = 2$. By Definition 5.2.2, $r^2 = r \cdot r = [(x_n^2)]$, and $2 = [(2)_n]$. Termwise, $(x_n^2) - (2)_n = (e_n)$. For $n \ge 1$,

$$0 < e_n \le \frac{1}{4n} \le \frac{1}{n+1},$$

where the last step is Proposition 4.1.2 (c), since $4n - (n+1) = 3n - 1 > 0$. The sequence $(1/(n+1))$ is defined for every $n \in \mathbb{N}$ and null in $\mathbb{Q}$ (Example (First Limits) (b)); so is $(0)_n$. The squeeze Proposition 4.4.4 (c) in $\mathbb{Q}$, with $N_0 = 1$, shows that (e_n) is null. (At $n = 0$ the bound $1/(4n)$ is undefined and $e_0 = 2$; the threshold $N_0 = 1$ skips it.) Hence $(x_n^2) \sim (2)_n$ and $r^2 = 2$.

r is not rational. Suppose $r = \kappa(q)$ for some $q \in \mathbb{Q}$. Then $\kappa(q^2) = \kappa(q)^2 = r^2 = \kappa(2)$ by Theorem 5.4.1 (a), so $q^2 = 2$ by injectivity. This contradicts Theorem 2.6.1. So $r \notin \kappa(\mathbb{Q})$.

$r = [(d_k)]$. By Example (The Truncations and the Newton Sequence Are Equivalent), $(d_k) \sim (x_k)$, so the classes are equal.

So $\mathbb{R}$ contains a positive element whose square is 2, and $\mathbb{Q}$ contains none. It is the number Chapter 4 promised, and the decimal truncations name the same number as the Newton sequence.

Second reading · *The lower Newton sequence names the same number*

Section 4.6 noted that the rationals $2/x_n$ approach the gap from below while the x_n approach it from above. They name the same real number. The sequence $(1/x_n)$ is in $\mathcal{C}$ by Proposition 4.5.6 with $c = 1$ and $N = 0$, since $|x_n| = x_n \ge 1$; so $(2/x_n) \in \mathcal{C}$ (Theorem 4.5.3 (b)). Termwise,

$$x_n - \frac{2}{x_n} = \frac{x_n^2 - 2}{x_n} = \frac{e_n}{x_n}, \qquad 0 < \frac{e_n}{x_n} \le e_n,$$

using $0 < 1/x_n$ (Proposition 2.5.2 (a)) and $1/x_n \le 1$ (Proposition 4.1.2 (c) with $a = 1$, $b = x_n$): then $e_n/x_n > 0$ as a product of positives (O4), and $e_n/x_n \le 1 \cdot e_n = e_n$ (Proposition 4.1.2 (d)). Since (e_n) is null, the squeeze makes (e_n/x_n) null, so $(2/x_n) \sim (x_n)$ and $[(2/x_n)] = r$.

Warning. r is **not** "the limit of (x_n) in $\mathbb{Q}$": there is no such limit (Theorem 4.6.1), and nothing in this section changes that. The statement $x_n \to r$ holds in $\mathbb{R}$, by Theorem 5.4.5; it is a statement about a different field, with different tolerances and a different set of candidate limits.

Quick check. Find a rational number strictly between $[(d_k)]$ and $[(d_k + 1/(k+1))]$.

Solution. There is none, and Theorem 5.4.3 does not apply. The two sequences differ by $(1/(k+1))$, which is null, so they are equivalent and the classes are equal. The hypothesis $x < y$ fails, and no number lies strictly between a number and itself.

5.4.7 Exercises

5.4.7.1 A. Check your understanding

Exercise. A1

(a) Show that $\kappa(q^k) = \kappa(q)^k$ for every $q \in \mathbb{Q}$ and $k \in \mathbb{N}$, the powers being those of Definition 4.2.1 in $\mathbb{Q}$ and in $\mathbb{R}$.

(b) Let r be as in Example (The Class of the Newton Sequence). Find the integer parts of r and of $10r$, and justify each from the terms d_k: use $d_1 = 7/5$, $d_2 = 141/100$, and the bounds of Example (The Truncations and the Newton Sequence Are Equivalent).

(c) True or false: if $a \in \mathcal{C}$, $q \in \mathbb{Q}$ and $a_n < q$ for every n, then $[a] < q$. Give a reason or a counterexample.

(d) Is $\kappa\colon \mathbb{Q} \to \mathbb{R}$ surjective?

Solution.

(a) By Logic and Proof, Axiom 4.2.1 on k, with base $P(0)$. Base: $\kappa(q^0) = \kappa(1) = 1_{\mathbb{R}} = \kappa(q)^0$, by Definition 4.2.1 (b) in $\mathbb{Q}$ and in $\mathbb{R}$ and Theorem 5.4.1 (a). Step: if $\kappa(q^k) = \kappa(q)^k$, then

$$\kappa(q^{k+1}) = \kappa(q^k q) = \kappa(q^k)\,\kappa(q) = \kappa(q)^k\,\kappa(q) = \kappa(q)^{k+1},$$

by the defining equation of powers, Theorem 5.4.1 (a), and the induction hypothesis. In particular $\kappa(q^2) = \kappa(q)^2$, the step used in Example (The Class of the Newton Sequence).

(b) By Example (The Truncations and the Newton Sequence Are Equivalent), (d_k) is non-decreasing, and $d_j < d_k + 1/10^k$ whenever $j \geq k$. With $k = 2$: for every $j \geq 2$,

$$\frac{7}{5} = d_1 \leq d_j < d_2 + \frac{1}{100} = \frac{142}{100} = \frac{71}{50}.$$

By Lemma 5.3.4 with $N_0 = 2$, and $r = [(d_k)]$ (Example (The Class of the Newton Sequence)), $7/5 \leq r \leq 71/50$. Since $1 \leq 7/5$ and $71/50 < 2$, $1 \leq r < 2$, so the integer part of r is 1. Multiplying by $10 > 0$ in $\mathbb{R}$ (Proposition 2.3.5 (c), with equality allowed), $14 \leq 10r \leq 71/5 < 15$, so the integer part of $10r$ is 14.

(c) False. Take $a_n := q - 1/(n+1)$. Here $a \in \mathcal{C}$: $(q)_n$ is constant and $(1/(n+1))$ is null (Example (First Limits) (b)), so both are Cauchy (Theorem 4.5.3 (a)), and so is their difference (Theorem 4.5.3 (b)). Then $a_n < q$ for every n, and $a - (q)_n = (-1/(n+1))$ is null (Theorem 4.4.3 (b)), so $[a] = \kappa(q)$, and $[a] < \kappa(q)$ fails. The correct conclusion is $[a] \leq q$ (Lemma 5.3.4); see the warning after it.

(d) No. By Example (The Class of the Newton Sequence), $r \in \mathbb{R}$ is not $\kappa(q)$ for any $q \in \mathbb{Q}$.

5.4.7.2 B. Practice

Exercise. B1

Deduce Theorem 5.4.3 from Theorem 4.2.5 and Theorem 5.4.2. Then say which hypothesis each of the two proofs uses that the other does not.

Solution. By Theorem 5.4.2, $\mathbb{R}$ is an Archimedean ordered field. So Theorem 4.2.5 applies with $F = \mathbb{R}$: for $x < y$ in $\mathbb{R}$ there is $q \in \mathbb{Q}$ with $x < \iota_{\mathbb{R}}(q) < y$. By Theorem 5.4.1 (d), $\iota_{\mathbb{R}}(q) = \kappa(q)$, which is the conclusion of Theorem 5.4.3.

The general theorem of Chapter 4 uses that F is **Archimedean**; it knows nothing about representatives, since an element of an arbitrary ordered field is not a class of sequences. The direct proof of Theorem 5.4.3 uses the **Cauchy property of the representatives** a and b, and positivity of $b - a$; it uses no Archimedean property, neither Theorem 2.5.3 nor Theorem 5.4.2.

Exercise. B2

Let $x \in \mathbb{R}$. For $k \in \mathbb{N}$, let m_k be the integer part of $(k+1)x$, and $q_k := m_k/(k+1) \in \mathbb{Q}$. Prove that $q_k \leq x < q_k + 1/(k+1)$ for every k, and that $q_k \to x$ in $\mathbb{R}$.

Solution. Fix $k \in \mathbb{N}$. By the definition of the integer part (after Theorem 5.4.2), $m_k \leq (k+1)x < m_k + 1$. Since $k+1 > 0$, also $1/(k+1) > 0$ (Proposition 2.5.2 (a)), and multiplying by it in $\mathbb{R}$ preserves both inequalities (Proposition 2.3.5 (c), with equality allowed on the left):

$$q_k = \frac{m_k}{k+1} \leq x < \frac{m_k + 1}{k+1} = q_k + \frac{1}{k+1}.$$

So $0 \leq x - q_k < 1/(k+1)$ for every k. In $\mathbb{R}$, the sequence $(1/(k+1))$, that is $(1/((k+1)\cdot 1_{\mathbb{R}}))$ by Theorem 5.4.1 (c), (d), is null, by Proposition 4.3.1 and Theorem 5.4.2. The squeeze Proposition 4.4.4 (c) in $\mathbb{R}$, with lower sequence (0), gives that $(x - q_k)$ is null. Then $(q_k - x)$ is null (Theorem 4.4.3 (b)), so $q_k \to x$ (Proposition 4.3.3).

Each q_k is given by a rule: the integer part of $(k+1)x$ is the **only** integer with the defining property, so nothing is chosen. Section 5.5 uses the same approximants, with a different real number x_k for each k.

Exercise. B3: Put the proof in order

The lines below, after the identification paragraph (so q stands for $\kappa(q)$), contain the proof of Theorem 5.4.5 for $a \in \mathcal{C}$, shuffled, together with two lines that do not belong. Put the correct lines in order, and say what is wrong with each of the other two.

(i) so $|[a] - a_k| = [(|a_m - a_k|)_m]$, by Proposition 5.3.5;
(ii) since $a \in \mathcal{C}$, fix N with $|a_m - a_k| < 1/(2n)$ for all $m, k \geq N$;
(iii) its terms are $< 1/(2n)$ for $m \geq N$, so $|[a] - a_k| < 1/(2n)$;
(iv) $1/(2n) < 1/n$; take $K = N$;
(v) let $k \geq N$;
(vi) let $n \in \mathbb{Z}_{>0}$; by Corollary 5.4.4 (b), it suffices to find K with $|[a] - a_k| < 1/n$ for every $k \geq K$, since $|a_k - [a]| = |[a] - a_k|$ (Proposition 4.1.3 (b));
(vii) $[a] - a_k$ is represented by $(a_m - a_k)_m$;
(viii) for each k, fix N_k with $|a_m - a_k| < 1/(2n)$ for every $m \geq N_k$;
(ix) its terms are $\leq 1/(2n)$ for $m \geq N$, so $|[a] - a_k| \leq 1/(2n)$, by Lemma 5.3.4.

Solution. *Order:* (vi), (ii), (v), (vii), (i), (ix), (iv). No other order is admissible: line (ii) introduces N, which line (v) needs; line (v) fixes the k that lines (vii), (i) and (ix) are about; line (i) rewrites the class that line (vii) names; line (ix) bounds that class; and line (iv) concludes from line (ix), so it comes last.

Line (iii) uses the strict version of Lemma 5.3.4, which is false: termwise $<$ passes to classes only as $\leq$ (Example (Termwise Order Is Not an Order on Classes)). That is why the proof aims at $1/(2n)$ and then uses $1/(2n) < 1/n$.

Line (viii) claims a threshold N_k for every k, and the Cauchy property does not supply one for an early k: a_k can be far from all later terms. For $a = (5, 0, 0, \dots) \in \mathcal{C}$, $n = 1$ and $k = 0$, $|a_m - a_0| = 5$ for every $m \geq 1$, so no N_0 exists. The Cauchy property gives $|a_m - a_k| < 1/(2n)$ only when both m and k are at least one N, which is why line (v) takes $k \geq N$. (If line (viii) did hold for every k, the remaining lines would give the bound for every k, so it is the claim that is unjustified, not the dependence on k.)

5.4.7.3 C. Going deeper

Exercise. C1

(a) Let $x \in \mathbb{R}$ with $|x| \leq 1/n$ for every $n \in \mathbb{Z}_{>0}$. Prove that $x = 0$.
(b) Which result used in your proof needs $\mathbb{R}$ to be Archimedean? Show that the statement of (a) fails in the ordered field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field).

Solution.

(a) Suppose that $|x| > 0$. By Corollary 5.4.4 (a) with $\varepsilon := |x|$, there is $N \in \mathbb{Z}_{>0}$ with $1/n < |x|$ for every $n \geq N$; in particular $1/N < |x|$. The hypothesis with $n = N$ gives $|x| \leq 1/N$. These contradict (O1). So $|x| > 0$ is false, by $\neg$I. Since $|x| \geq 0$ (Proposition 4.1.3 (a)), trichotomy leaves $|x| = 0$, and then $x = 0$ by Proposition 4.1.3 (a).
(b) Corollary 5.4.4 (a) rests on Theorem 5.4.2, through Proposition 4.2.3. In $\mathbb{Q}(t)$, Example

(A Non-Archimedean Ordered Field) shows $0 < 1/t < 1/(n \cdot 1)$ for every $n \geq 1$. So $x := 1/t$ satisfies $|x| = x \leq 1/(n \cdot 1)$ for every $n \geq 1$, and $x \neq 0$. The statement of (a) is a form of the Archimedean property, and it fails without it.

Exercise. C2

Let r be as in Example (The Class of the Newton Sequence). Prove that the set $\{x \in \mathbb{R} : x \notin \kappa(\mathbb{Q})\}$ of irrational real numbers is dense in $\mathbb{R}$: for all $x, y \in \mathbb{R}$ with $x < y$ there is an irrational z with $x < z < y$.

Solution (Hint 1). Use Theorem 5.4.3, and the irrational number r.

Solution (Hint 2). Let $q \in \mathbb{Q}$ with $x < q < y$ (Theorem 5.4.3). How small must r/k be for $q + r/k$ to stay below y?

Solution. Let $x < y$. By Theorem 5.4.3, fix $q \in \mathbb{Q}$ with $x < q < y$. Then $y - q > 0$, and $r > 0$ (Example (The Class of the Newton Sequence)), so $\varepsilon := (y - q)/r \in \mathbb{R}_{>0}$ (Proposition 2.5.2 (a), (O4)). By Corollary 5.4.4 (a), there is $k \in \mathbb{Z}_{>0}$ with $1/k < \varepsilon$. Multiplying by $r > 0$ (Proposition 2.3.5 (c)), $r/k < y - q$. Let $z := q + r/k$. Since $r/k > 0$,

$$x < q < z < q + (y - q) = y.$$

If z were rational, $z = p$ with $p \in \mathbb{Q}$, then $r = k(p - q)$ would be rational, since κ preserves differences and products (the identities after Theorem 5.4.1, and its part (a)); this contradicts Example (The Class of the Newton Sequence). So z is irrational, and lies strictly between x and y.

The same argument works with r replaced by any positive irrational number. Between any two reals there are therefore both rationals and irrationals.

5.5 Every Cauchy Sequence of Reals Converges

We built $\mathbb{R}$ to fill the holes in $\mathbb{Q}$: the Newton sequence is Cauchy in $\mathbb{Q}$ and has no rational limit (Theorem 4.6.1), and in $\mathbb{R}$ it converges to the class it represents (Theorem 5.4.5, Example (The Class of the Newton Sequence)). Could $\mathbb{R}$ have holes of its own, Cauchy sequences of *real* numbers crowding around nothing? If so, the construction would have to be repeated, with Cauchy sequences of reals in place of Cauchy sequences of rationals, and perhaps again after that. This section proves that it need not be: every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. The proof builds the limit from rationals, by a rule, and so uses no axiom of choice.

Throughout, $F = \mathbb{R}$ whenever a result of Chapter 4 is applied, which is legitimate because $\mathbb{R}$ is an ordered field (Theorem 5.3.3). We use the identification of Section 5.4: a rational q also denotes the real number $\kappa(q)$ (Definition 5.4.1). Where an inequality is carried from $\mathbb{R}$ back to

$\mathbb{Q}$, we write κ out, because that step is where the proof must cite Theorem 5.4.1.

5.5.1 The question

A sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}$ is Cauchy in $\mathbb{R}$ (Definition 4.5.1 with $F = \mathbb{R}$) if for every $\varepsilon \in \mathbb{R}_{>0}$ there is $K \in \mathbb{N}$ with $|x_j - x_k| < \varepsilon$ for all $j, k \geq K$. By Corollary 5.4.4 (b), it is enough to test the tolerances $1/n$ with $n \in \mathbb{Z}_{>0}$, and the same holds for convergence. So the tolerances may as well be rational; the terms x_k, however, are real numbers, that is, classes of Cauchy sequences of rationals (Definition 5.2.1).

We want a real number x with $x_k \to x$. A real number is a class $[a]$ with $a \in \mathcal{C}$ (Definition 5.1.1), so we need a Cauchy sequence of rationals that "is" the limit. Where would its terms come from? The data are the classes $x_0, x_1, x_2, \ldots$, each of them an infinite set of sequences.

5.5.2 The theorem

Theorem 5.5.1. The Constructed Reals Are Cauchy-Complete

Every Cauchy sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}$ (Definition 4.5.1, $F = \mathbb{R}$) converges in $\mathbb{R}$ (Definition 4.3.2, $F = \mathbb{R}$).

Try it first (ten minutes): given a Cauchy sequence of reals, which Cauchy sequence of rationals would you build, and why would its class be the limit?

Idea. The obvious candidate is a diagonal. Each x_k has representatives; take one, $a^{(k)} \in \mathcal{C}$ with $[a^{(k)}] = x_k$, for each k, and use the diagonal terms $a^{(k)}_k$. This attempt meets two separate obstacles.

- **No control.** A representative may begin with any finite list of rationals and still represent x_k (Example (Finite Changes and Shifts)). Its k-th term may be far from x_k, and the diagonal terms need not settle down at all. Exercise C2 below turns this into a counterexample.
- **No rule.** "Take one representative for each k" picks one element from each of infinitely many sets, with no rule saying which. That is an instance of the axiom of choice (Sets, Relations, and Cardinality, Axiom 3.3.1), which the construction of $\mathbb{R}$ has not used so far.

One device removes both: replace x_k by a single rational q_k within $1/(k+1)$ of it, named by a formula. The plan has three steps.

① **The q_k form a member of $\mathcal{C}$.** Compare q_j with q_k through the reals x_j and x_k: $|q_j - q_k| \leq |q_j - x_j| + |x_j - x_k| + |x_k - q_k|$. The outer two are small once j, k are large, by the choice of the approximant; the middle one is small because (x_k) is Cauchy. Give each a third of ε.

② **Its class is a candidate.** Let $x := [(q_k)_k]$. A term approximates its class (Theorem 5.4.5), so $q_k \to x$ in $\mathbb{R}$.

③ **The candidate is the limit.** x_k is near q_k, which is near x: give each half of the tolerance.

The tolerance in ① must be **rational**, since membership of $\mathcal{C}$ is tested with $\varepsilon \in \mathbb{Q}_{>0}$ (Def-

inition 5.1.1); the triangle inequality, however, is used in $\mathbb{R}$, where the x_k live. Step 1 of the proof ends by carrying the inequality back to $\mathbb{Q}$ through κ (Theorem 5.4.1). The formula: $q_k \coloneqq m_k/(k+1)$, where m_k is the integer part of $(k+1)x_k$ in the Archimedean field $\mathbb{R}$. *Before reading on: why must m_k be* ***the*** *integer part, and not just some integer with $m_k \le (k+1)x_k < m_k + 1$?*

Proof. Let (x_k) be a Cauchy sequence in $\mathbb{R}$.

Step 0: the approximants. For $k \in \mathbb{N}$, let $m_k \in \mathbb{Z}$ be the integer part of $(k+1)x_k$ in $\mathbb{R}$ (Proposition 4.2.4 with $F = \mathbb{R}$, which is Archimedean by Theorem 5.4.2; there $\iota_{\mathbb{R}} = \kappa$ by Theorem 5.4.1 (d)), and let $q_k \coloneqq m_k/(k+1) \in \mathbb{Q}$. Since the integer part is unique, $k \mapsto q_k$ is a function $\mathbb{N} \to \mathbb{Q}$ (Sets, Relations, and Cardinality, Definition 1.4.1); no use of Sets, Relations, and Cardinality, Axiom 3.3.1 is made. Multiplying $m_k \le (k+1)x_k < m_k + 1$ by the positive element $1/(k+1)$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)), and using $\kappa(m_k)\kappa(k+1)^{-1} = \kappa(q_k)$ (Theorem 5.4.1 (a), (c)), we get

$$0 \le x_k - \kappa(q_k) < \kappa\Big(\frac{1}{k+1}\Big) \qquad \text{for every } k \in \mathbb{N}. \tag{5.5.1}$$

Step 1: $(q_k) \in \mathcal{C}$. Let $\varepsilon \in \mathbb{Q}_{>0}$. By Corollary 2.5.4 there is $K_1 \in \mathbb{Z}_{>0}$ with $1/m < \varepsilon/3$ for every integer $m \ge K_1$; for $k \ge K_1$, the integer $k+1$ is $\ge K_1$, so $1/(k+1) < \varepsilon/3$. Next, $\kappa(\varepsilon/3) \in \mathbb{R}_{>0}$ by Theorem 5.4.1 (b), since $\varepsilon/3 > 0$ and $\kappa(0) = 0$. As (x_k) is Cauchy in $\mathbb{R}$, there is $K_2 \in \mathbb{N}$ with $|x_j - x_k| < \kappa(\varepsilon/3)$ for all $j, k \ge K_2$; fix one such K_2 [$\exists$E], one object for this one ε. Let $K \coloneqq \max\{K_1, K_2\}$; it depends on ε (and on the sequence) only.

Let $j, k \ge K$. By Eq. 5.5.1 and Theorem 5.4.1 (b), $|x_j - \kappa(q_j)| < \kappa(1/(j+1)) < \kappa(\varepsilon/3)$, and likewise for k. The triangle inequality in $\mathbb{R}$ (Proposition 4.1.3 (e), applied twice, with $|-y| = |y|$ from (b)) and Proposition 2.3.5 (b) give

$$\begin{aligned} |\kappa(q_j) - \kappa(q_k)| &\le |\kappa(q_j) - x_j| + |x_j - x_k| + |x_k - \kappa(q_k)| \\ &< \kappa(\varepsilon/3) + \kappa(\varepsilon/3) + \kappa(\varepsilon/3) = \kappa(\varepsilon), \end{aligned}$$

where the last equality is Theorem 5.4.1 (a). By Theorem 5.4.1 (a) and (c), $|\kappa(q_j) - \kappa(q_k)| = |\kappa(q_j - q_k)| = \kappa(|q_j - q_k|)$. So $\kappa(|q_j - q_k|) < \kappa(\varepsilon)$, and Theorem 5.4.1 (b) gives $|q_j - q_k| < \varepsilon$ in $\mathbb{Q}$. Hence (q_k) is Cauchy in $\mathbb{Q}$: $(q_k) \in \mathcal{C}$.

Step 2: the candidate. Let $x \coloneqq [(q_k)_k] \in \mathbb{R}$ (Definition 5.2.1). By Theorem 5.4.5, $\kappa(q_k) \to x$ in $\mathbb{R}$.

Step 3: $x_k \to x$. By Corollary 5.4.4 (b) it suffices to find, for each $n \in \mathbb{Z}_{>0}$, a $K \in \mathbb{N}$ with $|x_k - x| < \kappa(1/n)$ for every $k \ge K$. Let $n \in \mathbb{Z}_{>0}$. The tolerance $\kappa(1/(2n))$ is in $\mathbb{R}_{>0}$, so Step 2 gives $K_3 \in \mathbb{N}$ with $|\kappa(q_k) - x| < \kappa(1/(2n))$ for every $k \ge K_3$. For $k \ge 2n$ we have $k+1 > 2n$, so $1/(k+1) < 1/(2n)$ (Proposition 2.5.2 (b)), and Eq. 5.5.1 with Theorem 5.4.1 (b) gives $|x_k - \kappa(q_k)| < \kappa(1/(2n))$. Let $K \coloneqq \max\{K_3, 2n\}$; it depends on n only. For $k \ge K$,

$$\begin{aligned} |x_k - x| &\le |x_k - \kappa(q_k)| + |\kappa(q_k) - x| \\ &< \kappa\Big(\frac{1}{2n}\Big) + \kappa\Big(\frac{1}{2n}\Big) = \kappa\Big(\frac{1}{n}\Big), \end{aligned}$$

by Proposition 4.1.3 (e) and Theorem 5.4.1 (a). So $x_k \to x$ in $\mathbb{R}$. This proves the theorem.

■

The answer to the prompt about m_k: existence alone would give, for each k, a non-empty set of admissible integers, and taking one from each of these infinitely many sets, with no rule, is again a use of choice. Uniqueness makes each of these sets a singleton, so "the integer part of $(k+1)x_k$" names one integer for each k, and Step 0 defines a function outright. Existence of an integer part would not be enough; the proof needs the uniqueness half of Proposition 4.2.4.

Where is each hypothesis used? The Cauchy property of (x_k) is used once, for K_2 in Step 1; Step 3 uses only the approximants and Step 2. The Archimedean property of $\mathbb{R}$ is used in Step 0, through the integer part, and in Steps 2 and 3, through Corollary 5.4.4 (b) (cited in Step 3 and in the proof of Theorem 5.4.5); the Archimedean property of $\mathbb{Q}$ is used in Step 1, for K_1. Without the Cauchy hypothesis the conclusion fails: the sequence $x_k = k$ has approximants $q_k = k$, which do not form a member of $\mathcal{C}$, and it converges to nothing in $\mathbb{R}$, being unbounded (Proposition 4.3.5, with Theorem 5.4.2 for the unboundedness).

Real numbers have no holes of their own. Together with Proposition 4.5.1 (with $F = \mathbb{R}$), the theorem says that a sequence in $\mathbb{R}$ converges in $\mathbb{R}$ **if and only if** it is Cauchy in $\mathbb{R}$. In $\mathbb{Q}$ the "only if" half holds (Proposition 4.5.1, $F = \mathbb{Q}$), and the "if" half fails (Theorem 4.6.1). A second round of the construction, starting from Cauchy sequences of reals, would find every such sequence already convergent: there is no hole left for it to fill.

5.5.3 The approximant, by a rule

Step 0 is the heart of the proof, so we look at it again. By Theorem 5.4.2, $\mathbb{R}$ is an Archimedean ordered field. So Proposition 4.2.4 applies with $F = \mathbb{R}$, with κ in the role of $\iota_{\mathbb{R}}$ (Theorem 5.4.1 (d)): for every $y \in \mathbb{R}$ there is **exactly one** $m \in \mathbb{Z}$ with $m \le y < m+1$, the integer part of y.

For a real number x_k and the positive integer $k+1$, Step 0 applies this to $y = (k+1)x_k$ and divides by $k+1$:

$$\frac{m_k}{k+1} \le x_k < \frac{m_k}{k+1} + \frac{1}{k+1}.$$

The rational $q_k := m_k/(k+1)$ is the largest multiple of $1/(k+1)$ that does not exceed x_k, and it is within $1/(k+1)$ of x_k. Picture the multiples of $1/(k+1)$ as a ladder of rungs; q_k is the rung at or just below x_k, and the ladder gets finer as k grows.

The word that matters is **exactly**. For each k the set of integers m with $m \le (k+1)x_k < m+1$ has one element, so the relation "$q_k = m/(k+1)$ for the m with $m \le (k+1)x_k < m+1$" assigns exactly one rational to each $k \in \mathbb{N}$: it is a function $\mathbb{N} \to \mathbb{Q}$ (Sets, Relations, and Cardinality, Definition 1.4.1), given by a rule. Nothing is picked.

Quick check. Let r be the real number of Example (The Class of the Newton Sequence), so $r > 0$ and $r^2 = 2$, and let $x_k := r + 1/(k+1)$. Compute q_0, q_1 and q_2.

Solution. First locate r: $(7/5)^2 = 49/25 < 2 < 9/4 = (3/2)^2$. If $r < 7/5$, then $r^2 < 49/25 < 2$, and if $r \ge 3/2$, then $r^2 \ge 9/4 > 2$ (Proposition 2.3.5 (g), with $r > 0$); both contradict $r^2 = 2$, so $7/5 \le r < 3/2$.

Next, $(k+1)x_k = (k+1)r + 1$. For $k = 0$: $12/5 \le r + 1 < 5/2$, and $2 \le 12/5$, $5/2 < 3$, so $m_0 = 2$. For $k = 1$: $19/5 \le 2r + 1 < 4$, and $3 \le 19/5$, so $m_1 = 3$. For $k = 2$: $26/5 \le 3r + 1 < 11/2$, and $5 \le 26/5$, $11/2 < 6$, so $m_2 = 5$. Hence

$$q_0 = 2, \qquad q_1 = \frac{3}{2}, \qquad q_2 = \frac{5}{3}.$$

Each q_k is within $1/(k+1)$ below x_k; for instance $x_2 = r + 1/3 \in [26/15, 11/6)$ and $q_2 = 5/3 = 25/15$.

5.5.4 Choice, and why it was not needed

The proof used ∃E many times: to fix K_1, K_2, K_3. Each of those fixes **one** object, for one given ε or n, and is ordinary logic. The axiom of choice is something else: one pick from each of infinitely many sets at once, with no rule. The next example shows exactly where it would have entered.

Example. Picking Representatives Is a Use of Choice

Let (x_k) be a sequence in $\mathbb{R}$. A proof that says "for each k, pick a representative $a^{(k)} \in x_k$" asks for what? And does the proof of Theorem 5.5.1 ask for it?

Solution. Each x_k is a class, a non-empty subset of $\mathcal{C}$ (Sets, Relations, and Cardinality, Definition 2.5.2). The family $(x_k)_{k\in\mathbb{N}}$ is a family of non-empty sets indexed by $\mathbb{N}$, and "pick $a^{(k)} \in x_k$ for each k" asks for a function $c\colon \mathbb{N} \to \mathcal{C}$ with $c(k) \in x_k$ for every k: a choice function for the family. Its existence, with no rule given, is what Sets, Relations, and Cardinality, Axiom 3.3.1 provides. A single ∃E would give one representative of one x_k; it cannot give infinitely many in one step.

The proof of Theorem 5.5.1 never picks a representative of any x_k. It uses each x_k only as an element of the ordered field $\mathbb{R}$ (its arithmetic and order, in Steps 0, 1 and 3), never through a representative; the Cauchy condition on (x_k) enters only in Step 1. The only sequence it builds, (q_k), is given by a rule. So Sets, Relations, and Cardinality, Axiom 3.3.1 is not used.

In fact the theorem produces a rule for representatives as a by-product. Run its construction on the constant sequence $x_k = y$ for a fixed $y \in \mathbb{R}$, which is Cauchy in $\mathbb{R}$ (Theorem 4.5.3 (a)). Step 1 shows that the approximants $q_k = m_k/(k+1)$, with m_k the integer part of $(k+1)y$, form a member of $\mathcal{C}$, and Step 3 that the constant sequence y converges to $[(q_k)]$. It also converges to y (Example (First Limits) (a)), so $[(q_k)] = y$ by Theorem 4.3.4. Thus (q_k) is a representative of y, named by a formula in y. Choice is not needed for representatives at all; what goes wrong is picking them **without saying how**.

Warning. **The uncontrolled diagonal.** Even with representatives in hand, the diagonal term $a_k^{(k)}$ need not be near x_k. A representative of x_k may start with any finite list of rationals (Example (Finite Changes and Shifts)): for $x_k = 0$ for every k, the sequence equal to k at the indices $n \le k$ and to 0 afterwards represents 0, and its k-th term is k. Approximating x_k needs a **bound**, such as $|x_k - q_k| < 1/(k+1)$, and a term of an arbitrary representative carries none.

5.5.5 Exercises

5.5.5.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 5.5.1.

(a) Why $\varepsilon/3$ in Step 1? Which three distances does it bound, and in which field is each of them measured?

(b) Why $1/(2n)$ in Step 3, for each of the two distances there?

(c) Where is the Cauchy property of (x_k) used? Is it used in Step 3?

(d) Run the construction on the constant sequence $x_k = r$ (the real number of Example (The Class of the Newton Sequence)). Which class is x?

(e) Write "(x_k) converges in $\mathbb{R}$" in symbols, and its negation.

Solution.

(a) Step 1 bounds $|q_j - q_k|$ through x_j and x_k: the three distances are $|\kappa(q_j) - x_j|$ and $|x_k - \kappa(q_k)|$, each below $\kappa(1/(j+1))$ or $\kappa(1/(k+1))$ by the choice of approximant, and $|x_j - x_k|$, small because (x_k) is Cauchy. All three are measured in $\mathbb{R}$; only the conclusion $|q_j - q_k| < \varepsilon$ is carried back to $\mathbb{Q}$. Three thirds make one ε.

(b) Two distances are added: $|x_k - \kappa(q_k)| < \kappa(1/(2n))$ for $k \geq 2n$, from the approximant bound, and $|\kappa(q_k) - x| < \kappa(1/(2n))$ for $k \geq K_3$, from Theorem 5.4.5. Two halves make $1/n$.

(c) Only in Step 1, for the threshold K_2. Step 3 uses the approximant bound and Step 2, not the Cauchy property directly: by then it has been spent on making (q_k) a member of $\mathcal{C}$.

(d) The approximants are $q_k = m_k/(k+1)$, with m_k the integer part of $(k+1)r$, and $x := [(q_k)]$. By the theorem's proof, $x_k \to x$; the constant sequence r also tends to r, so $x = r$ by Theorem 4.3.4. Concretely, $q_0 = 1$, $q_1 = 1$ (since $2 \leq 2r < 3$), $q_2 = 4/3$ (since $4 \leq 3r < 5$): a third sequence of rationals, besides the Newton sequence of Example (A Rational Sequence Aimed at the Gap) and the truncations (d_k), both in the class r by Example (The Class of the Newton Sequence).

(e) Converges: $\exists x \in \mathbb{R}\ \forall \varepsilon \in \mathbb{R}_{>0}\ \exists K \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \geq K \Rightarrow |x_k - x| < \varepsilon)$. Negation (Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2, and $\neg(y < \varepsilon)$ read as $y \geq \varepsilon$ by (O1)):

$$\forall x \in \mathbb{R}\ \ \exists \varepsilon \in \mathbb{R}_{>0}\ \ \forall K \in \mathbb{N}\ \ \exists k \in \mathbb{N}\ \ (k \geq K \wedge |x_k - x| \geq \varepsilon).$$

Every candidate is beaten by some tolerance beyond every threshold.

5.5.5.2 B. Practice

Exercise. B1

Which results of Chapter 4, applied with $F = \mathbb{R}$, show that in $\mathbb{R}$ every convergent sequence

is Cauchy and has only one limit? Why may they be applied? Deduce that a sequence in $\mathbb{R}$ converges if and only if it is Cauchy.

Solution. Proposition 4.5.1 (convergent sequences are Cauchy) and Theorem 4.3.4 (uniqueness of limits). Both are proved in Chapter 4 for an arbitrary ordered field F, and $\mathbb{R}$ is an ordered field by Theorem 5.3.3, so they apply with $F = \mathbb{R}$ and tolerances in $\mathbb{R}_{>0}$.

If (x_k) converges in $\mathbb{R}$, it is Cauchy in $\mathbb{R}$ by the first; if it is Cauchy in $\mathbb{R}$, it converges by Theorem 5.5.1. So the two properties coincide in $\mathbb{R}$, and the limit is unique by the second.

Exercise. B2

Let (p_k) be a sequence of rationals such that $(\kappa(p_k))_k$ converges in $\mathbb{R}$ to some x. Prove that $(p_k) \in \mathcal{C}$ and that $x = [(p_k)]$.

Hint: a convergent sequence is Cauchy.

Solution. By Proposition 4.5.1 with $F = \mathbb{R}$, $(\kappa(p_k))$ is Cauchy in $\mathbb{R}$. Let $\varepsilon \in \mathbb{Q}_{>0}$. Then $\kappa(\varepsilon) \in \mathbb{R}_{>0}$ (Theorem 5.4.1 (b)), so there is K with $|\kappa(p_j) - \kappa(p_k)| < \kappa(\varepsilon)$ for $j, k \geq K$. As in Step 1 of the proof of Theorem 5.5.1, the left side is $\kappa(|p_j - p_k|)$ (Theorem 5.4.1 (a), (c)), and Theorem 5.4.1 (b) gives $|p_j - p_k| < \varepsilon$. So (p_k) is Cauchy in $\mathbb{Q}$, that is, $(p_k) \in \mathcal{C}$.

By Theorem 5.4.5, $\kappa(p_k) \to [(p_k)]$. Also $\kappa(p_k) \to x$. By Theorem 4.3.4 with $F = \mathbb{R}$, $x = [(p_k)]$.

So the real numbers that are limits of sequences of rationals in $\mathbb{R}$ are exactly the classes of those sequences; with Theorem 5.4.5, every real number is one.

5.5.5.3 C. Going deeper

Exercise. C1: Exit check: Cauchy completeness

Prove that every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$. Say where the axiom of choice could have entered your proof, and why it does not.

Solution. Let (x_k) be Cauchy in $\mathbb{R}$.

Approximants by a rule. For $k \in \mathbb{N}$ let m_k be the integer part of $(k+1)x_k$, unique by Proposition 4.2.4 ($\mathbb{R}$ is Archimedean, Theorem 5.4.2), and $q_k := m_k/(k+1)$. Then $0 \leq x_k - q_k < 1/(k+1)$, and $k \mapsto q_k$ is a function because the integer part is unique. This is where choice could have entered: "for each k take some rational within $1/(k+1)$ of x_k", or "take a representative of each x_k", makes infinitely many picks with no rule, a use of Sets, Relations, and Cardinality, Axiom 3.3.1. The integer part names one rational for each k, so no choice is made.

Cauchy in $\mathbb{Q}$. Let $\varepsilon \in \mathbb{Q}_{>0}$. Fix K_1 with $1/(k+1) < \varepsilon/3$ for $k \geq K_1$ (Corollary 2.5.4) and K_2 with $|x_j - x_k| < \varepsilon/3$ for $j, k \geq K_2$ (the Cauchy property, at the positive real $\kappa(\varepsilon/3)$). For $j, k \geq \max\{K_1, K_2\}$, the triangle inequality in $\mathbb{R}$ gives $|q_j - q_k| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$ in $\mathbb{R}$;

since κ preserves sums, absolute values and order and reflects order (Theorem 5.4.1 (a)–(c)), the same inequality holds in $\mathbb{Q}$. So $(q_k) \in \mathcal{C}$.

The limit. Let $x := [(q_k)]$. By Theorem 5.4.5, $q_k \to x$ in $\mathbb{R}$. Let $\varepsilon \in \mathbb{R}_{>0}$. By Corollary 5.4.4 (a) there is $n \in \mathbb{Z}_{>0}$ with $1/n < \varepsilon$. Fix K_3 with $|q_k - x| < 1/(2n)$ for $k \geq K_3$. For $k \geq \max\{K_3, 2n\}$, $|x_k - x| \leq |x_k - q_k| + |q_k - x| < 1/(2n) + 1/(2n) = 1/n < \varepsilon$. So $x_k \to x$.

Marking scheme. (10 points) - 2: an approximant $q_k \in \mathbb{Q}$ defined by a rule (unique integer part, or another explicit formula), with an explicit bound on $|x_k - q_k|$ tending to 0. - 3: $(q_k) \in \mathcal{C}$ by an $\varepsilon/3$ (or equivalent) estimate, with **rational** ε and the passage from $\mathbb{R}$ back to $\mathbb{Q}$ through κ justified. - 1: the candidate $x = [(q_k)]$ named. - 2: $q_k \to x$ in $\mathbb{R}$, by Theorem 5.4.5, or directly as in its proof: $|x - \kappa(q_k)| = [(|q_m - q_k|)_m]$ (Proposition 5.3.5), bounded through Lemma 5.3.4, which gives only $\leq$. Either route earns both points. - 2: $x_k \to x$, with the threshold chosen after the tolerance and depending on it only. Fatal errors (0 for the step): approximants or representatives "chosen" with no rule and no mention of choice; a real ε used to test membership of $\mathcal{C}$; "$q_k \to x$ because x is the limit of (q_k)" (circular: that is what Theorem 5.4.5 proves). A proof that uses choice openly (representatives, with a threshold for each) and says so proves the theorem, but earns at most 1 of the 2 points of the first step, since the question asks for a construction that avoids it.

Exercise. C2: Find the error

Here is an argument for Theorem 5.5.1. Is its conclusion true? Does the argument prove it? If not, name the first line that fails and why.

(1) Let (x_k) be a Cauchy sequence in $\mathbb{R}$.
(2) For each $k \in \mathbb{N}$, pick a representative $a^{(k)} \in \mathcal{C}$ of x_k.
(3) Let $q_k := a_k^{(k)}$, the k-th term of the k-th representative.
(4) The sequence (q_k) is Cauchy in $\mathbb{Q}$, and $|x_k - \kappa(q_k)| \to 0$ in $\mathbb{R}$, since each $a^{(k)}$ is Cauchy and (x_k) is Cauchy.
(5) Let $x := [(q_k)]$. Then $x_k \to x$, by an $\varepsilon/2$ argument through the terms q_k.

Solution (Hint 1). Ask whether $a_k^{(k)}$ has to be close to x_k at all.

Solution (Hint 2). Try the constant sequence $x_k = 0$ for every k, and representatives that are not 0 at their first few terms.

Solution. The conclusion is true: it is Theorem 5.5.1.

The failing line is (4). Each $a^{(k)}$ being Cauchy says that the terms of one representative settle down eventually, past a threshold that depends on k; nothing ties the k-th term of $a^{(k)}$ to x_k, and nothing ties different representatives to each other at index k. The Cauchy property of (x_k) is about the classes, and the line uses it as if it were about the diagonal terms.

Counterexample to the step. Let $x_k = 0$ for every k, a Cauchy sequence in $\mathbb{R}$. For each

k, let $a_n^{(k)} := k$ for $n \le k$ and $a_n^{(k)} := 0$ for $n > k$. Each $a^{(k)}$ is eventually equal to the zero sequence, so it is in $\mathcal{C}$ and represents 0 (Example (Finite Changes and Shifts)). Then $q_k = a_k^{(k)} = k$, and (k) is not Cauchy in $\mathbb{Q}$: consecutive terms satisfy $|q_{k+1} - q_k| = 1$, so the Cauchy condition fails for $\varepsilon = 1$. Nor does $|x_k - \kappa(q_k)| = \kappa(k)$ tend to 0: it is ≥ 1 for every $k \ge 1$. The premises of (4) hold and both parts of its conclusion fail.

A second, separate defect: line (2). "For each k, pick" makes one pick from each of infinitely many classes with no rule: an unflagged use of Sets, Relations, and Cardinality, Axiom 3.3.1 (Example (Picking Representatives Is a Use of Choice)). Even if (4) were repaired, the argument would use choice without saying so. Line (5) is correct given (4): $\kappa(q_k) \to x$ in $\mathbb{R}$ by Theorem 5.4.5, and $|x_k - x| \le |x_k - \kappa(q_k)| + |\kappa(q_k) - x|$. The bound on $|x_k - \kappa(q_k)|$ is needed: knowing only that (q_k) is Cauchy would not do (with $a_n^{(k)} = 1$ for $n \le k$ and 0 after, $q_k = 1$ for every k, yet $x_k = 0$ does not tend to $[(1)] = 1$).

Repair. Either keep representatives and control the diagonal (for each k, go far enough out in $a^{(k)}$ that its terms are within $1/(k+1)$ of x_k, which Theorem 5.4.5 makes possible, and use that term in place of $a_k^{(k)}$), at the price of a choice of representatives and of thresholds, flagged; or, better, drop representatives and use the integer-part approximant $q_k = m_k/(k+1)$ of the text, which needs no choice at all.

Marking scheme. (10 points) - 2: says the conclusion is true, and cites the theorem. - 3: identifies line (4) as the failing line, and says why: nothing ties $a_k^{(k)}$ to $a_j^{(j)}$ for $j \ne k$, or to x_k. - 3: a counterexample to the step, with every representative checked to lie in its class. - 2: names the use of choice in line (2), and a repair that replaces q_k by a rational with a bound on $|x_k - \kappa(q_k)|$ tending to 0 (a later term of $a^{(k)}$, or the integer-part approximant). Fatal errors (0 for the step): calling the conclusion false; a "counterexample" in which the sequence (x_k) is not Cauchy.

5.6 The Real Numbers Form a Complete Ordered Field

Section 5.5 proved one form of completeness: in $\mathbb{R}$ every Cauchy sequence converges (Theorem 5.5.1). *Analysis on the Real Line*, Chapter 1 starts from another form, about sets rather than sequences: every non-empty set of real numbers that is bounded above has a least upper bound. In $\mathbb{Q}$ both fail, the first by Theorem 4.6.1 and the second by Theorem 2.6.2. This section names the second property for an arbitrary ordered field, proves that an ordered field that is Archimedean and Cauchy-complete has it, and concludes that the constructed $\mathbb{R}$ is a complete ordered field. The proof bisects intervals, and one line of it, which we shall point at, cannot do without Archimedes.

Throughout, F is an ordered field (Definition 2.3.2). Its order $\le$ is a total order (Proposition 4.1.2 (a)), so the words "upper bound", "bounded above" and "supremum" of *Sets, Relations, and Cardinality*, Section 2.6 apply to subsets of F (Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5). When $F = \mathbb{R}$, we keep the identification of Section 5.4, writing q for $\kappa(q)$.

5.6.1 Complete ordered fields

Section 2.6 found a gap in $\mathbb{Q}$ without any sequence: the set $A = \{x \in \mathbb{Q} : x^2 < 2\}$ is non-empty and bounded above in $\mathbb{Q}$, and among its rational upper bounds there is no least one (Theorem 2.6.2). The upper bounds crowd down toward a place where $\mathbb{Q}$ has no element. A field without such gaps is one in which every set that could have a least upper bound does have one.

Every non-empty set that is bounded above has a least upper bound.

Definition 5.6.1. Complete Ordered Field

An ordered field F is **complete** if **every non-empty** subset $A \subseteq F$ that is **bounded above** in $(F, \leq)$ (Sets, Relations, and Cardinality, Definition 2.6.2) has a supremum in F (Sets, Relations, and Cardinality, Definition 2.6.5). In symbols, writing $\mathrm{ub}(u)$ for $\forall a \in A \;\; a \leq u$,

$$\forall A \subseteq F \;\; \big((A \neq \emptyset \wedge \exists u \in F \; \mathrm{ub}(u)) \Rightarrow \\ \exists s \in F \; (\mathrm{ub}(s) \wedge \forall u \in F \; (\mathrm{ub}(u) \Rightarrow s \leq u))\big).$$

An ordered field in which every Cauchy sequence (Definition 4.5.1) converges is called **Cauchy-complete**.

In words: the hypothesis asks for a set with at least one element and at least one upper bound in F; the conclusion asks for an element s of F that is an upper bound, clause (S1), and lies below every upper bound, clause (S2). The supremum need not belong to A.

Well-definedness. A set has at most one supremum: if s and s' both are, then (S2) for each, with the other as the upper bound, gives $s \leq s'$ and $s' \leq s$, so $s = s'$ by antisymmetry. So "the supremum of A", written $\sup A$, names one element when it exists.

Examples.

The negative elements. In every ordered field, $A \coloneqq \{x \in F : x < 0\}$ has supremum 0. It is non-empty ($-1 \in A$). The element 0 is an upper bound. If an upper bound b had $b < 0$, then $b < b/2 < 0$ (Proposition 4.1.2 (b)), so $b/2 \in A$ lies above b, which is impossible. So every upper bound b satisfies $b \geq 0$, and $\sup A = 0 \notin A$. No completeness was needed: some sets have suprema in every ordered field.

The rationals. $\mathbb{Q}$ is **not** complete: Theorem 2.6.2 exhibits a non-empty subset bounded above with no supremum in $\mathbb{Q}$.

The constructed reals. $\mathbb{R}$ is complete; this is the main theorem of the section, proved at its end.

Second reading · *A degenerate case, and non-examples by minimal change*

A singleton. $\sup\{a\} = a$, in every ordered field. This degenerate case matters below: a set may be bounded above by its own element.

Drop "non-empty": the empty set is bounded above by every element of F, so its upper bounds are all of F, and F has no least element ($x - 1 < x$). Drop "bounded above": F itself has no upper bound ($x < x + 1$), so no supremum. With either clause removed, no ordered field at all

would be complete.

Why this definition. It asks about suprema, not maxima: the set of negative elements has no greatest element in any ordered field, since $x < x/2 < 0$ for $x < 0$, and asking for maxima would make no field complete. It asks about sets, not sequences, the form *Analysis on the Real Line*, Chapter 1 takes as its axiom. "Complete" alone always means this property; the sequence property of Section 5.5 is "Cauchy-complete".

Warning. The supremum of a set is an upper bound **not necessarily in the set**, and "complete" is not "every non-empty bounded set has a **largest** element": the set of negative elements above has supremum 0 and no largest element. A proof that ends "so $\sup A \in A$" needs a reason, such as A being finite.

So $\mathbb{Q}$ is neither complete nor Cauchy-complete. Section 5.5 removed the second gap for $\mathbb{R}$; this section removes the first.

5.6.2 Powers of two in an ordered field

The bisection below halves an interval at each step, so after k steps its length is the first length divided by 2^k. We need 2^k to grow at least as fast as $k+1$, inside the field.

Lemma 5.6.1. A Lower Bound for Powers of Two

Let F be an ordered field, and let $(2 \cdot 1_F)^k$ be the power in F of Definition 4.2.1 (b). Then

$$(2 \cdot 1_F)^k \geq (k+1) \cdot 1_F \qquad \text{for every } k \in \mathbb{N}.$$

In particular $(2 \cdot 1_F)^k \geq 1_F > 0$.

Idea. Induction on k. Multiplying by 2 doubles the right side, $(k+1)+(k+1)$, and doubling gains at least 1, which is all the step needs.

Proof. We use Logic and Proof, Axiom 4.2.1, read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b), for $P(k)$: $(2 \cdot 1_F)^k \geq (k+1) \cdot 1_F$.

Base $P(0)$. $(2 \cdot 1_F)^0 = 1_F$ by Definition 4.2.1 (b), and $(0+1) \cdot 1_F = 1 \cdot 1_F = 0 \cdot 1_F + 1_F = 1_F$ by (a) of that definition, since $0 + 1 = 1$ (Definition 1.5.1 (c)). Equality holds.

Step. Assume $P(k)$, and let $y := (k+1) \cdot 1_F$. By Definition 4.2.1 (a) and Definition 1.5.1 (c), $2 \cdot 1_F = 1 \cdot 1_F + 1_F = 1_F + 1_F$, which is positive (Proposition 4.1.2 (b)). By Definition 4.2.1 (b), $P(k)$, and Proposition 2.3.5 (c) (or equality, if $P(k)$ holds with equality),

$$(2 \cdot 1_F)^{k+1} = (2 \cdot 1_F)^k (2 \cdot 1_F) \geq y(1_F + 1_F) = y + y,$$

the last step by distributivity (F9) and (F7). Since $k + 1 \geq 1$ (Theorem 1.6.1 (d)), Proposition 4.2.1 (b) gives $y \geq 1_F$, so $y + y \geq y + 1_F$ (Proposition 2.3.5 (b)). Finally $y + 1_F = ((k+1)+1) \cdot 1_F$ by Definition 4.2.1 (a). Hence $(2 \cdot 1_F)^{k+1} \geq ((k+1)+1) \cdot 1_F$, which is

$P(k+1)$.

The last sentence of the lemma follows, since $(k+1)\cdot 1_F \geq 1_F > 0$ (Proposition 4.2.1 (b), as $k+1 \geq 1$).

■

From here on, 2^k means $(2\cdot 1_F)^k$. By Proposition 4.2.1 (c) it equals $(2^k)\cdot 1_F$, the multiple of 1_F by the natural number 2^k, so the two readings agree; the proofs below use only the power in F and the lemma. Since $2^k > 0$, the quotient $D/2^k$ is defined for every $D \in F$.

5.6.3 From Cauchy-complete to complete

Theorem 5.6.2. Archimedean and Cauchy-Complete Implies Complete

Let F be an ordered field that is Archimedean (Definition 4.2.2) and Cauchy-complete (Definition 5.6.1). Then F is complete.

Idea. Given A, we want a point s that is an upper bound and below every upper bound. Trap it. Start with an interval $[l_0, u_0]$ whose right end is an upper bound of A and whose left end is not. Halve it: if the midpoint is an upper bound, it becomes the new right end; if not, the new left end. **Bisect and keep the invariant**: right end an upper bound, left end not.

The right ends decrease and the left ends increase, and both stay within the current interval, whose length is $D/2^k$. If that length goes to 0, the right ends form a Cauchy sequence, and Cauchy-completeness gives a limit s. Then s is an upper bound because every right end is, and it is the least one because left ends, which are not upper bounds, come arbitrarily close to it.

Step roadmap: ① the starting interval; ② the bisection, defined by a rule, and its invariants; ③ the lengths $D/2^k$ tend to 0; ④ (u_k) converges to some s; ⑤ s is the supremum. *Before reading on: which of these steps uses that F is Archimedean?*

Proof. Let $A \subseteq F$ be non-empty and bounded above.

Step 1: the starting interval. Fix $a_0 \in A$ [∃E] and an upper bound M of A [∃E]; each is one object, fixed once, and no choice is involved. Let $l_0 := a_0 - 1$, $u_0 := M$ and $D := u_0 - l_0$. Then $l_0 < a_0 \leq M = u_0$, so $D > 0$; u_0 is an upper bound of A; and l_0 is not, because $a_0 \in A$ and $a_0 > l_0$.

Step 2: the bisection. Define $g\colon \mathbb{N} \times (F \times F) \to F \times F$ by

$$g(k,(l,u)) := \begin{cases} (l,c) & \text{if } c \text{ is an upper bound of } A, \\ (c,u) & \text{otherwise,} \end{cases} \qquad c := \frac{l+u}{2}.$$

For each input exactly one case applies, so g is a function, given by a rule. By Sets, Relations, and Cardinality, Theorem 2.3.1 (read on $\mathbb{N} = \omega$ through Corollary 1.7.3 (b)), with $X = F \times F$ and starting value (l_0, u_0), there is exactly one sequence $k \mapsto (l_k, u_k)$ with $(l_{k+1}, u_{k+1}) =$

$g(k, (l_k, u_k))$ for every k.

We prove by induction on k that, for every $k \in \mathbb{N}$:

(I1) u_k is an upper bound of A;
(I2) l_k is not an upper bound of A;
(I3) $u_k - l_k = D/2^k$.

The base is Step 1, with $D/2^0 = D/1_F = D$. Suppose (I1)–(I3) hold for k, and let $c := (l_k + u_k)/2$. Then $c - l_k = u_k - c = (u_k - l_k)/2 = D/(2^k \cdot 2) = D/2^{k+1}$, by field arithmetic (Lemma 4.1.1 (b)) and Definition 4.2.1 (b). If c is an upper bound, then $(l_{k+1}, u_{k+1}) = (l_k, c)$: the right end c is an upper bound, the left end l_k is not, by (I2), and $c - l_k = D/2^{k+1}$. Otherwise $(l_{k+1}, u_{k+1}) = (c, u_k)$: the left end c is not an upper bound, the right end u_k is, by (I1), and $u_k - c = D/2^{k+1}$. This proves (I1)–(I3) for $k+1$.

By (I3), $u_k - l_k = D/2^k > 0$, so $l_k < c < u_k$ (Proposition 4.1.2 (b)); in either case $l_k \le l_{k+1}$ and $u_{k+1} \le u_k$. By induction on i, $l_k \le l_{k+i}$ and $u_{k+i} \le u_k$ for every $i \in \mathbb{N}$, and every $j \ge k$ is $k + i$ for some i (Proposition 1.6.3 (a)). So for $j \ge k$, $l_k \le l_j < u_j \le u_k$, and therefore

$$0 \le u_k - u_j \le u_k - l_k = \frac{D}{2^k} \qquad (j \ge k). \tag{5.6.1}$$

Step 3: the lengths tend to 0. Let $\varepsilon \in F_{>0}$. Since F is **Archimedean**, there is $n \in \mathbb{N}$ with

$$\frac{D}{\varepsilon} < n \cdot 1_F. \tag{5.6.2}$$

Let $k \ge n$. By Lemma 5.6.1 and Proposition 4.2.1 (b), $2^k \ge (k+1) \cdot 1_F > k \cdot 1_F \ge n \cdot 1_F > D/\varepsilon > 0$. By Proposition 2.5.2 (b), $1/2^k < (D/\varepsilon)^{-1} = \varepsilon/D$ (Lemma 4.1.1 (a)), and multiplying by $D > 0$ (Proposition 2.3.5 (c)) gives $D/2^k < \varepsilon$. So $|D/2^k - 0| < \varepsilon$ for every $k \ge n$: the sequence $(D/2^k)_k$ is null.

Consequently (u_k) is Cauchy in F: for ε and n as above and $j, k \ge n$, say $j \ge k$ (the other case is symmetric), Eq. 5.6.1 gives $|u_j - u_k| = |u_k - u_j| = u_k - u_j \le D/2^k < \varepsilon$ (Proposition 4.1.3 (b), Definition 4.1.1).

Steps 1–3 use only that F is an Archimedean ordered field. Step 4 is the only one that uses Cauchy-completeness.

Step 4: the limit. Since F is Cauchy-complete, (u_k) converges; fix $s \in F$ with $u_k \to s$ [∃E]. Since $l_k = u_k - D/2^k$ and $D/2^k \to 0$, Theorem 4.4.3 (b) gives $l_k \to s - 0 = s$.

Step 5: s is the supremum. (S1): let $a \in A$. By (I1), $a \le u_k$ for every k, so $a \le s$ by Proposition 4.4.4 (b). So s is an upper bound of A.

(S2): let b be an upper bound of A, and suppose, for ¬I, that $b < s$. Then $\varepsilon := s - b > 0$, and since $l_k \to s$ there is K with $|l_k - s| < \varepsilon$ for $k \ge K$; in particular $l_K > s - \varepsilon = b$ (Proposition 4.1.3 (d)). By (I2), l_K is not an upper bound: some $a \in A$ has $a \not\le l_K$ (Logic and Proof, Proposition 6.2.3 (a)), that is, $a > l_K$ by totality (Proposition 4.1.2 (a)). Then $a > l_K > b$, against b being an upper bound. So $b < s$ is false, and $s \le b$ by (O1).

By (S1) and (S2), $s = \sup A$ (Sets, Relations, and Cardinality, Definition 2.6.5). Since A was an arbitrary non-empty subset bounded above, F is complete.

■

The answer to the prompt: Step 3, and only the line Eq. 5.6.2. The rest of Step 3 is algebra in F and Lemma 5.6.1, which hold in every ordered field. Archimedes is what makes the halved lengths smaller than an arbitrary positive ε; without it, $D/2^k$ could stay above some positive element for every k, and (u_k) need not be Cauchy.

Each hypothesis has one job. Archimedes makes (u_k) Cauchy (Step 3); Cauchy-completeness supplies the limit (Step 4); "non-empty" supplies a left end that is not an upper bound, and "bounded above" a right end that is one. The invariants (I1) and (I2) are spent in Step 5, one on each clause of the supremum.

Warning. **A false conclusion: "every Cauchy-complete ordered field is complete".** Run the bisection without Archimedes. Steps 1, 2, 4 and 5 never mention it; the unjustified line is Eq. 5.6.2, "$D/\varepsilon < n \cdot 1_F$ for some n". In the non-Archimedean field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field) it fails for $D = 1$ and $\varepsilon = 1/t$: every $2^k = (2^k) \cdot 1$ lies below t, so $1/2^k > 1/t$ for every k (Proposition 2.5.2 (b)), and the bisection lengths never get below the tolerance $1/t$. (It shows only that the line fails, not that the conclusion does.)

Remark. **Not proved here.** There are ordered fields that are Cauchy-complete and not Archimedean, for instance a field of formal Laurent series in a positive infinitesimal t. So the conclusion of the warning is false, and the Archimedean hypothesis of Theorem 5.6.2 cannot be dropped. This book does not prove that such a field exists, and nothing in it depends on one; *Analysis on the Real Line*, Chapter 11 returns to the question when it compares the forms of completeness.

5.6.4 The constructed reals

Theorem 5.6.3. The Constructed Reals Form a Complete Ordered Field

The set $\mathbb{R}$ of Definition 5.2.1, with the operations of Definition 5.2.2 and the order of Definition 5.3.2, is a complete ordered field. The map $\kappa\colon \mathbb{Q} \to \mathbb{R}$ of Definition 5.4.1 is injective, preserves $+$, $\cdot$ and $<$, and has dense image: between any two real numbers $x < y$ there is $\kappa(q)$ with $q \in \mathbb{Q}$.

Proof. $\mathbb{R}$ is an ordered field by Theorem 5.3.3, Archimedean by Theorem 5.4.2, and Cauchy-complete by Theorem 5.5.1. By Theorem 5.6.2 it is complete. The statements about κ are Theorem 5.4.1 (a), (b) (injectivity is part of (b)) and Theorem 5.4.3.

■

This ends the construction begun with the axioms of set theory in Chapter 1: $\mathbb{N}$, then $\mathbb{Z}$ and $\mathbb{Q}$ as classes of pairs, now $\mathbb{R}$ as classes of Cauchy sequences. No property of $\mathbb{R}$ was assumed.

Quick check. Run the bisection of the proof of Theorem 5.6.2 in $\mathbb{R}$ on $A := \{q \in \mathbb{Q} : q^2 < 2\}$, viewed as a subset of $\mathbb{R}$, with $a_0 = 1$ and $M = 2$. List (l_k, u_k) for $k = 0, 1, 2, 3$.

Solution. $l_0 = a_0 - 1 = 0$ and $u_0 = 2$. Two tests decide the midpoints. An element $c > 0$ with $c^2 \geq 2$ is an upper bound of A: if some $q \in A$ had $q > c$, then $q^2 > c^2 \geq 2$ (Proposition 2.3.5 (g)), against $q^2 < 2$. An element exceeded by some member of A is not an upper bound; we use the member $7/5$, which is in A because $(7/5)^2 = 49/25 < 2$.

- $k = 0 \to 1$: $c = 1$; $7/5 \in A$ and $7/5 > 1$, so 1 is not an upper bound: $(l_1, u_1) = (1, 2)$.
- $k = 1 \to 2$: $c = 3/2$; $(3/2)^2 = 9/4 \geq 2$, so $3/2$ is an upper bound: $(l_2, u_2) = (1, 3/2)$.
- $k = 2 \to 3$: $c = 5/4$; $7/5 \in A$ and $7/5 > 5/4$, so $5/4$ is not an upper bound: $(l_3, u_3) = (5/4, 3/2)$.

So the pairs are $(0, 2)$, $(1, 2)$, $(1, 3/2)$, $(5/4, 3/2)$, and $u_3 - l_3 = 1/4 = D/2^3$ with $D = 2$, as (I3) says.

5.6.5 What we built, and what comes next

The table collects what the chapter proved about $\mathbb{R}$, beside the status of the same property in $\mathbb{Q}$.

Property	In $\mathbb{R}$	In $\mathbb{Q}$
ordered field	yes: Theorem 5.3.3	yes: Theorem 2.5.1
Archimedean	yes: Theorem 5.4.2	yes: Theorem 2.5.3
an element whose square is 2	yes: Example (The Class of the Newton Sequence)	no: Theorem 2.6.1
Cauchy-complete	yes: Theorem 5.5.1	no: Theorem 4.6.1
complete	yes: Theorem 5.6.3	no: Theorem 2.6.2

The first two rows are shared; the last three are what the construction was for.

Several chapters build on this one. Chapter 6 proves that any two complete ordered fields are isomorphic, by exactly one order-preserving field isomorphism; so "the" real numbers are well defined, and no theorem about them depends on how they were built. The optional Chapter 7 builds a complete ordered field from Dedekind cuts, sets of rationals instead of sequences, and by Chapter 6 it is a copy of this one. *Analysis on the Real Line*, Chapter 1 starts from the axioms of a complete ordered field; the theorem above gives those axioms a model, so they cannot contradict each other unless the set theory of Chapter 1, in which $\mathbb{Q}$ was built, already does. *Algebraic Preliminaries*, Chapter 1 builds the complex numbers on this $\mathbb{R}$. And *Metric and Topological Spaces*, Chapter 2 runs the construction of this chapter for an arbitrary metric space, as its completion.

↻ **Second reading** · *History*

Charles Méray published a construction of the irrational numbers from convergent sequences of rationals in 1869. Georg Cantor gave an equivalent construction from fundamental sequences,

our Cauchy sequences, in 1872, the year in which Richard Dedekind published his construction by cuts.

5.6.6 Exercises

5.6.6.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 5.6.2 and the definition of a complete ordered field.

(a) Where in the proof is the invariant "l_k is not an upper bound" used, and where "u_k is an upper bound"?

(b) Why does the proof start with $l_0 = a_0 - 1$, and not with $l_0 = a_0$?

(c) State the idea of the proof in two sentences.

(d) Write "F is complete" in symbols, and form its negation.

Solution.

(a) "Not an upper bound", (I2), is used in Step 5 for (S2): a left end l_K above a supposed smaller upper bound b yields an element of A above b. "Upper bound", (I1), is used in Step 5 for (S1): every element of A lies below every u_k, hence below their limit. (Both are also carried through the induction of Step 2, which is what keeps them true.)

(b) The left end must **not** be an upper bound, and a_0 might be one: for $A = \{a_0\}$, the element a_0 is an upper bound (even the supremum), and (I2) would fail at $k = 0$. Moving one unit left gives $a_0 - 1 < a_0 \in A$, which is never an upper bound.

(c) Halve an interval repeatedly by a rule that keeps its right end an upper bound of A and its left end not one; Archimedes makes the lengths tend to 0, so the right ends are Cauchy and converge. Their limit is an upper bound because the right ends are, and it is the least one because the left ends, which are not upper bounds, converge to it too.

(d) With $\mathrm{ub}(u)$ for $\forall a \in A\ a \leq u$, "complete" is the display in Definition 5.6.1. Its negation (Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2, Logic and Proof, Proposition 2.5.5 (d), and $\neg(x \leq y)$ as $y < x$ by (O1)):

$$\exists A \subseteq F\ \big(A \neq \emptyset \wedge \exists u \in F\ \mathrm{ub}(u) \wedge \forall s \in F \\ (\exists a \in A\ s < a\ \vee\ \exists u \in F\ (\mathrm{ub}(u) \wedge u < s))\big).$$

Some non-empty set bounded above defeats every candidate s, by an element above it or an upper bound below it.

5.6.6.2 B. Practice

Exercise. B1

Prove that every non-empty subset $A \subseteq \mathbb{R}$ that is bounded below has an infimum in $\mathbb{R}$

(Sets, Relations, and Cardinality, Definition 2.6.5), and that $\inf A = -\sup(-A)$, where $-A := \{-a : a \in A\}$.

Solution. Let b be a lower bound of A. The set $-A$ is non-empty, and $-b$ is an upper bound of it: for $a \in A$, $b \le a$ gives $-a \le -b$ (Proposition 2.3.5 (a)). By Theorem 5.6.3, $s := \sup(-A)$ exists.

We show that $-s$ is the infimum of A. *Lower bound:* for $a \in A$, $-a \in -A$, so $-a \le s$ and $-s \le a$. *Greatest:* let c be a lower bound of A. As for b, $-c$ is an upper bound of $-A$, so $s \le -c$ by (S2), that is, $c \le -s$. Hence $\inf A = -s = -\sup(-A)$.

Exercise. B2

Let $A := \{1 - 1/(n+1) : n \in \mathbb{N}\} \subseteq \mathbb{R}$. Prove that $\sup A = 1$, and decide whether $1 \in A$.

Solution. Upper bound. $1/(n+1) > 0$ (Proposition 2.5.2 (a)), so $1 - 1/(n+1) < 1$ for every n.

Least. Let $b < 1$. Then $1-b \in \mathbb{R}_{>0}$, and Corollary 5.4.4 (a) gives $N \in \mathbb{Z}_{>0}$ with $1/n < 1-b$ for every $n \ge N$. With $n := N$, the element $1-1/(N+1)$ of A satisfies $1/(N+1) < 1/N < 1-b$, so $1 - 1/(N+1) > b$, and b is not an upper bound. By (O1), every upper bound u satisfies $u \ge 1$.

So $\sup A = 1$, and $1 \notin A$, since every element of A is < 1. The supremum exists here without Theorem 5.6.3: we exhibited it.

5.6.6.3 C. Going deeper

Exercise. C1: Exit check: completeness from Cauchy completeness

Let F be an ordered field that is Archimedean and Cauchy-complete. Prove that every non-empty subset of F that is bounded above has a supremum in F, and name the line of your proof that uses the Archimedean property.

Solution. Let $A \neq \emptyset$ be bounded above; fix $a_0 \in A$ and an upper bound M. Put $l_0 := a_0 - 1$ (not an upper bound, since $a_0 > l_0$), $u_0 := M$ (an upper bound), $D := u_0 - l_0 > 0$.

Bisection. By the recursion theorem, with the step "replace u by $c = (l+u)/2$ if c is an upper bound, else replace l by c", define (l_k, u_k). By induction: u_k is an upper bound, l_k is not, and $u_k - l_k = D/2^k$. (Each case keeps both invariants and halves the length, as in Step 2 of the proof.) Hence l_k is non-decreasing, u_k is non-increasing, and $0 \le u_k - u_j \le D/2^k$ for $j \ge k$.

Cauchy. Let $\varepsilon \in F_{>0}$. **Archimedes** gives $n \in \mathbb{N}$ with $D/\varepsilon < n \cdot 1_F$ (this is the line). For $k \ge n$, $2^k \ge (k+1) \cdot 1_F > n \cdot 1_F > D/\varepsilon$ by Lemma 5.6.1, so $D/2^k < \varepsilon$, and $|u_j - u_k| < \varepsilon$ for $j, k \ge n$.

Limit. Cauchy-completeness gives s with $u_k \to s$; and $l_k = u_k - D/2^k \to s$, since $D/2^k \to 0$.

Supremum. For $a \in A$, $a \leq u_k$ for all k, so $a \leq s$ (Proposition 4.4.4 (b)). If an upper bound b had $b < s$, some $l_K > b$ (as $l_k \to s$, with tolerance $s - b$); l_K is not an upper bound, so some $a \in A$ exceeds l_K, hence exceeds b: impossible. So $s \leq b$ for every upper bound b, and $s = \sup A$.

Marking scheme. (10 points) - 2: a starting left end that is not an upper bound and a right end that is, each justified (in particular $a_0 - 1$, not a_0). - 2: the bisection defined by a rule, with the invariants stated and proved by induction. - 2: (u_k) Cauchy, with the Archimedean property named at the step where $D/2^k$ is made smaller than ε. - 2: the limit s is an upper bound, by the order-preservation of limits. - 2: s is the least upper bound, through $l_k \to s$ and "l_k is not an upper bound". Fatal errors (0 for the step): claiming $s \in A$; "$1/2^k \to 0$" asserted in a general F without Archimedes; midpoints or points of A "chosen" at each step with no rule. A route through monotone convergence earns the Cauchy and limit points only if monotone convergence is first proved from Cauchy-completeness and Archimedes in F. A dyadic-grid route ($u_k := m_k/2^k$, m_k the least integer making it an upper bound) maps point for point, with $u_k - 1/2^k$ for l_k and Archimedes also for the existence of m_k.

Exercise. C2: The rationals are not complete, by bisection

Run the bisection of the proof of Theorem 5.6.2 inside $F = \mathbb{Q}$, on $A := \{q \in \mathbb{Q} : q^2 < 2\}$ with $a_0 = 1$ and $M = 2$, so $l_0 = 0$ and $D = 2$.

(a) Explain why Steps 1–3 of that proof apply in $\mathbb{Q}$, so that (u_k) is Cauchy in $\mathbb{Q}$ and $D/2^k \to 0$ in $\mathbb{Q}$.

(b) Prove: if A had a supremum s in $\mathbb{Q}$, then $u_k \to s$ in $\mathbb{Q}$.

(c) Prove that (u_k) has no limit in $\mathbb{Q}$. *(Show $l_k^2 < 2 \leq u_k^2$.)*

(d) Conclude that A has no supremum in $\mathbb{Q}$. Why is "(u_k) is Cauchy with no rational limit" not enough on its own for this?

(e) In $\mathbb{R}$, show that $u_k \to r$, the real number of Example (The Class of the Newton Sequence).

Solution (Hint 1). For (b), compare s with u_k and with l_k, using (I1), (I2) and (S2). For (c), the left ends are below some element of A.

Solution (Hint 2). For $u_k^2 \geq 2$ in (c): suppose $u_k^2 < 2$, and look for a rational h with $0 < h \leq 1$ and $(u_k + h)^2 < 2$.

Solution.

(a) The proof states that Steps 1–3 use only that the field is an Archimedean ordered field, and $\mathbb{Q}$ is one (Theorem 2.5.1, Theorem 2.5.3). Lemma 5.6.1 holds in $\mathbb{Q}$ too. Step 1 applies: $1 \in A$ as $1^2 < 2$, and 2 bounds A, as $q > 2 \geq 0$ gives $q^2 > 4$ (Proposition 2.3.5 (g)). So the invariants (I1)–(I3) hold, $(D/2^k)$ is null in $\mathbb{Q}$, and (u_k) is Cauchy in $\mathbb{Q}$. Step 4 is where the proof would need Cauchy-completeness, which $\mathbb{Q}$ lacks.

(b) Suppose $s = \sup A$ in $\mathbb{Q}$. Each u_k is an upper bound (I1), so $s \leq u_k$ by (S2). Each l_k is

not an upper bound (I2), so some $a \in A$ has $l_k < a \le s$. Hence $0 \le u_k - s < u_k - l_k = D/2^k$ for every k. By the squeeze, Proposition 4.4.4 (c), with lower sequence 0 and the null upper sequence $(D/2^k)$, $u_k - s \to 0$, so $u_k \to s$ (Proposition 4.3.3).

(c) *Left ends.* $l_k \ge l_0 = 0$, since the left ends are non-decreasing (Step 2). By (I2) some $a \in A$ has $a > l_k \ge 0$, so $l_k^2 < a^2 < 2$ (Proposition 2.3.5 (g)).
Right ends. $u_k > l_k \ge 0$. Suppose $u_k^2 < 2$, and let $h := \min\{1, (2 - u_k^2)/(2(2u_k + 1))\}$, a positive rational $(2u_k + 1 > 0)$. Since $h \le 1$, $h^2 \le h$, and

$$(u_k + h)^2 = u_k^2 + 2u_k h + h^2 \le u_k^2 + (2u_k + 1)h$$
$$\le u_k^2 + \frac{2 - u_k^2}{2} < 2.$$

So $u_k + h \in A$ and $u_k + h > u_k$, against (I1). Hence $u_k^2 \ge 2$.
No limit. Suppose $u_k \to L \in \mathbb{Q}$. Then $l_k = u_k - D/2^k \to L$ (Theorem 4.4.3 (b)), and by the product law (c), $u_k^2 \to L^2$ and $l_k^2 \to L^2$. By Proposition 4.4.4 (b), $u_k^2 \ge 2$ gives $L^2 \ge 2$, and $l_k^2 < 2$ gives $L^2 \le 2$. So $L^2 = 2$, against Theorem 2.6.1.

(d) By (b) and (c), a supremum $s \in \mathbb{Q}$ would be a rational limit of (u_k), and there is none. So A has no supremum in $\mathbb{Q}$: $\mathbb{Q}$ is not complete. A Cauchy sequence with no rational limit shows only that $\mathbb{Q}$ is not **Cauchy**-complete; part (b) is the link between the two properties, and it uses the invariants of the bisection. Section 2.6 proved the same conclusion (Theorem 2.6.2 (a)) directly, by improving each rational candidate; here a sequence finds the gap.

(e) Run the bisection in $\mathbb{R}$ on A, viewed as a subset of $\mathbb{R}$. A rational c is an upper bound of A in $\mathbb{Q}$ if and only if it is one in $\mathbb{R}$ (κ preserves and reflects order, Theorem 5.4.1 (b)), so the midpoint tests give the same answers, and the sequence (l_k, u_k) is the same. By the proof of Theorem 5.6.2 in $\mathbb{R}$ (Steps 4 and 5), $u_k \to \sup A$ in $\mathbb{R}$. It remains to see $\sup A = r$. For $q \in \mathbb{Q}$: if $q \ge r$, then $q > 0$ and $q^2 \ge r^2 = 2$ (Proposition 2.3.5 (g) in $\mathbb{R}$, with $r > 0$), so $q \notin A$; hence r is an upper bound. If $t < r$ in $\mathbb{R}$, Theorem 5.4.3 gives a rational q with $\max\{t, 0\} < q < r$; then $0 < q < r$ gives $q^2 < 2$, so $q \in A$ and $q > t$: t is not an upper bound. So $\sup A = r$, and $u_k \to r$.

Marking scheme. (10 points) - 2: (a), with the Archimedean property of $\mathbb{Q}$ cited and Cauchy-completeness located in Step 4. - 3: (b), the two invariants used to get $l_k < s \le u_k$, and the squeeze. - 3: (c), both bounds $l_k^2 < 2 \le u_k^2$ justified (the second by a rational h), then the limit laws and Theorem 2.6.1. - 2: (d), the conclusion, with the difference between "not Cauchy-complete" and "not complete" stated. Part (e) is not marked. Fatal errors (0 for the step): using $\sqrt{2}$ as a rational or as a real number in (b)–(d); concluding (d) from (c) alone.

5.6.6.4 R. Review

Exercise. R1

Let F be an ordered field, (a_n) a sequence in F, and $L \in F$ with $L > 0$ and $a_n \to L$. Prove that $a_n \ge L/2$ for every sufficiently large n.

Solution. By Definition 4.3.2 with $\varepsilon := L/2 \in F_{>0}$ (Proposition 4.1.2 (b)), there is N with $|a_n - L| < L/2$ for every $n \ge N$. For such n, Proposition 4.1.3 (d) gives $-L/2 < a_n - L$, so $a_n > L - L/2 = L/2$; in particular $a_n \ge L/2$.

This is the second half of Lemma 4.4.2. It is also the shape of Theorem 4.5.4, in the case where the limit is known: a sequence with a positive limit is eventually above a fixed positive element, which is what "positive" meant for Cauchy sequences in Section 5.3.

Exercise. R2

On $X := \mathbb{Z} \times \mathbb{Z}_{>0}$, let $(p, q) \sim (p', q')$ mean $pq' = p'q$. Show that the condition "$p \geq q$" on a pair gives a well-defined property of classes: if $(p, q) \sim (p', q')$, then $p \geq q$ if and only if $p' \geq q'$. Then show that "$p \geq 2$" does **not** give a well-defined property of classes.

Solution. These are the classes of Definition 2.4.1, and the move is **representatives, then independence**: a property stated on a representative is a property of the class only if equivalent representatives agree. By Theorem 2.2.1 (c), with $S := \{(p, q) \in X : p \geq q\}$, it suffices to check one direction (the relation is symmetric).

Let $pq' = p'q$ with $q, q' > 0$, and suppose $p \geq q$. Multiplying by $q' > 0$ (Proposition 2.3.5 (c), or equality), $pq' \geq qq'$, that is, $p'q \geq q'q$. If $p' < q'$, then $p'q < q'q$ by Proposition 2.3.5 (c), since $q > 0$, a contradiction by (O1). So $p' \geq q'$. In $\mathbb{Q}$, the property says $[(p, q)] \geq 1$, by the comparison formula of Theorem 2.5.1 (a).

"$p \geq 2$" is not well defined: $(1, 2) \sim (2, 4)$, since $1 \cdot 4 = 2 \cdot 2$, but $1 \not\geq 2$ while $2 \geq 2$. One class, two verdicts: the hypothesis of Theorem 2.2.1 (c) fails. (The test "$p \geq 1$" would pass, but only because for an integer it says $p > 0$ (Theorem 2.3.4 (c)), the positivity test of Definition 2.5.1.)

Exercise. R3

Prove that $\mathbb{R}$ is uncountable. For $s \in \{0, 1\}^{\mathbb{N}}$, consider the rationals

$$S_k(s) := \sum_{i=0}^{k} \frac{2s_i}{3^{i+1}} \qquad (k \in \mathbb{N}),$$

and show that $s \mapsto [(S_k(s))_k]$ is an injection $\{0, 1\}^{\mathbb{N}} \to \mathbb{R}$.

Solution. (i) The sums. The powers 3^i are those of Definition 4.2.1 in $\mathbb{Q}$, with $3^{i+1} = 3^i \cdot 3$. The sum is defined by Sets, Relations, and Cardinality, Theorem 2.3.1 with $X = \mathbb{Q}$: $S_0(s) := 2s_0/3$ and $S_{k+1}(s) := S_k(s) + 2s_{k+1}/3^{k+2}$.

(ii) A tail bound. For $m \geq k$,

$$0 \leq S_m(s) - S_k(s) \leq \frac{1}{3^{k+1}} - \frac{1}{3^{m+1}}. \tag{5.6.3}$$

By Logic and Proof, Axiom 4.2.1 on $m - k$ (base $P(0)$, $m = k$: $0 \leq 0 \leq 0$): passing from m to $m + 1$ adds $2s_{m+1}/3^{m+2}$, which lies between 0 and $2/3^{m+2}$, and $-1/3^{m+1} + 2/3^{m+2} = -3/3^{m+2} + 2/3^{m+2} = -1/3^{m+2}$.

So $|S_m(s) - S_k(s)| < 1/3^{k+1}$ for $m \geq k$. Also $3^{k+1} \geq 2^{k+1} \geq (k+2)\cdot 1$: the first by induction ($3^0 = 2^0$, and $3^{i+1} = 3^i \cdot 3 \geq 2^i \cdot 2$ by Proposition 4.1.2 (d), with $2^i > 0$), the second by Lemma 5.6.1 with $F = \mathbb{Q}$. So $1/3^{k+1} < 1/(k+1)$. Given $\varepsilon \in \mathbb{Q}_{>0}$, Corollary 2.5.4 gives $N \in \mathbb{Z}_{>0}$ with $1/n < \varepsilon$ for $n \geq N$; for $m, k \geq N$, say $m \geq k$, $|S_m(s) - S_k(s)| < 1/(k+1) < \varepsilon$. So $(S_k(s))_k \in \mathcal{C}$.

(iii) Different sequences give different classes. Let $s \neq t$. The set $\{i \in \mathbb{N} : s_i \neq t_i\}$ is non-empty and bounded below by 0, so it has a least element j (Proposition 2.3.6). Say $s_j = 1$ and $t_j = 0$. Since $s_i = t_i$ for $i < j$, induction gives $S_{j-1}(s) = S_{j-1}(t)$ when $j \geq 1$, so $S_j(s) - S_j(t) = 2/3^{j+1}$ (also when $j = 0$). For $k \geq j$, by Eq. 5.6.3 applied to s and to t,

$$\begin{aligned} S_k(s) - S_k(t) &= \frac{2}{3^{j+1}} + (S_k(s) - S_j(s)) - (S_k(t) - S_j(t)) \\ &\geq \frac{2}{3^{j+1}} + 0 - \frac{1}{3^{j+1}} = \frac{1}{3^{j+1}}. \end{aligned}$$

So $S(s) - S(t)$ is eventually $\geq \delta := 1/3^{j+1} \in \mathbb{Q}_{>0}$: it is positive (Definition 5.3.1), so $[S(t)] < [S(s)]$ (Definition 5.3.2, Definition 5.2.2), and in particular $[S(s)] \neq [S(t)]$. The map $\Phi\colon s \mapsto [(S_k(s))_k]$ is injective.

(iv) Uncountable. If $\mathbb{R}$ were countable, there would be an injection $g\colon \mathbb{R} \to \mathbb{N}$ (Sets, Relations, and Cardinality, Theorem 3.2.2, (i) $\Rightarrow$ (ii)). Then $g \circ \Phi\colon \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ would be injective (Sets, Relations, and Cardinality, Theorem 1.4.4 (a)), so $\{0,1\}^{\mathbb{N}}$ would be countable ((ii) $\Rightarrow$ (i)), against Sets, Relations, and Cardinality, Theorem 3.4.1.

The method is that of *Sets, Relations, and Cardinality*, Chapter 3: an injection from a set known to be uncountable. (The route through Sets, Relations, and Cardinality, Proposition 3.5.2 is not taken: its ($\Leftarrow$) direction uses choice.)

Exercise. R4

Let $x \in \mathbb{R}$. Prove that there is a **strictly increasing** sequence (q_k) of rationals with $q_k \to x$ in $\mathbb{R}$.

Solution. The tool is Sets, Relations, and Cardinality, Theorem 2.3.1 with a step given by a rule, on the set $X := \{q \in \mathbb{Q} : q < x\}$. Integer parts (Proposition 4.2.4 with $F = \mathbb{R}$, Archimedean by Theorem 5.4.2; unique) supply the rule.

Start. Let p be the integer part of x, and $q_0 := p - 1$. Then $q_0 < p \leq x$, so $q_0 \in X$.

Step. For $k \in \mathbb{N}$ and $q \in X$, let $L := \max\{q, x - 1/(k+1)\}$, a real number with $L < x$. Let n be the integer part of $2/(x-L)$, plus 1: a positive integer with $n > 2/(x-L)$, so $2/n < x - L$ (Proposition 2.5.2 (b)). Let m be the integer part of nx, and put $g(k,q) := (m-1)/n$. From $m \leq nx < m+1$, dividing by n,

$$L < x - \frac{2}{n} < \frac{m-1}{n} \leq x - \frac{1}{n} < x.$$

So $g(k,q) \in X$, and $g\colon \mathbb{N} \times X \to X$ is a function: each step is a formula, made unambiguous by the uniqueness of integer parts. By Sets, Relations, and Cardinality, Theorem 2.3.1 there is exactly one sequence with $q_{k+1} = g(k, q_k)$.

Properties. By the display, $q_{k+1} > L \geq q_k$, so the sequence is strictly increasing; and

$x - 1/(k+1) \le L < q_{k+1} < x$, so $|q_{k+1} - x| < 1/(k+1)$. Given $\varepsilon \in \mathbb{R}_{>0}$, Corollary 5.4.4 (a) gives N with $1/j < \varepsilon$ for $j \ge N$; for $k \ge N+1$, $|q_k - x| < 1/k < \varepsilon$. So $q_k \to x$.

Why the look-alikes fail. The terms of a representative of x converge to x (Theorem 5.4.5) but need not increase. The approximants $m_k/(k+1)$ of Section 5.5 are $\le x$ but not always strictly increasing (for $x = 0$ they are all 0). The rational between two reals from the proof of Theorem 5.4.3 is a midpoint of terms of representatives at thresholds fixed by ∃E: it is not given by a rule in the data (k, q), so it cannot serve as a recursion step. And "at each step take some rational between L and x" makes one pick for each k with no rule, an unflagged use of Sets, Relations, and Cardinality, Axiom 3.3.1 (compare Example (Picking Representatives Is a Use of Choice)).

Chapter 6

The Complete Ordered Field Is Unique

Chapter 5 ended with one complete ordered field, whose elements are classes of Cauchy sequences of rationals. The optional Chapter 7 builds another, whose elements are sets of rationals, and *Analysis on the Real Line*, Chapter 1 builds none: it assumes that a complete ordered field exists and calls it $\mathbb{R}$. Are these the same $\mathbb{R}$? The axioms of an ordered field alone do not decide it. The field $\mathbb{R}(x)$ of rational functions, compared at large arguments, satisfies all of them and has an element above every natural number; the field $\mathbb{Q}(\sqrt{2})$ is even Archimedean, and still 3 has no square root in it. This chapter proves that completeness is the property that pins $\mathbb{R}$ down: any two complete ordered fields are linked by an order-preserving field isomorphism, and by only one. So "the" real numbers are well defined, and no theorem about them depends on how they were built.

The tool is one idea. In an Archimedean ordered field an element is determined by **the rationals below it**, and a map that respects sums, products, 1 and the order must send each element to one with the same rationals below it. In a complete field the supremum of the copies of those rationals is such an element. So from an Archimedean field into a complete one there is exactly one order-preserving field map, and each identity it satisfies is checked by comparing rationals below the two sides, with no arithmetic of suprema. When both fields are complete, maps run both ways, and the word "exactly" makes them inverse to each other without a computation. The same uniqueness leaves a complete ordered field no symmetry: a map of it to itself that respects only sums, products and 1 is the identity, because there the positive elements are the non-zero squares. The two non-examples show what each hypothesis buys: $\mathbb{Q}(\sqrt{2})$ has a symmetry exchanging $\sqrt{2}$ and $-\sqrt{2}$, which breaks the order, and $\mathbb{R}(x)$ has one sending x to $x+1$, which keeps it. The chapter, the last of the volume's main line, ends by comparing the two ways to begin analysis, assuming the axioms or building a model, and by listing what *Analysis on the Real Line* takes from this volume.

What you need. Chapters 4 and 5 above all, and Chapter 2; *Sets, Relations, and Cardinality*, Chapters 1–3, chiefly Section 2.6; and *Logic and Proof*, Chapters 2–4. From Chapter 4: the rules of an ordered field and the absolute value (Proposition 4.1.2, Lemma 4.1.1, Definition 4.1.1, Proposition 4.1.3); powers and the one copy ι_F of $\mathbb{Q}$ in an ordered field, with its uniqueness clause (Definition 4.2.1, Proposition 4.2.1, Theorem 4.2.2); the Archimedean property, density of $\mathbb{Q}$, and the non-Archimedean field $\mathbb{Q}(t)$ (Definition 4.2.2, Proposition 4.2.3, Theorem 4.2.5, Example (A Non-Archimedean Ordered Field)); and limits (Definition 4.3.2, Proposition 4.3.1), with Cauchy sequences in the exercises of Section 6.5 (Definition 4.5.1). From Chapter 5: complete ordered fields (Definition 5.6.1), the theorem that the Cauchy model $\mathbb{R}$ is one (Theorem 5.6.3),

the map $\kappa = \iota_{\mathbb{R}}$ (Theorem 5.4.1), the class of the Newton sequence, a positive real number whose square is 2 (Example (The Class of the Newton Sequence)), and, for contrast, completeness from Archimedes and Cauchy-completeness (Theorem 5.6.2). From Chapter 2: the axioms (F1)–(F9) and (O1)–(O4) (Definition 2.3.1, Definition 2.3.2) and their rules (Proposition 2.3.2, Proposition 2.3.5, Proposition 2.5.2), $\mathbb{Q}$ Archimedean (Theorem 2.5.3), least elements (Proposition 2.3.6), and no rational square equal to 2 (Theorem 2.6.1, Theorem 2.6.2). Chapter 1 supplies one analogy (Theorem 1.7.1) and, in a review solution, the order arithmetic of ω (Proposition 1.6.3); Chapter 3 supplies even and odd integers (Corollary 3.1.3), used in Section 6.4, and one review exercise (Theorem 3.4.3, Theorem 3.5.2). From *Sets, Relations, and Cardinality*: injective and bijective maps and two-sided inverses, Sections 1.4 and 1.5 (Sets, Relations, and Cardinality, Definition 1.4.3, Sets, Relations, and Cardinality, Definition 1.4.5, Sets, Relations, and Cardinality, Theorem 1.5.4); finite sums, Section 2.3 (Sets, Relations, and Cardinality, Definition 2.3.1); equivalence relations and quotient sets, Section 2.5 (Sets, Relations, and Cardinality, Definition 2.5.1, Sets, Relations, and Cardinality, Definition 2.5.2, Sets, Relations, and Cardinality, Definition 2.5.4); upper bounds and suprema, Section 2.6 (Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5); and, for exercises of Sections 6.4 and 6.5, countable and uncountable sets, Chapter 3 (Sets, Relations, and Cardinality, Theorem 3.2.2, Sets, Relations, and Cardinality, Theorem 3.4.1). From *Logic and Proof*: the contrapositive, Chapter 2 (Logic and Proof, Proposition 2.3.2); induction, Chapter 3 (Logic and Proof, Axiom 4.2.1); and the negation of quantifiers, unbounded and bounded, Chapter 4 (Logic and Proof, Proposition 6.2.1, Logic and Proof, Proposition 6.2.3).

Time. About 20 hours on the core track (all of A, one B and one C per section, the exit checks); about 30 hours for everything.

6.0.1 Roadmap

- **What Completeness Gives at Once.** Order-preserving field maps, which reflect the order and fix the rationals; the theorem that a complete ordered field is Archimedean, by a supremum of the copy of $\mathbb{N}$ that cannot exist; the rationals below an element, $\mathbb{Q}_F(x)$, which determine x when F is Archimedean; and square roots in a complete ordered field, built as a supremum and unique.
- **Building the Map from the Rationals Below.** The rationals below a sum, and below a product of positive elements, computed from those below each term; and the main theorem in its one-sided form: from an Archimedean ordered field F into a complete one L there is exactly one order-preserving field map, which sends x to the supremum in L of the copies of the rationals below x and keeps the rationals below every element.
- **Two Complete Ordered Fields, One Isomorphism.** The uniqueness theorem, with the inverse map obtained from the uniqueness clause instead of computed; the relation $F \cong L$, every complete ordered field a copy of the Cauchy model, and the Dedekind cut of Chapter 7 read as a set of rationals below; and the automorphisms of a complete ordered field, of which there is one, the identity, even among the maps that respect only sums, products and 1.
- **Two Non-Examples.** Polynomial functions compared at large arguments, where the top term decides the sign; the ordered field $\mathbb{R}(x)$ of rational functions, which is not Archimedean and in which x is not a square; the field $\mathbb{Q}(\sqrt{2})$, Archimedean and not complete, with conjugation as a symmetry that breaks the order; the shift $x \mapsto x + 1$, a symmetry of $\mathbb{R}(x)$ that keeps it; and a table of which of $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{R}(x)$ and $\mathbb{R}$ are Archimedean, complete, have square roots of positive elements, and have no field automorphism except the identity.
- **Assume the Axioms or Build a Model.** The two ways to begin analysis and the two questions, existence and uniqueness, that assuming the axioms leaves open; what an

isomorphism carries from one complete ordered field to another, with limits as the worked case, and what it does not carry, shown by a model whose zero is the empty set; a table of what *Analysis on the Real Line*, Chapter 1 assumes and where this series supplies it; and review exercises that return to Chapter 5's real numbers, Chapter 3's $\mathbb{Z}/5\mathbb{Z}$, equivalence relations and suprema.

6.0.2 Named moves

- **Rationals separate.** To show $a = b$ in an Archimedean ordered field, show that the same rationals lie below each; otherwise a rational would lie strictly between them. It is proved in Section 6.1 and then proves every equation between elements of the complete field in the main theorem of Section 6.2: the values of the map on the rationals, sums, products and uniqueness (Steps 3, 5, 6 and 7); the exit check C1 of Section 6.2 runs it alone. In $\mathbb{R}(x)$, where Archimedes fails, the elements $1/x$ and $2/x$ have the same rationals below them (Section 6.4).
- **Construct a point as a supremum.** To produce an element you cannot name, take the supremum of a well-chosen set and refute the two wrong cases of trichotomy, one with an element of the set above the supremum, the other with a smaller upper bound. It builds square roots in Section 6.1 and the map of Section 6.2. Exercise B2 of Section 6.2 (cube roots) leaves the last step to you, and Exercise C3 of Section 6.4 (a bounded increasing sequence converges to the supremum of its terms) the last half.
- **Uniqueness gives inverses.** If the only order-preserving field map $F \to F$ is the identity, then two maps $F \to L \to F$ compose to the identity without computing anything. It proves in Section 6.3 that the map between two complete ordered fields is bijective, and the same uniqueness clause shows there that a complete ordered field has no order-preserving automorphism except the identity; once square roots recover the order, also no field automorphism.

6.1 What Completeness Gives at Once

Chapter 5 ended with one complete ordered field, the classes of Cauchy sequences of rationals (Theorem 5.6.3). The optional Chapter 7 builds another from Dedekind cuts, and *Analysis on the Real Line* simply assumes that one exists. This chapter asks whether these are the same field. Before that question can be asked we need a precise meaning for "the same": a map between ordered fields that respects everything. This section fixes that meaning, and then collects what completeness gives at once: the Archimedean property, a way to tell elements apart by the rationals below them, and square roots.

Standing setting. Throughout the chapter, F and L are ordered fields (Definition 2.3.2), so $\leq$ is a total order on each (Proposition 4.1.2 (a)) and upper bounds and suprema in the sense of *Sets, Relations, and Cardinality*, Section 2.6 apply to their subsets (Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5). "Complete" is always the property of Definition 5.6.1: every non-empty subset that is bounded above has a supremum. We do not redefine it.

Each ordered field F carries its own copy of $\mathbb{Q}$: the map $\iota_F \colon \mathbb{Q} \to F$ of Theorem 4.2.2, the only map with $\iota_F(1) = 1_F$ that preserves sums and products. It is injective, strictly increasing and reflects order, and $\iota_F(n) = n \cdot 1_F$ for $n \in \mathbb{N}$. For the constructed reals, $\iota_{\mathbb{R}}$ is the map κ of Chapter 5 (Theorem 5.4.1 (d)). When only one field is in play, a proof may write q for $\iota_F(q)$, as Chapter 4 did; statements keep ι_F visible, and so does every argument that moves between two fields.

6.1.1 Maps that respect everything

To say that two ordered fields are "the same", we want a dictionary between them: a map that turns sums into sums, products into products, 1 into 1, and smaller into smaller. An isomorphism of Peano systems in Chapter 1 (Theorem 1.7.1) was such a dictionary for 0 and the successor; here the structure is larger.

A map that respects addition, multiplication, the unit and the order.

Definition 6.1.1. Order-Preserving Field Map

Let F and L be ordered fields. A map $\chi\colon F \to L$ is an **order-preserving field map** if, **for all** $a, b \in F$,

$$\chi(a+b) = \chi(a) + \chi(b), \qquad \chi(ab) = \chi(a)\,\chi(b), \qquad \chi(1_F) = 1_L,$$

and $a < b$ implies $\chi(a) < \chi(b)$. An **order-preserving field isomorphism** is an order-preserving field map that is **bijective** (Sets, Relations, and Cardinality, Definition 1.4.5). An **order-preserving field automorphism** of F is an order-preserving field isomorphism $F \to F$.

In words: the first two equations say that computing in F and then translating gives the same answer as translating and then computing in L; the third pins down the unit; the last says that χ is **strictly** increasing.

Two consequences come free. A strictly increasing map is injective: if $a \neq b$, then $a < b$ or $b < a$ by (O1), and the images differ in the same way. And it **reflects order**: if $\chi(a) < \chi(b)$, then $a < b$. For $a = b$ would give $\chi(a) = \chi(b)$, and $b < a$ would give $\chi(b) < \chi(a)$; both contradict $\chi(a) < \chi(b)$ by (O1), so only $a < b$ is left. (This is the contrapositive of "$b \leq a$ implies $\chi(b) \leq \chi(a)$", Logic and Proof, Proposition 2.3.2, read with trichotomy.) So an order-preserving field map satisfies $a < b \iff \chi(a) < \chi(b)$.

Examples.

The identity. id_F satisfies all four clauses trivially; it is an order-preserving field automorphism of F.

The copy of the rationals. For every ordered field F, $\iota_F\colon \mathbb{Q} \to F$ is an order-preserving field map, by Theorem 4.2.2. It is injective and usually not surjective: $\iota_{\mathbb{R}} = \kappa$ misses the class r of the Newton sequence, which is not $\kappa(q)$ for any rational q (Example (The Class of the Newton Sequence)).

Doubling. On $\mathbb{Q}$, $\chi(x) \coloneqq 2x$ preserves sums and is strictly increasing, but $\chi(1) = 2 \neq 1$, and $\chi(1 \cdot 1) = 2 \neq 4 = \chi(1)\chi(1)$. It is not a field map: the clauses about multiplication and the unit fail, the additive ones survive.

↻ **Second reading** · *A degenerate case*

The zero map $\chi(a) \coloneqq 0_L$ preserves sums and products, since $0+0=0$ and $0\cdot 0 = 0$. It fails the unit clause, because $1_L \neq 0_L$ in a field, and it fails strict increase. In Definition 6.1.1 the unit clause in fact follows from the other three: additivity gives $\chi(0) = \chi(0) + \chi(0)$, so $\chi(0) = 0$; strict increase then gives $\chi(1) > \chi(0) = 0$; and $\chi(1) = \chi(1)^2$ with $\chi(1) \neq 0$ forces $\chi(1) = 1$. It is

stated anyway because Proposition 6.1.1, Theorem 6.3.4 and the remark after its proof assume sums, products and the unit but no order. Without the unit hypothesis, the zero map would be a counterexample to each of them.

Why this definition. Preserving sums, products and the unit is what makes χ a dictionary for the field operations; strict increase makes it a dictionary for the order. Section 6.4 shows a map of $\mathbb{Q}(\sqrt{2})$ to itself that respects sums, products and the unit but sends a positive element to a negative one, so the order clause does not follow from the others in general. Section 6.3 shows that for complete ordered fields it does.

Quick check. Which clauses of Definition 6.1.1 does the map $\chi\colon \mathbb{Q} \to \mathbb{Q}$, $\chi(x) \coloneqq -x$, satisfy?

Solution. Only the sum clause: $-(a+b) = (-a)+(-b)$. The unit clause fails, $\chi(1) = -1 \neq 1$, and so does the product clause, since $\chi(1 \cdot 1) = -1$ while $\chi(1)\chi(1) = (-1)(-1) = 1$; and the order is reversed, $0 < 1$ but $\chi(1) = -1 < 0 = \chi(0)$.

The first result about these maps needs no order hypothesis at all. A map that preserves sums, products and the unit cannot move a rational number.

Proposition 6.1.1. Field Maps Fix the Rationals

Let F and L be ordered fields, and let $\chi\colon F \to L$ satisfy $\chi(a+b) = \chi(a) + \chi(b)$, $\chi(ab) = \chi(a)\chi(b)$ for all $a, b \in F$, and $\chi(1_F) = 1_L$. Then

$$\chi \circ \iota_F = \iota_L, \qquad \text{that is,} \qquad \chi(\iota_F(q)) = \iota_L(q) \quad \text{for every } q \in \mathbb{Q}.$$

Proof. The composite $\chi \circ \iota_F\colon \mathbb{Q} \to L$ sends 1 to $\chi(1_F) = 1_L$, and for $x, y \in \mathbb{Q}$,

$$\chi(\iota_F(x+y)) = \chi(\iota_F(x) + \iota_F(y)) = \chi(\iota_F(x)) + \chi(\iota_F(y)),$$

by Theorem 4.2.2 for F and the hypothesis on χ; products are the same with $\cdot$ in place of $+$. So $\chi \circ \iota_F$ has the three properties that, by the uniqueness clause of Theorem 4.2.2 for L, only ι_L has. Hence $\chi \circ \iota_F = \iota_L$. ∎

This is the remark after Theorem 4.2.2, now with a label. Read with the identifications $q = \iota_F(q)$ and $q = \iota_L(q)$, it says $\chi(q) = q$: such a map **fixes the rationals**. If χ is also order-preserving, more follows. For $q \in \mathbb{Q}$ and $x \in F$, $\iota_F(q) < x$ holds if and only if $\chi(\iota_F(q)) < \chi(x)$, that is, if and only if $\iota_L(q) < \chi(x)$. The rationals below x in F are exactly the rationals below $\chi(x)$ in L. That observation drives the whole chapter.

6.1.2 A complete ordered field is Archimedean

Chapter 4 called F Archimedean if every element lies below some $n \cdot 1_F$ (Definition 4.2.2), and showed that this can fail: in $\mathbb{Q}(t)$ the element t lies above every $n \cdot 1$ (Example (A Non-Archimedean Ordered Field)). Chapter 5 assumed it, together with Cauchy-completeness, to prove completeness (Theorem 5.6.2). For the supremum form of completeness, no assumption is needed: it implies the Archimedean property. We write $\iota_F(\mathbb{N}) \coloneqq \{\iota_F(n) : n \in \mathbb{N}\} = \{n \cdot 1_F : n \in \mathbb{N}\}$.

Theorem 6.1.2. A Complete Ordered Field Is Archimedean

Let F be a complete ordered field. Then F is Archimedean (Definition 4.2.2).

Idea. If some element bounded $\iota_F(\mathbb{N})$ above, completeness would give a **least** upper bound s. Step down by 1: $s - 1$ is not an upper bound, so some $n \cdot 1_F$ is above it, and one more step, $(n+1) \cdot 1_F$, is above s. The set $\iota_F(\mathbb{N})$ is closed under adding 1, so it cannot have a least upper bound, and hence no upper bound at all.

Proof. Suppose, for ¬I, that $\iota_F(\mathbb{N})$ is bounded above, and fix an upper bound $b \in F$ [∃E], so that $n \cdot 1_F \leq b$ for every $n \in \mathbb{N}$. The set $\iota_F(\mathbb{N})$ is non-empty, since $0 \cdot 1_F \in \iota_F(\mathbb{N})$. Since F is complete, $\iota_F(\mathbb{N})$ has a supremum; fix $s \coloneqq \sup \iota_F(\mathbb{N})$ [∃E].

Since $s - 1_F < s$ and s is the least upper bound, $s - 1_F$ is not an upper bound of $\iota_F(\mathbb{N})$. So some $n \in \mathbb{N}$ has $n \cdot 1_F \not\leq s - 1_F$ (Logic and Proof, Proposition 6.2.3 (a)), that is, $n \cdot 1_F > s - 1_F$ by totality (Proposition 4.1.2 (a)). Then, by Definition 4.2.1 (a) and (O3),

$$(n+1) \cdot 1_F = n \cdot 1_F + 1_F > (s - 1_F) + 1_F = s,$$

and $(n+1) \cdot 1_F \in \iota_F(\mathbb{N})$. This contradicts that s is an upper bound of $\iota_F(\mathbb{N})$. Hence $\iota_F(\mathbb{N})$ is not bounded above [¬I].

So no $x \in F$ is an upper bound of $\iota_F(\mathbb{N})$ (Logic and Proof, Proposition 6.2.1). For $x \in F$, as above, some $n \in \mathbb{N}$ has $n \cdot 1_F \not\leq x$ (Logic and Proof, Proposition 6.2.3 (a)), that is, $x < n \cdot 1_F$ by totality. Hence F is Archimedean (Definition 4.2.2).

■

Both hypotheses of completeness were used: "non-empty" through $0 \cdot 1_F$, "bounded above" through b. The converse fails: $\mathbb{Q}$ is Archimedean (Theorem 2.5.3) and not complete (Theorem 2.6.2). And dropping completeness can lose the conclusion: $\mathbb{Q}(t)$ is an ordered field that is not Archimedean.

The proof never claimed that s is an element of $\iota_F(\mathbb{N})$; a supremum need not belong to its set. Exercise C2 shows how tempting that claim is.

Warning. Completeness here is about **subsets**, not sequences. Chapter 5 proved that the constructed $\mathbb{R}$ is complete by going through Cauchy sequences, and Theorem 5.6.2 needed the Archimedean property as a separate hypothesis to get from "every Cauchy sequence converges" to "every bounded non-empty set has a supremum". In this chapter "complete"

is the supremum property of Definition 5.6.1, and the theorem above shows that it brings the Archimedean property with it. Do not replace it by Cauchy-completeness in the proofs below: Chapter 5 records, without proof, that a Cauchy-complete ordered field need not be Archimedean.

Every complete ordered field therefore has the consequences of the Archimedean property that Chapter 4 proved. The one we use most is density (Theorem 4.2.5): in an Archimedean ordered field F, for all $x < y$ there is $q \in \mathbb{Q}$ with $x < \iota_F(q) < y$. We cite it, not reprove it; Section 6.2 uses it in two fields at once.

Here is the Archimedean property at work on a specific supremum. The pattern has two halves: check the candidate is an upper bound, then show that anything smaller is not one.

Example. A Supremum Computed by Archimedes

Let F be a complete ordered field. Show that $\sup\{\iota_F(1 - 1/n) : n \in \mathbb{Z}_{>0}\} = 1_F$.

Solution. Let $A := \{\iota_F(1 - 1/n) : n \in \mathbb{Z}_{>0}\}$.

Upper bound. For $n \in \mathbb{Z}_{>0}$, $1/n > 0$ in $\mathbb{Q}$, so $1 - 1/n < 1$, and $\iota_F(1 - 1/n) < \iota_F(1) = 1_F$ because ι_F is strictly increasing (Theorem 4.2.2). So 1_F is an upper bound of A.

Least. Let b be an upper bound of A, and suppose, for contradiction, that $b < 1_F$, so $1_F - b > 0$. By Theorem 6.1.2, F is Archimedean, so Proposition 4.2.3 (with $\varepsilon := 1_F - b$, applied at $n = N$) gives $N \in \mathbb{Z}_{>0}$ with $1/(N \cdot 1_F) < 1_F - b$. Since ι_F preserves products and $\iota_F(N) = N \cdot 1_F$, we have $\iota_F(1/N) \cdot (N \cdot 1_F) = \iota_F(1) = 1_F$, so $\iota_F(1/N) = 1/(N \cdot 1_F)$ by uniqueness of inverses (Lemma 4.1.1 (a)). Since ι_F preserves sums, it preserves negatives, and adding $b - 1/(N \cdot 1_F)$ to both sides of $1/(N \cdot 1_F) < 1_F - b$ (O3) gives

$$\iota_F(1 - 1/N) = 1_F - \frac{1}{N \cdot 1_F} > b.$$

This element of A lies above b, against b being an upper bound. So $1_F \leq b$ by totality, for every upper bound b, and $1_F = \sup A$ (Sets, Relations, and Cardinality, Definition 2.6.5).

The supremum 1_F is not an element of A: every element is strictly below it. Completeness entered only through the Archimedean property; in an ordered field that is Archimedean but not complete, such as $\mathbb{Q}$ (Theorem 2.5.3), the same two halves give the same answer. Exercises B1 and B2 follow this pattern.

6.1.3 The rationals below an element

How can we tell two elements of a field apart? In an Archimedean field, by a rational between them. That suggests recording, for each element, which rationals lie below it.

An element is measured by the rationals below it.

Definition 6.1.2. The Rationals below an Element

Let F be an ordered field and $x \in F$. The **rationals below** x form the set

$$\mathbb{Q}_F(x) := \{q \in \mathbb{Q} : \iota_F(q) < x\} \subseteq \mathbb{Q}.$$

In words: $\mathbb{Q}_F(x)$ is a set of **rational numbers**, not of elements of F; a rational q belongs to it when its copy $\iota_F(q)$ lies **strictly** below x.

Two properties hold in every ordered field and are used without comment. The set is **closed downward**: if $q \in \mathbb{Q}_F(x)$ and $p < q$, then $\iota_F(p) < \iota_F(q) < x$, so $p \in \mathbb{Q}_F(x)$. And if $x \le y$, then $\mathbb{Q}_F(x) \subseteq \mathbb{Q}_F(y)$, by transitivity.

Examples.

A rational element. For $r \in \mathbb{Q}$, $\mathbb{Q}_F(\iota_F(r)) = \{q \in \mathbb{Q} : q < r\}$. Indeed $\iota_F(q) < \iota_F(r)$ holds if and only if $q < r$, because ι_F is strictly increasing and reflects order. In particular $r \notin \mathbb{Q}_F(\iota_F(r))$: the inequality is strict.

Example. The Rationals below the Square Root of Two

Let $r \in \mathbb{R}$ be the class of the Newton sequence, so that $r > 0$ and $r^2 = 2$ (Example (The Class of the Newton Sequence)). Show that

$$\mathbb{Q}_{\mathbb{R}}(r) = \{q \in \mathbb{Q} : q < 0\} \cup \{q \in \mathbb{Q} : q \ge 0 \text{ and } q^2 < 2\}.$$

Solution. Let $q \in \mathbb{Q}$, and recall $\iota_{\mathbb{R}} = \kappa$ (Theorem 5.4.1 (d)). If $q < 0$, then $\kappa(q) < \kappa(0) = 0 < r$, so $q \in \mathbb{Q}_{\mathbb{R}}(r)$. Let $q \ge 0$, so $\kappa(q) \ge 0$. Since also $r > 0$, Proposition 2.3.5 (g) in $\mathbb{R}$ gives

$$\kappa(q) < r \iff \kappa(q)^2 < r^2 = 2.$$

Now $\kappa(q)^2 = \kappa(q^2)$ and $2 = \kappa(2)$ (Theorem 5.4.1 (a)), and $\kappa(q^2) < \kappa(2)$ if and only if $q^2 < 2$ (Theorem 5.4.1 (b)). So for $q \ge 0$, $q \in \mathbb{Q}_{\mathbb{R}}(r)$ if and only if $q^2 < 2$. Together the two cases give the displayed set.

The set of Section 2.6, $\{x \in \mathbb{Q} : x^2 < 2\}$ (Theorem 2.6.2), is **not** $\mathbb{Q}_{\mathbb{R}}(r)$: it omits -2, whose square is 4. That set is not closed downward, and a rationals-below set always is.

Quick check. Let F be an ordered field and $x := \iota_F(3/2)$. Describe $\mathbb{Q}_F(x)$, and say whether $3/2 \in \mathbb{Q}_F(x)$.

Solution. By the first example, $\mathbb{Q}_F(x) = \{q \in \mathbb{Q} : q < 3/2\}$. It does not contain $3/2$, since $\iota_F(3/2) < \iota_F(3/2)$ is false by (O1).

The point of the definition is that in an Archimedean field, these sets lose no information.

Lemma 6.1.3. Rationals Separate

Let F be an Archimedean ordered field, and let $x, y \in F$. Then

$$x \leq y \iff \mathbb{Q}_F(x) \subseteq \mathbb{Q}_F(y).$$

Consequently $x = y \iff \mathbb{Q}_F(x) = \mathbb{Q}_F(y)$.

Proof. ($\Rightarrow$) This is the second property after Definition 6.1.2: if $\iota_F(q) < x$ and $x \leq y$, then $\iota_F(q) < y$.

($\Leftarrow$) We prove the contrapositive (Logic and Proof, Proposition 2.3.2). Suppose that $x \leq y$ fails. Then $y < x$ by totality (Proposition 4.1.2 (a)). By Theorem 4.2.5 there is $q \in \mathbb{Q}$ with $y < \iota_F(q) < x$. Then $q \in \mathbb{Q}_F(x)$, and $q \notin \mathbb{Q}_F(y)$, since $\iota_F(q) < y$ and $y < \iota_F(q)$ cannot both hold (O1). So $\mathbb{Q}_F(x) \not\subseteq \mathbb{Q}_F(y)$.

For the consequence: $x = y$ holds if and only if $x \leq y$ and $y \leq x$, by antisymmetry of the total order $\leq$; by the equivalence just proved, that is $\mathbb{Q}_F(x) \subseteq \mathbb{Q}_F(y)$ and $\mathbb{Q}_F(y) \subseteq \mathbb{Q}_F(x)$, that is, $\mathbb{Q}_F(x) = \mathbb{Q}_F(y)$.

■

This is the first named move of the chapter, **rationals separate**: to prove two elements of an Archimedean field equal, show that the same rationals lie below each. A rational strictly between them would otherwise exist. Section 6.2 uses it in four of the seven steps of its main proof (Steps 3, 5, 6 and 7), each time to prove that two elements of the complete field are equal.

Second reading · *Without Archimedes the sets cannot tell elements apart*

The direction ($\Rightarrow$) and the remark that equal elements have equal sets hold in every ordered field. The direction ($\Leftarrow$) needs density, and density needs Archimedes. In the non-Archimedean field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field) (school polynomial algebra, as flagged there; nothing below depends on it), every rational q satisfies $q < n$ for some $n \in \mathbb{N}$ (Theorem 2.5.3), and $n \cdot 1 < t$, so $\iota_{\mathbb{Q}(t)}(q) < t < t + 1$. Hence $\mathbb{Q}_{\mathbb{Q}(t)}(t) = \mathbb{Q} = \mathbb{Q}_{\mathbb{Q}(t)}(t + 1)$, although $t \neq t+1$. Both elements lie above all of the rationals, and the rationals cannot see the difference.

6.1.4 Square roots

Completeness lets us build elements we cannot name in advance, as the supremum of a well-chosen set. We show the method once, in full, on square roots; it is the second named move, **construct a point as a supremum**. Exercise B2 of Section 6.2 runs it again on cube roots.

Proposition 6.1.4. Square Roots in a Complete Ordered Field

Let F be a complete ordered field and $a \in F$ with $a \geq 0$. Then there is **exactly one** $b \in F$ with $b \geq 0$ and $b^2 = a$.

Idea. The square root of a should be the top of the set S of non-negative y with $y^2 < a$. Let $s \coloneqq \sup S$. Two of the three cases of trichotomy must be ruled out.

If $s^2 < a$, there is room above s: we look for $h > 0$ with $(s+h)^2 < a$, which puts $s+h$ in S above its supremum. Working backwards, for $0 < h \leq 1$ we have $h^2 \leq h$, so

$$(s+h)^2 = s^2 + 2sh + h^2 \leq s^2 + h(2s+1),$$

and it is enough that $h(2s+1)$ be at most half the room $a - s^2$. So $h \coloneqq \min\{1, (a - s^2)/(2(2s+1))\}$.

If $s^2 > a$, then s overshoots: we look for $h > 0$ with $(s-h)^2 \geq a$ still, which makes $s - h$ a smaller upper bound. Since $(s-h)^2 \geq s^2 - 2sh$, it is enough that $2sh = s^2 - a$, so $h \coloneqq (s^2 - a)/(2s)$.

Proof. Existence. If $a = 0$, then $b \coloneqq 0$ works. Let $a > 0$, and let

$$S \coloneqq \{y \in F : y \geq 0 \text{ and } y^2 < a\}.$$

Then $0 \in S$, since $0^2 = 0 < a$. And $a+1$ is an upper bound of S: if $y > a+1$, then $y > 1 > 0$, so $y^2 = y \cdot y > y \cdot 1 = y > a$ (Proposition 2.3.5 (c)), and $y \notin S$; so every $y \in S$ satisfies $y \leq a+1$ by totality. Since F is complete, fix $s \coloneqq \sup S$ [∃E]. Since $0 \in S$, $s \geq 0$.

Case $s^2 < a$. Let $h \coloneqq \min\{1, (a - s^2)/(2(2s+1))\}$. It is defined and positive, since $2s + 1 \geq 1 > 0$ and $a - s^2 > 0$ (Proposition 2.5.2 (a), (O4)). Since $0 < h \leq 1$, $h^2 \leq h$ (Proposition 2.3.5 (c)), so

$$\begin{aligned}(s+h)^2 = s^2 + 2sh + h^2 &\leq s^2 + h(2s+1) \\ &\leq s^2 + \frac{a - s^2}{2} < a,\end{aligned}$$

where the second inequality uses $h \leq (a - s^2)/(2(2s+1))$ and $2s + 1 > 0$, and the last uses $(a - s^2)/2 < a - s^2$ (Proposition 4.1.2 (b)). So $s + h \geq 0$ lies in S, and $s + h > s$, against s being an upper bound of S.

Case $s^2 > a$. Then $s \neq 0$, since $0^2 = 0 < a$; so $s > 0$. Let $h \coloneqq (s^2 - a)/(2s)$. It is positive, since $s^2 - a > 0$ and $2s > 0$ (Proposition 2.5.2 (a), (O4)). Then $s - h = (s^2 + a)/(2s) > 0$ by the same reasons, since $s^2 + a > 0$; and

$$(s-h)^2 = s^2 - 2sh + h^2 = s^2 - (s^2 - a) + h^2 = a + h^2 > a.$$

We claim that $s - h$ is an upper bound of S. Let $y > s - h$. Then $y > 0$, and Proposition 2.3.5 (g) gives $y^2 > (s-h)^2 > a$, so $y \notin S$. Hence every element of S is $\leq s - h$. But $s - h < s$, against s being the **least** upper bound.

Both cases are impossible, so $s^2 = a$ by trichotomy (O1), and $b \coloneqq s \geq 0$ works.

Uniqueness. Let $b, c \geq 0$ with $b^2 = c^2 = a$. Then $(b-c)(b+c) = b^2 - c^2 = 0$, so $b = c$ or $b + c = 0$ (Proposition 2.3.5 (f)). If $b + c = 0$, then $b = -c \leq 0 \leq b$, so $b = 0$ and $c = -b = 0$. In either case $b = c$.

■

For $a \geq 0$ in a complete ordered field we write $\sqrt{a}$ for this b. In the constructed reals, the class r of Example (The Class of the Newton Sequence) has $r > 0$ and $r^2 = 2$, so $r = \sqrt{2}$ by the uniqueness clause, and Example (The Rationals below the Square Root of Two) describes $\mathbb{Q}_{\mathbb{R}}(\sqrt{2})$.

The constants follow the Idea: each h depends on s and a only, and each was found backwards and then placed first. Completeness was used once, to produce s; the rest is arithmetic in an ordered field. Without completeness the statement fails: in $\mathbb{Q}$, the element $2 \geq 0$ has no square root (Theorem 2.6.1). Section 6.4 meets a field that is Archimedean and still lacks some square roots.

6.1.5 Exercises

6.1.5.1 A. Check your understanding

Exercise. A1

(a) True or false, with a reason: in **every** ordered field F, if $\mathbb{Q}_F(x) \neq \mathbb{Q}_F(y)$ then $x \neq y$. Is the converse true in every ordered field?

(b) Write "F is Archimedean" in symbols, and state its negation with the quantifiers moved to the front (Logic and Proof, Proposition 6.2.1).

(c) The eight lines (i)–(viii) below are about a complete ordered field F; six of them, in a suitable order, prove Theorem 6.1.2. Write an admissible order, and say why each of the other two lines is wrong.

(i) Then $\iota_F(n+1) = \iota_F(n) + 1 > s$, and $n+1 \in \mathbb{N}$.
(ii) For every $n \in \mathbb{N}$, $\iota_F(n) > s - 1$.
(iii) The set $\iota_F(\mathbb{N})$ is non-empty ($0 \in \mathbb{N}$), so $s := \sup \iota_F(\mathbb{N})$ exists.
(iv) This contradicts that s is an upper bound of $\iota_F(\mathbb{N})$.
(v) Suppose that $\iota_F(\mathbb{N})$ is bounded above.
(vi) So $\iota_F(n) > s - 1$ for some $n \in \mathbb{N}$.
(vii) So $s \in \iota_F(\mathbb{N})$, say $s = \iota_F(m)$.
(viii) Since $s - 1 < s$, the element $s - 1$ is not an upper bound of $\iota_F(\mathbb{N})$.

Solution.

(a) True. It is the contrapositive (Logic and Proof, Proposition 2.3.2) of "if $x = y$ then $\mathbb{Q}_F(x) = \mathbb{Q}_F(y)$", which holds because equal elements have the same rationals below them. The converse, "if $x \neq y$ then $\mathbb{Q}_F(x) \neq \mathbb{Q}_F(y)$", is the direction ($\Leftarrow$) of Lemma 6.1.3, and it needs the Archimedean property: in $\mathbb{Q}(t)$, $t \neq t+1$ and both have every rational below them (the second-reading note after Lemma 6.1.3).

(b) $\forall x \in F\ \exists n \in \mathbb{N}\ \ x < n \cdot 1_F$. Negation: $\exists x \in F\ \forall n \in \mathbb{N}\ \ n \cdot 1_F \leq x$, using $\neg(x < y) \iff y \leq x$ (totality and (O1)). In words: some element lies above or at every $n \cdot 1_F$.

(c) (v), (iii), (viii), (vi), (i), (iv). No other order is admissible: each line uses the one before it ((iii) needs the bound of (v) for the supremum to exist, (viii) needs s, (vi) unpacks (viii), (i) uses the n of (vi), and (iv) is the contradiction with (i)). The contradiction in (iv) refutes (v), so $\iota_F(\mathbb{N})$ is not bounded above; unpacked with the negation laws, as at the end of the proof of Theorem 6.1.2, that is the statement of (b).

Line (vii) is wrong: a supremum need not belong to the set (in every ordered field, the set of negative elements has supremum 0, which is not negative: the first example after Definition 5.6.1). Line (ii) has the wrong quantifier: "$s-1$ is not an upper bound" gives **some** n with $\iota_F(n) > s-1$, by Logic and Proof, Proposition 6.2.3, not every n (and for $n=0$ it may be false).

6.1.5.2 B. Practice

Exercise. B1

Let F be a complete ordered field. Prove that $\inf\{\iota_F(1/n) : n \in \mathbb{Z}_{>0}\} = 0$.

Solution. Let $T := \{\iota_F(1/n) : n \in \mathbb{Z}_{>0}\}$. For $n \in \mathbb{Z}_{>0}$, $1/n > 0$ in $\mathbb{Q}$, so $\iota_F(1/n) > 0$: 0 is a lower bound of T.

Let c be a lower bound of T, and suppose, for contradiction, that $c > 0$. By Theorem 6.1.2, F is Archimedean, so Proposition 4.2.3 (with $\varepsilon := c$, applied at $n = N$) gives $N \in \mathbb{Z}_{>0}$ with $1/(N \cdot 1_F) < c$. Since ι_F preserves products and $\iota_F(N) = N \cdot 1_F$, we have $\iota_F(1/N) \cdot (N \cdot 1_F) = \iota_F(1) = 1_F$, so $\iota_F(1/N) = 1/(N \cdot 1_F)$ by uniqueness of inverses (Lemma 4.1.1 (a)). So the element $\iota_F(1/N)$ of T lies below c, against c being a lower bound. Hence every lower bound c satisfies $c \leq 0$, and 0 is the greatest lower bound (Sets, Relations, and Cardinality, Definition 2.6.5).

Exercise. B2

Let F be an Archimedean ordered field and $x \in F$. Prove that the set $\iota_F(\mathbb{Q}_F(x)) = \{\iota_F(q) : q \in \mathbb{Q}_F(x)\}$ has supremum x in F.

Solution. Upper bound. Every element $\iota_F(q)$ of the set satisfies $\iota_F(q) < x$ by the definition of $\mathbb{Q}_F(x)$.

Least. Let u be an upper bound, and suppose $u < x$. By Theorem 4.2.5 there is $q \in \mathbb{Q}$ with $u < \iota_F(q) < x$. Then $q \in \mathbb{Q}_F(x)$, so $\iota_F(q)$ is in the set and exceeds u, against u being an upper bound. So $x \leq u$ by totality, for every upper bound u. Hence $x = \sup \iota_F(\mathbb{Q}_F(x))$.

No completeness was used: the supremum exists because we exhibited it.

6.1.5.3 C. Going deeper

Exercise. C1: An Archimedean field is no larger than the subsets of the rationals

Let F be an Archimedean ordered field. Prove that the map $F \to \mathcal{P}(\mathbb{Q})$, $x \mapsto \mathbb{Q}_F(x)$, is injective. What does this say about the size of F?

Solution. Let $x, y \in F$ with $\mathbb{Q}_F(x) = \mathbb{Q}_F(y)$. By Lemma 6.1.3, $x = y$. So the map is injective (Sets, Relations, and Cardinality, Definition 1.4.3).

An injection $F \to \mathcal{P}(\mathbb{Q})$ says that F has at most as many elements as $\mathcal{P}(\mathbb{Q})$, in the sense of *Sets, Relations, and Cardinality*, Chapter 3. In particular every Archimedean ordered field, and so (by Theorem 6.1.2) every complete ordered field, is no larger than the set of all sets of rationals. The non-Archimedean field $\mathbb{Q}(t)$ shows that the hypothesis is needed for injectivity: t and $t+1$ have the same image.

Exercise. C2: Find the error

Here is a "proof" that a complete ordered field F is Archimedean.

> Suppose that $\iota_F(\mathbb{N})$ is bounded above, and let s be its supremum. Since s is the least upper bound of $\iota_F(\mathbb{N})$, it is the largest element of $\iota_F(\mathbb{N})$, so $s = \iota_F(m)$ for some $m \in \mathbb{N}$. But $\iota_F(m+1) > s$, a contradiction.

Is the conclusion true? Find the line that fails, give a counterexample to that line, and repair the proof.

Solution. The conclusion is true: it is Theorem 6.1.2.

The illegal step is "the least upper bound of a set is its largest element". A supremum is an upper bound that lies below every other upper bound (Sets, Relations, and Cardinality, Definition 2.6.5); nothing says it belongs to the set. Counterexample to the step: in the constructed $\mathbb{R}$, $A := \{1 - 1/n : n \in \mathbb{Z}_{>0}\}$ has supremum 1, and $1 \notin A$. Indeed every element is < 1; and if $b < 1$, Archimedes in $\mathbb{R}$ (Theorem 5.4.2, with Proposition 4.2.3) gives $n \in \mathbb{Z}_{>0}$ with $1/n < 1 - b$, so $1 - 1/n > b$ and b is not an upper bound. So the step fails for this set, though the theorem's conclusion is not affected.

Repair: replace the second and third sentences by the argument of the proof of Theorem 6.1.2. Since $s - 1 < s$, $s - 1$ is not an upper bound, so $\iota_F(n) > s - 1$ for some n; then $\iota_F(n+1) = \iota_F(n) + 1 > s$, against s being an upper bound. The repair uses that s is **least** only through "$s-1$ is not an upper bound", which hands us an element of the set just below s, not s itself.

Marking scheme. (10 points) - 3: says the conclusion is true, with the reason (the theorem). - 3: names the line "the supremum is the largest element" and says which property of suprema it wrongly assumes. - 2: a counterexample **to that step**: a set whose supremum is not an element of it, with the supremum verified. - 2: a correct repair that uses that s is **least**, without assuming s lies in the set: for example, $s - 1$ is not an upper bound, so some $\iota_F(n) > s - 1$ and then $\iota_F(n+1) > s$; or, equivalently, if $\iota_F(n+1) \leq s$ for all n then $s - 1$ is a smaller upper bound (the case $n = 0$ uses $1 \leq s$). Fatal errors (0 for the step): declaring the conclusion false; a counterexample to the theorem rather than to the step.

6.2 Building the Map from the Rationals Below

Section 6.1 gave every element x of an ordered field F a record, the set $\mathbb{Q}_F(x)$ of rationals below it, and showed that in an Archimedean field the record determines the element (Lemma 6.1.3). This section turns the record into a map. Given an Archimedean field F and a complete field L, we send $x \in F$ to the element of L with the same record. The result is the main theorem of the

chapter in its one-sided form: there is exactly one order-preserving field map $F \to L$. Section 6.3 applies it twice, in both directions, to compare two complete ordered fields.

Two fields are in play throughout, so every proof here keeps ι_F and ι_L visible. Only the two lemmas, each about one field, write q for $\iota_F(q)$.

6.2.1 What the map must be

Suppose $\varphi\colon F \to L$ is an order-preserving field map (Definition 6.1.1). By Proposition 6.1.1 it fixes the rationals, $\varphi(\iota_F(q)) = \iota_L(q)$, and it preserves and reflects order. So for $q \in \mathbb{Q}$ and $x \in F$,

$$\iota_F(q) < x \iff \varphi(\iota_F(q)) < \varphi(x) \iff \iota_L(q) < \varphi(x).$$

In the notation of Section 6.1: $\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x)$. *The image of x has exactly the rationals below it that x has.*

That already shows there can be at most one such map when L is Archimedean: two candidates for $\varphi(x)$ would have the same rationals below them, and rationals separate. For existence, we need an element of L with a prescribed set of rationals below it. When L is complete there is a natural candidate: the least element of L that lies above all the copies $\iota_L(q)$, $q \in \mathbb{Q}_F(x)$, that is, their supremum.

The check that this candidate respects sums and products goes through the record, not through suprema. For that we need to know how the rationals below $a+b$ and below ab are made from those below a and b.

6.2.2 Sums and products, measured by rationals

A rational q lies below $a+b$ exactly when it splits into a rational below a and a rational below b.

Lemma 6.2.1. Rationals below a Sum

Let F be an Archimedean ordered field, $a, b \in F$ and $q \in \mathbb{Q}$. Then $\iota_F(q) < a+b$ if and only if there is $r \in \mathbb{Q}$ with

$$\iota_F(r) < a \qquad \text{and} \qquad \iota_F(q-r) < b.$$

In other words, $q \in \mathbb{Q}_F(a+b) \iff \exists r \in \mathbb{Q}\ \big(r \in \mathbb{Q}_F(a) \wedge q - r \in \mathbb{Q}_F(b)\big)$.

Idea. One direction is addition of inequalities. For the other, split q as $r + (q-r)$ with r a rational just below a: we need r in the gap between $q-b$ and a, which is non-empty exactly because $q < a+b$, and density supplies it.

Proof. We write q for $\iota_F(q)$; since ι_F preserves sums, it preserves differences, so $\iota_F(q-r) = q - r$ in F.

($\Leftarrow$) If $r < a$ and $q - r < b$, then $q = r + (q-r) < a + b$ by Proposition 2.3.5 (b).

($\Rightarrow$) Let $q < a+b$. Then $q - b < a$ (Proposition 2.3.5 (a)). By Theorem 4.2.5 there is $r \in \mathbb{Q}$ with $q - b < r < a$. Then $r < a$, and $q - r < b$ by Proposition 2.3.5 (a) again.

■

For products the same idea works for positive elements and positive rationals, with q/r in place of $q - r$.

Lemma 6.2.2. Rationals below a Product

Let F be an Archimedean ordered field, $a, b \in F$ with $a > 0$ and $b > 0$, and $q \in \mathbb{Q}$ with $q > 0$. Then $\iota_F(q) < ab$ if and only if there is $r \in \mathbb{Q}$ with

$$r > 0, \qquad \iota_F(r) < a \qquad \text{and} \qquad \iota_F(q/r) < b.$$

For $q \leq 0$, $\iota_F(q) < ab$ holds always.

In words: below a product of positives, a positive rational splits as a product of a rational below each factor. Exercise B1 asks for the proof; the fold below has it.

Second reading · *The proof of the product lemma (try Exercise B1 first)*

Idea. As for sums, with division in place of subtraction: split q as $r \cdot (q/r)$ with r a rational in the gap between q/b and a, a gap of positive elements.

Proof. We write q for $\iota_F(q)$; since ι_F preserves products and $\iota_F(r) \neq 0$ for $r \neq 0$, it preserves quotients, so $\iota_F(q/r) = q/r$ in F.

If $q \leq 0$, then $q \leq 0 < ab$ by (O4). Let $q > 0$.

($\Leftarrow$) Let $r > 0$ with $r < a$ and $q/r < b$. Since $q/r > 0$ (Proposition 2.5.2 (a), (O4)), Proposition 2.3.5 (c) gives $q = r \cdot (q/r) < a \cdot (q/r)$; since $a > 0$, it gives $a \cdot (q/r) < ab$. So $q < ab$.

($\Rightarrow$) Let $q < ab$. Multiplying by $b^{-1} > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)) gives $q/b < a$, and $q/b > 0$ by (O4). By Theorem 4.2.5 there is $r \in \mathbb{Q}$ with $q/b < r < a$. Then $r > q/b > 0$, so $r > 0$ in $\mathbb{Q}$, since ι_F reflects order. Multiplying $q/b < r$ by $b > 0$ gives $q < rb$, and multiplying by $r^{-1} > 0$ gives $q/r < b$.

■

Quick check. In $F = \mathbb{Q}$, with $a := 3/4$, $b := 1/2$ and $q := 1$, find a rational r as in Lemma 6.2.1. Which rationals r work?

Solution. We need $r < 3/4$ and $1 - r < 1/2$, that is, $1/2 < r < 3/4$. For example $r = 2/3$: then $1 - 2/3 = 1/3 < 1/2$. Exactly the rationals of the open gap between $q - b = 1/2$ and

$a = 3/4$ work, as the proof predicts.

6.2.3 The map into a complete field

Here is the theorem. Its proof is the first key proof of the chapter.

Theorem 6.2.3. Embedding an Archimedean Field in a Complete One

Let F be an Archimedean ordered field and L a complete ordered field. Then there is **exactly one** order-preserving field map $\varphi\colon F \to L$. It is given by

$$\varphi(x) \coloneqq \sup_L \{\iota_L(q) : q \in \mathbb{Q}_F(x)\} \qquad (x \in F),$$

the supremum taken in L, and it satisfies

$$\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x) \qquad \text{for every } x \in F. \tag{$\star$}$$

Try it first (ten minutes): if φ exists and fixes $\mathbb{Q}$, which element of L must $\varphi(x)$ be?

Idea. The natural first attempt defines φ on the copy of $\mathbb{Q}$, by $\varphi(\iota_F(q)) \coloneqq \iota_L(q)$, and then tries to "extend". It stalls at once: an element $x \in F$ that is not a rational has no formula, and nothing in F names it. What we do know, from the previous subsection, is its record: $\varphi(x)$ must have exactly the rationals $\mathbb{Q}_F(x)$ below it. So we **construct a point as a supremum**: $\varphi(x)$ is the least upper bound in L of the copies of those rationals, and $(\star)$ says the construction has the right record. After that, sums, products and uniqueness are checked the same way: compute the record of both sides (for sums and products by the two lemmas) and let rationals separate in L. Monotonicity needs only one rational between x and y.

Step roadmap: ① $\varphi(x)$ exists; ② $(\star)$; ③ φ fixes the rationals; ④ φ is strictly increasing; ⑤ sums; ⑥ products and the unit; ⑦ uniqueness. *Before reading on: where is completeness of L used, and where the Archimedean property of F?*

Proof. By Theorem 6.1.2, L is Archimedean. So Theorem 4.2.5, Lemma 6.1.3, Lemma 6.2.1 and Lemma 6.2.2 apply in L as well as in F. For $x \in F$ write $A_x \coloneqq \{\iota_L(q) : q \in \mathbb{Q}_F(x)\} \subseteq L$.

Step 1: $\varphi(x)$ is defined. Let $x \in F$. Since F is Archimedean, there is $n \in \mathbb{N}$ with $-x < n \cdot 1_F$, so $\iota_F(-n) = -(n \cdot 1_F) < x$ (Proposition 2.3.5 (a)), and $-n \in \mathbb{Q}_F(x)$: the set A_x is non-empty. Applied to x, the Archimedean property gives $m \in \mathbb{N}$ with $x < m \cdot 1_F = \iota_F(m)$. For $q \in \mathbb{Q}_F(x)$, $\iota_F(q) < x < \iota_F(m)$, so $q < m$ because ι_F reflects order, and $\iota_L(q) < \iota_L(m)$. So $\iota_L(m)$ is an upper bound of A_x. Since L is complete, A_x has a supremum in L, and it is unique; $\varphi(x)$ is that supremum. So $\varphi\colon F \to L$ is a function.

Step 2: $(\star)$. Let $x \in F$ and $q \in \mathbb{Q}$.

$(\subseteq)$ Let $q \in \mathbb{Q}_F(x)$. By density in F there is $r \in \mathbb{Q}$ with $\iota_F(q) < \iota_F(r) < x$. Then $r \in \mathbb{Q}_F(x)$, so $\iota_L(r) \in A_x$ and $\iota_L(r) \le \varphi(x)$, since $\varphi(x)$ is an upper bound. Also $q < r$ (reflection in F), so $\iota_L(q) < \iota_L(r) \le \varphi(x)$, and $q \in \mathbb{Q}_L(\varphi(x))$.

$(\supseteq)$ Let $\iota_L(q) < \varphi(x)$. Since $\varphi(x)$ is the **least** upper bound of A_x, $\iota_L(q)$ is not an upper

bound, so there is $r \in \mathbb{Q}_F(x)$ with $\iota_L(q) < \iota_L(r)$ (totality). Then $q < r$ (reflection in L), so $\iota_F(q) < \iota_F(r) < x$, and $q \in \mathbb{Q}_F(x)$.

Step 3: φ fixes the rationals. Let $r \in \mathbb{Q}$. By $(\star)$ and the first example after Definition 6.1.2, used in F and in L,

$$\mathbb{Q}_L(\varphi(\iota_F(r))) = \mathbb{Q}_F(\iota_F(r)) = \{q \in \mathbb{Q} : q < r\} = \mathbb{Q}_L(\iota_L(r)).$$

By Lemma 6.1.3 in L, $\varphi(\iota_F(r)) = \iota_L(r)$. In particular $\varphi(0_F) = 0_L$ and $\varphi(1_F) = 1_L$.

Step 4: φ is strictly increasing. Let $x < y$ in F. By density in F there is $r \in \mathbb{Q}$ with $x < \iota_F(r) < y$. Then $r \notin \mathbb{Q}_F(x)$ (by (O1)), so $r \notin \mathbb{Q}_L(\varphi(x))$ by $(\star)$, and $\varphi(x) \le \iota_L(r)$ by totality. And $r \in \mathbb{Q}_F(y) = \mathbb{Q}_L(\varphi(y))$, so $\iota_L(r) < \varphi(y)$. Hence $\varphi(x) \le \iota_L(r) < \varphi(y)$.

Step 5: sums. Let $x, y \in F$ and $q \in \mathbb{Q}$. Then

$$\begin{aligned} q \in \mathbb{Q}_L(\varphi(x+y)) &\iff q \in \mathbb{Q}_F(x+y) \\ &\iff \exists r \in \mathbb{Q}\ (r \in \mathbb{Q}_F(x) \wedge q - r \in \mathbb{Q}_F(y)) \\ &\iff \exists r \in \mathbb{Q}\ (r \in \mathbb{Q}_L(\varphi(x)) \wedge q - r \in \mathbb{Q}_L(\varphi(y))) \\ &\iff q \in \mathbb{Q}_L(\varphi(x) + \varphi(y)), \end{aligned}$$

by $(\star)$, Lemma 6.2.1 in F, $(\star)$ twice, and Lemma 6.2.1 in L. By Lemma 6.1.3 in L, $\varphi(x+y) = \varphi(x) + \varphi(y)$.

Step 6: products and the unit. First let $x > 0$ and $y > 0$. By Steps 3 and 4, $\varphi(x) > \varphi(0_F) = 0_L$ and $\varphi(y) > 0_L$; and $xy > 0$, $\varphi(x)\varphi(y) > 0$ by (O4). Let $q \in \mathbb{Q}$. If $q \le 0$, then $q \in \mathbb{Q}_F(xy) = \mathbb{Q}_L(\varphi(xy))$, and $q \in \mathbb{Q}_L(\varphi(x)\varphi(y))$, by the last clause of Lemma 6.2.2 in F and in L. If $q > 0$, then

$$\begin{aligned} q \in \mathbb{Q}_L(\varphi(xy)) &\iff q \in \mathbb{Q}_F(xy) \\ &\iff \exists r \in \mathbb{Q}_{>0}\ (r \in \mathbb{Q}_F(x) \wedge q/r \in \mathbb{Q}_F(y)) \\ &\iff \exists r \in \mathbb{Q}_{>0}\ (r \in \mathbb{Q}_L(\varphi(x)) \wedge q/r \in \mathbb{Q}_L(\varphi(y))) \\ &\iff q \in \mathbb{Q}_L(\varphi(x)\varphi(y)), \end{aligned}$$

by $(\star)$, Lemma 6.2.2 in F, $(\star)$ twice, and Lemma 6.2.2 in L (with $\varphi(x), \varphi(y) > 0$). By Lemma 6.1.3 in L, $\varphi(xy) = \varphi(x)\varphi(y)$ for $x, y > 0$.

Now the signs. By Steps 5 and 3, $\varphi(x) + \varphi(-x) = \varphi(0_F) = 0_L$, so $\varphi(-x) = -\varphi(x)$ (Proposition 2.3.2 (a)). If $x = 0$ or $y = 0$, both sides of $\varphi(xy) = \varphi(x)\varphi(y)$ are 0_L (Step 3 and Proposition 2.3.2 (b)). If $x < 0 < y$, then $-x > 0$ (Proposition 2.3.5 (a)) and $xy = -((-x)y)$ (Proposition 2.3.2 (c)), so

$$\varphi(xy) = -\varphi((-x)y) = -(\varphi(-x)\varphi(y)) = -(-\varphi(x))\varphi(y) = \varphi(x)\varphi(y).$$

The case $x > 0 > y$ follows by (F6), and if $x, y < 0$, then $xy = (-x)(-y)$ and $\varphi(xy) = (-\varphi(x))(-\varphi(y)) = \varphi(x)\varphi(y)$. Finally $\varphi(1_F) = 1_L$ by Step 3. With Steps 4 and 5, φ is an order-preserving field map.

Step 7: uniqueness. Let $\chi\colon F \to L$ be any order-preserving field map, and let $x \in F$, $q \in \mathbb{Q}$. By Proposition 6.1.1, $\chi(\iota_F(q)) = \iota_L(q)$. If $\iota_F(q) < x$, then $\iota_L(q) = \chi(\iota_F(q)) < \chi(x)$. If $\iota_L(q) < \chi(x)$, that is, $\chi(\iota_F(q)) < \chi(x)$, then $\iota_F(q) < x$, because χ reflects order (the remark after Definition 6.1.1). So

$$\mathbb{Q}_L(\chi(x)) = \mathbb{Q}_F(x) = \mathbb{Q}_L(\varphi(x)),$$

the second equality by $(\star)$. By Lemma 6.1.3 in L, $\chi(x) = \varphi(x)$. Since x was arbitrary, $\chi = \varphi$. This proves the theorem.

■

The answer to the prompt. Completeness of L is used in Step 1, where the supremum must exist, and in Step 2 ($\supseteq$), where its leastness turns "below $\varphi(x)$" into "below some $\iota_L(r)$". The Archimedean property of F is used in Step 1 for the bounds $-n$ and m, in Steps 2 and 4 through density in F, and in Steps 5 and 6 through the two lemmas applied in F. Completeness of L is used a third time, indirectly: it makes L Archimedean (Theorem 6.1.2), which Steps 3, 5, 6 and 7 need for rationals to separate in L and for the lemmas in L.

Remark. Step 7 is the whole of "uniqueness". Its idea, comparing the rationals below two candidate values, needs only that L is Archimedean, not that it is complete: Exercise C1 runs it for any two order-preserving field maps $F \to L$, with no appeal to $(\star)$. Section 6.3 leans on the uniqueness clause throughout: to show that the maps between two complete fields are inverse to each other, that the isomorphism they give is the only one, and that a complete ordered field has no order-preserving field automorphism other than the identity.

Remark. Neither hypothesis can be dropped from the existence half. *Without completeness of L:* take $F = \mathbb{R}$, which is Archimedean (Theorem 5.4.2), and $L = \mathbb{Q}$. An order-preserving field map $\varphi\colon \mathbb{R} \to \mathbb{Q}$ would fix 2 (Proposition 6.1.1), so for $\sqrt{2} \in \mathbb{R}$ (the element of Example (The Class of the Newton Sequence), as in Section 6.1), $\varphi(\sqrt{2})^2 = \varphi((\sqrt{2})^2) = \varphi(2) = 2$ in $\mathbb{Q}$, against Theorem 2.6.1. *Without the Archimedean property of F:* take $F = \mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field) and any complete L. An order-preserving field map φ would fix each $n \cdot 1$ and be strictly increasing, so $n \cdot 1_L = \varphi(n \cdot 1) < \varphi(t)$ for every $n \in \mathbb{N}$, against the Archimedean property of L (Theorem 6.1.2).

Warning. The supremum in the definition of φ is taken **in** L, of the copies $\iota_L(q)$. It cannot be taken in $\mathbb{Q}$, or in F. In $\mathbb{Q}$: the set $\mathbb{Q}_{\mathbb{R}}(\sqrt{2})$ of Example (The Rationals below the Square Root of Two) (with $\sqrt{2}$ as in Section 6.1) has no supremum in $\mathbb{Q}$. It contains the set $A = \{p \in \mathbb{Q} : p^2 < 2\}$ of Theorem 2.6.2, and its other elements are negative, so below $0 \in A$; hence the two sets have the same upper bounds in $\mathbb{Q}$, and A has no least one. In F: the supremum of $\iota_F(\mathbb{Q}_F(x))$ always exists and is x itself, since x is an upper bound and, by density in F (Theorem 4.2.5), every $u < x$ lies below some $\iota_F(q) < x$. So it is an element of the wrong field.

Quick check. Take $L \coloneqq F$, a complete ordered field. Which map $F \to F$ does Theorem 6.2.3 produce, and why, without computing a single supremum?

Solution. The identity. The theorem says there is exactly one order-preserving field map $F \to F$ (F is Archimedean by Theorem 6.1.2), and id_F is one (the first example after Definition 6.1.1). So $\varphi = \mathrm{id}_F$. This is the argument Section 6.3 builds on.

6.2.4 Exercises

6.2.4.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 6.2.3.

(a) Which steps use that L is complete, and which that F is Archimedean? Which steps use that L is Archimedean, and where does that come from?

(b) Run the definition of φ with $F = \mathbb{Q}$ and $L = \mathbb{R}$, the constructed reals. What is φ?

(c) State the idea of the proof in two sentences.

(d) Why may Step 1 not define $\varphi(x)$ as a supremum taken in F?

Solution.

(a) L complete: Step 1 (the supremum exists) and Step 2 ($\supseteq$) (its leastness). F Archimedean: Step 1 (the bounds $-n$, m), Steps 2 and 4 (density in F), and Steps 5 and 6 (Lemma 6.2.1 and Lemma 6.2.2 applied in F). L Archimedean: Steps 3, 5, 6 and 7, through Lemma 6.1.3 in L and the two lemmas in L. It comes from the completeness of L, by Theorem 6.1.2.

(b) $\varphi = \iota_{\mathbb{R}} = \kappa$. By Step 3, $\varphi(\iota_{\mathbb{Q}}(r)) = \iota_{\mathbb{R}}(r)$ for every $r \in \mathbb{Q}$, and $\iota_{\mathbb{Q}}$ is the identity (Chapter 4, after Theorem 4.2.2). So $\varphi(r) = \iota_{\mathbb{R}}(r)$, which is $\kappa(r)$ by Theorem 5.4.1 (d). Concretely, $\varphi(r)$ is the supremum in $\mathbb{R}$ of $\{\kappa(q) : q < r\}$, and that supremum is $\kappa(r)$.

(c) An order-preserving map that fixes the rationals must send x to an element with the same rationals below it, and in a complete L the supremum of their copies is such an element. Every field identity for φ is then checked by comparing the rationals below the two sides, which the sum and product lemmas compute, and rationals separate in L.

(d) In F the supremum of $\iota_F(\mathbb{Q}_F(x))$ always exists, since F is Archimedean, and it is x itself (Exercise B2 of Section 6.1): an element of F, not of L. So it does not give a map into L. The supremum must be of the copies $\iota_L(q)$, taken in L.

6.2.4.2 B. Practice

Exercise. B1

Prove Lemma 6.2.2: let F be an Archimedean ordered field, $a, b \in F$ with $a, b > 0$, and $q \in \mathbb{Q}$ with $q > 0$; show that $\iota_F(q) < ab$ if and only if some rational $r > 0$ has $\iota_F(r) < a$ and $\iota_F(q/r) < b$. Show also that $\iota_F(q) < ab$ for every rational $q \leq 0$.

Hint: imitate the proof of Lemma 6.2.1, with division in place of subtraction.

Solution. Write q for $\iota_F(q)$. Since ι_F preserves products, $\iota_F(r)\,\iota_F(r^{-1}) = \iota_F(1) = 1$ for $r \neq 0$, so $\iota_F(r^{-1}) = \iota_F(r)^{-1}$ (Lemma 4.1.1 (a)) and $\iota_F(q/r) = q/r$ in F.

If $q \leq 0$, then $q \leq 0 < ab$ by (O4).

Let $q > 0$. ($\Leftarrow$) Let $r > 0$, $r < a$, $q/r < b$. Then $q/r > 0$ (Proposition 2.5.2 (a), (O4)), so

Proposition 2.3.5 (c) gives $q = r(q/r) < a(q/r)$, and, as $a > 0$, $a(q/r) < ab$. Hence $q < ab$.

($\Rightarrow$) Let $q < ab$. Since $b^{-1} > 0$, Proposition 2.3.5 (c) gives $q/b < a$, and $q/b > 0$. By Theorem 4.2.5 there is $r \in \mathbb{Q}$ with $q/b < \iota_F(r) < a$. Then $\iota_F(r) > 0 = \iota_F(0)$, so $r > 0$ since ι_F reflects order. From $q/b < r$ and $b > 0$, $q < rb$; multiplying by $r^{-1} > 0$, $q/r < b$. This r is as required.

Exercise. B2: Cube roots

Let F be a complete ordered field and $a \in F$ with $a > 0$. Let $S := \{y \in F : y \geq 0 \text{ and } y^3 < a\}$. The following steps prove that some $s \in F$ has $s^3 = a$.

(1) S is non-empty ($0 \in S$) and bounded above by $1 + a$: if $y > 1 + a$, then $y > 1$, so $y^3 > y > a$. Let $s := \sup S$. Then $s > 0$, since $m := \min\{1, a/2\}$ has $0 < m \leq 1$, so $m^3 \leq m \leq a/2 < a$, and $m \in S$.

(2) For $y, z \in F$, $z^3 - y^3 = (z - y)(z^2 + zy + y^2)$ and $z^2 + zy + y^2 = (z + y/2)^2 + (3/4)y^2$, which is > 0 unless $y = z = 0$ (if $y \neq 0$, then $(3/4)y^2 > 0$ by Proposition 2.3.5 (e), (c) and the square is ≥ 0, so the sum is positive by (b); if $y = 0$ and $z \neq 0$, the expression is $z^2 > 0$ by (e)). Hence $y < z \iff y^3 < z^3$, for all $y, z \in F$.

(3) The case $s^3 < a$ is impossible. Let $h := \min\{1, (a - s^3)/(2(3s^2 + 3s + 1))\} > 0$. Since $0 < h \leq 1$ and $s \geq 0$, $(s + h)^3 = s^3 + 3s^2h + 3sh^2 + h^3 \leq s^3 + h(3s^2 + 3s + 1) \leq s^3 + (a - s^3)/2 < a$, so $s + h \in S$, against s being an upper bound.

(4) The case $s^3 > a$ is impossible. *Name the move, then complete it.*

(5) Hence $s^3 = a$.

Then explain how to get a cube root of $a = 0$ and of $a < 0$, and why every $a \in F$ has **exactly one** cube root.

Solution. The move is **construct a point as a supremum**: s is built as $\sup S$, and each of the two wrong cases of trichotomy is refuted by exhibiting a point that contradicts one clause of the supremum. In step (3) it was an element of S above s, against (S1). In step (4) it is an upper bound below s, against (S2).

(4) Suppose $s^3 > a$. Let $h := (s^3 - a)/(3s^2)$; it is positive, since $s > 0$ by (1). Also $h < s$: this says $s^3 - a < 3s^3$, which holds because $a > 0$ and $2s^3 > 0$. So $s - h > 0$, and

$$(s - h)^3 - a = (s^3 - a) - 3s^2h + 3sh^2 - h^3 = h^2(3s - h) > 0,$$

since $3s^2h = s^3 - a$, $h^2 > 0$ and $3s - h > 2s > 0$. Let $y > s - h$. By (2), $y^3 > (s - h)^3 > a$, so $y \notin S$. Hence every element of S is $\leq s - h$, and $s - h$ is an upper bound of S smaller than s, against s being the least upper bound.

(5) Steps (3) and (4) leave only $s^3 = a$, by trichotomy (O1).

For $a = 0$, $0^3 = 0$. For $a < 0$, $-a > 0$ has a cube root c by the above, and $(-c)^3 = -c^3 = a$ (Proposition 2.3.2 (c)). Uniqueness: if $b^3 = c^3$, then neither $b < c$ nor $c < b$ can hold, since by (2) either would make the cubes differ; so $b = c$. So every element of a complete ordered field has exactly one cube root.

6.2.4.3 C. Going deeper

Exercise. C1: Exit check: at most one map into an Archimedean field

Let F and L be Archimedean ordered fields, and let $\chi_1, \chi_2 \colon F \to L$ be order-preserving field maps. Prove that $\chi_1 = \chi_2$.

Solution (Hint 1). Compare the rationals below $\chi_1(x)$ and below $\chi_2(x)$.

Solution (Hint 2). By Proposition 6.1.1, $\chi_i(\iota_F(q)) = \iota_L(q)$ for $i = 1, 2$ and every $q \in \mathbb{Q}$.

Solution. Let $x \in F$ and $i \in \{1, 2\}$. By Proposition 6.1.1 (which needs no order hypothesis), $\chi_i(\iota_F(q)) = \iota_L(q)$ for every $q \in \mathbb{Q}$. Since χ_i is strictly increasing and reflects order (the remark after Definition 6.1.1),

$$\iota_F(q) < x \iff \chi_i(\iota_F(q)) < \chi_i(x) \iff \iota_L(q) < \chi_i(x).$$

So $\mathbb{Q}_L(\chi_1(x)) = \mathbb{Q}_F(x) = \mathbb{Q}_L(\chi_2(x))$. Since L is Archimedean, Lemma 6.1.3 gives $\chi_1(x) = \chi_2(x)$. As x was arbitrary, $\chi_1 = \chi_2$. (This is Step 7 of Theorem 6.2.3 with χ_2 in place of φ; completeness of L is not needed, and neither is the Archimedean property of F.)

Why L must be Archimedean (remark, not proved here). The last step separates $\chi_1(x)$ from $\chi_2(x)$ by a rational strictly between them, which is density of $\iota_L(\mathbb{Q})$ in L, and density needs Archimedes (Theorem 4.2.5). In the non-Archimedean field $\mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field), the elements $1/t$ and $2/t$ are different and have the same rationals below them, namely all $q \leq 0$: a positive rational q has $q > 1/n$ in $\mathbb{Q}$ for some $n \in \mathbb{Z}_{>0}$ (Theorem 2.5.3), so $\iota_{\mathbb{Q}(t)}(q) > \iota_{\mathbb{Q}(t)}(1/n) = 1/(n \cdot 1) > 2/t$, because $t > 2n \cdot 1$ (Example (A Non-Archimedean Ordered Field)). So the separating step fails in such a field. For the stronger statement proved above, with F arbitrary, the hypothesis on L cannot be dropped: on $\mathbb{Q}(t)$, the substitution $t \mapsto t + 1$ keeps degrees and leading coefficients, so it is an order-preserving field map $\mathbb{Q}(t) \to \mathbb{Q}(t)$ different from the identity (school polynomial algebra, as flagged in Example (A Non-Archimedean Ordered Field)). An example with F Archimedean needs more algebra than this volume has, and we give none.

Marking scheme. (10 points) - 2: each χ_i fixes $\mathbb{Q}$, with the reason (Proposition 6.1.1, or directly: induction for $\mathbb{N}$, then negatives and quotients). - 3: links the rationals below x to those below $\chi_i(x)$, using that χ_i fixes $\mathbb{Q}$ and reflects order. Either gives the full equivalence $\iota_F(q) < x \iff \iota_L(q) < \chi_i(x)$, or, for the squeeze route of the next item, applies reflection by χ_1 and χ_2 to a rational strictly between $\chi_1(x)$ and $\chi_2(x)$. - 3: concludes $\chi_1(x) = \chi_2(x)$ by rationals separating, or by a rational strictly between the two, invoking the Archimedean property **of** L. - 2: states the conclusion for every $x \in F$. Fatal errors (0 for the step): using density in L without a hypothesis that gives it; assuming χ_i surjective.

Exercise. C2: Dense image, and when the map is onto

Let F be an Archimedean ordered field, L a complete ordered field, and $\varphi \colon F \to L$ the map of Theorem 6.2.3.

(a) Prove that $\varphi(F)$ is dense in L: for all $u < v$ in L there is $x \in F$ with $u < \varphi(x) < v$.
(b) Prove that if F is complete, then φ is surjective.
(c) Prove that if φ is surjective, then F is complete.

Solution (Hint 1). For (b), given $y \in L$, look at the set $S \coloneqq \mathbb{Q}_L(y)$ and its copy $\iota_F(S)$ in F.

Solution (Hint 2). For (b): $\iota_F(S)$ is non-empty and bounded above in F; let $x \coloneqq \sup_F \iota_F(S)$, and show $\mathbb{Q}_F(x) = S$.

Solution. By Theorem 6.1.2, L is Archimedean.

(a) Let $u < v$ in L. By Theorem 4.2.5 in L there is $q \in \mathbb{Q}$ with $u < \iota_L(q) < v$. By Step 3 of the proof of Theorem 6.2.3, $\iota_L(q) = \varphi(\iota_F(q))$, so $x \coloneqq \iota_F(q)$ works.
(b) Let F be complete and $y \in L$. Let $S \coloneqq \mathbb{Q}_L(y)$. Since L is Archimedean, there are $n, m \in \mathbb{N}$ with $-y < \iota_L(n)$ and $y < \iota_L(m)$. So $-n \in S$, and every $q \in S$ has $\iota_L(q) < y < \iota_L(m)$, hence $q < m$ and $\iota_F(q) < \iota_F(m)$. So $\iota_F(S)$ is non-empty and bounded above in F, and $x \coloneqq \sup_F \iota_F(S)$ exists.
We show $\mathbb{Q}_F(x) = S$. ($\supseteq$) Let $q \in S$. By density in L there is $r \in \mathbb{Q}$ with $\iota_L(q) < \iota_L(r) < y$; then $r \in S$ and $q < r$, so $\iota_F(q) < \iota_F(r) \leq x$. ($\subseteq$) Let $\iota_F(q) < x$. Then $\iota_F(q)$ is not an upper bound of $\iota_F(S)$, so $\iota_F(q) < \iota_F(r)$ for some $r \in S$; hence $q < r$ and $\iota_L(q) < \iota_L(r) < y$, so $q \in S$.
By ($\star$), $\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x) = S = \mathbb{Q}_L(y)$, so $\varphi(x) = y$ by Lemma 6.1.3 in L.
(c) Let φ be surjective. It is injective (strictly increasing), so bijective, with an inverse $\varphi^{-1}\colon L \to F$ (Sets, Relations, and Cardinality, Theorem 1.5.4). Let $A \subseteq F$ be non-empty with an upper bound b. Then $\varphi(A)$ is non-empty and bounded above by $\varphi(b)$, since $a \leq b$ gives $\varphi(a) \leq \varphi(b)$. As L is complete, $s \coloneqq \sup \varphi(A)$ exists; let $c \coloneqq \varphi^{-1}(s)$. *Upper bound:* for $a \in A$, $\varphi(a) \leq s = \varphi(c)$, so $a \leq c$, because φ reflects order (and is injective, for the case of equality). *Least:* if u is an upper bound of A, then $\varphi(u)$ is an upper bound of $\varphi(A)$, so $\varphi(c) = s \leq \varphi(u)$, and $c \leq u$ as before. So $c = \sup A$ (Sets, Relations, and Cardinality, Definition 2.6.5), and F is complete.

So φ is an order-preserving field isomorphism exactly when F is complete. Section 6.3 reaches the "if" direction by a different route, which computes no supremum at all.

Marking scheme. (10 points) - 2: (a) L is Archimedean (Theorem 6.1.2), so some $\iota_L(q)$ lies strictly between u and v, and $\iota_L(q) = \varphi(\iota_F(q))$ (Step 3). - 2: (b) $\iota_F(S)$ is non-empty and bounded above in F (Archimedes in L, transferred through ι_F), so $x = \sup_F \iota_F(S)$ exists by completeness of F. - 2: (b) $\mathbb{Q}_F(x) = S$, both inclusions; ($\supseteq$) uses density in L. - 1: (b) $\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x) = \mathbb{Q}_L(y)$ by ($\star$), hence $\varphi(x) = y$ by Lemma 6.1.3 in L. - 3: (c) for non-empty $A \subseteq F$ bounded above by b, $\varphi(A)$ is bounded above by $\varphi(b)$, and $\varphi^{-1}(\sup \varphi(A))$ is an upper bound of A and the least one, both using that φ reflects order and is bijective. Fatal errors (0 for the step): using density of $\iota_L(\mathbb{Q})$ in L without the Archimedean property of L; in (b), taking a supremum in F before checking that the set is non-empty and bounded above; in (c), assuming φ^{-1} preserves order without justification.

6.3 Two Complete Ordered Fields, One Isomorphism

Section 6.2 measured an element of an Archimedean ordered field F by the rationals below it and used the measurement to build a map into a complete ordered field L: by Theorem 6.2.3 there is **exactly one** order-preserving field map $\varphi\colon F \to L$, and it keeps the rationals below each element, $\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x)$ (property $(\star)$ of that theorem). This section takes both fields complete. Then maps run in both directions, and the word "exactly" turns them into inverses of each other without a single computation. The result is the theorem the chapter is named for: any two complete ordered fields are copies of each other, by one isomorphism and no other. The section ends with automorphisms: the only map of a complete ordered field to itself that respects sums, products and 1 is the identity, even when the map is asked neither to respect the order nor to be onto.

Throughout, "order-preserving field map" has the meaning of Definition 6.1.1: a map $\chi\colon F \to L$ between ordered fields with $\chi(a+b) = \chi(a) + \chi(b)$, $\chi(ab) = \chi(a)\chi(b)$, $\chi(1) = 1$, and $a < b \Rightarrow \chi(a) < \chi(b)$ for all $a, b \in F$. Such a map also **reflects** order: if $\chi(a) < \chi(b)$, then $a < b$, since $a = b$ would give $\chi(a) = \chi(b)$ and $b < a$ would give $\chi(b) < \chi(a)$, both excluded by trichotomy (O1) in L.

6.3.1 The uniqueness theorem

Section 6.2 gave us, for complete F and L, a map $\varphi\colon F \to L$. It did not say that φ is onto. A map can preserve everything in sight and still miss most of its target: the map $\iota_{\mathbb{R}}\colon \mathbb{Q} \to \mathbb{R}$ preserves sums, products and order (Theorem 4.2.2), and the real number r of Example (The Class of the Newton Sequence), whose square is 2, is not in its image. What rules this out when the source is complete as well?

Theorem 6.3.1. Uniqueness of the Complete Ordered Field

Let F and L be complete ordered fields (Definition 5.6.1). Then there is **exactly one** order-preserving field isomorphism $F \to L$ (Definition 6.1.1). It is the map φ of Theorem 6.2.3.

Idea. Section 6.2 runs in both directions. Since F is complete, it is Archimedean (Theorem 6.1.2), so it is a legal source; since L is complete, it is a legal target. That gives $\varphi\colon F \to L$. Swapping the roles gives $\psi\colon L \to F$.

To show that ψ undoes φ, we could try to compute $\psi(\varphi(x))$ from the two suprema. We do not have to. The composite $\psi \circ \varphi\colon F \to F$ is again an order-preserving field map, and so is the identity of F. The uniqueness clause of Theorem 6.2.3, applied with F as source **and** as target, says there is only one such map. So the two are equal. **Uniqueness gives inverses.**

Before reading on: self-explain why $\psi \circ \varphi$ is order-preserving. Which of the four conditions of Definition 6.1.1 needs more than one line?

Proof. Let F and L be complete ordered fields. By Theorem 6.1.2, each of them is Archimedean.

The two maps. Since F is Archimedean and L is complete, Theorem 6.2.3 gives an order-preserving field map $\varphi\colon F \to L$. Since L is Archimedean and F is complete, the same theorem, with the roles of the two fields exchanged, gives an order-preserving field map $\psi\colon L \to F$.

The composite $\psi \circ \varphi$ is an order-preserving field map $F \to F$. For $a, b \in F$, additivity of φ and then of ψ gives

$$\psi(\varphi(a+b)) = \psi(\varphi(a) + \varphi(b)) = \psi(\varphi(a)) + \psi(\varphi(b)),$$

and the same two steps with products give $\psi(\varphi(ab)) = \psi(\varphi(a))\,\psi(\varphi(b))$. Also $\psi(\varphi(1)) = \psi(1) = 1$. If $a < b$, then $\varphi(a) < \varphi(b)$, since φ is strictly increasing, and then $\psi(\varphi(a)) < \psi(\varphi(b))$, since ψ is.

The composite is the identity. The identity map id_F is an order-preserving field map $F \to F$: each of the four conditions reads $u = u$ or $a < b \Rightarrow a < b$. Now apply Theorem 6.2.3 with source F, which is Archimedean, and target F, which is complete: there is **exactly one** order-preserving field map $F \to F$. Both $\psi \circ \varphi$ and id_F are such maps, so $\psi \circ \varphi = \mathrm{id}_F$.

Exchanging the roles of F and L, the same argument gives $\varphi \circ \psi = \mathrm{id}_L$.

Existence. So ψ is a two-sided inverse of φ, and φ is bijective by Sets, Relations, and Cardinality, Theorem 1.5.4. Hence φ is an order-preserving field isomorphism $F \to L$.

Uniqueness. Let $\chi\colon F \to L$ be any order-preserving field isomorphism. Forgetting that it is bijective, χ is an order-preserving field map $F \to L$, and by the uniqueness clause of Theorem 6.2.3 there is only one such map. Hence $\chi = \varphi$. This proves the theorem.

■

The answer to the prompt: three of the four conditions are one line each, because each of φ, ψ satisfies it and the condition passes through a composite. The order condition is the same, in two steps: $a < b$ gives $\varphi(a) < \varphi(b)$, which is a strict inequality in L, the kind of hypothesis ψ accepts. Nothing harder is hiding there. The weight of the proof sits in the uniqueness clause, which the proof uses three times: once for $\psi \circ \varphi$, once for $\varphi \circ \psi$, and once for χ.

The proof never evaluates $\psi(\varphi(x))$. A direct proof that φ is onto exists: given $y \in L$, take the supremum in F of the rationals below y, and check that φ sends it to y. It needs the measuring argument of Section 6.2 once more, now from L back to F. The route above gets the same conclusion from a sentence about how many maps there are.

Quick check. Why could the proof not simply **define** $\psi := \varphi^{-1}$, instead of building ψ from Theorem 6.2.3?

Solution. The inverse $\varphi^{-1}\colon L \to F$ exists only if φ is bijective (Sets, Relations, and Cardinality, Theorem 1.5.4), and bijectivity is what the theorem is trying to prove. Section 6.2 gives φ as an injective map (it is strictly increasing), with no claim that it is onto. The map ψ is built independently of φ, and only afterwards shown to be its inverse.

Each hypothesis has a job. Completeness of L makes L a target, so that φ exists. Completeness

of F makes F a target, so that ψ exists. The uniqueness clause with F on both sides is cited as stated, for a complete target, but by the remark after its proof in Section 6.2 it needs only that F is Archimedean. Completeness of each field also gives the Archimedean property that makes it a source. Drop the completeness of F, and the conclusion fails: $\mathbb{Q}$ is Archimedean, $\iota_{\mathbb{R}} : \mathbb{Q} \to \mathbb{R}$ is the only order-preserving field map (Theorem 6.2.3), and it is not onto, as the opening of this section noted.

6.3.2 One real line

The theorem says that two complete ordered fields have the same shape, with a unique dictionary between them. We want a short way to write the first half of that sentence.

Two ordered fields are the same up to relabeling when an order-preserving bijection respects the arithmetic.

For ordered fields F and L, write

$$F \cong L$$

when there is **an** order-preserving field isomorphism $F \to L$ (Definition 6.1.1). Read it "F is isomorphic to L", with "as ordered fields" understood.

Examples.

- $\mathbb{R} \cong \mathbb{R}$, by $\mathrm{id}_{\mathbb{R}}$. Any ordered field is isomorphic to itself in the same way.
- $F \cong \mathbb{R}$ for every complete ordered field F, where $\mathbb{R}$ is the field of Chapter 5. This is the corollary below.
- $\mathbb{Q} \ncong \mathbb{R}$. An order-preserving field isomorphism $\chi : \mathbb{Q} \to \mathbb{R}$ would satisfy $\chi = \chi \circ \iota_{\mathbb{Q}} = \iota_{\mathbb{R}}$ (Proposition 6.1.1 with $F = \mathbb{Q}$, and $\iota_{\mathbb{Q}} = \mathrm{id}_{\mathbb{Q}}$ by the uniqueness clause of Theorem 4.2.2, since $\mathrm{id}_{\mathbb{Q}}$ preserves sums and products and fixes 1). But $\iota_{\mathbb{R}}$ is not onto: the element r of Example (The Class of the Newton Sequence) has $r^2 = 2$ and is not $\iota_{\mathbb{R}}(q)$ for any $q \in \mathbb{Q}$.

Non-example by minimal change. Drop "order-preserving" and keep the rest: a bijection that respects sums, products and 1. Such a map can send a positive element to a negative one; the warning below names one. So the order condition is a genuine extra demand, and $\cong$ in this book always includes it.

Corollary 6.3.2. Every Complete Ordered Field Is a Copy of the Reals

Let F be a complete ordered field, and let $\mathbb{R}$ be the field of Chapter 5. Then $F \cong \mathbb{R}$: there is **exactly one** order-preserving field isomorphism $F \to \mathbb{R}$.

Proof. By Theorem 5.6.3, $\mathbb{R}$ is a complete ordered field. Apply Theorem 6.3.1 with $L = \mathbb{R}$. ∎

So "the real numbers" names one ordered field up to a unique relabeling. Informally, a statement about a complete ordered field that speaks only of its sums, products and order is true in F exactly when it is true in $\mathbb{R}$, since the isomorphism carries one situation to the other. Section 6.5 explains this pattern and why we do not state it as a theorem; nothing here depends on it.

Remark. On any set $\mathcal{S}$ of ordered fields, $\cong$ is an equivalence relation (Sets, Relations, and Cardinality, Definition 2.5.1): it is reflexive, symmetric and transitive. No proof in this

section uses its symmetry or transitivity; the remark records that the word "isomorphic" behaves as the word "equal" suggests. Section 6.5 uses the symmetric clause: the inverse of the isomorphism of Theorem 6.3.1 is again one.

Second reading · *Why isomorphism is an equivalence relation*

Reflexive. id_F is an order-preserving field isomorphism $F \to F$.

Symmetric. Let $\chi\colon F \to L$ be an order-preserving field isomorphism, and let $\chi^{-1}\colon L \to F$ be its two-sided inverse (Sets, Relations, and Cardinality, Theorem 1.5.4); it is bijective by the same theorem. For $u, v \in L$, additivity of χ gives

$$\chi(\chi^{-1}(u) + \chi^{-1}(v)) = \chi(\chi^{-1}(u)) + \chi(\chi^{-1}(v)) = u + v = \chi(\chi^{-1}(u + v)),$$

and since χ is injective, $\chi^{-1}(u + v) = \chi^{-1}(u) + \chi^{-1}(v)$. The same lines with products give $\chi^{-1}(uv) = \chi^{-1}(u)\chi^{-1}(v)$, and $\chi(1) = 1$ gives $\chi^{-1}(1) = 1$. Finally, let $u < v$. Then $\chi(\chi^{-1}(u)) < \chi(\chi^{-1}(v))$, and since χ reflects order, $\chi^{-1}(u) < \chi^{-1}(v)$.

Transitive. If $\chi\colon F \to L$ and $\chi'\colon L \to M$ are order-preserving field isomorphisms, then $\chi' \circ \chi$ is bijective (Sets, Relations, and Cardinality, Theorem 1.4.4), and it is an order-preserving field map by the computation in the proof of Theorem 6.3.1 for $\psi \circ \varphi$, which used nothing but the four conditions for each factor.

Example. Two Models, One Square Root of Two

Let M be a complete ordered field; Chapter 7 builds one whose elements are sets of rationals. Let $\varphi\colon \mathbb{R} \to M$ be the isomorphism of Theorem 6.3.1, and r the real number of Example (The Class of the Newton Sequence), with $r > 0$ and $r^2 = 2$. Find $\varphi(r)$ in terms of M alone, and describe the set $\mathbb{Q}_M(\varphi(r))$.

Solution. The image of r. Since φ is multiplicative and fixes the rationals (Proposition 6.1.1), $\varphi(r)^2 = \varphi(r^2) = \varphi(\iota_{\mathbb{R}}(2)) = \iota_M(2)$. Since φ is strictly increasing and $r > 0$, $\varphi(r) > \varphi(0) = 0$. By Proposition 6.1.4 in the complete field M, there is exactly one $b \geq 0$ in M with $b^2 = \iota_M(2)$. So $\varphi(r)$ is that b: the square root of 2 in M, however M was built.

The rationals below it. By $(\star)$ of Theorem 6.2.3, $\mathbb{Q}_M(\varphi(r)) = \mathbb{Q}_{\mathbb{R}}(r)$. By Example (The Rationals below the Square Root of Two) this is the set of rationals q with $q < 0$, or with $q \geq 0$ and $q^2 < 2$. The same set of rationals sits below $\sqrt{2}$ in both models.

In Chapter 7 a real number **is** a set of rationals: non-empty, not all of $\mathbb{Q}$, closed downward, and with no largest element. The set $\mathbb{Q}_F(x)$ of an element of an Archimedean ordered field has each of these four properties (the first two by the Archimedean property, the third by transitivity of $<$, the fourth by density, Theorem 4.2.5), and the isomorphism of that chapter, from $\mathbb{R}$ to the cuts, sends each $x \in \mathbb{R}$ to the set $\mathbb{Q}_{\mathbb{R}}(x)$. So the Dedekind cut is the rationals-below set of Section 6.1, made into the definition of a number.

Warning. "Isomorphic" without "order-preserving" is a weaker statement in general. Section 6.4 shows that the ordered field $\mathbb{Q}(\sqrt{2})$ has a bijection onto itself that respects sums and products and sends the positive element $\sqrt{2}$ to the negative element $-\sqrt{2}$. For complete ordered fields the two notions coincide: the last subsection shows that a map out of a complete ordered field that respects sums, products and 1 is automatically order-preserving.

6.3.3 Automorphisms

An isomorphism from a field to itself is a symmetry of the field: a relabeling of its elements that leaves the arithmetic unchanged. The field $\mathbb{Q}(\sqrt{2})$ of the warning above has a symmetry that exchanges $\sqrt{2}$ and $-\sqrt{2}$. The uniqueness theorem rules out such a symmetry in a complete ordered field as long as it respects the order; the theorem at the end of this subsection rules it out altogether.

Corollary 6.3.3. A Complete Ordered Field Has Only the Trivial Order Automorphism

Let F be a complete ordered field. The only order-preserving field automorphism of F is id_F. In particular, this holds for the field $\mathbb{R}$ of Chapter 5.

Proof. F is Archimedean by Theorem 6.1.2 and complete by hypothesis, so by Theorem 6.2.3 with source and target F there is exactly one order-preserving field map $F \to F$. The identity is one, and every order-preserving field automorphism is one. So every such automorphism is id_F. For $\mathbb{R}$, use Theorem 5.6.3. ∎

The corollary assumes that the map respects the order. Can a map that only respects the arithmetic move anything?

Theorem 6.3.4. Field Maps of a Complete Ordered Field to Itself

Let F be a complete ordered field, and let $\sigma\colon F \to F$ satisfy, for all $a, b \in F$,

$$\sigma(a+b) = \sigma(a) + \sigma(b), \qquad \sigma(ab) = \sigma(a)\sigma(b), \qquad \sigma(1) = 1.$$

Then $\sigma = \mathrm{id}_F$. No order condition and no bijectivity is assumed.

Try it first (ten minutes): why must such a σ send positive elements to positive elements? Look for a description of "positive" that uses only multiplication.

Idea. The natural first attempt is to copy Corollary 6.3.3, and it stops at once: that proof needs σ to be strictly increasing, and nothing in the hypotheses mentions $<$. A second attempt computes directly: σ fixes every rational (Proposition 6.1.1, which needs no order hypothesis), and one would like to squeeze $\sigma(x)$ between rationals close to x. But to squeeze, we must know that $q < x$ gives $\sigma(q) < \sigma(x)$, which is the order condition again.

So the order condition has to be **recovered** from the arithmetic. In a complete ordered field, positivity has a description in terms of multiplication alone: an element is positive exactly

when it is a non-zero square (Proposition 6.1.4 gives the square root). A map that respects products sends squares to squares, and squares in an ordered field are never negative. Then additivity turns "positives go to positives" into "σ is strictly increasing", and the uniqueness clause of Theorem 6.2.3 finishes, as in the corollary's proof (which never uses bijectivity).

Plan: ① positive to positive, through a square root; ② strictly increasing, through differences; ③ the uniqueness clause.

Before reading on: which line of the proof below fails for the field $\mathbb{Q}(\sqrt{2})$ of the warning above?

Proof. **Step 1: σ sends positive elements to positive elements.** Let $a \in F$ with $a > 0$. By Proposition 6.1.4 there is $b \in F$ with $b \geq 0$ and $b^2 = a$. Then $b \neq 0$, since $0^2 = 0 \neq a$ (Proposition 2.3.2 (b)). So b has an inverse b^{-1}, and

$$\sigma(b)\,\sigma(b^{-1}) = \sigma(bb^{-1}) = \sigma(1) = 1.$$

If $\sigma(b) = 0$, the left side would be 0 (Proposition 2.3.2 (b)), and $0 \neq 1$ in the field F (Definition 2.3.1). Hence $\sigma(b) \neq 0$, and

$$\sigma(a) = \sigma(bb) = \sigma(b)^2 > 0$$

by Proposition 2.3.5 (e): a non-zero square is positive.

Step 2: σ is strictly increasing. First, $\sigma(-c) = -\sigma(c)$ for every $c \in F$: additivity gives $\sigma(c) + \sigma(-c) = \sigma(c + (-c)) = \sigma(0)$, and $\sigma(0) = \sigma(0 + 0) = \sigma(0) + \sigma(0)$ gives $\sigma(0) = 0$; so $\sigma(-c)$ is the negative of $\sigma(c)$ (Proposition 2.3.2 (a)). Now let $x < y$ in F. Then $y - x > 0$ (Proposition 2.3.5 (a)), so by Step 1

$$\sigma(y) - \sigma(x) = \sigma(y) + \sigma(-x) = \sigma(y - x) > 0,$$

and $\sigma(x) < \sigma(y)$ by Proposition 2.3.5 (a).

Step 3: $\sigma = \mathrm{id}_F$. By Step 2 and the hypotheses, σ is an order-preserving field map $F \to F$ (Definition 6.1.1). The field F is Archimedean by Theorem 6.1.2, and complete by hypothesis. By the uniqueness clause of Theorem 6.2.3, with source and target F, there is exactly one order-preserving field map $F \to F$; the identity is one. Hence $\sigma = \mathrm{id}_F$.

■

The answer to the prompt: the line that fails in $\mathbb{Q}(\sqrt{2})$ is the first line of Step 1, "there is b with $b^2 = a$". In $\mathbb{Q}(\sqrt{2})$ the positive element $\sqrt{2}$ has no square root. Section 6.4 proves this, proves that $\mathbb{Q}(\sqrt{2})$ is Archimedean, and shows that the symmetry exchanging $\pm\sqrt{2}$ sends $\sqrt{2}$ to a negative element. Every later line of Steps 1 and 2 works in any ordered field. Step 3 uses completeness, as the next paragraph explains.

Completeness entered twice. It supplied square roots in Step 1, and it supplied the Archimedean property, through Theorem 6.1.2, for the uniqueness clause in Step 3. Step 3 also cites completeness of F as the target of Theorem 6.2.3, but by the remark after that proof in Section 6.2 the uniqueness clause needs only an Archimedean target, so this is the same use again. Neither use can be dropped on its own. In $\mathbb{Q}(\sqrt{2})$ only the square roots are missing, and it has a symmetry that is not the identity. In the other direction, an ordered field in which every positive element

is a square need not be Archimedean, and then an order-preserving automorphism can move elements; we do not construct such a field here.

Remark. Steps 1 and 2 used two things only: square roots of positive elements in the **source**, and the rule "non-zero squares are positive" in the **target**. If F is a complete ordered field, L is any ordered field, and $\chi\colon F \to L$ satisfies $\chi(a+b) = \chi(a) + \chi(b)$, $\chi(ab) = \chi(a)\chi(b)$ and $\chi(1) = 1$, then χ is strictly increasing, that is, an order-preserving field map. In particular, for complete ordered fields F and L, every bijection $F \to L$ that respects sums, products and 1 is the isomorphism of Theorem 6.3.1.

Quick check. Let $\sigma\colon \mathbb{R} \to \mathbb{R}$ respect sums and products, with $\sigma(1) = 1$, and let r be the real number of Example (The Class of the Newton Sequence) ($r > 0$, $r^2 = 2$). Without quoting Theorem 6.3.4, show that $\sigma(r) = r$.

Solution. $\sigma(r)^2 = \sigma(r^2) = \sigma(2) = 2$, since σ fixes the rationals (Proposition 6.1.1). By Proposition 6.1.4 there is $s \geq 0$ with $s^2 = r$, and $s \neq 0$ since $r \neq 0$. So s has an inverse, and $\sigma(s)\sigma(s^{-1}) = \sigma(ss^{-1}) = \sigma(1) = 1$, which forces $\sigma(s) \neq 0$. Hence $\sigma(r) = \sigma(s)^2 > 0$ by Proposition 2.3.5 (e). So $\sigma(r)$ and r are both non-negative square roots of 2, and the uniqueness part of Proposition 6.1.4 gives $\sigma(r) = r$.

6.3.4 Exercises

6.3.4.1 A. Check your understanding

Exercise. A1

Parts (a)–(c) are about the proof of Theorem 6.3.1, parts (d)–(f) about the proof of Theorem 6.3.4.

(a) Where is the completeness of F used? Name every place.

(b) Why is $\psi \circ \varphi$ order-preserving?

(c) Suppose F is only assumed to be an Archimedean ordered field, and L is complete. Which step of the proof fails first?

(d) Where is completeness of F used in the proof of Theorem 6.3.4?

(e) Why is $\sigma(b) \neq 0$ in Step 1?

(f) Why does σ fix every rational, although no order hypothesis is made?

Solution.

(a) In two ways. First, through Theorem 6.1.2, completeness makes F Archimedean, so F is a legal **source** for Theorem 6.2.3: for $\varphi\colon F \to L$, and for the uniqueness clause applied to maps $F \to F$ such as $\psi \circ \varphi$. Second, it makes F a legal **target**, so that $\psi\colon L \to F$ exists. The uniqueness clause for $\psi \circ \varphi$ is cited with target F, so as stated it asks for F complete; but by the remark after its proof in Section 6.2 only the Archimedean property of F is used there, and the first use already supplies it (as in (d)).

(b) If $a < b$, then $\varphi(a) < \varphi(b)$ because φ is strictly increasing, and this strict inequality in L is a hypothesis ψ accepts, so $\psi(\varphi(a)) < \psi(\varphi(b))$. A composite of strictly increasing maps is strictly increasing.

(c) The construction of $\psi\colon L \to F$: Theorem 6.2.3 needs a complete target, and F is not one. (The map φ still exists.) The conclusion fails too: for $F = \mathbb{Q}$ and $L = \mathbb{R}$, the only order-preserving field map is $\iota_{\mathbb{R}}$, which is not onto.

(d) Twice in substance. In Step 1, Proposition 6.1.4 needs F complete to give the square root b of $a > 0$, and that is how $\sigma(a) > 0$ is proved. In Step 3, Theorem 6.1.2 makes F Archimedean, which the uniqueness clause of Theorem 6.2.3 needs of its target (the clause separates elements by rationals in the target, here F, Lemma 6.1.3). Step 3 also cites the completeness of F, because that theorem is stated for a complete target; by the remark after its proof in Section 6.2, the uniqueness clause needs only an Archimedean target, so this is the same use again. Square roots alone would not do: in a non-Archimedean ordered field in which every positive element is a square, Steps 1 and 2 still make σ order-preserving, but σ need not be the identity. (We give no example.)

(e) $b \neq 0$ has an inverse, and $\sigma(b)\sigma(b^{-1}) = \sigma(1) = 1$. If $\sigma(b)$ were 0, the left side would be 0, and $0 \neq 1$ in a field.

(f) By Proposition 6.1.1, whose hypotheses are additivity, multiplicativity and $\sigma(1) = 1$ only: the composite $\sigma \circ \iota_F\colon \mathbb{Q} \to F$ respects sums and products and fixes 1, so it is ι_F by the uniqueness clause of Theorem 4.2.2. Order plays no part.

6.3.4.2 B. Practice

Exercise. B1

Prove that the only order-preserving field map $\mathbb{Q} \to \mathbb{Q}$ is the identity, using Proposition 6.1.1 and Theorem 4.2.2 but not Section 6.2. Does your proof use the order condition?

Solution. Let $\chi\colon \mathbb{Q} \to \mathbb{Q}$ be an order-preserving field map. It respects sums and products and fixes 1, so Proposition 6.1.1 with $F = L = \mathbb{Q}$ gives $\chi \circ \iota_{\mathbb{Q}} = \iota_{\mathbb{Q}}$. The identity $\mathrm{id}_{\mathbb{Q}}$ respects sums and products and fixes 1, so by the uniqueness clause of Theorem 4.2.2, $\iota_{\mathbb{Q}} = \mathrm{id}_{\mathbb{Q}}$. Hence $\chi = \chi \circ \mathrm{id}_{\mathbb{Q}} = \mathrm{id}_{\mathbb{Q}}$.

The order condition was never used: every map $\mathbb{Q} \to \mathbb{Q}$ that respects sums and products and fixes 1 is the identity. In $\mathbb{Q}$ there is nothing besides rationals to move.

Exercise. B2

Let F and L be complete ordered fields, and $\varphi\colon F \to L$ the isomorphism of Theorem 6.3.1. Let $A \subseteq F$ be non-empty and bounded above. Prove that $\varphi(A) := \{\varphi(a) : a \in A\}$ has a supremum in L and that $\varphi(\sup A) = \sup \varphi(A)$ (Sets, Relations, and Cardinality, Definition 2.6.5).

Solution. Since F is complete, $s := \sup A$ exists. We show that $\varphi(s)$ satisfies (S1) and (S2) for $\varphi(A)$ in L; then it is the supremum.

(S1). Let $a \in A$. Then $a \leq s$, so $\varphi(a) \leq \varphi(s)$: equal if $a = s$, and smaller if $a < s$, since φ is

strictly increasing.

(S2). Let $u \in L$ be an upper bound of $\varphi(A)$. Since φ is bijective, $u = \varphi(v)$ for some $v \in F$. For $a \in A$, $\varphi(a) \leq \varphi(v)$, so $a \leq v$: if $v < a$, then $\varphi(v) < \varphi(a)$, against (O1). So v is an upper bound of A, hence $s \leq v$ by (S2) for A, and $\varphi(s) \leq \varphi(v) = u$.

So $\varphi(s) = \sup \varphi(A)$. The proof used that φ is onto (in (S2)) and that φ reflects order; for an order-preserving field map that is not onto, the image of a supremum can fail to be the supremum of the image, because the image may have new upper bounds that are not images. For example, take $\iota_{\mathbb{Q}(t)} \colon \mathbb{Q} \to \mathbb{Q}(t)$ of Example (A Non-Archimedean Ordered Field) and $A = \{1 - 1/n : n \in \mathbb{Z}_{>0}\}$. Then $\sup A = 1$ in $\mathbb{Q}$, but $1 - 1/t$ is an upper bound of the image that lies below 1, since $0 < 1/t < 1/(n \cdot 1)$ for every $n \geq 1$. (For $\iota_{\mathbb{R}}$ this cannot happen: $\mathbb{Q}$ is dense in the Archimedean field $\mathbb{R}$, so $\iota_{\mathbb{R}}$ keeps every supremum that exists in $\mathbb{Q}$.)

6.3.4.3 C. Going deeper

Exercise. C1: Exit check: field maps of a complete ordered field

Let F be a complete ordered field and $\sigma \colon F \to F$ a map that respects sums and products and has $\sigma(1) = 1$. Using only Proposition 6.1.4, Proposition 6.1.1, Lemma 6.1.3, Theorem 6.1.2 and Theorem 6.2.3 (and the rules of an ordered field), prove that $\sigma = \mathrm{id}_F$. Write the proof without looking at the one in this section.

Solution. σ fixes 0 and negatives. $\sigma(0) = \sigma(0 + 0) = \sigma(0) + \sigma(0)$, so $\sigma(0) = 0$; then $\sigma(c) + \sigma(-c) = \sigma(0) = 0$, so $\sigma(-c) = -\sigma(c)$.

Positive to positive. Let $a > 0$. By Proposition 6.1.4, $a = b^2$ for some $b \geq 0$, and $b \neq 0$ since $a \neq 0$. Then $\sigma(b)\sigma(b^{-1}) = \sigma(1) = 1 \neq 0$, so $\sigma(b) \neq 0$, and $\sigma(a) = \sigma(b)^2 > 0$ because non-zero squares are positive.

Strictly increasing. If $x < y$, then $y - x > 0$, so $\sigma(y) - \sigma(x) = \sigma(y - x) > 0$.

Conclusion. σ is an order-preserving field map $F \to F$. By Theorem 6.1.2, F is Archimedean; it is complete. Theorem 6.2.3 with source and target F gives exactly one order-preserving field map $F \to F$, and id_F is one. So $\sigma = \mathrm{id}_F$.

(Alternatively, the last paragraph can be done by hand: σ fixes the rationals by Proposition 6.1.1, and a strictly increasing map that fixes the rationals keeps $\mathbb{Q}_F(x)$; since F is Archimedean, $\mathbb{Q}_F(\sigma(x)) = \mathbb{Q}_F(x)$ gives $\sigma(x) = x$ by Lemma 6.1.3.)

Marking scheme. (10 points) - 2: $\sigma(0) = 0$ and $\sigma(-c) = -\sigma(c)$ from additivity (on the by-hand route, also that σ fixes the rationals, with a reason that uses no order hypothesis). - 3: $a > 0 \Rightarrow \sigma(a) > 0$ through a square root of a, **including** the proof that $\sigma(b) \neq 0$. - 2: strict increase from additivity, through $\sigma(y) - \sigma(x) = \sigma(y - x)$. - 3: concludes $\sigma = \mathrm{id}_F$ by the uniqueness clause, or directly by comparing the rationals below x and below $\sigma(x)$, and in either case says that F is Archimedean by Theorem 6.1.2. Fatal errors (0 for the step): assuming σ order-preserving; using the uniqueness clause or the rationals-below comparison without saying why F is Archimedean; taking a square root of a negative element; dividing by $\sigma(b)$ without showing $\sigma(b) \neq 0$.

Exercise. C2: Find the error

Here is an argument that the map $\varphi\colon F \to L$ of Theorem 6.2.3 is additive. For subsets $A, B \subseteq \mathbb{Q}$, write $A + B \coloneqq \{a + b : a \in A, b \in B\}$.

(1) Let $x, y \in F$ and $q \in \mathbb{Q}_F(x + y)$.
(2) Then $q/2 < x$ and $q/2 < y$.
(3) So $q = q/2 + q/2 \in \mathbb{Q}_F(x) + \mathbb{Q}_F(y)$.
(4) Conversely, if $r < x$ and $s < y$, then $r + s < x + y$; so $\mathbb{Q}_F(x) + \mathbb{Q}_F(y) \subseteq \mathbb{Q}_F(x + y)$.
(5) Hence $\mathbb{Q}_F(x + y) = \mathbb{Q}_F(x) + \mathbb{Q}_F(y)$.
(6) Taking suprema in L, $\varphi(x + y) = \varphi(x) + \varphi(y)$.

Is the conclusion true? Name the first line that fails, give a counterexample **to that line**, and name a second gap. Then repair the argument with results of Section 6.2.

Solution. The conclusion is true: φ is additive by Theorem 6.2.3. The argument does not prove it.

The first illegal line is (2). From $q < x + y$ it does not follow that $q/2 < x$. Counterexample to the step, in $F = \mathbb{Q}$: $x = 0$, $y = 2$, $q = 1$. Then $q = 1 < 2 = x + y$, but $q/2 = 1/2$ is not below $x = 0$. The line would be right if $x = y$; it splits q evenly while $x + y$ may be split unevenly.

The second gap is (6). It uses, without saying so, that the supremum of a sum set is the sum of the suprema: $\sup_L \iota_L(A + B) = \sup_L \iota_L(A) + \sup_L \iota_L(B)$. That is true in a complete ordered field, but this book has not proved it, and the route of Section 6.2 was chosen to avoid it.

Repair. Line (5) is in fact true, for another reason: by Lemma 6.2.1 in the Archimedean field F, $q < x + y$ if and only if there is $r \in \mathbb{Q}$ with $r < x$ and $q - r < y$; then $q = r + (q - r)$ with $r \in \mathbb{Q}_F(x)$ and $q - r \in \mathbb{Q}_F(y)$. So replace (2) and (3) by Lemma 6.2.1, which chooses the split r to fit x and y. Replace (6) by the rationals-below argument: by $(\star)$, $\mathbb{Q}_L(\varphi(x + y)) = \mathbb{Q}_F(x + y)$ and $\mathbb{Q}_L(\varphi(x)) = \mathbb{Q}_F(x)$, $\mathbb{Q}_L(\varphi(y)) = \mathbb{Q}_F(y)$; by Lemma 6.2.1 in L (Archimedean by Theorem 6.1.2), $\mathbb{Q}_L(\varphi(x) + \varphi(y)) = \mathbb{Q}_L(\varphi(x)) + \mathbb{Q}_L(\varphi(y))$. So $\varphi(x + y)$ and $\varphi(x) + \varphi(y)$ have the same rationals below them, and they are equal by Lemma 6.1.3 in L.

Marking scheme. (10 points) - 3: names line (2) as the first failure and says what it would need (an even split of q, which $x + y$ does not guarantee). - 1: names line (6) as a second gap, because it assumes the sum-of-suprema rule. - 3: a counterexample to the step itself (values of x, y, q with $q < x + y$ and $q/2 \not< x$), not to the conclusion. - 3: a correct repair through Lemma 6.2.1 and the rationals below, with $(\star)$ and Lemma 6.1.3 in L; or a proof of the sum-of-suprema rule, if the reader supplies one. Fatal errors (0 for the item): declaring the conclusion false.

6.4 Two Non-Examples

The uniqueness theorem of Section 6.3 used three hypotheses in three different ways. The Archimedean property let rationals separate elements (Lemma 6.1.3). Completeness built the

map as a supremum (Theorem 6.2.3), and it supplied square roots (Proposition 6.1.4). The order condition was either assumed of the map or, for complete fields, recovered from the square roots (Theorem 6.3.4). This section builds two ordered fields that are not $\mathbb{R}$, each missing something. The field $\mathbb{R}(x)$ of rational functions obeys every axiom of an ordered field and fails the Archimedean property; it has an order-preserving symmetry that moves x. The field $\mathbb{Q}(\sqrt{2})$ is Archimedean and not complete; it has a symmetry that does not preserve the order. A table at the end compares the four fields.

Throughout, $\mathbb{R}$ is the complete ordered field of Chapter 5 (Theorem 5.6.3), and we write q for $\iota_{\mathbb{R}}(q)$ when $q \in \mathbb{Q}$, as Section 5.4 did. Powers t^i of a real number t are those of Definition 4.2.1, and finite sums $\sum_{i=0}^{d} c_i$ of real numbers are defined by recursion, as in *Sets, Relations, and Cardinality*, Section 2.3 (Sets, Relations, and Cardinality, Definition 2.3.1), with $\mathbb{R}$ as the set of values; the empty sum is 0.

6.4.1 Polynomial functions

Chapter 4 met a field of quotients of polynomials, $\mathbb{Q}(t)$ (Example (A Non-Archimedean Ordered Field)), and stated without proof that it is an ordered field. Here proofs depend on such a field, so we build it from things we can check: functions on $\mathbb{R}$, compared at large arguments, with no algebra of formal polynomials.

For $d \in \mathbb{N}$ and a list $(a_0, a_1, \ldots, a_d)$ of real numbers, the function

$$p\colon \mathbb{R} \to \mathbb{R}, \qquad p(t) := \sum_{i=0}^{d} a_i t^i,$$

is the **polynomial function of the list**. A **polynomial function** is a function $\mathbb{R} \to \mathbb{R}$ that is the polynomial function of **at least one** list. When $a_d \neq 0$, the list has **top degree** d and **top coefficient** a_d.

The phrase "of at least one list" matters. Appending zeros to a list does not change its function: $(1, 2)$ and $(1, 2, 0)$ both give $t \mapsto 1 + 2t$. So degree and top coefficient belong to a list, not to a function, until we know more.

Examples. The **constant** $c \in \mathbb{R}$ is the polynomial function of the list (c). The function $t \mapsto t$ is that of $(0, 1)$, and $t \mapsto t^2 - 100t$ is that of $(0, -100, 1)$, with top degree 2 and top coefficient 1. The list $(0, 0, 0)$ gives the zero function and has no top degree.

We compare such functions at large arguments. A statement $S(t)$ about real numbers t holds **for all large** t if there is $T \in \mathbb{R}$ such that $S(t)$ holds **for every** $t > T$. If $S_1(t)$ holds for $t > T_1$ and $S_2(t)$ for $t > T_2$, both hold for $t > \max\{T_1, T_2\}$; so finitely many statements that each hold for all large t hold together for all large t. The finiteness matters: infinitely many thresholds may have no maximum.

The basic fact is that far out, the top term wins.

Lemma 6.4.1. The Top Term Decides the Sign

Let $d \in \mathbb{N}$, let $a_0, \ldots, a_d \in \mathbb{R}$ with $a_d \neq 0$, and let $p(t) = \sum_{i=0}^{d} a_i t^i$. Put

$$T := 1 + \sum_{i=0}^{d-1} \frac{|a_i|}{|a_d|}.$$

Then for every $t > T$, $p(t) \neq 0$, and $p(t)$ has the sign of a_d: $p(t) > 0$ if $a_d > 0$, and $p(t) < 0$

if $a_d < 0$.

Idea. Split off the top term: $p(t) = a_d t^d + R(t)$ with $R(t) = \sum_{i<d} a_i t^i$. For $t \geq 1$ every power t^i with $i < d$ is at most t^{d-1}, so $|R(t)|$ is at most $(\sum_{i<d}|a_i|)t^{d-1}$. That is smaller than $|a_d|t^d = |a_d|t \cdot t^{d-1}$ as soon as $|a_d|t > \sum_{i<d}|a_i|$, which is what $t > T$ guarantees. The $1+$ in T makes $t \geq 1$.

Proof. *Case* $d = 0$. Then $p(t) = a_0$ for every t, with $a_0 \neq 0$, and the claim holds for every t (here $T = 1$, the sum being empty).

Case $d \geq 1$. Let $t > T$. Since $T \geq 1$, $t > 1 > 0$, so $t^i > 0$ and $1 \leq t^i \leq t^{d-1}$ for $0 \leq i \leq d-1$ (power laws in the ordered field $\mathbb{R}$, recorded in the ledger). Let $R(t) := \sum_{i=0}^{d-1} a_i t^i$ and $A := \sum_{i=0}^{d-1}|a_i| \geq 0$. By the triangle inequality Proposition 4.1.3 (e), applied term by term (induction on the number of terms), and $|a_i t^i| = |a_i|t^i$ (Proposition 4.1.3 (b)),

$$|R(t)| \leq \sum_{i=0}^{d-1}|a_i|t^i \leq A\, t^{d-1}.$$

From $t > T > A/|a_d|$ we get $|a_d|t > A$, and multiplying by $t^{d-1} > 0$ (Proposition 2.3.5 (c)),

$$A\, t^{d-1} < |a_d|t^d = |a_d t^d|.$$

So $|R(t)| < |a_d t^d|$. Now $p(t) = a_d t^d + R(t)$, where $t^d > 0$. If $a_d > 0$, then $a_d t^d > 0$ and $p(t) \geq a_d t^d - |R(t)| > 0$ (Proposition 4.1.3 (c)). If $a_d < 0$, then $a_d t^d < 0$ and $p(t) \leq a_d t^d + |R(t)| = -|a_d t^d| + |R(t)| < 0$. In both cases $p(t) \neq 0$ and has the sign of a_d.

■

So a polynomial function given by a list with a non-zero entry has, far out, a constant non-zero sign.

Quick check. For $p(t) = t^2 - 100t$, compute the threshold T of Lemma 6.4.1, and check the conclusion at $t = 102$. Is T the smallest threshold that works?

Solution. The list is $(0, -100, 1)$, so $T = 1 + (0 + 100)/1 = 101$. At $t = 102$, $p(102) = 102 \cdot 2 = 204 > 0$, the sign of the top coefficient 1. The threshold is not the smallest: $p(t) = t(t - 100) > 0$ already for every $t > 100$. The lemma promises a threshold, not the best one.

Three facts about these functions are all the algebra the field $\mathbb{R}(x)$ needs: sums and products of polynomial functions are polynomial functions, the top coefficient of a product is the product of the top coefficients, and a list whose function vanishes far out is all zeros.

Lemma 6.4.2. Sums, Products and the Identity Principle

Let $p(t) = \sum_{i=0}^{d} a_i t^i$ and $r(t) = \sum_{j=0}^{e} b_j t^j$, for **arbitrary** lists $(a_0, \ldots, a_d)$ and $(b_0, \ldots, b_e)$ of real numbers (zero entries, and lists of zeros, are allowed).

(a) The function $p + r$ is the polynomial function of the list $(a_k + b_k)_k$, where the shorter list is first extended by zeros to the length of the longer.

(b) The function $p \cdot r$ is the polynomial function of a list $(c_0, \ldots, c_{d+e})$ whose last entry is $c_{d+e} = a_d b_e$. If moreover $a_d \neq 0$ and $b_e \neq 0$, then $a_d b_e \neq 0$: the product list has top degree $d + e$ and top coefficient $a_d b_e$.

(c) *(Identity principle.)* If $p(t) = 0$ for all large t, then $a_i = 0$ for every i.

Part (b) gives no formula for the other entries c_k; nothing below uses one. A consequence used in the next subsection: if each of two lists has a non-zero entry, the product of their functions is given by a list with a non-zero entry. Indeed, cut each list after its last non-zero entry (this drops only zero terms, so the function is unchanged); then (b) gives a list for the product whose last entry is a product of two non-zero reals, which is non-zero (Proposition 2.3.5 (f)).

Proof.

(c) Suppose some $a_i \neq 0$. The indices $i \leq d$ with $a_i \neq 0$ form a non-empty set of integers bounded above by d, so it has a greatest element k (Proposition 2.3.6). Dropping the terms with $i > k$, which are 0, gives $p(t) = \sum_{i=0}^{k} a_i t^i$ with $a_k \neq 0$. By Lemma 6.4.1 there is T with $p(t) \neq 0$ for every $t > T$. If also $p(t) = 0$ for every $t > T'$, then $t := \max\{T, T'\} + 1$ gives $p(t) \neq 0$ and $p(t) = 0$, a contradiction. So every $a_i = 0$.

■

Second reading · *Proof of the sum and product rules*

Idea. Sums regroup term by term. For products, peel off the lowest coefficient of r: $r = b_0 + t \cdot r'$, where r' has a shorter list, and multiplying by t only shifts a list. Induction on the length of the list of r then keeps track of the last entry.

Proof.

(a) Extending a list by zeros adds only terms $0 \cdot t^i = 0$, so it does not change the function. After extending, both lists have the same length $m+1$, and regrouping a finite sum term by term (commutativity and associativity of $+$, by induction on m) and distributivity give $\sum_{k=0}^{m} a_k t^k + \sum_{k=0}^{m} b_k t^k = \sum_{k=0}^{m} (a_k + b_k) t^k$.

(b) Two special products come first. For $c \in \mathbb{R}$, $c \cdot p$ is the polynomial function of $(ca_0, \ldots, ca_d)$, by distributivity over a finite sum (induction on the number of terms). And $t \cdot p(t) = \sum_{i=0}^{d} a_i t^{i+1}$, since $t \cdot t^i = t^{i+1}$ (Definition 4.2.1 (b) and commutativity); re-indexing, $t \mapsto t \cdot p(t)$ is the polynomial function of the **shifted** list $(0, a_0, \ldots, a_d)$, of length $d + 2$ with the same last entry a_d.

Now we prove (b) by induction on e (Logic and Proof, Axiom 4.2.1, base $P(0)$), for every list $(a_0, \ldots, a_d)$ at once. For $e = 0$, $p \cdot r = b_0 \cdot p$, the function of $(b_0 a_0, \ldots, b_0 a_d)$, of length $d + 1$ and last entry $a_d b_0$. For the step, let $e \geq 1$ and let r' be the function of $(b_1, \ldots, b_e)$, a

list with last index $e-1$. Re-indexing as above, $r(t) = b_0 + t \cdot r'(t)$, so

$$p(t)\, r(t) = b_0\, p(t) + t \cdot \big(p(t)\, r'(t)\big).$$

By the induction hypothesis, $p \cdot r'$ is the function of a list with last index $d+e-1$ and last entry $a_d b_e$. Shifting gives a list for $t \mapsto t\, p(t) r'(t)$ with last index $d+e$ and last entry $a_d b_e$. The list of $b_0 \cdot p$ has last index $d \le d+e-1$, so in the sum (a) its extension has 0 in position $d+e$, and the last entry of the sum list is $a_d b_e + 0 = a_d b_e$. This is the claim for e.

Nothing here used a non-zero entry, so the statement holds for arbitrary lists. Only the last sentence of (b) uses that $\mathbb{R}$ has no zero divisors: $a_d \neq 0$ and $b_e \neq 0$ give $a_d b_e \neq 0$ (Proposition 2.3.5 (f)).

■

6.4.2 The field of rational functions

A rational function should be a quotient p/q of polynomial functions. As a function on $\mathbb{R}$ it is undefined where q vanishes, so we only ever look far out, where Lemma 6.4.1 keeps q away from 0, and we count two quotients as equal when they agree far out.

A rational function is a quotient of polynomial functions, seen only at large arguments.

Definition 6.4.1. The Ordered Field of Rational Functions

A **fraction** is a pair (p, q) of polynomial functions such that q is the polynomial function of a list with **a non-zero** entry. By Lemma 6.4.1, applied to that list cut after its last non-zero entry, there is $T_q \in \mathbb{R}$ with $q(t) \neq 0$ for every $t > T_q$; the **value** of the fraction at such t is the real number $p(t)/q(t)$.

(a) Two fractions (p_1, q_1) and (p_2, q_2) are **equivalent** if $p_1(t)/q_1(t) = p_2(t)/q_2(t)$ **for all large** t. The set $\mathbb{R}(x)$ of **rational functions** is the set of equivalence classes (Sets, Relations, and Cardinality, Definition 2.5.4); the class of (p, q) is written $[p/q]$.

(b) **Sum and product:**

$$[p_1/q_1] + [p_2/q_2] \coloneqq [(p_1 q_2 + p_2 q_1)/(q_1 q_2)],$$
$$[p_1/q_1] \cdot [p_2/q_2] \coloneqq [(p_1 p_2)/(q_1 q_2)].$$

(c) **Positive set:** $P \coloneqq \{[p/q] \in \mathbb{R}(x) : p(t)/q(t) > 0 \text{ for all large } t\}$, and $f < g$ means $g - f \in P$.

(d) $x \coloneqq [\mathrm{id}_{\mathbb{R}}/1]$, where $\mathrm{id}_{\mathbb{R}} : t \mapsto t$ is the polynomial function of $(0, 1)$ and 1 is the constant.

In words: (a) a rational function is known only through its values far out; (b) adds and multiplies by the school rules for fractions; (c) calls it positive if its values are eventually positive; (d) names the simplest non-constant one.

Well-definedness. Every condition in the definition is a statement about values at all large t, and a value at a single t is computed in the field $\mathbb{R}$. In particular, sums and products are **pointwise at large** t: beyond the thresholds of the two denominators, the value of the sum fraction is the sum of the two values, and the value of the product fraction is their product

(Lemma 4.1.1 (b) in $\mathbb{R}$). That is why equivalence of fractions is an equivalence relation, why the formulas give fractions, and why sum, product and P do not depend on the representatives; the fold checks each point.

↻ **Second reading** · *The well-definedness checks*

- *Equivalence relation* (Sets, Relations, and Cardinality, Definition 2.5.1). Reflexive: take $T = T_q$. Symmetric: the definition is symmetric. Transitive: if the values of (p_1, q_1), (p_2, q_2) agree for $t > T$ and those of (p_2, q_2), (p_3, q_3) for $t > T'$, then those of the first and third agree for $t > \max\{T, T'\}$.
- *The formulas give fractions.* $p_1q_2 + p_2q_1$ and p_1p_2 are polynomial functions by Lemma 6.4.2 (a), (b), and q_1q_2 is given by a list with a non-zero entry, by the consequence stated after that lemma.
- *Values of a sum and a product.* For $t > \max\{T_{q_1}, T_{q_2}\}$, $q_1(t)q_2(t) \neq 0$, and by the fraction rules in $\mathbb{R}$ (Lemma 4.1.1 (b)) the value of the sum fraction is $p_1(t)/q_1(t) + p_2(t)/q_2(t)$, and that of the product fraction is the product of the values. Sums and products are **pointwise at large** t.
- *Independence of representatives.* If (p_1, q_1) is replaced by an equivalent fraction, its values change at no large t, so by the previous point the values of the sum and product fractions change at no large t: the new sum and product are equivalent to the old. The same holds for the second argument. By Theorem 2.2.1 (b), sum and product are well-defined functions $\mathbb{R}(x) \times \mathbb{R}(x) \to \mathbb{R}(x)$. In the same way, "positive for all large t" does not depend on the representative, so P is well defined (Theorem 2.2.1 (c)).

Examples. The constant $c \in \mathbb{R}$ gives the class $[c/1]$, with value c at every t. The class $x = [\mathrm{id}_{\mathbb{R}}/1]$ has value t. The fraction whose numerator is $t \mapsto t$ and whose denominator is $t \mapsto t^2$ is equivalent to the fraction $(1, \mathrm{id}_{\mathbb{R}})$: their values t/t^2 and $1/t$ agree for every $t > 0$. This is the usual cancellation, now a statement about values.

Proposition 6.4.3. Rational Functions Form an Ordered Field

(a) $\mathbb{R}(x)$, with the sum, product and order of Definition 6.4.1, is an ordered field (Definition 2.3.2), with zero $[0/1]$, unit $[1/1]$, and positive elements exactly the elements of P.

(b) The map $\mathbb{R} \to \mathbb{R}(x)$, $c \mapsto [c/1]$, is an order-preserving field map (Definition 6.1.1).

Idea. Every field identity holds at each large t, because it holds in $\mathbb{R}$ and the operations are pointwise there; so it holds in $\mathbb{R}(x)$. The two places where something more is needed are where an existence claim is made or a sign must be decided: the inverse of a non-zero class, and trichotomy. Both rest on Lemma 6.4.1: a fraction's numerator, unless its list is all zeros, has an eventually constant sign.

↻ **Second reading** · *Proof that the rational functions form an ordered field*

Proof. Commutative ring. Each of (F1), (F2), (F5), (F6), (F9) of Definition 2.3.1 is an identity between classes built from given classes by sums and products. By the pointwise rule, its two sides have, at every large t, the values of the two sides of the same identity among real numbers, which holds in $\mathbb{R}$ (Theorem 5.2.3). So the two sides are equivalent fractions. For (F3), (F7) and (F4): $[0/1]$ and $[1/1]$ have values 0 and 1, so $f + [0/1] = f$ and $f \cdot [1/1] = f$; and $[(-p)/q]$, where $-p = (-1) \cdot p$ is a polynomial function by Lemma 6.4.2 (b), has value $-p(t)/q(t)$, so it is a negative of $[p/q]$. The table at the end of this proof writes out two of the identities.

$1 \neq 0$. If $[1/1] = [0/1]$, then $1 = 0$ at every $t > T$ for some T, and $t = T + 1$ gives $1 = 0$ in $\mathbb{R}$, which is false.

Inverses (F8). Let $f = [p/q] \neq [0/1]$. If every entry of a list of p were 0, then $p(t) = 0$ for all t, and f would have value 0 at all large t, so $f = [0/1]$. Hence p is the function of a list with a non-zero entry, (q, p) is a fraction, and for $t > \max\{T_p, T_q\}$ the value of $f \cdot [q/p]$ is $(p(t)/q(t))(q(t)/p(t)) = 1$. So $[q/p]$ is an inverse of f.

Positive set. We check (P1) and (P2) of Proposition 2.3.3. (P1): if $f, g \in P$, with values positive for $t > T$ and for $t > T'$ respectively, then for $t > \max\{T, T'\}$ the values of $f + g$ and fg are a sum and a product of two positive reals, hence positive by Proposition 2.3.5 (b) and (O4) in $\mathbb{R}$. (P2): let $f = [p/q]$. If a list of p is all zeros, then $f = [0/1]$, and neither f nor $-f$ is in P, since their values are 0 far out. Otherwise, cut the lists of p and q after their last non-zero entries; by Lemma 6.4.1, each of $p(t)$ and $q(t)$ has a constant non-zero sign for all large t, so $p(t)/q(t)$ has a constant non-zero sign for all large t (Proposition 2.5.2 (a) and Proposition 2.3.5 (c), (d)). If that sign is positive, $f \in P$; if negative, $-f \in P$. In this case $f \neq [0/1]$, since a value that is non-zero for all large t is not 0 for all large t. And $f \in P$, $-f \in P$ cannot both hold: at t beyond both thresholds the value would be positive and negative. So exactly one of $f \in P$, $f = 0$, $-f \in P$ holds.

By Proposition 2.3.3, $\mathbb{R}(x)$ with $f < g \Leftrightarrow g - f \in P$ is an ordered commutative ring whose positive elements are those of P. With (F8) and $1 \neq 0$, it is an ordered field. This proves (a).

(b) The value of $[c/1]$ is c at every t. By the pointwise rule, $[(c+c')/1] = [c/1] + [c'/1]$ and $[cc'/1] = [c/1][c'/1]$, and $[1/1]$ is the unit. If $c < c'$, then $[c'/1] - [c/1] = [(c'-c)/1]$ has the positive value $c' - c$ at every t, so it lies in P, and $[c/1] < [c'/1]$.

■

Two field identities written out.

Identity	Value of the left side at large t	Value of the right side	Why equal in $\mathbb{R}$
(F9) $f(g+h) = fg + fh$	$u(t)\big(v(t) + w(t)\big)$	$u(t)v(t) + u(t)w(t)$	(F9) in $\mathbb{R}$
(F2) $(f+g)+h = f+(g+h)$	$\big(u(t) + v(t)\big) + w(t)$	$u(t) + \big(v(t) + w(t)\big)$	(F2) in $\mathbb{R}$

Here $u(t)$, $v(t)$, $w(t)$ are the values of chosen representatives of f, g, h, and "large" means beyond the maximum of the finitely many thresholds T_q involved. Each value column is computed with the pointwise rule for sums and products; the representatives produced by the formulas of

Definition 6.4.1 (b) for the two sides are different fractions, but their values agree, and that is equivalence. Exercise B1 writes out one more identity, (F5), and the closure of P under sums.

From now on we write c for the class $[c/1]$ of a real number c, as (b) permits. Then for $q \in \mathbb{Q}$, $\iota_{\mathbb{R}(x)}(q) = q$: the composite $q \mapsto [\iota_{\mathbb{R}}(q)/1]$ is $\iota_{\mathbb{R}(x)}$ by Proposition 6.1.1, applied to the map of (b). Also, by induction on i, x^i is the class of $t \mapsto t^i$ (the product is pointwise), so the polynomial function $p(t) = \sum a_i t^i$ has class $\sum a_i x^i$, and $[p/q] = p(x)/q(x)$, a quotient in the field $\mathbb{R}(x)$. The notation for rational functions is now the familiar one.

Warning. The element x of $\mathbb{R}(x)$ is **not** a real number, and the order of $\mathbb{R}(x)$ is **not** "compare values at one t". At $t = 1$, the value of x is 1 and that of x^2 is also 1, yet $x < x^2$, since $t^2 - t > 0$ for every $t > 1$. A comparison made at one argument says nothing; a comparison made at all large arguments is the order.

The order is unusual in one way that matters for this chapter: x lies above every natural number.

Proposition 6.4.4. Rational Functions Are Not Archimedean

In $\mathbb{R}(x)$, $n \cdot 1 < x$ for every $n \in \mathbb{N}$. Hence $\mathbb{R}(x)$ is not Archimedean (Definition 4.2.2), and it is not complete (Definition 5.6.1).

Proof. Let $n \in \mathbb{N}$. By Theorem 4.2.2, $n \cdot 1 = \iota_{\mathbb{R}(x)}(n)$, which is the class of the constant n. The difference $x - n$ has value $t - n$, positive for every $t > n$. So $x - n \in P$, that is, $n \cdot 1 < x$.

So the element x has no $n \in \mathbb{N}$ with $x < n \cdot 1$, which is the negation of the Archimedean property (Logic and Proof, Proposition 6.2.1). If $\mathbb{R}(x)$ were complete, it would be Archimedean by Theorem 6.1.2. So it is not complete.

■

The set $\{n \cdot 1 : n \in \mathbb{N}\}$ is non-empty and bounded above by x; in fact it has no supremum, and Exercise C1 asks for a direct proof of that, with no appeal to Theorem 6.1.2.

Above every natural number sits x; below every $1/n$ sits its reciprocal.

Example. An Element Smaller Than Every $1/n$

Show that $0 < 1/x < 1/n$ in $\mathbb{R}(x)$ for every $n \in \mathbb{Z}_{>0}$. Deduce that the sequence $(1/(n+1))_{n\in\mathbb{N}}$ does **not** converge to 0 in $\mathbb{R}(x)$ (Definition 4.3.2), and that "$0 \le a < 1/n$ for every $n \ge 1$" does not force $a = 0$.

Solution. The class $1/x = [1/\mathrm{id}_{\mathbb{R}}]$ has value $1/t$, positive for $t > 0$; so $1/x > 0$. The class $1/n - 1/x$ has value $1/n - 1/t$, positive for every $t > n$ (Proposition 2.5.2 (b)); so $1/x < 1/n$.

Take $\varepsilon := 1/x \in \mathbb{R}(x)_{>0}$. For every $n \in \mathbb{N}$, $|1/(n+1) - 0| = 1/(n+1) > 1/x = \varepsilon$, so no N has $|1/(n+1) - 0| < \varepsilon$ for $n \ge N$; the sequence does not converge to 0. This is the sequence of Proposition 4.3.1, which is null exactly in Archimedean fields, and $\mathbb{R}(x)$ is not one. Finally,

$a := 1/x$ satisfies $0 \le a < 1/n$ for every $n \ge 1$, and $a \neq 0$.

A step about sequences in an ordered field that "obviously" holds is often the Archimedean property in disguise; $\mathbb{R}(x)$ is where to test it. It also shows why the rationals cannot separate elements without Archimedes (Lemma 6.1.3): the elements $1/x$ and $2/x$ are different, and the same rationals lie below each, namely the $q \le 0$.

Example. A Positive Element That Is Not a Square

Show that $x > 0$ in $\mathbb{R}(x)$, but that there is no $f \in \mathbb{R}(x)$ with $f^2 = x$.

Solution. $x > 0$ because its value t is positive for $t > 0$.

Suppose $x = f^2$ with $f = [p/q]$. Comparing values, $t = p(t)^2/q(t)^2$ for all large t, so

$$h(t) := t \cdot q(t)^2 - p(t)^2 = 0 \qquad \text{for all large } t.$$

By Lemma 6.4.2 (a), (b), h is the polynomial function of a list: q^2 and p^2 are products, $t \cdot q^2$ is a product with the function of $(0, 1)$, $-p^2$ is a product with the constant -1, and h is a sum. We now find a non-zero entry of such a list, against the identity principle (c).

Cut the list of q after its last non-zero entry b_e (it has one, since (p, q) is a fraction). By (b), q^2 has a list with top degree $2e$ and top coefficient b_e^2, and $t \cdot q^2$ has a list with top degree $2e + 1$ and top coefficient $b_e^2 \neq 0$.

If a list of p is all zeros, then $p^2 = 0$, so $t \cdot q(t)^2 = 0$ for all large t; by (c), every entry of the list of $t \cdot q^2$ is 0, against $b_e^2 \neq 0$.

Otherwise, cut the list of p after its last non-zero entry a_d. By (b), p^2 has a list with top degree $2d$ and top coefficient a_d^2, so $-p^2$ has one with top degree $2d$ and top coefficient $-a_d^2$. Now $2d \neq 2e + 1$: an integer is not both even and odd (Corollary 3.1.3 (a)). If $2e + 1 > 2d$, the sum list (a) for h has entry $b_e^2 + 0 \neq 0$ at index $2e + 1$; if $2d > 2e + 1$, it has entry $0 - a_d^2 \neq 0$ at index $2d$. Either way (c) is contradicted.

So x is not a square in $\mathbb{R}(x)$, while in $\mathbb{R}$ every positive element is one (Proposition 6.1.4).

Quick check. Is $-x$ a square in $\mathbb{R}(x)$? Is $x^2 - 2x + 1$? Answer each in one line, without the argument above.

Solution. $-x$ is not: $-x < 0$ because $x > 0$, and in any ordered field every square is ≥ 0 (Proposition 2.3.5 (e)). $x^2 - 2x + 1$ is: it equals $(x - 1)^2$ by Proposition 2.3.2 (d), since $\mathbb{R}(x)$ is a commutative ring. A sign test can only rule out negative elements; for the positive element x the example needed the top degrees.

6.4.3 The field $\mathbb{Q}(\sqrt{2})$

The second non-example sits inside $\mathbb{R}$. By Proposition 6.1.4 in the complete field $\mathbb{R}$, there is exactly one $r \geq 0$ in $\mathbb{R}$ with $r^2 = 2$; we write $\sqrt{2} \coloneqq r$. (It is the element of Example (The Class of the Newton Sequence), by the uniqueness in Proposition 6.1.4.)

Definition 6.4.2. The Field Generated by the Square Root of Two

$$\mathbb{Q}(\sqrt{2}) \coloneqq \{a + b\sqrt{2} : a, b \in \mathbb{Q}\} \subseteq \mathbb{R},$$

with the sum, product and order of $\mathbb{R}$.

Proposition 6.4.5. An Archimedean Ordered Field Inside the Reals

(a) *(Unique representation.)* If $a + b\sqrt{2} = a' + b'\sqrt{2}$ with $a, b, a', b' \in \mathbb{Q}$, then $a = a'$ and $b = b'$.

(b) $\mathbb{Q}(\sqrt{2})$ is closed under sums, negatives and products, contains 0 and 1, and contains the inverse of each of its non-zero elements:

$$(a + b\sqrt{2})^{-1} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}, \qquad a^2 - 2b^2 \neq 0.$$

(c) With the operations and order of $\mathbb{R}$, $\mathbb{Q}(\sqrt{2})$ is an Archimedean ordered field.

Idea. If two representations differed in the coefficient of $\sqrt{2}$, we could solve for $\sqrt{2}$ as a rational, which Theorem 2.6.1 forbids. Inverses come from the school trick of multiplying by $a - b\sqrt{2}$. The Archimedean property is inherited from $\mathbb{R}$, because $n \cdot 1$ is the same element in both fields.

Proof.

(a) Suppose $b \neq b'$. Then $\sqrt{2} = (a - a')/(b' - b)$, a rational s. So $\iota_{\mathbb{R}}(s^2) = \iota_{\mathbb{R}}(s)^2 = 2 = \iota_{\mathbb{R}}(2)$, and $s^2 = 2$ since $\iota_{\mathbb{R}}$ is injective (Theorem 4.2.2), against Theorem 2.6.1. Hence $b = b'$, and then $a = a'$.

(b) Sums and negatives are clear. For products, with $(\sqrt{2})^2 = 2$,

$$(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}.$$

$0 = 0 + 0\sqrt{2}$ and $1 = 1 + 0\sqrt{2}$. Let $a + b\sqrt{2} \neq 0$. If $b = 0$, then $a \neq 0$ and $a^2 - 2b^2 = a^2 \neq 0$. If $b \neq 0$ and $a^2 = 2b^2$, then $(a/b)^2 = 2$, against Theorem 2.6.1. So $a^2 - 2b^2 \neq 0$ in either case, and $(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2$ gives the displayed inverse, which lies in $\mathbb{Q}(\sqrt{2})$.

(c) The field axioms (F1)–(F9) and the order axioms (O1)–(O4) are statements "for all elements"; they hold for all elements of $\mathbb{R}$ (Theorem 5.6.3), hence for all elements of the subset $\mathbb{Q}(\sqrt{2})$, which by (b) is closed under the operations and contains 0, 1, negatives and inverses. So $\mathbb{Q}(\sqrt{2})$ is an ordered field, with the same 1 and the same sums as $\mathbb{R}$; in particular $n \cdot 1$ is the same element in both. Let $y \in \mathbb{Q}(\sqrt{2})$. Since $\mathbb{R}$ is complete, it is Archimedean (Theorem 6.1.2), so $y < n \cdot 1$ for some $n \in \mathbb{N}$. This inequality holds in $\mathbb{Q}(\sqrt{2})$, whose order is that of $\mathbb{R}$. So $\mathbb{Q}(\sqrt{2})$ is Archimedean.

■

Being Archimedean is inherited by every ordered subfield of $\mathbb{R}$. Completeness is not.

Proposition 6.4.6. The Field Generated by the Square Root of Two Is Not Complete

The element 3 of $\mathbb{Q}(\sqrt{2})$ is positive and has no square root in $\mathbb{Q}(\sqrt{2})$. Hence $\mathbb{Q}(\sqrt{2})$ is not complete.

Idea. A square root $a + b\sqrt{2}$ of 3 would, by unique representation, have $2ab = 0$ and $a^2 + 2b^2 = 3$. So either $a^2 = 3$ or $(2b)^2 = 6$, with a, $2b$ rational. Neither 3 nor 6 is a rational square, by the descent that showed 2 is not (Theorem 2.6.1). The descent there used $1 < y < 2$ for the would-be root y; for 3 we use $1 < y < 2$ again, and for 6, $2 < y < 3$.

We isolate the descent as a claim; its proof is that of Theorem 2.6.1 with 2 replaced by n, and it sits in the fold.

Claim. Let $n, k \in \mathbb{N}$ with $k^2 < n < (k+1)^2$. Then no $y \in \mathbb{Q}$ has $y^2 = n$.

↻ **Second reading** · *Proof of the claim: the descent*

Proof. Suppose $y \in \mathbb{Q}$ and $y^2 = n$. Then $y \neq 0$, since $n > k^2 \geq 0$. Replacing y by $-y$ if necessary, as in Theorem 2.6.1, we may take $y > 0$. By Proposition 2.4.4 (b), $y = a/b$ with $a \in \mathbb{Z}$ and $b \in \mathbb{Z}_{>0}$; then $a > 0$ by Theorem 2.5.1 (a), and $y^2 = n$ says $a^2 = nb^2$, as in Theorem 2.6.1. So the set

$$D := \{v \in \mathbb{Z}_{>0} : u^2 = nv^2 \text{ for some } u \in \mathbb{Z}_{>0}\}$$

contains b and is bounded below by 0. By Proposition 2.3.6 it has a least element v; fix $u \in \mathbb{Z}_{>0}$ with $u^2 = nv^2$.

$kv < u$. Otherwise $0 \leq u \leq kv$, so $u^2 \leq k^2v^2$ (Proposition 2.3.5 (g), or equality), and $k^2v^2 < nv^2$ since $v^2 > 0$ (Proposition 2.3.5 (c), (e)). Then $u^2 < nv^2$, against $u^2 = nv^2$.

$u < (k+1)v$. Otherwise $0 \leq (k+1)v \leq u$, so $(k+1)^2v^2 \leq u^2 = nv^2 < (k+1)^2v^2$, a contradiction.

Descent. Put $v' := u - kv$ and $u' := nv - ku$. By the two inequalities, $0 < v' < v$. Multiplying $kv < u$ by $u > 0$ gives $v \cdot ku < u^2 = v \cdot nv$, so $ku < nv$ (if $ku \geq nv$, multiplying by $v > 0$ would give the reverse), and $u' > 0$. Expanding,

$$\begin{aligned} u'^2 - nv'^2 &= (n^2v^2 - 2nkuv + k^2u^2) - (nu^2 - 2nkuv + nk^2v^2) \\ &= (k^2 - n)(u^2 - nv^2) = 0. \end{aligned}$$

So $v' \in D$ and $v' < v$, against the choice of v. This proves the claim.

■

Proof. Since $1 < 3 < 4$ and $4 < 6 < 9$, the claim with $(n, k) = (3, 1)$ and $(6, 2)$ shows that neither 3 nor 6 is the square of a rational.

Now suppose $c = a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, with $a, b \in \mathbb{Q}$, has $c^2 = 3$. By the product formula of Proposition 6.4.5 (b),

$$(a^2 + 2b^2) + 2ab\sqrt{2} = 3 + 0\sqrt{2},$$

so by unique representation, $2ab = 0$ and $a^2+2b^2 = 3$. Hence $a = 0$ or $b = 0$ (Proposition 2.3.5 (f), as $2 \neq 0$). If $b = 0$, then $a^2 = 3$; if $a = 0$, then $2b^2 = 3$, so $(2b)^2 = 6$. Both are excluded by the claim. So 3, which is positive, has no square root in $\mathbb{Q}(\sqrt{2})$.

If $\mathbb{Q}(\sqrt{2})$ were complete, Proposition 6.1.4 would give a square root of 3 in it. So $\mathbb{Q}(\sqrt{2})$ is not complete.

■

The proof found no explicit set without a supremum; it found a missing square root and let Proposition 6.1.4 supply the set. The set is easy to name: $\{y \in \mathbb{Q}(\sqrt{2}) : y \geq 0,\ y^2 < 3\}$ is non-empty and bounded above by 2, and in $\mathbb{R}$ its supremum is the square root $\sqrt{3}$ of 3 given by Proposition 6.1.4, which is not in $\mathbb{Q}(\sqrt{2})$.

Second reading · *Why the supremum is the square root of three*

The element $\sqrt{3}$ bounds the set above: $y \geq 0$ and $y^2 < 3 = (\sqrt{3})^2$ give $y < \sqrt{3}$ (Proposition 2.3.5 (g)). If $s < \sqrt{3}$, then, $\mathbb{R}$ being Archimedean (Theorem 6.1.2), Theorem 4.2.5 gives a rational q with $\max\{s, 0\} < q < \sqrt{3}$; then $q \geq 0$ and $q^2 < 3$, so q lies in the set and s is not an upper bound. Hence $\sqrt{3}$ is the least upper bound.

Quick check. Why does the descent claim not show that 4 is not a rational square? Which hypothesis fails?

Solution. There is no $k \in \mathbb{N}$ with $k^2 < 4 < (k+1)^2$: $k^2 < 4$ forces $k \leq 1$ (for $k \geq 2$, $k^2 \geq 4$), and then $(k+1)^2 \leq 4$. The descent needs the root $y = u/v$ to lie strictly between k and $k+1$, so that $0 < v' < v$; the natural choice $k = 2$ (for which $k^2 < n$ fails) gives $v' = u - 2v = 0$ when $y = 2$, and the descent stops. And indeed $4 = 2^2$.

The field $\mathbb{Q}(\sqrt{2})$ has a symmetry that the uniqueness theorem forbids in $\mathbb{R}$.

Example. Conjugation in $\mathbb{Q}(\sqrt{2})$

Define $\sigma\colon \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$ by $\sigma(a+b\sqrt{2}) \coloneqq a-b\sqrt{2}$ for $a, b \in \mathbb{Q}$. Show that σ is a bijection that respects sums and products and has $\sigma(1) = 1$, that it is not the identity, and that it is not order-preserving. Then find the line of the proof of Theorem 6.3.4 that fails for σ.

Solution. Well defined. By unique representation (Proposition 6.4.5 (a)), each element has exactly one pair (a, b), so the formula defines a function.

Sums, products, 1. σ respects sums, since $(a-b\sqrt{2})+(c-d\sqrt{2}) = (a+c)-(b+d)\sqrt{2}$. For products, by the product formula,

$$\sigma((ac+2bd)+(ad+bc)\sqrt{2}) = (ac+2bd)-(ad+bc)\sqrt{2} = (a-b\sqrt{2})(c-d\sqrt{2}).$$

And $\sigma(1) = \sigma(1+0\sqrt{2}) = 1$.

Bijective. $\sigma(\sigma(a+b\sqrt{2})) = a+b\sqrt{2}$, so σ is its own two-sided inverse (Sets, Relations, and Cardinality, Theorem 1.5.4).

Not the identity, not order-preserving. $\sqrt{2} > 0$, since it is a non-zero element with $\sqrt{2} \geq 0$. But $\sigma(\sqrt{2}) = -\sqrt{2} < 0 = \sigma(0)$. So $\sigma(\sqrt{2}) \neq \sqrt{2}$, and $0 < \sqrt{2}$ while $\sigma(0) > \sigma(\sqrt{2})$.

The failing line. Step 1 of the proof of Theorem 6.3.4 begins: "there is $b \in F$ with $b \geq 0$ and $b^2 = a$". For $a = \sqrt{2}$ there is no such b in $\mathbb{Q}(\sqrt{2})$. Indeed, if $c \in \mathbb{Q}(\sqrt{2})$ had $c^2 = \sqrt{2}$, then $\sigma(c)^2 = \sigma(c^2) = \sigma(\sqrt{2}) = -\sqrt{2} < 0$; but $\sigma(c)^2 \geq 0$, as every square in the ordered field $\mathbb{R}$ is (Proposition 2.3.5 (e)). So the step "$\sigma(a) = \sigma(b)^2 > 0$" cannot start, and σ indeed sends the positive element $\sqrt{2}$ to a negative one. Step 2 works in any ordered field. Step 3 as written cites completeness of F, for the Archimedean property and for the target of Theorem 6.2.3, but its conclusion still holds in $\mathbb{Q}(\sqrt{2})$: if σ were order-preserving, then σ followed by the inclusion would be an order-preserving field map $\mathbb{Q}(\sqrt{2}) \to \mathbb{R}$, hence the inclusion by the uniqueness clause of Theorem 6.2.3 (source $\mathbb{Q}(\sqrt{2})$, Archimedean by Proposition 6.4.5; target $\mathbb{R}$), and σ would be the identity. Only Step 1 genuinely fails.

Two maps into $\mathbb{R}$. Since $\mathbb{Q}(\sqrt{2})$ is Archimedean and $\mathbb{R}$ complete, Theorem 6.2.3 gives exactly one order-preserving field map $\mathbb{Q}(\sqrt{2}) \to \mathbb{R}$; it is the inclusion. The composite of σ with the inclusion is a second map $\mathbb{Q}(\sqrt{2}) \to \mathbb{R}$ that respects sums, products and 1. The uniqueness clause is not contradicted, because this second map is not order-preserving.

6.4.4 A symmetry that keeps the order

The conjugation of $\mathbb{Q}(\sqrt{2})$ escapes Theorem 6.3.4 by breaking the order. The field $\mathbb{R}(x)$ escapes Corollary 6.3.3 in the other way: it has a symmetry that keeps the order and still moves elements. The symmetry shifts the argument by 1.

Example. Shifting the Argument

For a polynomial function p, write p^+ for the function $t \mapsto p(t+1)$. Show that $\sigma([p/q]) \coloneqq [p^+/q^+]$ defines an order-preserving field automorphism σ of $\mathbb{R}(x)$ with $\sigma(x) = x+1 \neq x$.

Second reading · *The checks for the shift*

Solution. (i) p^+ is a polynomial function. By induction on the length of a list of p (Logic and Proof, Axiom 4.2.1). A list of length 1 gives a constant p, and $p^+ = p$. For a list $(a_0, \dots, a_d)$ with $d \geq 1$, let p' be the function of $(a_1, \dots, a_d)$; re-indexing the sum, $p(t) = a_0 + t \cdot p'(t)$, so

$$p(t+1) = a_0 + (1+t) \cdot p'(t+1).$$

By the induction hypothesis p'^+ is a polynomial function, $t \mapsto 1+t$ is that of $(1,1)$, and Lemma 6.4.2 (b), (a) make the right side a polynomial function.

(ii) Fractions go to fractions. If (p,q) is a fraction, then $q(t) \neq 0$ for every $t > T_q$, so $q^+(t) \neq 0$ for every $t > T_q - 1$. If every entry of a list of q^+ were 0, then q^+ would vanish everywhere; so a list of q^+ has a non-zero entry, and (p^+, q^+) is a fraction. Its value at t is the value of (p,q) at $t+1$.

(iii) Well defined, additive, multiplicative. If (p_1, q_1) and (p_2, q_2) have equal values for every $t > T$, the shifted fractions have equal values for every $t > T - 1$; so σ is well defined (Theorem 2.2.1 (a)). Since sums and products are pointwise at large t, and shifting the argument commutes with pointwise operations, $\sigma(f+g)$ and $\sigma(f) + \sigma(g)$ have the same values at all large t, and likewise for products. And $\sigma(1) = 1$, since a constant is unchanged by the shift.

(iv) Order. If $f \in P$, with positive values for $t > T$, then $\sigma(f)$ has positive values for $t > T-1$, so $\sigma(f) \in P$. For $f < g$, $\sigma(g) - \sigma(f) = \sigma(g-f) \in P$ (additivity gives $\sigma(-f) = -\sigma(f)$, as in Step 2 of the proof of Theorem 6.3.4), so $\sigma(f) < \sigma(g)$.

(v) Bijective. Shifting by -1 in the same way defines τ with $\tau(\sigma(f)) = f = \sigma(\tau(f))$: the values are shifted by 1 and back. By Sets, Relations, and Cardinality, Theorem 1.5.4, σ is bijective.

(vi) It moves x. $\sigma(x)$ has value $t+1$, which is the value of $x+1$; so $\sigma(x) = x+1$. And $x + 1 \neq x$, since $1 \neq 0$.

Remark. So Corollary 6.3.3 needs its hypothesis. Its proof went through Theorem 6.2.3, whose uniqueness clause rests on Lemma 6.1.3, which needs the Archimedean property. In $\mathbb{R}(x)$ that lemma fails exactly here: every rational lies below x and below $x+1$, so x and $x+1$ have the same rationals below them, all of $\mathbb{Q}$, and nothing can tell them apart. An order-preserving map that fixes the rationals is free to send one to the other. No proof uses this example; only the table below cites it.

6.4.5 What each hypothesis buys

The table collects the four fields of this chapter. Each entry names where it is proved; no entry rests on an exercise.

	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{2})$	$\mathbb{R}(x)$	$\mathbb{R}$
Archimedean	yes: Theorem 2.5.3	yes: Proposition 6.4.5	no: Proposition 6.4.4	yes: Theorem 5.6.3, Theorem 6.1.2
complete	no: see below	no: Proposition 6.4.6	no: Proposition 6.4.4	yes: Theorem 5.6.3
every positive element a square	no: Theorem 2.6.1	no: Proposition 6.4.6	no: Example (A Positive Element That Is Not a Square)	yes: Proposition 6.1.4
only field automorphism the identity	yes: see below	no: Example (Conjugation in $\mathbb{Q}(\sqrt{2})$)	no: Example (Shifting the Argument)	yes: Theorem 6.3.4

Two entries need a line. $\mathbb{Q}$ is not complete: if it were, Proposition 6.1.4 would give a rational square root of 2, against Theorem 2.6.1. And a map $\sigma\colon \mathbb{Q} \to \mathbb{Q}$ that respects sums and products and fixes 1 is the identity, with no order hypothesis: by Proposition 6.1.1 with $F = L = \mathbb{Q}$, $\sigma = \sigma \circ \iota_{\mathbb{Q}} = \iota_{\mathbb{Q}} = \mathrm{id}_{\mathbb{Q}}$, the last equality by the uniqueness clause of Theorem 4.2.2, since $\mathrm{id}_{\mathbb{Q}}$ is such a map.

Read across, the table shows what the uniqueness theorem used. $\mathbb{Q}(\sqrt{2})$ has the Archimedean property and lacks square roots, and its symmetry breaks the order. $\mathbb{R}(x)$ lacks the Archimedean property, and its symmetry keeps the order. Of these three, only $\mathbb{R}$, which has both, has no symmetry at all. Read down, the column of $\mathbb{Q}$ shows that the last row does not characterize $\mathbb{R}$: a field can be rigid for a cheaper reason, having nothing besides rationals to move.

6.4.6 Exercises

6.4.6.1 A. Check your understanding

Exercise. A1

(a) Is $x^2 - 100x$ positive in $\mathbb{R}(x)$? Justify your answer.

(b) Here is an argument. "$\mathbb{Q}(\sqrt{2})$ is Archimedean, so by Theorem 6.2.3 it has an order-preserving field map into the complete ordered field $\mathbb{R}$. Hence $\mathbb{Q}(\sqrt{2})$ is complete." Is the conclusion true? Which step fails?

(c) Put $1/x$, $1/x^2$ and $1/10^6$ in increasing order in $\mathbb{R}(x)$.

Solution.

(a) Yes. Its value $t^2 - 100t = t(t - 100)$ is positive for every $t > 100$. (Or: the list $(0, -100, 1)$ has top coefficient $1 > 0$, and Lemma 6.4.1 applies with $T = 101$.)

(b) The conclusion is false (Proposition 6.4.6). The illegal step is the last: completeness of the target says nothing about suprema in the source. A set $A \subseteq \mathbb{Q}(\sqrt{2})$ that is bounded above has a supremum in $\mathbb{R}$, but it may lie outside $\mathbb{Q}(\sqrt{2})$: for $A = \{y \in \mathbb{Q}(\sqrt{2}) : y \geq 0,\ y^2 < 3\}$ it is $\sqrt{3}$, which is not in $\mathbb{Q}(\sqrt{2})$ (the remark after Proposition 6.4.6).

(c) $0 < 1/x^2 < 1/x < 1/10^6$. The value of $1/x - 1/x^2$ is $(t-1)/t^2 > 0$ for $t > 1$, and that of $1/10^6 - 1/x$ is $1/10^6 - 1/t > 0$ for $t > 10^6$.

6.4.6.2 B. Practice

Exercise. B1

Prove that multiplication in $\mathbb{R}(x)$ is associative and that the positive set P is closed under sums. In each, choose representatives, name a threshold T beyond which every denominator involved is non-zero (and, for P, every value involved is positive), compute the values of both sides at $t > T$ with the pointwise rule, and say which fact about $\mathbb{R}$ makes them equal or positive.

Solution. Associativity. Let $f = [p_1/q_1]$, $g = [p_2/q_2]$, $h = [p_3/q_3]$, and $T := \max\{T_{q_1}, T_{q_2}, T_{q_3}\}$. For $t > T$, the denominators of all fractions involved are non-zero, and by the pointwise rule the value of $(fg)h$ is $(u(t)v(t))w(t)$ and that of $f(gh)$ is $u(t)(v(t)w(t))$, where u, v, w are the values of the three fractions. These are equal by (F5) in $\mathbb{R}$. So the two fractions produced by the formula have equal values for every $t > T$: they are equivalent, and $(fg)h = f(gh)$.

Closure of P under sums. Let $f, g \in P$: the value of f is positive for $t > T_1$ and that of g for $t > T_2$ (these thresholds may be taken $\geq T_{q_1}, T_{q_2}$). For $t > \max\{T_1, T_2\}$, the value of $f + g$ is $u(t) + v(t) > 0$ by (O3) in $\mathbb{R}$ (Proposition 2.3.5 (b)). So $f + g \in P$. The finiteness of the list of thresholds is what lets us take a maximum.

Exercise. B2

Show that no rational number squares to 5 or to 10. Then show that 5 has no square root in $\mathbb{Q}(\sqrt{2})$.

Solution. Since $2^2 = 4 < 5 < 9 = 3^2$ and $3^2 = 9 < 10 < 16 = 4^2$, the claim stated with Proposition 6.4.6, with $(n, k) = (5, 2)$ and $(10, 3)$, shows that neither is a rational square. (The descent there checks $u'^2 - nv'^2 = (k^2 - n)(u^2 - nv^2)$ for these values too; the identity holds for all n, k.)

Now let $c = a + b\sqrt{2}$ with $a, b \in \mathbb{Q}$ and $c^2 = 5$. By the product formula, $(a^2 + 2b^2) + 2ab\sqrt{2} = 5 + 0\sqrt{2}$, so by unique representation (Proposition 6.4.5 (a)), $2ab = 0$ and $a^2 + 2b^2 = 5$. If $b = 0$, then $a^2 = 5$; if $a = 0$, then $2b^2 = 5$, so $(2b)^2 = 10$. Both are excluded. Hence 5 has no square root in $\mathbb{Q}(\sqrt{2})$.

Exercise. B3

Show that $x^3 + 1$ is positive in $\mathbb{R}(x)$ but is not a square there.

Hint: the argument of Example (A Positive Element That Is Not a Square), with $t^3 + 1$ in place of t.

Solution. Positive. $t \mapsto t^3+1$ is the function of $(1,0,0,1)$, with top coefficient $1>0$; by Lemma 6.4.1 its value is positive for all large t.

Not a square. Suppose $x^3+1=[p/q]^2$. Then $h(t) \coloneqq (t^3+1)q(t)^2 - p(t)^2 = 0$ for all large t, and h is the function of a list by Lemma 6.4.2 (a), (b). Cut q after its last non-zero entry b_e. By (b), $(t^3+1)\cdot q^2$ has a list with top degree $3+2e = 2e+3$, odd, and top coefficient $1\cdot b_e^2 \neq 0$. If a list of p is all zeros, then $(t^3+1)q(t)^2 = 0$ for all large t, and (c) contradicts $b_e^2 \neq 0$. Otherwise cut p after its last non-zero entry a_d; then $-p^2$ has a list with top degree $2d$, even, and top coefficient $-a_d^2 \neq 0$. Since $2d \neq 2e+3$ (Corollary 3.1.3 (a)), the sum list for h has the non-zero entry b_e^2 or $-a_d^2$ at index $\max\{2d, 2e+3\}$, against Lemma 6.4.2 (c). So x^3+1 is not a square.

6.4.6.3 C. Going deeper

Exercise. C1: Exit check: the supremum property fails in the rational functions

Show directly from Definition 5.6.1, without Theorem 6.1.2, that $\mathbb{R}(x)$ is not complete.

Solution. Let $A \coloneqq \{n\cdot 1 : n \in \mathbb{N}\} \subseteq \mathbb{R}(x)$. It is non-empty ($0\cdot 1 \in A$), and it is bounded above by x: the value of $x-n$ is $t-n>0$ for $t>n$ (Proposition 6.4.4).

Suppose $s \in \mathbb{R}(x)$ is a supremum of A. For every $n\in\mathbb{N}$, $n+1\in\mathbb{N}$, so $(n+1)\cdot 1 \leq s$ by (S1), that is, $n\cdot 1+1 \leq s$ (Definition 4.2.1 (a)), and $n\cdot 1 \leq s-1$. So $s-1$ is an upper bound of A, and (S2) gives $s \leq s-1$, that is, $1\leq 0$, which is false in an ordered field (Proposition 2.3.5 (e)). So A has no supremum, and $\mathbb{R}(x)$ is not complete.

Marking scheme. (10 points) - 2: exhibits a non-empty set bounded above, with the bound verified (here x, by a threshold for the value of $x-n$, or by Lemma 6.4.1). - 3: supposes a supremum s, and uses (S1) on $(n+1)\cdot 1$. - 3: plays leastness against $s-1$: either shows that $s-1$ is an upper bound, from $(n+1)\cdot 1 \leq s \Rightarrow n\cdot 1 \leq s-1$ for every n, or notes $s-1<s$ so that (S2) gives n with $n\cdot 1 > s-1$. - 2: reaches the contradiction (with (S2), or with (S1) via $(n+1)\cdot 1 > s$), concluding that no supremum exists. Fatal errors (0 for the step): proving only "not Archimedean" and stopping; treating x as a real number (for instance, "take $n>x$").

Exercise. C2: A complete ordered field is uncountable

Let F be a complete ordered field. Write 3^n for the power $(3\cdot 1_F)^n$ of Definition 4.2.1. For $s=(s_n)_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$, let

$$S_k(s) \coloneqq \sum_{n=0}^{k} \frac{2s_n}{3^{n+1}} \quad (k\in\mathbb{N}), \qquad \Phi(s) \coloneqq \sup_{k\in\mathbb{N}} S_k(s).$$

(a) Show that $\Phi(s)$ exists.

(b) Show that $\Phi\colon \{0,1\}^{\mathbb{N}} \to F$ is injective.

(c) Deduce that F is uncountable (Sets, Relations, and Cardinality, Theorem 3.4.1).

Solution (Hint 1). Think of ternary expansions that use only the digits 0 and 2. For (a), find a single upper bound for all $S_k(s)$ by comparing with $s_n = 1$ for every n.

Solution (Hint 2). For (b), let j be the least index where s and s' differ (it exists by least elements), with $s_j = 1$ and $s'_j = 0$, and put $c := S_{j-1}(s) = S_{j-1}(s')$ (with $c := 0$ if $j = 0$).

Solution. Tail bound. For $k \le m$, by induction on m (Logic and Proof, Axiom 4.2.1), exactly as in Eq. 5.6.3 but in F:

$$0 \le S_m(s) - S_k(s) \le \frac{1}{3^{k+1}} - \frac{1}{3^{m+1}},$$

since passing from m to $m+1$ adds $2s_{m+1}/3^{m+2} \in [0, 2/3^{m+2}]$, and $-1/3^{m+1} + 2/3^{m+2} = -1/3^{m+2}$. In particular $S_k(s)$ is non-decreasing in k.

(a) The set $\{S_k(s) : k \in \mathbb{N}\}$ is non-empty, and bounded above by 1: $S_k(s) \le S_0(s) + 1/3 - 1/3^{k+1} \le 2/3 + 1/3 = 1$ by the tail bound with $k = 0$. Since F is complete, $\Phi(s)$ exists.

(b) Let $s \neq s'$. The set of indices where they differ is a non-empty subset of $\mathbb{N}$, so it has a least element j (Proposition 2.3.6); say $s_j = 1$ and $s'_j = 0$. Let $c := S_{j-1}(s) = S_{j-1}(s')$, the sums agreeing up to $j-1$ (with $c := 0$ if $j = 0$). Then $\Phi(s) \ge S_j(s) = c + 2/3^{j+1}$ by (S1). For $k \ge j$, the tail bound gives $S_k(s') \le S_j(s') + 1/3^{j+1} = c + 1/3^{j+1}$; for $k < j$, $S_k(s') \le S_j(s') = c$. So $c + 1/3^{j+1}$ is an upper bound of the $S_k(s')$, and $\Phi(s') \le c + 1/3^{j+1}$ by (S2). Hence $\Phi(s) - \Phi(s') \ge 1/3^{j+1} > 0$, and $\Phi(s) \neq \Phi(s')$.

(c) If F were countable, there would be an injection $g\colon F \to \mathbb{N}$ (Sets, Relations, and Cardinality, Theorem 3.2.2), and $g \circ \Phi$ would be an injection $\{0,1\}^{\mathbb{N}} \to \mathbb{N}$ (Sets, Relations, and Cardinality, Theorem 1.4.4), making $\{0,1\}^{\mathbb{N}}$ countable, against Sets, Relations, and Cardinality, Theorem 3.4.1.

Alternatively, by Corollary 6.3.2 there is a bijection $F \to \mathbb{R}$, and $\mathbb{R}$ is uncountable by Exercise (R3). The route above shows that uncountability follows from the supremum property directly, with no construction behind it.

Marking scheme. (10 points) - 2: (a), a common upper bound for the partial sums, and completeness invoked for the supremum. - 2: the tail bound, by induction, in F. - 4: (b), the first differing index j by least elements, the lower bound $c + 2/3^{j+1}$ for $\Phi(s)$ by (S1), and the upper bound $c + 1/3^{j+1}$ for $\Phi(s')$ by (S2). - 2: (c), the composite injection and the uncountability of $\{0,1\}^{\mathbb{N}}$. The second route, by Corollary 6.3.2 and the uncountability of $\mathbb{R}$, also earns these 2 points; (a) and (b) are still marked as above. Fatal errors (0 for the step): treating $\Phi(s)$ as an infinite sum computed term by term; using decimal or ternary expansions of real numbers as known facts.

Exercise. C3: Increasing and bounded

Let F be a complete ordered field and $(a_n)_{n\in\mathbb{N}}$ a sequence in F with $a_n \le a_{n+1}$ for every n, whose set of values is bounded above. Prove that $a_n \to s$ in F (Definition 4.3.2), where s is the supremum of the set of values. Here is the start of a proof. *Name the move, then complete it.*

"Let $s := \sup\{a_n : n \in \mathbb{N}\}$, which exists because the set is non-empty and bounded above and F is complete. Let $\varepsilon \in F_{>0}$. ..."

Solution. The move is **construct a point as a supremum, and use its leastness through "$s-\varepsilon$ is not an upper bound"**, which hands us a witness. It built square roots in Proposition 6.1.4; here it builds a limit.

Completion. Since $s-\varepsilon < s$ and s is the least upper bound, $s-\varepsilon$ is not an upper bound of the values, so $a_N > s-\varepsilon$ for some $N \in \mathbb{N}$. For $n \geq N$, $a_N \leq a_n$, by induction on $n-N$ from $a_m \leq a_{m+1}$; and $a_n \leq s$ by (S1). So

$$s-\varepsilon < a_N \leq a_n \leq s < s+\varepsilon,$$

that is, $|a_n - s| < \varepsilon$ for every $n \geq N$ (Proposition 4.1.3 (d)). Since ε was arbitrary, $a_n \to s$. Here N depends on ε.

6.5 Assume the Axioms or Build a Model

Section 6.3 showed that two complete ordered fields are always linked, and by exactly one order-preserving field isomorphism (Theorem 6.3.1). This last section asks what that theorem means for the rest of the series. There are two honest ways to begin analysis: assume the axioms of a complete ordered field, or build a model of them. We compare the two, show what an isomorphism carries from one model to another and what it does not, and list what *Analysis on the Real Line* takes from this volume.

6.5.1 Two ways to begin analysis

(A) Assume the axioms. Declare that $\mathbb{R}$ is a set with an addition, a multiplication and a relation $<$ satisfying the field axioms (F1)–(F9) and $1 \neq 0$ of Definition 2.3.1, the order axioms (O1)–(O4) of Definition 2.3.2, and completeness, Definition 5.6.1: every non-empty subset that is bounded above has a supremum. Then prove theorems from these axioms and nothing else. *Analysis on the Real Line*, Chapter 1 begins this way. It assumes an ordered field containing $\mathbb{Q}$ in which the least upper bound axiom holds, and every theorem of that volume about real numbers is proved from this assumption, with the logic and set theory of the earlier volumes.

(B) Build a model. Start from the axioms of set theory and construct the objects one by one. This volume did that: the natural numbers as the set ω in Chapter 1, the integers and rationals as classes of pairs in Chapter 2, and the real numbers as classes of Cauchy sequences of rationals in Chapter 5. The result is a complete ordered field (Theorem 5.6.3), and no property of it was assumed.

Route (A) is shorter, and it is the route of most analysis courses. On its own, though, it leaves two questions open.

(Q1) *Is there a complete ordered field at all?* A list of axioms can be contradictory: a list of demands that nothing meets. Theorems proved from a contradictory list would all be worthless.

(Q2) *If there are several, which one is* $\mathbb{R}$*?* Route (A) says "let $\mathbb{R}$ be **a** complete ordered field". The article matters.

Route (B) answers (Q1). The Cauchy model of Chapter 5 is a complete ordered field, built inside the set theory of Chapter 1. So if a contradiction could be derived from the axioms of a complete ordered field, the same derivation, run in the Cauchy model, would be a contradiction about one particular set built in that set theory. In short, the axioms are consistent if the set theory of Chapter 1 is. That is the most a construction can give: it moves the question of consistency from the axioms of $\mathbb{R}$ to the axioms of set theory.

Route (B) alone does not answer (Q2). A second builder who uses Dedekind cuts, as the optional Chapter 7 does, obtains a different set, whose elements are sets of rationals rather than classes of sequences. The answer to (Q2) is the theorem of Section 6.3, and the rest of this section explains why it is the right answer.

6.5.2 What uniqueness adds

Throughout this part, F and L are complete ordered fields and $\varphi\colon F \to L$ is the order-preserving field isomorphism of Theorem 6.3.1. By definition (Definition 6.1.1), φ is a bijection that respects $+$, $\cdot$, 1 and $<$. Everything else we need follows in a line or two. Within this section, (P1)–(P6) name the six facts below; they are not the positive-set clauses (P1)–(P2) of Proposition 2.3.3 used in Section 6.4.

(P1) $\varphi(0) = 0$, $\varphi(-x) = -\varphi(x)$ and $\varphi(x - y) = \varphi(x) - \varphi(y)$. Indeed $\varphi(0) = \varphi(0 + 0) = \varphi(0) + \varphi(0)$, and adding $-\varphi(0)$ gives $\varphi(0) = 0$; then $\varphi(x) + \varphi(-x) = \varphi(0) = 0$.

(P2) $\varphi(x^{-1}) = \varphi(x)^{-1}$ for $x \neq 0$, since $\varphi(x)\varphi(x^{-1}) = \varphi(1) = 1$.

(P3) $x < y$ **if and only if** $\varphi(x) < \varphi(y)$: the forward direction is the definition, and the backward one is the reflection of order shown just after Definition 6.1.1. With (P1), $x > 0$ if and only if $\varphi(x) > 0$.

(P4) $\varphi(|x|) = |\varphi(x)|$. If $x \geq 0$, then $\varphi(x) \geq 0$ by (P3), and both sides are $\varphi(x)$. If $x < 0$, then $\varphi(x) < 0$, and both sides are $\varphi(-x) = -\varphi(x)$ (Definition 4.1.1, (P1)).

(P5) $\varphi(\iota_F(q)) = \iota_L(q)$ for every $q \in \mathbb{Q}$ (Proposition 6.1.1): each rational goes to the same rational.

(P6) The inverse $\varphi^{-1}\colon L \to F$ is again an order-preserving field isomorphism (the remark after Corollary 6.3.2). So everything said about φ holds for φ^{-1} too.

The map acts on larger objects through their elements. A subset $A \subseteq F$ goes to its image $\varphi(A) = \{\varphi(a) : a \in A\}$; a sequence (a_n) in F goes to $(\varphi(a_n))$; a function $f\colon F \to F$ goes to $\varphi \circ f \circ \varphi^{-1}\colon L \to L$. Because φ is a bijection, $x \in A$ if and only if $\varphi(x) \in \varphi(A)$, and every subset, sequence or function of L arises in this way from exactly one of F.

An isomorphism is a renaming: it changes what the elements are called and keeps every relation between them.

Here is the renaming at work on a statement with three quantifiers.

Example: limits are carried across. Let (a_n) be a sequence in F and $a \in F$. Then $a_n \to a$ in F (Definition 4.3.2) if and only if $\varphi(a_n) \to \varphi(a)$ in L.

For ($\Rightarrow$), let $\varepsilon \in L_{>0}$, and put $\varepsilon' := \varphi^{-1}(\varepsilon)$. Then $\varepsilon' > 0$ by (P3), since $\varphi(\varepsilon') = \varepsilon > 0$. Since $a_n \to a$, there is $N \in \mathbb{N}$, depending on ε' and so on ε, with $|a_n - a| < \varepsilon'$ for every $n \geq N$. For such n, by (P1), (P4) and (P3),

$$|\varphi(a_n) - \varphi(a)| = \varphi(|a_n - a|) < \varphi(\varepsilon') = \varepsilon.$$

So $\varphi(a_n) \to \varphi(a)$. For ($\Leftarrow$), apply ($\Rightarrow$) to φ^{-1}, which is allowed by (P6).

The step that needs φ to be **onto** is the first one: the tolerance ε is an element of L, and the hypothesis speaks only about tolerances in F. Onto is also what carries "converges to some limit" back, since that limit must have a preimage. The map $\kappa = \iota_{\mathbb{R}} \colon \mathbb{Q} \to \mathbb{R}$ (Theorem 5.4.1 (d)) is an order-preserving field map that is not onto, and it fails here: the Newton sequence (x_n) has no limit in $\mathbb{Q}$ (Theorem 4.6.1), while $(\kappa(x_n))$ converges in $\mathbb{R}$ (Theorem 5.4.5) to $r = [(x_n)]$, which is not κ of any rational (Example (The Class of the Newton Sequence)).

The example shows the general pattern. Take a statement about a complete ordered field that is phrased with $+$, $\cdot$, $<$, elements, and subsets, sequences and functions built from elements, with quantifiers ranging over these. Rename every object by φ. Each basic piece keeps its truth value: an equation because φ respects $+$ and $\cdot$ and is injective, with (P1), (P2), (P4) and (P5) for terms that contain $-$, $^{-1}$, $|\cdot|$ or rationals; an inequality by (P3) as well; and a membership because φ is a bijection, as noted above. Quantifiers match as well: "for every element of F" matches "for every element of L" because φ is onto; the same holds for subsets, sequences and functions. So the statement is true in F exactly when it is true in L. We do not state this as a theorem. Making "statement" precise is the business of formal languages, which *Logic and Proof*, Chapter 4 begins with its first-order formulas, and the proof would be an induction on how statements are built. A particular instance is checked as the example was: one renaming for each object, and one preservation fact for each basic piece.

Quick check. Let F, L and φ be as above, and suppose some $x \in F$ satisfies $x^3 = x + 1$. Show that some $y \in L$ satisfies $y^3 = y + 1$.

Solution. Take $y := \varphi(x)$. By Definition 4.2.1 (b), $x^3 = ((1 \cdot x)x)x$, and φ respects 1 and products, so $\varphi(x^3) = \varphi(x)^3$. Hence $y^3 = \varphi(x^3) = \varphi(x + 1) = \varphi(x) + \varphi(1) = y + 1$.

Theorems need no uniqueness; truth does. One consequence of route (A) holds for any list of axioms, with no uniqueness theorem: a statement **proved** from the axioms is true in every structure satisfying them, since the proof can be run in each. The ordered-field axioms alone have the models $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$ (Definition 6.4.2) and $\mathbb{R}(x)$ (Definition 6.4.1), and every result proved from those axioms, such as the rules of Proposition 2.3.5, holds in all three. Yet the three disagree about statements the axioms do not settle:

- "some x has $x \cdot x = 1 + 1$" is true in $\mathbb{Q}(\sqrt{2})$ and false in $\mathbb{Q}$ (Theorem 2.6.1);
- "some y lies above $n \cdot 1$ for every $n \in \mathbb{N}$" is true in $\mathbb{R}(x)$ (witness $y = x$, Proposition 6.4.4) and false in $\mathbb{Q}$ (Theorem 2.5.3).

For complete ordered fields no such disagreement can occur. Every statement of the kind described above has the same truth value in all complete ordered fields, **whether or not** anybody has a proof of it from the axioms. That is what uniqueness adds to route (A): its axioms describe one structure, not a family of structures that agree only on what the axioms prove.

6.5.3 What is not carried across

The renaming keeps relations between elements. It does not keep what the elements are made of. In the Cauchy model, the zero is $0_{\mathbb{R}} = [(0)_n]$ (Definition 5.2.2), a class that contains the constant sequence $(0)_n$; so the zero of $\mathbb{R}$ is not the empty set. In another complete ordered field the zero can be the empty set, because the elements of a model can be chosen almost at will.

A model with any elements we like. Let F be a complete ordered field, X a set, and $\beta\colon F \to X$ a bijection. Define, for $u, v \in X$,

$$u \oplus v \coloneqq \beta(\beta^{-1}(u) + \beta^{-1}(v)), \qquad u \odot v \coloneqq \beta(\beta^{-1}(u)\,\beta^{-1}(v)),$$
$$u \prec v \text{ if and only if } \beta^{-1}(u) < \beta^{-1}(v).$$

Then X, with $\oplus$, $\odot$ and $\prec$, is a complete ordered field, and β is an order-preserving field isomorphism. Indeed, with $u = \beta(a)$ and $v = \beta(b)$, the definitions read $\beta(a+b) = \beta(a) \oplus \beta(b)$, $\beta(ab) = \beta(a) \odot \beta(b)$, and $a < b$ if and only if $\beta(a) \prec \beta(b)$. Each axiom for X, rewritten in this way, becomes the same axiom for F: for (F1), $u \oplus v = \beta(a+b) = \beta(b+a) = v \oplus u$; the zero and one of X are $\beta(0)$ and $\beta(1)$, and $\beta(1) \neq \beta(0)$ because β is injective and $1 \neq 0$ in F; (F2)–(F9) and (O1)–(O4) go the same way. For completeness, let $B \subseteq X$ be non-empty and bounded above by w. Then $\beta^{-1}(B)$ is non-empty and bounded above by $\beta^{-1}(w)$, so it has a supremum s in F, and $\beta(s)$ is a supremum of B, because β and β^{-1} both preserve the order by the definition of $\prec$.

Apply this to the Cauchy model with $X \coloneqq (\mathbb{R} \setminus \{0_{\mathbb{R}}\}) \cup \{\emptyset\}$, $\beta(0_{\mathbb{R}}) \coloneqq \emptyset$ and $\beta(x) \coloneqq x$ otherwise. Every element of $\mathbb{R}$ is a class $[a]$, which contains a and so is not empty; hence $\emptyset \notin \mathbb{R}$, and β is a bijection. The zero of X is $\beta(0_{\mathbb{R}}) = \emptyset$. So the statement "the zero element is the empty set" is false in $\mathbb{R}$ and true in X, although the two are isomorphic. The statement is about what an element is made of, and no renaming can be expected to keep that.

Warning. **Isomorphic is not equal.** The isomorphism links two models; it does not make them one set, and it does not make $\mathbb{Q}$ a subset of $\mathbb{R}$. In the Cauchy model the rational number 2 is a class of pairs of integers (Definition 2.4.1), while the real number 2 is $\kappa(2) = [(2)_n]$, a class of sequences (Definition 5.4.1); they are different sets, and, as Section 5.2 noted after Definition 5.2.1, the rational 2 is not an element of $\mathbb{R}$. "$\mathbb{Q} \subseteq \mathbb{R}$" is shorthand for the embedding $\iota_{\mathbb{R}}$, the one copy of $\mathbb{Q}$ that Theorem 4.2.2 provides, and every isomorphism respects that copy by (P5).

So "which construction is $\mathbb{R}$?" has no mathematical content. Every construction yields a complete ordered field; any two are linked by exactly one isomorphism; and that isomorphism decides, in the same way in both, every question posed with $+$, $\cdot$ and $<$. Questions such as "is $\sqrt{2}$ a set of rationals?" do depend on the model. In the Cauchy model, $\sqrt{2}$ is the element r of Example (The Class of the Newton Sequence), a set of sequences; in the cut model of Chapter 7 it is the set of rationals below it. Nothing in analysis depends on the answer, and from *Analysis on the Real Line* on the question is never asked again.

6.5.4 What *Analysis on the Real Line* takes from here

The analysis volume uses route (A). The table matches each of its assumptions with the place where this series supplies or justifies it.

Assumed in *Analysis on the Real Line*, Chapter 1	Supplied or justified here
$\mathbb{R}$ is an ordered field, axioms (F1)–(F9), (O1)–(O4)	the axioms: Definition 2.3.1, Definition 2.3.2, in the same numbering (in *Analysis on the Real Line*, (F7) includes $1 \neq 0$; here it is a separate clause of the field definition); a model: Theorem 5.3.3
$\mathbb{R}$ contains $\mathbb{Q}$	every ordered field contains exactly one copy of $\mathbb{Q}$: Theorem 4.2.2; isomorphisms respect it: Proposition 6.1.1
the least upper bound axiom	the property: Definition 5.6.1; a model: Theorem 5.6.3
"the" real numbers	any two models are isomorphic, by exactly one isomorphism: Theorem 6.3.1, Corollary 6.3.2
logic and set theory	*Logic and Proof*, *Sets, Relations, and Cardinality*, and Chapter 1 of this volume

Some results are proved in both volumes, from different starting points. *Analysis on the Real Line*, Chapter 1 proves from its axioms that $\mathbb{N}$ is not bounded above, that $\mathbb{Q}$ is dense in $\mathbb{R}$, and that positive numbers have square roots. Here the same three facts were proved for an arbitrary complete ordered field: Theorem 6.1.2, Theorem 4.2.5 (which needs Archimedes, supplied by the first) and Proposition 6.1.4. The two volumes start from different places, and neither volume's proofs cite the other's.

Remark. **Not proved here.** The supremum property speaks about **all subsets** of the field, and that is why it pins $\mathbb{R}$ down. Properties that speak only about elements do not. For instance, there are ordered fields that are not Archimedean, yet in which, as in $\mathbb{R}$, every positive element is a square and every polynomial function of odd degree with coefficients in the field has a zero. The distinction is between statements whose quantifiers range over elements only, the first-order formulas of *Logic and Proof*, Chapter 4 (Logic and Proof, Definition 5.2.2), and statements that also quantify over subsets. This volume does not prove that such fields exist, and nothing in it depends on them.

This ends the main line of the volume. The optional Chapter 7 builds a second complete ordered field from Dedekind cuts; by Theorem 6.3.1 it is a copy of the Cauchy model, and the chapter can be read as a second look at the rationals-below sets of Section 6.1.

6.5.5 Exercises

6.5.5.1 A. Check your understanding

Exercise. A1

Throughout, F and L are complete ordered fields and $\varphi\colon F \to L$ is the order-preserving field isomorphism.

(a) For each statement, decide whether its truth could depend on which complete ordered field is taken as $\mathbb{R}$, and justify your answer: (i) $\sqrt{2}+\sqrt{3} > 3$; (ii) the rational number

0 is an element of the set $\sqrt{2}$; (iii) $\mathbb{R}$ is uncountable.

(b) True or false: "The ordered-field axioms prove every statement that is true in $\mathbb{Q}$."

(c) Which step of the example "limits are carried across" uses that φ is onto?

(d) Why does a model of the axioms, built inside set theory, show that the axioms of a complete ordered field cannot contradict each other unless that set theory does?

Solution.

(a) (i) No. Write b and c for the non-negative square roots of 2 and 3 in F, which exist and are unique by Proposition 6.1.4. Then $\varphi(b) \geq 0$ by (P3) and $\varphi(b)^2 = \varphi(b^2) = \varphi(2) = 2$ by (P5), so $\varphi(b)$ is the non-negative square root of 2 in L, by the uniqueness in Proposition 6.1.4; likewise for c. By (P3) and (P5), $b + c > 3$ if and only if $\varphi(b) + \varphi(c) > 3$. (The statement is in fact true: both sides are positive, $(b + c)^2 = 5 + 2bc$, and bc is the non-negative square root of 6, so $5 + 2bc > 9$ if and only if $bc > 2$, if and only if $6 > 4$, by Proposition 2.3.5 (g).)

(ii) Yes. In the Cauchy model, $\sqrt{2} = r$ is a class of sequences $\mathbb{N} \to \mathbb{Q}$ (Example (The Class of the Newton Sequence)). By Sets, Relations, and Cardinality, Definition 1.4.1 each such sequence is the pair $(G, \mathbb{Q})$ of its graph and its codomain, that is, the set $\{\{G\}, \{G, \mathbb{Q}\}\}$ of Definition 1.2.1, with at most two elements. The rational 0 is the class $0_{\mathbb{Q}} = \{(0, s) : s \in \mathbb{Z}_{>0}\}$ of Definition 2.4.1; its elements $(0, 1)$, $(0, 2)$, $(0, 3)$ are pairwise different by Theorem 1.2.1, so it has more than two elements. So $0_{\mathbb{Q}}$ is not a sequence, and $0_{\mathbb{Q}} \notin r$.
Now apply the construction above with $F \coloneqq \mathbb{R}$, $X \coloneqq (\mathbb{R} \setminus \{r\}) \cup \{\{0_{\mathbb{Q}}\}\}$, $\beta(r) \coloneqq \{0_{\mathbb{Q}}\}$ and $\beta(x) \coloneqq x$ otherwise. Every element of $\mathbb{R}$ is a class $[a]$, which contains the sequence a; the only element of $\{0_{\mathbb{Q}}\}$ is $0_{\mathbb{Q}}$, which is not a sequence. So $\{0_{\mathbb{Q}}\} \notin \mathbb{R}$, and β is a bijection. By the argument of (i) with $\varphi \coloneqq \beta$, the non-negative square root of 2 in the complete ordered field X is $\beta(r) = \{0_{\mathbb{Q}}\}$, which contains the rational 0. (In the cut model of the optional Chapter 7, $\sqrt{2}$ contains $0_{\mathbb{Q}}$ as well.)

(iii) No. A set is countable exactly when it injects into $\mathbb{N}$ (Sets, Relations, and Cardinality, Theorem 3.2.2). If L were countable, an injection $g\colon L \to \mathbb{N}$ composed with the bijection φ would give an injection $g \circ \varphi\colon F \to \mathbb{N}$; and symmetrically with φ^{-1}. So F is countable exactly when L is.

(b) False. "No x has $x \cdot x = 1 + 1$" is true in $\mathbb{Q}$ (Theorem 2.6.1) and false in the ordered field $\mathbb{Q}(\sqrt{2})$ (Proposition 6.4.5), which contains $\sqrt{2} = 0 + 1\sqrt{2}$ (Definition 6.4.2) with $\sqrt{2} \cdot \sqrt{2} = 2 = 1 + 1$ by the choice of $\sqrt{2}$ (Proposition 6.1.4). A statement proved from the ordered-field axioms is true in every ordered field (for these first-order statements this is Logic and Proof, Theorem 7.3.3), so this one is not provable from them.

(c) The first: the tolerance ε is given in L, and the hypothesis $a_n \to a$ speaks only about tolerances in F. Surjectivity provides $\varepsilon' = \varphi^{-1}(\varepsilon) \in F$.

(d) A contradiction derived from the axioms could be run in the model, where it would become a contradiction about one set built in the set theory; so the set theory would be inconsistent. In this volume the model is the Cauchy model (Theorem 5.6.3) and the set theory is that of Chapter 1.

6.5.5.2 B. Practice

Exercise. B1

Let F and L be complete ordered fields and $\varphi\colon F \to L$ the order-preserving field isomorphism. Let $a \in F$ with $a \geq 0$, and let b be the non-negative square root of a. Show that $\varphi(a) \geq 0$ and that $\varphi(b)$ is the non-negative square root of $\varphi(a)$.

Solution. Since $0 \leq a$, (P1) and the order-preservation of φ (or equality) give $0 = \varphi(0) \leq \varphi(a)$. So $\varphi(a)$ has exactly one non-negative square root in L, by Proposition 6.1.4 applied in the complete field L. Now $\varphi(b) \geq \varphi(0) = 0$ in the same way, and $\varphi(b)^2 = \varphi(b^2) = \varphi(a)$, since φ respects products. By the uniqueness in Proposition 6.1.4, $\varphi(b)$ is the non-negative square root of $\varphi(a)$.

The argument used only that φ is an order-preserving field map into a complete ordered field; it needs neither bijectivity nor completeness of F, beyond the existence of b.

Exercise. B2

Let F, L and φ be as in B1, and let (a_n) be a sequence in F. Prove that (a_n) is a Cauchy sequence in F (Definition 4.5.1) if and only if $(\varphi(a_n))$ is a Cauchy sequence in L.

Solution. ($\Rightarrow$) Let $\varepsilon \in L_{>0}$, and put $\varepsilon' := \varphi^{-1}(\varepsilon)$. Since $\varphi(\varepsilon') = \varepsilon > 0 = \varphi(0)$, and φ reflects order, $\varepsilon' > 0$. Since (a_n) is Cauchy, there is $N \in \mathbb{N}$, depending on ε, with $|a_m - a_n| < \varepsilon'$ for all $m, n \geq N$. For such m, n,

$$|\varphi(a_m) - \varphi(a_n)| = \varphi(|a_m - a_n|) < \varphi(\varepsilon') = \varepsilon,$$

by (P1), (P4) and (P3). So $(\varphi(a_n))$ is Cauchy in L.

($\Leftarrow$) The inverse φ^{-1} is an order-preserving field isomorphism $L \to F$ (P6), so the same argument with the roles of F and L exchanged shows that $(\varphi^{-1}(\varphi(a_n))) = (a_n)$ is Cauchy in F.

6.5.5.3 C. Going deeper

Exercise. C1: Which hypothesis does the work

(a) Prove that there is no map $\sigma\colon \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}$ with $\sigma(1) = 1$ that respects $+$ and $\cdot$. In particular $\mathbb{Q}$ and $\mathbb{Q}(\sqrt{2})$ are not isomorphic, even as fields.

(b) Prove that there is no order-preserving field isomorphism $\mathbb{R} \to \mathbb{R}(x)$.

(c) For each of the two pairs, say which hypothesis of Theorem 6.3.1 fails, and give a statement phrased with $+$, $\cdot$ and $<$ that is true in one field of the pair and false in the other.

Solution (Hint 1). In (a), ask where σ sends $\sqrt{2}$. In (b), apply the inverse map to the element x.

Solution (Hint 2). In (b), the inverse $\chi^{-1}\colon \mathbb{R}(x) \to \mathbb{R}$ of a supposed isomorphism χ is an order-preserving field map, and it fixes the copies of $\mathbb{N}$.

Solution.

(a) Suppose σ is such a map. Then $\sigma(2) = \sigma(1+1) = 1+1 = 2$, and

$$\sigma(\sqrt{2})^2 = \sigma(\sqrt{2}\cdot\sqrt{2}) = \sigma(2) = 2,$$

so the rational $\sigma(\sqrt{2})$ squares to 2, against Theorem 2.6.1. A field isomorphism $\mathbb{Q}(\sqrt{2}) \to \mathbb{Q}$ would be such a map, so there is none. A field isomorphism $\mathbb{Q} \to \mathbb{Q}(\sqrt{2})$ would have such a map as its inverse (the inverse respects 1, + and ·, by the argument for the symmetric clause in the remark after Corollary 6.3.2, which does not use the order), so there is none in that direction either.

(b) Suppose $\chi\colon \mathbb{R} \to \mathbb{R}(x)$ is an order-preserving field isomorphism. Its inverse $\psi := \chi^{-1}\colon \mathbb{R}(x) \to \mathbb{R}$ is an order-preserving field map (the remark after Corollary 6.3.2). By Proposition 6.1.1, $\psi(n\cdot 1) = \psi(\iota_{\mathbb{R}(x)}(n)) = \iota_{\mathbb{R}}(n) = n\cdot 1_{\mathbb{R}}$ for every $n \in \mathbb{N}$ (Theorem 4.2.2). In $\mathbb{R}(x)$, $n\cdot 1 < x$ for every $n \in \mathbb{N}$ (Proposition 6.4.4). Applying the strictly increasing ψ, $n\cdot 1_{\mathbb{R}} < \psi(x)$ for every $n \in \mathbb{N}$. But $\mathbb{R}$ is complete (Theorem 5.6.3), hence Archimedean (Theorem 6.1.2), so some n has $\psi(x) < n\cdot 1_{\mathbb{R}}$ (Definition 4.2.2), and (O1) is violated. So χ does not exist.

(c) $\mathbb{Q}$ *and* $\mathbb{Q}(\sqrt{2})$. Both are Archimedean ordered fields (Theorem 2.5.3, Proposition 6.4.5), and **neither is complete**: $\mathbb{Q}(\sqrt{2})$ by Proposition 6.4.6, and $\mathbb{Q}$ because a complete field would contain a square root of 2 (Proposition 6.1.4), against Theorem 2.6.1. The hypothesis that fails is completeness, for both fields. Separating statement: "some x has $x\cdot x = 1+1$", true in $\mathbb{Q}(\sqrt{2})$, false in $\mathbb{Q}$.

$\mathbb{R}$ *and* $\mathbb{R}(x)$. $\mathbb{R}$ is complete, and $\mathbb{R}(x)$ is not (Proposition 6.4.4): completeness fails for one field of the pair. Separating statement: "some y satisfies $n\cdot 1 < y$ for every $n \in \mathbb{N}$", true in $\mathbb{R}(x)$, false in $\mathbb{R}$.

In both pairs the fields are ordered fields; Archimedes alone, as in the first pair, does not rescue the conclusion of Theorem 6.3.1.

Marking scheme. (10 points) - 3: (a), $\sigma(2) = 2$ from additivity and $\sigma(1) = 1$, then $\sigma(\sqrt{2})^2 = 2$ and Theorem 2.6.1. - 4: (b), Archimedes or completeness carried across the supposed isomorphism, either through χ^{-1} (an order-preserving field map that fixes every $n\cdot 1$) or through χ and its surjectivity, with the reason $\mathbb{R}$ is complete and Archimedean, and the contradiction with Proposition 6.4.4. - 3: (c), the failing hypothesis named for each pair with its source, and a correct separating statement for each. Fatal errors (0 for the step): in (a), assuming σ preserves order or is onto; in (b), assuming (rather than deriving from the supposed isomorphism) that $\mathbb{R}(x)$ is Archimedean, or asserting "$\mathbb{R}(x)$ is not complete, so not isomorphic" without showing that completeness or Archimedes is carried by the isomorphism.

6.5.5.4 R. Review

Exercise. R1

In the field $\mathbb{R}$ of classes of Cauchy sequences of rationals, for $n \in \mathbb{N}$ let k_n be the largest

natural number k with $k^2 \le 2(2^n)^2$, and let $q_n := k_n/2^n \in \mathbb{Q}$.

(a) Show that k_n exists and that $1 \le q_n < 2$.

(b) Show that $(q_n) \in \mathcal{C}$.

(c) Show that $[(q_n)]^2 = 2$ in $\mathbb{R}$, and that $[(q_n)]$ is the non-negative square root of 2 in $\mathbb{R}$.

Solution. This is the binary version of Example (The Truncations and the Newton Sequence Are Equivalent), and every estimate is in $\mathbb{Q}$. Here 2^n is the power $(2\cdot 1_{\mathbb{Q}})^n$ of Definition 4.2.1; by Proposition 4.2.1 (c), (d) it is the natural number 2^n, and $2^n \ge n+1 \ge 1$ by Lemma 5.6.1 with $F = \mathbb{Q}$. The first values are $q_0, \dots, q_5 = 1, 1, 5/4, 11/8, 11/8, 45/32$.

(a) Let $S_n := \{m \in \mathbb{N} : m^2 > 2(2^n)^2\}$. It is bounded below by 0, and non-empty: $(2\cdot 2^n)^2 = 4(2^n)^2 > 2(2^n)^2$, as $(2^n)^2 > 0$. By Proposition 2.3.6 it has a least element m^*. Since $0 \notin S_n$, $m^* \ge 1$ by discreteness (Theorem 2.3.4 (c)), so $k_n := m^* - 1 \in \mathbb{N}$. It is not in S_n, so $k_n^2 \le 2(2^n)^2$; and every natural $k > k_n$ has $k \ge m^*$, so $k^2 \ge (m^*)^2 > 2(2^n)^2$ (Proposition 2.3.5 (g)). So k_n is the largest such k, and it is unique. In particular

$$\left(q_n + \frac{1}{2^n}\right)^2 = \frac{(k_n+1)^2}{(2^n)^2} > 2, \qquad q_n^2 \le 2. \tag{$*$}$$

Since $k = 2^n$ has $k^2 \le 2(2^n)^2$, $k_n \ge 2^n$ and $q_n \ge 1$. Since $q_n \ge 0$ and $q_n^2 \le 2 < 4 = 2^2$, $q_n < 2$ (Proposition 2.3.5 (g)).

(b) *Monotone.* By $2^{n+1} = 2^n \cdot 2$ (Definition 4.2.1 (b)), $(2k_n)^2 = 4k_n^2 \le 8(2^n)^2 = 2(2^{n+1})^2$, so $2k_n \le k_{n+1}$ and $q_{n+1} = k_{n+1}/2^{n+1} \ge 2k_n/(2^n\cdot 2) = q_n$. By Logic and Proof, Axiom 4.2.1 (base $P(0)$: $q_n \le q_{n+0}$), $q_n \le q_{n+i}$ for every $i \in \mathbb{N}$, and every $m \ge n$ is $n+i$ for some i (Proposition 1.6.3 (a)).
A gap bound. Let $m \ge n$. If $q_m \ge q_n + 1/2^n$, then, both sides being non-negative, $q_m^2 \ge (q_n + 1/2^n)^2 > 2$ by Proposition 2.3.5 (g) and $(*)$, against $q_m^2 \le 2$. So, with Proposition 2.5.2 (b) and $2^n \ge n+1$,

$$0 \le q_m - q_n < \frac{1}{2^n} \le \frac{1}{n+1} \qquad (m \ge n).$$

Cauchy. Let $\varepsilon \in \mathbb{Q}_{>0}$. Since $\mathbb{Q}$ is Archimedean (Theorem 2.5.3), Proposition 4.2.3 with $F = \mathbb{Q}$ gives $N \in \mathbb{Z}_{>0}$, depending on ε, with $1/j < \varepsilon$ for every $j \ge N$. For $m, n \ge N$, say $m \ge n$ (the other case is symmetric), $|q_m - q_n| < 1/(n+1) < \varepsilon$, since $n+1 \ge N$. So $(q_n) \in \mathcal{C}$ (Definition 5.1.1).

(c) Expanding $(*)$, $2 < q_n^2 + 2q_n/2^n + 1/(2^n)^2$. With $q_n < 2$ and $0 < 1/2^n \le 1$,

$$0 \le 2 - q_n^2 < \frac{1}{2^n}\left(2q_n + \frac{1}{2^n}\right) \le \frac{5}{2^n} \le \frac{5}{n+1}.$$

The sequence $(1/(n+1))$ is null in $\mathbb{Q}$ (Example (First Limits) (b)), so $(5/(n+1))$ is null (Theorem 4.4.3 (b)). By the squeeze Proposition 4.4.4 (c), between $(0)_n$ and $(5/(n+1))$, the sequence $(2-q_n^2)$ is null, and so is $(q_n^2 - 2)$ (Theorem 4.4.3 (b), $c = -1$). Both (q_n^2) and $(2)_n$ are in $\mathcal{C}$ (Theorem 4.5.3 (a), (c)), so $(q_n^2) \sim (2)_n$ (Definition 5.1.2), and by Definition 5.2.2 and Definition 5.4.1,

$$[(q_n)]^2 = [(q_n^2)] = [(2)_n] = \kappa(2),$$

which is 2 in $\mathbb{R}$ under the identification of q with $\kappa(q)$. Since $q_n \ge 1$ for every n, Lemma 5.3.4 gives $[(q_n)] \ge [(1)_n] = 1_{\mathbb{R}} > 0$. So $[(q_n)]$ is a non-negative element of

$\mathbb{R}$ whose square is 2. The element r of Example (The Class of the Newton Sequence) is another, and $\mathbb{R}$ is complete (Theorem 5.6.3), so the uniqueness in Proposition 6.1.4 gives $[(q_n)] = r$: the binary truncations, the decimal truncations and the Newton sequence all name the same real number.

Exercise. R2

Show that no relation $<$ makes the field $\mathbb{Z}/5\mathbb{Z}$ an ordered field. More generally, show that if F is a field in which $n \cdot 1 = 0$ for some $n \in \mathbb{N}$ with $n \geq 1$, then no relation makes F an ordered field.

Solution. The general statement first. Suppose $<$ makes F an ordered field (Definition 2.3.2). Since $1 \neq 0$ in a field, Proposition 2.3.5 (h) gives $n \cdot 1 > 0$ for every $n \geq 1$. If $n \cdot 1 = 0$, this says $0 > 0$, against (O1). So no such relation exists.

$\mathbb{Z}/5\mathbb{Z}$ is a field because 5 is prime (Theorem 3.5.2), and $5 \cdot [1] = [5] = [0]$ by Theorem 3.4.3 (d). The general statement with $n = 5$ applies. (For the ring, this is the case $n = 5$ of Theorem 3.4.3 (e).) So $\mathbb{Z}/5\mathbb{Z}$ satisfies every field axiom and admits no order, while $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$ and $\mathbb{R}(x)$ admit one: the field axioms do not determine the order axioms.

Exercise. R3

Let X be the complete ordered field of this section whose zero is $\emptyset$, and let $\mathcal{S} := \{\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}(x), \mathbb{R}, X\}$, with $\mathbb{R}$ the Cauchy model. Isomorphism of ordered fields, $\cong$, is an equivalence relation on $\mathcal{S}$. Find its equivalence classes, and the number of elements of the quotient set $\mathcal{S}/\cong$. For each two fields in different classes, name a property that one has and the other lacks, and prove that an order-preserving field isomorphism carries it across.

Solution. By the remark after Corollary 6.3.2, $\cong$ is an equivalence relation on $\mathcal{S}$ (Sets, Relations, and Cardinality, Definition 2.5.1).

Two transfer facts. Let $\chi\colon F \to L$ be an order-preserving field isomorphism. By the remark after Corollary 6.3.2, $\chi^{-1}\colon L \to F$ is one too, so each fact below also holds with F and L exchanged. Recall $\chi(n \cdot 1_F) = \chi(\iota_F(n)) = \iota_L(n) = n \cdot 1_L$ for $n \in \mathbb{N}$ (Proposition 6.1.1).

(i) *Square roots of $n \cdot 1$.* If $b \in F$ has $b^2 = n \cdot 1_F$, then $\chi(b)^2 = \chi(b^2) = \chi(n \cdot 1_F) = n \cdot 1_L$, since χ respects products. So if $n \cdot 1$ is a square in F, it is a square in L, and conversely.

(ii) *Archimedes.* Let F be Archimedean (Definition 4.2.2), and let $y \in L$. Take $n \in \mathbb{N}$ with $\chi^{-1}(y) < n \cdot 1_F$. Since χ is strictly increasing, $y = \chi(\chi^{-1}(y)) < \chi(n \cdot 1_F) = n \cdot 1_L$. So L is Archimedean, and conversely.

The classes. The map $\beta\colon \mathbb{R} \to X$ of this section is an order-preserving field isomorphism, so $\mathbb{R} \cong X$. The other pairs are not isomorphic:

- $\mathbb{R}(x)$ is not Archimedean (Proposition 6.4.4), while $\mathbb{Q}$ (Theorem 2.5.3), $\mathbb{Q}(\sqrt{2})$ (Proposition 6.4.5), $\mathbb{R}$ and X (complete, by Theorem 5.6.3 and by the construction of X, hence Archimedean by Theorem 6.1.2) are. By (ii), $\mathbb{R}(x)$ is isomorphic to none of the four.

- 2 is not a square in $\mathbb{Q}$ (Theorem 2.6.1), while it is one in $\mathbb{Q}(\sqrt{2})$ ($\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, Definition 6.4.2) and in the complete fields $\mathbb{R}$ and X (Proposition 6.1.4). By (i) with $n = 2$, $\mathbb{Q}$ is isomorphic to none of these three.
- 3 is not a square in $\mathbb{Q}(\sqrt{2})$ (Proposition 6.4.6), while it is one in $\mathbb{R}$ and X (Proposition 6.1.4). By (i) with $n = 3$, $\mathbb{Q}(\sqrt{2}) \ncong \mathbb{R}$ and $\mathbb{Q}(\sqrt{2}) \ncong X$.

So the equivalence classes (Sets, Relations, and Cardinality, Definition 2.5.2), $[K] = \{K' \in \mathcal{S} : K' \cong K\}$, are $\{\mathbb{Q}\}$, $\{\mathbb{Q}(\sqrt{2})\}$, $\{\mathbb{R}(x)\}$ and $\{\mathbb{R}, X\}$, and the quotient set $\mathcal{S}/\!\cong$ (Sets, Relations, and Cardinality, Definition 2.5.4) has four elements. The two complete fields share a class, as Theorem 6.3.1 requires; the other three fall into three classes.

Exercise. R4

Let F be a complete ordered field, and for $n \in \mathbb{N}$ write 10^{-n} for $\left((10 \cdot 1_F)^n\right)^{-1}$. Prove that

$$\sup\{1 - 10^{-n} : n \in \mathbb{N}\} = 1.$$

Solution. The tool is the Archimedean property of F, which holds by Theorem 6.1.2, together with a lower bound for the powers $(10 \cdot 1_F)^n$ of Definition 4.2.1.

The bound. We show $(10 \cdot 1_F)^n \geq (1 + 9n) \cdot 1_F$ for every $n \in \mathbb{N}$, by Logic and Proof, Axiom 4.2.1 with base $P(0)$: $(10 \cdot 1_F)^0 = 1_F = (1 + 0) \cdot 1_F$. For the step, $10 \cdot 1_F > 0$ (Proposition 4.2.1 (b)), so by Definition 4.2.1 (b), $P(n)$, Proposition 2.3.5 (c) (or equality), and Proposition 4.2.1 (a), (b),

$$\begin{aligned}(10 \cdot 1_F)^{n+1} &= (10 \cdot 1_F)^n (10 \cdot 1_F) \geq ((1 + 9n) \cdot 1_F)(10 \cdot 1_F) \\ &= (10 + 90n) \cdot 1_F \geq (10 + 9n) \cdot 1_F = \big(1 + 9(n + 1)\big) \cdot 1_F,\end{aligned}$$

using $10 + 9n \leq 10 + 90n$ in $\mathbb{N}$. In particular $(10 \cdot 1_F)^n \geq 1_F > 0$, so 10^{-n} is defined and positive (Proposition 2.5.2 (a)).

Upper bound. $10^{-n} > 0$ gives $1 - 10^{-n} < 1$ for every n.

Least. Let $b \in F$ with $b < 1$. Then $1 - b > 0$, and Proposition 4.2.3 gives $N \in \mathbb{Z}_{>0}$, depending on b, with $1/(N \cdot 1_F) < 1 - b$. Since $(10 \cdot 1_F)^N \geq (1 + 9N) \cdot 1_F \geq N \cdot 1_F > 0$ (Proposition 4.2.1 (b)), Proposition 2.5.2 (b) (or equality) gives $10^{-N} \leq 1/(N \cdot 1_F) < 1 - b$, so the element $1 - 10^{-N}$ of the set exceeds b, and b is not an upper bound. By (O1), every upper bound is ≥ 1. So the supremum is 1.

Why the look-alike tools fail. Reading $1 - 10^{-n}$ as 0.9, 0.99, ... and summing a geometric series uses infinite decimals, which this volume never defined; and an argument by limits would need the limit of (10^{-n}), which is the same Archimedean estimate with more machinery. The estimate is also where Archimedes cannot be skipped: in the non-Archimedean $\mathbb{R}(x)$ of Definition 6.4.1, $(10 \cdot 1)^n = (10^n) \cdot 1$ (Proposition 4.2.1 (c)) lies below x for every n (Proposition 6.4.4), so $10^{-n} > 1/x > 0$ (Proposition 2.5.2), and $1 - 1/x$ is an upper bound of the set that lies below 1.

Chapter 7

Constructing the Real Numbers by Dedekind Cuts

This chapter is optional: nothing outside it depends on it, and the volume's main line ended with Chapter 6. Chapter 5 built a real number as a class of Cauchy sequences of rationals, many rational approximations identified when they agree in the limit. Dedekind's idea, published in 1872, is shorter to state: a real number is the set of all rationals **below** it. The gap that Section 2.6 found in $\mathbb{Q}$ splits the rationals into a lower piece, those that are negative or have square below 2, and the rest, and Dedekind takes the lower piece **as** the square root of 2. The chapter makes this precise. It says which sets of rationals qualify (the cuts), orders them by inclusion, adds and multiplies them, and proves that they form a complete ordered field $\mathbb{R}_{\mathrm{D}}$ in which the rationals sit as the rational cuts and the cut of the gap squares to 2.

The question behind the chapter is: *if two people build the real numbers in different ways, in what sense do they have the same numbers?* On this road the costs fall the other way round from Chapter 5. The least upper bound of a bounded set of cuts is its union, checked in a few lines with no arithmetic at all, while negatives need a margin and multiplication needs a detour through positive cuts and signs; on the Cauchy road the arithmetic was termwise and completeness took two sections. The last section applies the uniqueness theorem of Chapter 6 to the two models. There is exactly one order-preserving field isomorphism from Chapter 5's field to the cuts, and it can be written down: it sends a Cauchy real to the set of rationals below it. So the two roads give the same real numbers, and Dedekind's definition is what the Cauchy model sees when it looks at one of its elements through the rationals.

What you need. Chapter 2 above all, and Chapter 4; for Sections 7.5 and 7.6, Chapters 5 and 6; *Sets, Relations, and Cardinality*, Chapters 1 and 2, chiefly Section 2.6; and *Logic and Proof*, Chapters 4 and 5. From Chapter 2: $\mathbb{Q}$ as an ordered field, with the axioms (F1)–(F9) and (O1)–(O4) and their rules (Theorem 2.5.1, Definition 2.3.1, Definition 2.3.2, Proposition 2.3.2, Proposition 2.3.5, Proposition 2.5.2); $\mathbb{Q}$ Archimedean (Theorem 2.5.3, Corollary 2.5.4); least elements and discreteness of the integers, for the walk up by ε (Proposition 2.3.6, Theorem 2.3.4); and the gap at $\sqrt{2}$ (Theorem 2.6.1, Theorem 2.6.2). From Chapter 4: halves and fractions in an ordered field and the absolute value (Proposition 4.1.2, Lemma 4.1.1, Definition 4.1.1); the one copy ι_F of $\mathbb{Q}$ in an ordered field, with its uniqueness clause (Theorem 4.2.2); the Archimedean property and density of $\mathbb{Q}$ (Definition 4.2.2, Theorem 4.2.5); and, for exercises of Section 7.6, limits and Cauchy sequences (Definition 4.3.2, Definition 4.5.1). From Chapter 5: complete ordered fields (Definition 5.6.1), the Cauchy model and the theorem that it is one (Definition 5.2.1, Theorem 5.6.3), its copy of $\mathbb{Q}$ (Theorem 5.4.1), and the Newton class whose square is 2 (Example (The Class of the Newton Sequence)). From Chapter 6: order-preserving field maps and

the rationals below an element (Definition 6.1.1, Definition 6.1.2, Proposition 6.1.1), and the rationals below a sum (Lemma 6.2.1), for the remark after the proof in Section 7.6 and Exercise B1; the uniqueness theorem (Theorem 6.3.1), which Section 7.6 applies; that complete ordered fields are Archimedean (Theorem 6.1.2), for the second proof in Section 7.5; and the embedding of an Archimedean field into a complete one, with its property ($\star$) and uniqueness clause (Theorem 6.2.3), together with the fold after Corollary 6.3.2 on inverses of isomorphisms, for the inverse map in Section 7.6. Chapter 1 supplies the axioms that make the set of all cuts a set (Axiom 1.1.1). From *Sets, Relations, and Cardinality*: power sets, unions of families and inclusion, Sections 1.1 and 1.2 (Sets, Relations, and Cardinality, Definition 1.2.2, Sets, Relations, and Cardinality, Definition 1.2.4, Sets, Relations, and Cardinality, Proposition 1.1.1, Sets, Relations, and Cardinality, Proposition 1.1.2); sums of sets of numbers, Section 1.6 (Sets, Relations, and Cardinality, Example (Reflections, scalings and sums as images)); partial orders, upper bounds and suprema, Section 2.6 (Sets, Relations, and Cardinality, Definition 2.6.1, Sets, Relations, and Cardinality, Definition 2.6.2, Sets, Relations, and Cardinality, Definition 2.6.5); and, for one review exercise, countable sets, Chapter 3 (Sets, Relations, and Cardinality, Theorem 3.2.2). From *Logic and Proof*: the quantifier rules and the negation of quantified statements, Chapter 4 (Logic and Proof, Definition 5.4.1, Logic and Proof, Definition 5.5.1, Logic and Proof, Definition 5.6.1, Logic and Proof, Proposition 6.2.3, Logic and Proof, Proposition 6.2.2), and the catalogue of illegal steps, Chapter 5, for the find-the-error exercises (Logic and Proof, Example (What is not allowed)).

Time. About 25 hours on the core track (all of A, one B and one C per section, the exit checks); about 30 hours for everything.

7.0.1 Roadmap

- **Cuts.** The set below the gap at $\sqrt{2}$; the definition of a cut by three conditions, non-empty and not all of $\mathbb{Q}$, closed downward, no largest element, and the set $\mathbb{R}_{\mathrm{D}}$ of all cuts; the rational cuts D_q and the cut σ of the gap; why each condition is there, with a set that has a largest element as the cautionary case; and the two lemmas the rest of the chapter rests on, separation of inside and outside and the walk up by ε.
- **Order by Inclusion, and Least Upper Bounds by Unions.** Cuts ordered by inclusion, which recovers the order of $\mathbb{Q}$ on rational cuts and is total because of separation; and the theorem that a non-empty set of cuts bounded above has a supremum, its union, with each hypothesis shown to be needed.
- **Addition and Negatives.** The sum of two cuts as the set of sums, and the zero cut D_0; the natural recipe for a negative, which fails on the rational cuts, and its repair by a positive margin; the key proof that a cut plus its negative is D_0, by walking up; and the cuts as an ordered abelian group.
- **Multiplication.** Why the set of all products fails; the product of positive cuts, its laws and the walk up in multiplicative form; the extension to all cuts by absolute value and sign, through one lemma of sign rules; distributivity, the reciprocal, and the theorem that the cuts form an ordered field.
- **The Rationals Inside the Cuts, and a Complete Ordered Field.** The rational cuts as the copy of $\mathbb{Q}$, with its order and arithmetic; density of the rational cuts and the Archimedean property; the cut σ squares to D_2 and is not rational; and the cuts as a complete ordered field, by the union theorem of Section 7.2.
- **The Two Roads Give the Same Real Numbers.** The unique order-preserving field isomorphism from Chapter 5's field to the cuts, computed explicitly as "the rationals below", and its inverse as a supremum; the two square roots of 2 matched; a table of what each construction pays for; and review exercises that return to separation by rationals (Chapter 6), limits and Cauchy sequences in an ordered field (Chapter 4), and countability

(*Sets, Relations, and Cardinality*, Chapter 3).

7.0.2 Named moves

- **Walk up by ε.** From a point inside a cut, add ε until you leave; the last step inside and the first step outside are ε apart (Lemma 7.1.3, Section 7.1). It proves the hard $\supseteq$ halves of the chapter: in full for the additive inverse in Section 7.3, and in a multiplicative form (Lemma 7.4.2) for the reciprocal in Section 7.4 and for $\sigma\sigma = D_2$ in Section 7.5. Exercise B2 of Section 7.4 leaves you its last step, and Exercise C2 of Section 7.5 (a cube root of 2) its last half.
- **A cut is a downset with an open top.** Inside is closed downward, outside is closed upward, everything inside lies below everything outside, and the inside has no largest element (Lemma 7.1.1, Section 7.1). Most "is it a cut?" checks are this sentence read clause by clause; it also makes inclusion a total order in Section 7.2 and bounds sums and products from above in Sections 7.3 and 7.4.
- **Reduce to the positive case.** Define and prove for positive cuts, then extend to all cuts by signs once, through one lemma (Lemma 7.4.4, Section 7.4), never axiom by axiom. It gives the product and the reciprocal of Section 7.4, distributivity by moving the negative summand to the other side, and the products of rational cuts in Section 7.5.

7.1 Cuts

This section opens Chapter 7, which is optional: nothing outside the chapter depends on it, and the volume's main line ended with Chapter 6. Chapter 5 built a real number as a class of Cauchy sequences of rationals, and Chapter 6 proved that any two complete ordered fields are linked by exactly one order-preserving field isomorphism. Here we begin a second construction, Dedekind's, in which a real number is a single set of rationals: the set of all rationals **below** it. This section defines these sets, called cuts, tests the definition on the rationals and on the gap that Section 2.6 found, and proves two lemmas on which the rest of the chapter rests. The inside and the outside of a cut are separated, and one can walk up through a cut in steps of any fixed size until one steps out.

Throughout the chapter, lower-case Latin letters such as p, q, r, s stand for rationals and m, n, k for natural numbers; Section 7.6, which brings back the field of Chapter 5, says where that changes. Nothing about real numbers is assumed: they are what is being built.

7.1.1 The set below the gap

Section 2.6 proved that the set $A := \{x \in \mathbb{Q} : x^2 < 2\}$ is non-empty and bounded above in $\mathbb{Q}$, and has no supremum in $\mathbb{Q}$ (Theorem 2.6.2 (a)). Its rational upper bounds crowd down toward a place where $\mathbb{Q}$ has no element. Chapter 5 filled that place with a new object, a class of sequences. Dedekind noticed that something already present in $\mathbb{Q}$ describes the place exactly: the set of rationals that lie below it.

That set is not quite A. The set A contains -1, since $(-1)^2 = 1 < 2$, but not -2, since $(-2)^2 = 4 > 2$; it is a window of rationals around 0, not "everything below" anything. Adding the non-positive rationals repairs this. Let

$$\sigma := \{p \in \mathbb{Q} : p \leq 0 \text{ or } p^2 < 2\}.$$

This set has the same upper bounds in $\mathbb{Q}$ as A. It contains A, so every upper bound of σ bounds A. Conversely, let u be an upper bound of A; since $1 \in A$, $u \geq 1 > 0$, and every element p of σ

outside A has $p \le 0 < u$. The conditions (S1) and (S2) of a supremum mention only the set of upper bounds (Sets, Relations, and Cardinality, Definition 2.6.5), so σ, like A, has no supremum in $\mathbb{Q}$; this is the argument of Theorem 2.6.2 (b).

So no rational sits at the top of σ, yet σ itself is a perfectly good set of rationals. Dedekind's move is to **take the set as the number**. The rest of the chapter makes this precise: it says which sets of rationals qualify, orders them, adds and multiplies them, and proves that they form a complete ordered field.

7.1.2 Dedekind cuts

Which subsets of $\mathbb{Q}$ should count as "the set of rationals below a number"? Three features of σ suggest the answer. It is neither empty nor all of $\mathbb{Q}$: a number should have rationals below it and rationals above it. With any rational it contains every smaller one. And it has no largest element, as Example (The cut of the gap) below shows. The third feature is the least obvious, and the paragraph "Why this definition" after the examples explains it.

A real number is the set of rationals below it.

Definition 7.1.1. Dedekind Cut

A subset $\alpha \subseteq \mathbb{Q}$ is a **cut** (or **Dedekind cut**) if

(C1) $\alpha \neq \emptyset$ and $\alpha \neq \mathbb{Q}$;
(C2) **for every** $p \in \alpha$ and **every** $q \in \mathbb{Q}$ with $q < p$, $q \in \alpha$;
(C3) α has **no largest element**: **for every** $p \in \alpha$ **there exists** $p' \in \alpha$ with $p' > p$.

The set of all cuts is

$$\mathbb{R}_{\mathrm{D}} := \{\alpha \in \mathcal{P}(\mathbb{Q}) : \alpha \text{ satisfies (C1), (C2) and (C3)}\},$$

a subset of the power set $\mathcal{P}(\mathbb{Q})$ (Sets, Relations, and Cardinality, Definition 1.2.2).

In symbols, with bounded quantifiers,

$$\begin{aligned} &\text{(C1)} \quad \exists p \in \mathbb{Q}\ (p \in \alpha) \ \wedge\ \exists q \in \mathbb{Q}\ (q \notin \alpha), \\ &\text{(C2)} \quad \forall p \in \alpha\ \forall q \in \mathbb{Q}\ (q < p \Rightarrow q \in \alpha), \\ &\text{(C3)} \quad \forall p \in \alpha\ \exists p' \in \alpha\ (p' > p). \end{aligned}$$

In words, clause by clause. (C1) says that the cut is a proper, non-empty part of $\mathbb{Q}$: some rational is in it and some rational is not. (C2) says that the cut is **closed downward**: whatever rational it contains, it contains every smaller one too. (C3) says that it has **no top**: above each of its elements it has another. The letters $\alpha, \beta, \gamma, \delta$ are reserved for cuts in this chapter.

Well-definedness. The collection $\mathbb{R}_{\mathrm{D}}$ is a set: it is carved out of the set $\mathcal{P}(\mathbb{Q})$ by a condition, which (Z6) of Axiom 1.1.1 allows, and $\mathcal{P}(\mathbb{Q})$ exists by (Z5). A real number in the sense of this chapter is therefore literally an element of $\mathcal{P}(\mathbb{Q})$, a subset of $\mathbb{Q}$.

Negations. We shall often show that a set is **not** a cut, so the three negations are worth writing out once. By Logic and Proof, Proposition 6.2.3 and Logic and Proof, Proposition 6.2.2, a subset $\alpha \subseteq \mathbb{Q}$ fails

- (C1) if $\alpha = \emptyset$ or $\alpha = \mathbb{Q}$;

- (C2) if $\exists p \in \alpha\ \exists q \in \mathbb{Q}\ (q < p \wedge q \notin \alpha)$: some element has a smaller rational missing;
- (C3) if $\exists p \in \alpha\ \forall p' \in \alpha\ (p' \leq p)$: α has a **largest element**.

In the last line, "not $p' > p$" became $p' \leq p$ by trichotomy in $\mathbb{Q}$ (Proposition 2.3.5 (i)).

The first examples are the rationals themselves. For a rational q, the rationals below q form a cut, and this is how the old numbers will sit among the new ones.

Definition 7.1.2. Rational Cut

For $q \in \mathbb{Q}$, let

$$D_q := \{r \in \mathbb{Q} : r < q\}.$$

It is a cut (checked below), the **rational cut** of q. A cut α is **rational** if $\alpha = D_q$ for some $q \in \mathbb{Q}$.

D_q is a cut. (C1): $q - 1 \in D_q$, since $q - (q-1) = 1 > 0$ (Proposition 2.3.5 (e), as $1 \neq 0$ in the field $\mathbb{Q}$, Theorem 2.5.1), so $q - 1 < q$ by Proposition 2.3.5 (a); and $q \notin D_q$, since $q < q$ fails by (O1). (C2): if $p \in D_q$ and $r < p$, then $r < p < q$, so $r < q$ by (O2), and $r \in D_q$. (C3): if $r \in D_q$, the midpoint $(r+q)/2$ satisfies $r < (r+q)/2 < q$ (Section 2.5, "Between two rationals"), so it lies in D_q and is larger than r.

The letter D is for Dedekind.

Example. The cut of the gap

Show that $\sigma = \{p \in \mathbb{Q} : p \leq 0 \text{ or } p^2 < 2\}$ is a cut.

Solution. (C1): $0 \in \sigma$, since $0 \leq 0$. And $2 \notin \sigma$: $2 > 0$, and $2^2 = 4 > 2$.

(C2): let $p \in \sigma$ and $q \in \mathbb{Q}$ with $q < p$. If $q \leq 0$, then $q \in \sigma$. Otherwise $0 < q < p$. Then $p > 0$, so $p \leq 0$ fails and $p^2 < 2$. Since $0 \leq q < p$, Proposition 2.3.5 (g) gives $q^2 < p^2$, so $q^2 < 2$ by (O2), and $q \in \sigma$.

(C3): let $p \in \sigma$. If $p \leq 0$, then $1 \in \sigma$ (as $1^2 = 1 < 2$) and $1 > 0 \geq p$. If $p > 0$, then $p^2 < 2$. Put

$$d := 2 - p^2, \qquad h := \frac{d}{2p+2}.$$

Here $d > 0$, and $d < 2$ because $p^2 > 0$ (Proposition 2.3.5 (e)); also $2p + 2 > 2$, since $p > 0$. So $0 < h < 1$ (Proposition 2.5.2 (a) for $h > 0$; $d < 2p+2$ for $h < 1$). From $0 < h < 1$, $h^2 < h$ by Proposition 2.3.5 (c). Hence

$$\begin{aligned}(p+h)^2 &= p^2 + 2ph + h^2 < p^2 + 2ph + h \\ &< p^2 + h(2p+2) = p^2 + d = 2,\end{aligned}$$

where the second inequality adds $h > 0$ (O3). So $p + h \in \sigma$, and $p + h > p$.

The step size h depends on p: the closer p^2 is to 2, the smaller d and so h. That is unavoidable, since a fixed step would jump out of σ from points near the top. Section 7.5 proves that σ is not a rational cut, $\sigma \neq D_q$ for every $q \in \mathbb{Q}$, and Exercise C1 below asks for a direct proof. So σ is a genuinely new number: the cut that will square to 2.

Quick check. Which of the following subsets of $\mathbb{Q}$ are cuts? For each that is not, name a condition that fails.

(a) $\{p \in \mathbb{Q} : p < 0 \text{ or } p^2 < 4\}$;
(b) $\{p \in \mathbb{Q} : p^2 < 2\}$;
(c) $\{p \in \mathbb{Q} : p^3 < 5\}$;
(d) $\emptyset$.

Solution.

(a) It is a cut, because it equals D_2. If $p < 0$, then $p < 2$. If $p \geq 0$ and $p^2 < 4 = 2^2$, then $p < 2$ by Proposition 2.3.5 (g). Conversely, if $p < 2$, either $p < 0$, or $0 \leq p < 2$ and then $p^2 < 4$ by (g).

(b) Not a cut: (C2) fails. The set contains 1, and $-2 < 1$, but $(-2)^2 = 4 > 2$, so -2 is missing.

(c) A cut. Call the set E. *Cubes are monotone on non-negatives:* for $0 \leq q < p$, $p^3 - q^3 = (p-q)(p^2+pq+q^2)$, where $p - q > 0$ and $p^2 + pq + q^2 \geq p^2 > 0$ (Proposition 2.3.5 (e), with $pq, q^2 \geq 0$); so $p^3 - q^3 > 0$ by (O4). (C1): $0 \in E$, and $2 \notin E$, as $2^3 = 8 > 5$. (C2): let $q < p \in E$. If $q \leq 0$, then $q^3 \leq 0 < 5$ ($q^3 = q \cdot q^2$, a non-positive times a non-negative). If $q > 0$, then $0 < q < p$ and $q^3 < p^3 < 5$. (C3): let $p \in E$. If $p \leq 0$, then $1 \in E$ is larger. If $p > 0$, put $d := 5 - p^3$ and $h := d/(3p^2 + 3p + 5)$. Then $0 < d < 5 < 3p^2 + 3p + 5$, so $0 < h < 1$, hence $h^3 < h^2 < h$ by Proposition 2.3.5 (c), and

$$\begin{aligned}(p+h)^3 &= p^3 + 3p^2h + 3ph^2 + h^3 \\ &< p^3 + h(3p^2 + 3p + 1) < p^3 + d = 5.\end{aligned}$$

So $p + h \in E$ and $p + h > p$.

(d) Not a cut: (C1) fails, since $\emptyset$ has no element.

Why this definition. Each clause excludes something that would spoil the arithmetic to come.

- Without (C1), $\emptyset$ and $\mathbb{Q}$ would be numbers, below and above every rational. Section 7.3 adds cuts as sets, and $\mathbb{Q} + D_0 = \mathbb{Q} = \mathbb{Q} + D_1$ there would make cancellation fail.
- Without (C2), $\{p : p^2 < 2\}$ of the Quick check would be a number. It contains 1 but not the smaller rational -2, so it is not the set of **all** rationals below anything.
- Without (C3), each rational q would be represented twice, by D_q and by $\{r \in \mathbb{Q} : r \leq q\}$. The next example shows that this is more than untidy: the sum of sets cannot tell the two apart.

Example. Two candidates for one rational

Optional: nothing later depends on this.

Let $E_1 := \{r \in \mathbb{Q} : r \leq 1\}$. Show that E_1 satisfies (C1) and (C2) but not (C3), and that, with the sum of sets $X + Y := \{x + y : x \in X,\ y \in Y\}$ of *Sets, Relations, and Cardinality*, Section 1.6 (Sets, Relations, and Cardinality, Example (Reflections, scalings and sums as

images)),

$$E_1 + D_1 = D_2 = D_1 + D_1, \qquad \text{although } E_1 \neq D_1.$$

Solution. The conditions. (C1): $0 \in E_1$ and $2 \notin E_1$. (C2): $q < p \leq 1$ gives $q < 1$. (C3) fails: $1 \in E_1$, and every $p' \in E_1$ has $p' \leq 1$, so 1 is a largest element. And $E_1 \neq D_1$, since $1 \in E_1 \setminus D_1$.

$E_1 + D_1 = D_2$. ($\subseteq$): if $a \leq 1$ and $b < 1$, then $a + b < 2$ by Proposition 2.3.5 (b). ($\supseteq$): if $s < 2$, then $s = 1 + (s - 1)$ with $1 \in E_1$ and $s - 1 < 1$, so $s - 1 \in D_1$.

$D_1 + D_1 = D_2$. ($\subseteq$): $a < 1$ and $b < 1$ give $a + b < 2$, by (b) again. ($\supseteq$): if $s < 2$, then $s = s/2 + s/2$, and multiplying $s < 2$ by $1/2 > 0$ (Proposition 2.3.5 (c)) gives $s/2 < 1$, so $s/2 \in D_1$.

So if E_1 were admitted as a number, the equation $\xi + D_1 = D_2$ would have two solutions, E_1 and D_1, and addition could not be canceled, so the numbers could not form a group under addition.

Condition (C3) chooses between the two sets $\{r \in \mathbb{Q} : r < 1\}$ and $\{r \in \mathbb{Q} : r \leq 1\}$ that describe "the rationals up to 1": it keeps the one that leaves 1 out.

Warning. A cut is a set of rationals, not a rational. "$\sigma < 2$" compares a set with a number and means nothing yet. Section 7.2 orders cuts, so that $\sigma < D_2$ acquires a meaning, and Section 7.5 shows that D_2 plays the part of the rational 2 among the cuts. Until then, every comparison in this chapter is between rationals, or between sets by inclusion.

7.1.3 Inside and outside

The defining conditions speak about the inside of a cut. The first lemma says what they imply about the outside: everything outside lies above everything inside.

Lemma 7.1.1. Separation

Let α be a cut, $p \in \alpha$ and $q \in \mathbb{Q} \setminus \alpha$. Then $p < q$. Consequently the complement $\mathbb{Q} \setminus \alpha$ is **closed upward**: if $q \notin \alpha$ and $q' \in \mathbb{Q}$ with $q' > q$, then $q' \notin \alpha$.

Proof. Suppose $p < q$ fails. By trichotomy (Proposition 2.3.5 (i)), $q \leq p$. If $q = p$, then $q \in \alpha$; if $q < p$, then $q \in \alpha$ by (C2). Either way $q \in \alpha$, against $q \notin \alpha$. So $p < q$.

For the consequence, let $q \notin \alpha$ and $q' > q$, and suppose $q' \in \alpha$. The first part, with q' in the role of p, gives $q' < q$, which together with $q < q'$ contradicts (O1). So $q' \notin \alpha$.

■

So a cut and its complement split $\mathbb{Q}$ into a lower piece and an upper piece, every element of the lower below every element of the upper. In particular every rational outside a cut is an upper bound of it. Together with (C3), this gives the picture to keep in mind, and one of the chapter's named moves:

A cut is a downset with an open top.

Inside is closed downward, outside is closed upward, everything inside lies below everything outside, and the inside has no largest element. The outside may have a least element: the complement of D_q is $\{r \in \mathbb{Q} : r \geq q\}$, with least element q. Most "is it a cut?" checks in this chapter are this sentence, read clause by clause.

The second result locates a rational relative to a cut by comparing two sets. It is the bridge between a rational q and its cut D_q.

Proposition 7.1.2. Where a Rational Lies

Let α be a cut and $q \in \mathbb{Q}$. Then

(a) $q \in \alpha$ if and only if $D_q \subsetneq \alpha$;
(b) $q \notin \alpha$ if and only if $\alpha \subseteq D_q$.

Idea. If q is in α, closure downward brings everything below q along, and q itself is the extra element. If q is outside, separation puts all of α below q. The two converses come from the same two facts.

Proof.

(a) ($\Rightarrow$) Let $q \in \alpha$. For $r \in D_q$, $r < q$, so $r \in \alpha$ by (C2); thus $D_q \subseteq \alpha$. The inclusion is proper, because $q \in \alpha$ and $q \notin D_q$ (as $q < q$ fails, (O1)). ($\Leftarrow$) Let $D_q \subsetneq \alpha$. Then some $r \in \alpha$ has $r \notin D_q$, that is, $r < q$ fails, so $q \leq r$ (Proposition 2.3.5 (i)). If $q = r$, then $q \in \alpha$; if $q < r$, then $q \in \alpha$ by (C2).

(b) ($\Rightarrow$) Let $q \notin \alpha$. For $r \in \alpha$, Lemma 7.1.1 gives $r < q$, so $r \in D_q$; thus $\alpha \subseteq D_q$. ($\Leftarrow$) Let $\alpha \subseteq D_q$. Since $q \notin D_q$, also $q \notin \alpha$.

■

Since every rational is either in α or not, exactly one of $D_q \subsetneq \alpha$ and $\alpha \subseteq D_q$ holds. Section 7.2 rewrites (a) as $q \in \alpha \iff D_q < \alpha$, once inclusion has become the order on cuts.

7.1.4 Walking up through a cut

Many proofs in this chapter need two rationals, one inside a cut and one outside, that are **close together**: within a prescribed rational distance $\varepsilon > 0$. Separation says that the inside lies below the outside, but not that the two come close. They do, and the reason is a walk: start inside, climb in steps of ε, and stop at the first step that leaves the cut.

Lemma 7.1.3. The Step Lemma

Let α be a cut, $a \in \alpha$, and $\varepsilon \in \mathbb{Q}$ with $\varepsilon > 0$. Then there is $n \in \mathbb{N}$ with

$$a + n\varepsilon \in \alpha \qquad \text{and} \qquad a + (n+1)\varepsilon \notin \alpha.$$

Idea. Picture the rungs $a, a+\varepsilon, a+2\varepsilon, \ldots$. The bottom rung a is inside. Some rung is outside: pick any rational b outside α; the rationals are Archimedean, so finitely many steps of size ε climb past b, and everything above b is outside by separation. Among the rungs outside there is a lowest one, because a non-empty set of natural numbers has a least element. It is not the bottom rung, so the rung just below it exists and is inside. The n found depends on α, a and ε.

Proof. By (C1), there is $b \in \mathbb{Q}$ with $b \notin \alpha$. By Theorem 2.5.3, applied to the rational $(b-a)/\varepsilon$, there is $m \in \mathbb{N}$ with $(b-a)/\varepsilon < m$. Multiplying by $\varepsilon > 0$ (Proposition 2.3.5 (c)) gives $b - a < m\varepsilon$, so $b < a + m\varepsilon$ by (O3). Since $b \notin \alpha$, Lemma 7.1.1 gives $a + m\varepsilon \notin \alpha$.

Let

$$K \coloneqq \{k \in \mathbb{N} : a + k\varepsilon \notin \alpha\}.$$

Then $m \in K$, so K is a non-empty set of integers, and 0 is a lower bound of it. By Proposition 2.3.6, K has a least element k_0. Since $a + 0 \cdot \varepsilon = a \in \alpha$, $0 \notin K$, so $k_0 \neq 0$ and $k_0 > 0$. By discreteness (Theorem 2.3.4 (c)), $k_0 \geq 1$. Put $n \coloneqq k_0 - 1$. Then $n \geq 0$ by (O3), so $n \in \mathbb{N}$ (Theorem 2.3.4 (b)). Since $n < k_0$ and k_0 is least in K, $n \notin K$, that is, $a + n\varepsilon \in \alpha$. And $a + (n+1)\varepsilon = a + k_0\varepsilon \notin \alpha$, since $k_0 \in K$. This proves the lemma.

■

Since $n \geq 0$ and $\varepsilon > 0$, the point $p \coloneqq a + n\varepsilon$ found by the lemma satisfies $p \geq a$. So the lemma hands over a pair $p \in \alpha$, $p + \varepsilon \notin \alpha$: a point inside and a point outside, exactly ε apart, with the inside point as high as we like above a given element. This is another of the chapter's named moves.

Walk up by ε: from a point inside a cut, add ε until you leave; the last step inside and the first step outside are ε apart.

It drives the hardest "$\supseteq$" halves to come: the additive inverse in Section 7.3 and, in its multiplicative form of Section 7.4, the multiplicative inverse in Section 7.4 and the square of σ in Section 7.5. Note what it needs from $\mathbb{Q}$: the Archimedean property, to know that the walk leaves the cut at all, and least elements and discreteness of the integers, to find the first step out and the step before it.

Here is one walk, through σ from $a = 1$ in steps of $\varepsilon = 1/10$. The rungs $1, 11/10, 12/10, \ldots$ are positive, so a rung r lies in σ exactly when $r^2 < 2$. Now $(14/10)^2 = 196/100 < 2$, because $196 < 200$, and the lower positive rungs have smaller squares (Proposition 2.3.5 (g)), so the rungs for $k = 0, \ldots, 4$ are inside. But $(15/10)^2 = 225/100 > 2$. So the least k with $1 + k/10 \notin \sigma$ is $k_0 = 5$, the n of the lemma is 4, and the pair is $7/5 \in \sigma$, $3/2 \notin \sigma$.

Quick check. Continue the walk with a finer step: start at $a = 7/5 \in \sigma$ and take $\varepsilon = 1/100$. Which pair $p \in \sigma$, $p + \varepsilon \notin \sigma$ does the lemma give?

Solution. The rungs are $140/100, 141/100, 142/100, \ldots$. Since $141^2 = 19881 < 20000$, $(141/100)^2 < 2$, so $141/100 \in \sigma$. Since $142^2 = 20164 > 20000$, $142/100 \notin \sigma$. So $k_0 = 2$, $n = 1$, and the pair is $p = 141/100$, $p + \varepsilon = 142/100 = 71/50$.

Each walk traps the top of σ between two rationals ε apart, and a smaller ε traps it more tightly. No walk ever lands on the top, since σ has none.

7.1.5 Exercises

7.1.5.1 A. Check your understanding

Exercise. A1

(a) State (C1)–(C3) for a subset $\alpha \subseteq \mathbb{Q}$, and write the negation of (C3) in symbols.

(b) For each set, name every condition among (C1)–(C3) that fails: (i) $\{p \in \mathbb{Q} : p \leq 3\}$; (ii) $\{p \in \mathbb{Q} : p > 0\}$; (iii) $\{p \in \mathbb{Q} : p^2 \leq 0\}$; (iv) $\{p \in \mathbb{Q} : p < p + 1\}$.

(c) Let $q, r \in \mathbb{Q}$. Is $D_q \cup D_r$ a cut? Is $D_q \cap D_r$? Identify each.

Solution.

(a) (C1) $\alpha \neq \emptyset$ and $\alpha \neq \mathbb{Q}$; (C2) for every $p \in \alpha$ and every $q \in \mathbb{Q}$ with $q < p$, $q \in \alpha$; (C3) for every $p \in \alpha$ there is $p' \in \alpha$ with $p' > p$. The negation of (C3) is $\exists p \in \alpha\, \forall p' \in \alpha\, (p' \leq p)$: α has a largest element.

(b) (i) Only (C3) fails: 3 is a largest element. (C1) holds (0 in, 4 out) and (C2) holds ($q < p \leq 3$ gives $q < 3$). (ii) Only (C2) fails: 1 is in, and $0 < 1$ is not. (C1) holds (1 in, 0 out), and (C3) holds ($p+1 > p > 0$). (iii) The set is $\{0\}$, since $p \neq 0$ gives $p^2 > 0$ (Proposition 2.3.5 (e)). (C2) fails ($-1 < 0$, -1 missing) and (C3) fails (0 is largest); (C1) holds. (iv) The set is $\mathbb{Q}$, since $0 < 1$ gives $p < p + 1$ for every p (O3). (C1) fails; (C2) and (C3) hold.

(c) Both are cuts. If $q \leq r$, then $D_q \subseteq D_r$ ($s < q \leq r$ gives $s < r$), so $D_q \cup D_r = D_r$ and $D_q \cap D_r = D_q$; if $r \leq q$, the roles swap. So $D_q \cup D_r = D_{\max(q,r)}$ and $D_q \cap D_r = D_{\min(q,r)}$, rational cuts by Definition 7.1.2.

7.1.5.2 B. Practice

Exercise. B1

Prove that $\{p \in \mathbb{Q} : 3p < 2\}$ is a cut, and that it equals $D_{2/3}$.

Solution. For $p \in \mathbb{Q}$, $3p < 2$ implies $p < 2/3$, by multiplying with $1/3 > 0$ (Proposition 2.5.2 (a), Proposition 2.3.5 (c)); and $p < 2/3$ implies $3p < 2$, by multiplying with $3 > 0$. So the set is $\{p \in \mathbb{Q} : p < 2/3\} = D_{2/3}$, and it is a cut because every rational cut is one (Definition 7.1.2, with the check after it).

Exercise. B2

Prove that $\{p \in \mathbb{Q} : p < 0 \text{ or } p^2 < 3\}$ is a cut. For (C3), give an explicit larger element.

Solution. Call the set β. (C1): $0 \in \beta$, since $0^2 = 0 < 3$; and $2 \notin \beta$, since $2 \geq 0$ and $2^2 = 4 > 3$.

(C2): let $q < p \in \beta$. If $q < 0$, then $q \in \beta$. Otherwise $0 \leq q < p$, so $p > 0$, $p < 0$ fails, and $p^2 < 3$. By Proposition 2.3.5 (g), $q^2 < p^2 < 3$, so $q \in \beta$.

(C3): let $p \in \beta$. If $p < 0$, then $0 \in \beta$ and $0 > p$. If $p \geq 0$, then $p^2 < 3$. Put

$$d \coloneqq 3 - p^2, \qquad h \coloneqq \frac{d}{2p+4}.$$

Then $0 < d \leq 3 < 4 \leq 2p + 4$, so $0 < h < 1$ and $h^2 < h$ (Proposition 2.3.5 (c)). Hence

$$\begin{aligned}(p+h)^2 &= p^2 + 2ph + h^2 < p^2 + h(2p+1) \\ &< p^2 + h(2p+4) = p^2 + d = 3,\end{aligned}$$

so $p + h \in \beta$ and $p + h > p$. (The denominator is chosen to exceed d, so that $h < 1$, and to be at least $2p + 1$; any such denominator works.)

7.1.5.3 C. Going deeper

Exercise. C1: The cut of the gap is not rational

Prove directly from the definitions that $\sigma \neq D_q$ for every $q \in \mathbb{Q}$.

Solution (Hint 1). If $\sigma = D_q$, where would q itself sit relative to σ? What does that say about q and q^2?

Solution (Hint 2). You will reach $q > 0$ and $q^2 \geq 2$. If $q^2 > 2$, look for a smaller positive rational whose square still exceeds 2; Section 2.6 has one.

Solution. Suppose $\sigma = D_q$ for some $q \in \mathbb{Q}$. Since $q \notin D_q$, $q \notin \sigma$: so $q \leq 0$ fails, and $q^2 < 2$ fails. That is, $q > 0$ and $q^2 \geq 2$.

Case $q^2 = 2$. Impossible, by Theorem 2.6.1.

Case $q^2 > 2$. Let $q' \coloneqq (q^2 + 2)/(2q)$, the rational $t(q)$ of Section 2.6. It is positive, as a quotient of positives (Proposition 2.5.2 (a), (O4)). By the identities Eq. 2.6.1,

$$q - q' = \frac{q^2 - 2}{2q} > 0, \qquad q'^2 - 2 = \frac{(q^2-2)^2}{4q^2} > 0,$$

the first a quotient of positives, the second because $q^2 - 2 \neq 0$ (Proposition 2.3.5 (e)). So $q' < q$, that is, $q' \in D_q = \sigma$. But $q' > 0$ and $q'^2 > 2$, so $q' \notin \sigma$. This contradiction shows that $\sigma \neq D_q$.

Marking scheme. (10 points) - 3: from $q \notin D_q = \sigma$, derives both $q > 0$ and $q^2 \geq 2$. - 2: excludes $q^2 = 2$, citing that no rational squares to 2. - 3: for $q^2 > 2$, exhibits a rational q'

with $0 < q' < q$ and $q'^2 > 2$, with both inequalities verified (any such q' earns the points). - 2: concludes: q' lies in D_q but not in σ. Fatal errors (0 for the step): "$q = \sqrt{2}$" or any use of a square root of 2 as a number; a candidate q' whose square is not checked to exceed 2.

7.2 Order by Inclusion, and Least Upper Bounds by Unions

This section belongs to the optional Chapter 7; nothing outside the chapter depends on it. Section 7.1 defined a cut as a set of rationals that is non-empty and not all of $\mathbb{Q}$, closed downward, and without a largest element (Definition 7.1.1), and it called the set of all cuts $\mathbb{R}_{\mathrm{D}}$. Before cuts can be added or multiplied, they can already be compared, and the comparison costs nothing: one cut is below another when it is contained in it. This section shows that this order is total, and then proves the property that the Cauchy road of Chapter 5 reached only at its end. Every non-empty set of cuts that is bounded above has a least upper bound, and that least upper bound is simply the union of the set.

7.2.1 Comparing cuts

For two rationals p and q, "p is at most q" says the same as "every rational below p is below q". In the language of Section 7.1, $p \le q$ holds exactly when $D_p \subseteq D_q$: if $p \le q$ and $r < p$, then $r < q$; and if $q < p$, then $q \in D_p \setminus D_q$. So on rational cuts, inclusion is the order of $\mathbb{Q}$ in disguise. We take it as the order of all cuts.

One cut is below another when every rational in the first is also in the second.

Definition 7.2.1. Order of Cuts

For cuts α and β, we write $\alpha \le \beta$ if $\alpha \subseteq \beta$, and $\alpha < \beta$ if $\alpha \subsetneq \beta$, that is, if $\alpha \subseteq \beta$ **and** $\alpha \neq \beta$. As usual, $\beta \ge \alpha$ means $\alpha \le \beta$, and $\beta > \alpha$ means $\alpha < \beta$.

The relation $\le$ is a partial order on $\mathbb{R}_{\mathrm{D}}$ (Sets, Relations, and Cardinality, Definition 2.6.1): it is reflexive, since $\alpha \subseteq \alpha$; antisymmetric, since $\alpha \subseteq \beta$ and $\beta \subseteq \alpha$ give $\alpha = \beta$ (Sets, Relations, and Cardinality, Proposition 1.1.1); and transitive (Sets, Relations, and Cardinality, Proposition 1.1.2). Its strict part, $\alpha \le \beta$ and $\alpha \neq \beta$, is exactly $<$. These are the facts that make inclusion a partial order on all of $\mathcal{P}(\mathbb{Q})$ (Sets, Relations, and Cardinality, Example (Inclusion is a partial order, rarely total)), restricted to the cuts.

Two translations. First, the order of $\mathbb{Q}$ is recovered on rational cuts: for $p, q \in \mathbb{Q}$,

$$p < q \iff D_p < D_q, \qquad p = q \iff D_p = D_q.$$

Indeed, Proposition 7.1.2 (a) with $\alpha = D_q$ says $p \in D_q \iff D_p \subsetneq D_q$, and $p \in D_q$ means $p < q$. For the second equivalence, if $p \neq q$, then $p < q$ or $q < p$, and either way $D_p \neq D_q$ by the first. Second, Proposition 7.1.2 now reads, for a cut α and $q \in \mathbb{Q}$,

$$q \in \alpha \iff D_q < \alpha, \qquad q \notin \alpha \iff \alpha \le D_q.$$

So "the rational q lies in the cut α" and "the rational cut of q is strictly below α" are the same statement. This is the precise sense in which a cut is the set of rationals below it.

Quick check. Show that $D_1 < \sigma < D_{3/2}$, where σ is the cut of Example (The cut of the gap).

Solution. By the second translation, $D_1 < \sigma$ because $1 \in \sigma$ ($1^2 = 1 < 2$), and $\sigma \leq D_{3/2}$ because $3/2 \notin \sigma$ ($3/2 > 0$ and $(3/2)^2 = 9/4 > 2$). The inclusion $\sigma \subseteq D_{3/2}$ is proper: $29/20$ lies in $D_{3/2}$, since $29 \cdot 2 = 58 < 60 = 3 \cdot 20$, but not in σ, since $29/20 > 0$ and $(29/20)^2 = 841/400 > 2$ (as $841 > 800$). So $\sigma < D_{3/2}$.

Inclusion is rarely a total order: in $\mathcal{P}(\mathbb{Q})$, neither of $\{0\}$ and $\{1\}$ contains the other. Among cuts it is total, and the reason is separation.

Theorem 7.2.1. Cuts Are Totally Ordered

For all cuts α and β, $\alpha \leq \beta$ or $\beta \leq \alpha$. Hence exactly one of $\alpha < \beta$, $\alpha = \beta$, $\beta < \alpha$ holds.

Idea. If α is not inside β, then α has a point p outside β. Everything in β lies below every outside point of β, so below p; and α, being closed downward, contains everything below its point p. So α contains all of β.

Proof. Let α, β be cuts. If $\alpha \subseteq \beta$, then $\alpha \leq \beta$ and we are done; so suppose $\alpha \not\subseteq \beta$. Then some $p \in \alpha$ has $p \notin \beta$. Let $q \in \beta$ be arbitrary. Since $q \in \beta$ and $p \notin \beta$, Lemma 7.1.1 gives $q < p$. Since $p \in \alpha$ and $q < p$, (C2) for α gives $q \in \alpha$. As q was arbitrary, $\beta \subseteq \alpha$, that is, $\beta \leq \alpha$.

For the second statement: if $\alpha \leq \beta$, then $\alpha = \beta$ or $\alpha < \beta$; otherwise $\beta \leq \alpha$ and $\beta \neq \alpha$, that is, $\beta < \alpha$. So at least one of the three holds. No two hold together: $\alpha < \beta$ and $\beta < \alpha$ would give $\alpha = \beta$ by antisymmetry, against $\alpha \neq \beta$; and $\alpha < \beta$ excludes $\alpha = \beta$ by definition, as $\beta < \alpha$ does.

■

The proof used (C2), directly and through Lemma 7.1.1, whose proof uses only (C2) and trichotomy in $\mathbb{Q}$; it never used (C1) or (C3). So any two subsets of $\mathbb{Q}$ that are closed downward are comparable by inclusion. The pair $\{0\}$, $\{1\}$ fails (C2), and that is why it could be incomparable.

With Theorem 7.2.1, $(\mathbb{R}_D, \leq)$ is a totally ordered set, and the vocabulary of *Sets, Relations, and Cardinality*, Section 2.6 applies to it: a set $A \subseteq \mathbb{R}_D$ is **bounded above** if some cut β has $\alpha \leq \beta$ for every $\alpha \in A$ (Sets, Relations, and Cardinality, Definition 2.6.2), and a cut γ is its **supremum** if it is an upper bound of A, clause (S1), and $\gamma \leq \delta$ for every upper bound δ of A, clause (S2) (Sets, Relations, and Cardinality, Definition 2.6.5). A supremum is unique when it exists (Sets, Relations, and Cardinality, Proposition 2.6.1, applied to the set of upper bounds).

7.2.2 Least upper bounds are unions

In $\mathbb{Q}$, the set $\{x \in \mathbb{Q} : x^2 < 2\}$ is non-empty and bounded above, and has no supremum (Theorem 2.6.2). The cuts were built so that no such gap remains, and here is the theorem that

says so.

Theorem 7.2.2. Least Upper Bounds of Cuts

Let $A \subseteq \mathbb{R}_{\mathrm{D}}$ be **non-empty** and **bounded above**: some cut β has $\alpha \leq \beta$ for every $\alpha \in A$. Then

$$\gamma := \bigcup A = \{p \in \mathbb{Q} : p \in \alpha \text{ for some } \alpha \in A\}$$

(Sets, Relations, and Cardinality, Definition 1.2.4, the family indexed by A itself) is a cut, and γ is the supremum of A in $(\mathbb{R}_{\mathrm{D}}, \leq)$.

Try it first (ten minutes): which cut would you propose as the supremum of a set of cuts? Check your proposal against (C1)–(C3), and then against (S1) and (S2).

Idea. The word "bounds" invites a first guess: the intersection $\bigcap A$, the rationals common to all members. It sits on the wrong side. The intersection lies **inside** every member, so it is a candidate for a lower bound, the infimum's job, and it can even fail to be a cut (try it in Exercise C1). An upper bound must **contain** every member, and the smallest set that does is the union. In $\mathcal{P}(\mathbb{Q})$ the union is the supremum: it contains every member, and any set that contains every member contains the union ((S1) and (S2) in the proof below check this for cuts). So everything comes down to one question: is the union a cut? Conditions (C2) and (C3) pass from the members to the union one element at a time, and (C1) needs both hypotheses, one for each half. *Before reading on: which of the three conditions will use "bounded above"?*

Proof. Let β be an upper bound of A. The set γ is a subset of $\mathbb{Q}$, since every member of A is.

(C1). Since A is non-empty, fix $\alpha_0 \in A$; since $\alpha_0 \neq \emptyset$ by (C1) for α_0, fix $p_0 \in \alpha_0$. Then $p_0 \in \gamma$, so $\gamma \neq \emptyset$. Next, $\gamma \subseteq \beta$: if $p \in \gamma$, then $p \in \alpha$ for some $\alpha \in A$ [$\exists$E], and $\alpha \subseteq \beta$ because β is an upper bound, so $p \in \beta$. By (C1) for β, some $b \in \mathbb{Q}$ has $b \notin \beta$; then $b \notin \gamma$, so $\gamma \neq \mathbb{Q}$.

(C2). Let $p \in \gamma$ and $q \in \mathbb{Q}$ with $q < p$. Then $p \in \alpha$ for some $\alpha \in A$ [$\exists$E]. By (C2) for α, $q \in \alpha$, so $q \in \gamma$.

(C3). Let $p \in \gamma$, say $p \in \alpha$ with $\alpha \in A$ [$\exists$E]. By (C3) for α, some $p' \in \alpha$ has $p' > p$, and $p' \in \gamma$.

So γ is a cut. *(S1).* For each $\alpha \in A$, every element of α lies in γ, so $\alpha \subseteq \gamma$, that is, $\alpha \leq \gamma$. *(S2).* Let δ be an upper bound of A. If $p \in \gamma$, then $p \in \alpha$ for some $\alpha \in A$ [$\exists$E], and $\alpha \subseteq \delta$, so $p \in \delta$. Thus $\gamma \subseteq \delta$, that is, $\gamma \leq \delta$. Hence γ is the supremum of A.

■

The answer to the prompt: "bounded above" is used once, in the second half of (C1), to show $\gamma \neq \mathbb{Q}$. "Non-empty" is used once too, in the first half of (C1). Conditions (C2) and (C3) of γ, and the clauses (S1) and (S2), hold for **every** set of cuts; it is only being a cut at all that needs the hypotheses.

Compare the Cauchy road. There, completeness took two sections: Section 5.5 proved that every Cauchy sequence of real numbers converges, by building rational approximants, and Section 5.6

bisected intervals, with one line that needed the Archimedean property, to reach least upper bounds. Here it took one union and five short checks, (C1)–(C3) and (S1)–(S2). The proof above uses no arithmetic of $\mathbb{Q}$ at all, only its order, through (C2) and (C3) of the members; in particular it does not use the Archimedean property. The price is paid elsewhere: cuts have no sum or product yet, and Sections 7.3 and 7.4 build them, multiplication with some effort.

Warning. Both hypotheses of Theorem 7.2.2 are needed for the union to be a cut. Drop "bounded above": the rational cuts D_n, $n \in \mathbb{N}$, have union $\mathbb{Q}$, since every rational r lies below some $n \in \mathbb{N}$ (Theorem 2.5.3) and so in D_n; and $\mathbb{Q}$ fails (C1). Drop "non-empty": $\bigcup \emptyset = \emptyset$, which fails (C1) too.

Second reading · *The empty set and unbounded sets have no supremum among cuts*

The warning shows that the **proof** breaks without each hypothesis. In fact the **conclusion** fails too: without either hypothesis, there is no supremum at all.

The empty set. Every cut is an upper bound of $\emptyset$, vacuously, so a supremum of $\emptyset$ would be a least cut. There is none: for a cut α, pick $a \in \alpha$ (C1); then $D_a < \alpha$ by Proposition 7.1.2 (a), and D_a is a cut, so α is not least.

An unbounded set. A set with no upper bound has no supremum, since a supremum is an upper bound by (S1). For example $\{D_n : n \in \mathbb{N}\}$ has none: if β bounded it, then $\mathbb{Q} = \bigcup_{n\in\mathbb{N}} D_n \subseteq \beta$, against (C1) for β.

Quick check. Find $\sup\{D_{1-1/n} : n \in \mathbb{Z}_{>0}\}$ in $(\mathbb{R}_{\mathrm{D}}, \leq)$. Is it an element of the set?

Solution. The set is non-empty and bounded above by D_1, since $1/n > 0$ (Proposition 2.5.2 (a)), so $1 - 1/n < 1$ and $D_{1-1/n} \subseteq D_1$. By Theorem 7.2.2 its supremum is the union $U := \bigcup_{n\geq 1} D_{1-1/n}$. We show $U = D_1$. ($\subseteq$): if $r < 1 - 1/n$ for some n, then $r < 1$. ($\supseteq$): let $r < 1$. Then $1 - r > 0$, and Corollary 2.5.4 gives $n \in \mathbb{Z}_{>0}$ with $1/n < 1 - r$, that is, $r < 1 - 1/n$, so $r \in U$. Hence the supremum is D_1. It is not in the set: $D_{1-1/n} = D_1$ would force $1 - 1/n = 1$ by the first translation, and $1/n \neq 0$.

So $(\mathbb{R}_{\mathrm{D}}, \leq)$ is a totally ordered set in which every non-empty subset that is bounded above has a supremum. Once Sections 7.3 and 7.4 have made $\mathbb{R}_{\mathrm{D}}$ an ordered field with this order, Section 7.5 reads Theorem 7.2.2 as the statement that it is a **complete** ordered field (Definition 5.6.1).

7.2.3 Exercises

7.2.3.1 A. Check your understanding

Exercise. A1

These questions are about Theorem 7.2.2 and its proof.

(a) Let $A := \{D_q : q \in \sigma\}$, with σ the cut of Example (The cut of the gap). Check that A satisfies the hypotheses, and find the cut γ of the proof, arguing directly from (C2) and (C3) of σ.

(b) State the idea of the proof in two sentences.

(c) Which line of the proof fails if $A = \emptyset$, and what is γ then?

(d) In the check of (C3) for γ, which cut's condition (C3) is used?

Solution.

(a) A is non-empty, since $0 \in \sigma$ gives $D_0 \in A$. It is bounded above by σ: for $q \in \sigma$, $D_q < \sigma$ by Proposition 7.1.2 (a). The proof's cut is $\gamma = \bigcup_{q\in\sigma} D_q$, and $\gamma = \sigma$. ($\subseteq$): each D_q with $q \in \sigma$ is contained in σ, by (C2) of σ. ($\supseteq$): let $q \in \sigma$; by (C3) of σ there is $q' \in \sigma$ with $q' > q$, so $q \in D_{q'} \subseteq \gamma$. So $\sup A = \sigma$: the cut of the gap is the supremum of the rational cuts below it.

(b) The union of the cuts contains each of them, and lies inside every upper bound, so it is the least upper bound once it is known to be a cut. It is one because (C2) and (C3) pass from the members to the union element by element, while (C1) comes from $A \neq \emptyset$ (some element) and from an upper bound $\beta \neq \mathbb{Q}$ (some non-element).

(c) The first line of (C1): with $A = \emptyset$ there is no $\alpha_0 \in A$ to fix. Then $\gamma = \bigcup \emptyset = \emptyset$, which is not a cut.

(d) Condition (C3) of the member $\alpha \in A$ that contains the given $p \in \gamma$. Different points of γ may use different members.

7.2.3.2 B. Practice

Exercise. B1

Let $q, r \in \mathbb{Q}$. Prove that $\sup\{D_q, D_r\} = D_{\max(q,r)}$ in $(\mathbb{R}_D, \leq)$.

Solution. Let $m := \max(q, r)$. Since $q \leq m$ and $r \leq m$, $D_q \subseteq D_m$ and $D_r \subseteq D_m$ (if $s < q \leq m$, then $s < m$); so $\{D_q, D_r\}$ is non-empty and bounded above by D_m. By Theorem 7.2.2 its supremum is $D_q \cup D_r$. This union contains D_m, because D_m is one of D_q, D_r; and it is contained in D_m, by the two inclusions above. Hence $\sup\{D_q, D_r\} = D_q \cup D_r = D_m$.

Exercise. B2: Put the proof in order

Here are lines of a proof that for cuts α and β, $\alpha \leq \beta$ or $\beta \leq \alpha$, shuffled, together with two lines that do not belong. Without looking back, write the lines of a correct proof in a correct order, and say what is wrong with each line you leave out.

(i) Let $q \in \beta$ be arbitrary.

(ii) Since $q < p$ and $p \in \alpha$, $q \in \beta$.

(iii) As q was arbitrary, $\beta \subseteq \alpha$.

(iv) Let α, β be cuts with $\alpha \not\subseteq \beta$ (if $\alpha \subseteq \beta$, there is nothing to prove).

(v) Since $p \in \alpha$ and $q < p$, $q \in \alpha$, because α is closed downward.

(vi) Since $q \in \beta$ and $p \notin \beta$, $p < q$.
(vii) Then some $p \in \alpha$ has $p \notin \beta$.
(viii) Since $q \in \beta$ and $p \notin \beta$, $q < p$, by separation.

Solution. An admissible order is

$$\text{(iv)},\ \text{(vii)},\ \text{(i)},\ \text{(viii)},\ \text{(v)},\ \text{(iii)}.$$

Line (i) may also come before (vii), since it only introduces a name for an arbitrary element of β, and (vii) does not mention q. Otherwise the order is forced: (vii) unwinds the assumption of (iv); (viii) needs p from (vii) and q from (i); (v) needs $q < p$ from (viii); and (iii) closes the argument about the arbitrary q.

Line (vi) reads separation backwards. Lemma 7.1.1 puts a point **inside** a cut below a point **outside** it. Here q is inside β and p is outside, so separation gives $q < p$, as in (viii); line (vi) asserts the opposite, and the two together contradict (O1).

Line (ii) applies closure to the wrong cut. Closure downward of α turns "$p \in \alpha$ and $q < p$" into $q \in \alpha$, not into $q \in \beta$; the fact that $p \in \alpha$ says nothing about β. (Its conclusion $q \in \beta$ is true, but only because it was assumed in (i).)

7.2.3.3 C. Going deeper

Exercise. C1: Infima

Let $A \subseteq \mathbb{R}_{\mathrm{D}}$ be non-empty and bounded below.

(a) Show that $\bigcap A$ need not be a cut, by computing $\bigcap A_0$ for $A_0 := \{D_{1/n} : n \in \mathbb{Z}_{>0}\}$.
(b) Prove that A nevertheless has an infimum in $(\mathbb{R}_{\mathrm{D}}, \leq)$: it is the supremum of the set of lower bounds of A.
(c) Find $\inf A_0$.

Solution (Hint 1). For (b), the set L of lower bounds of A is non-empty by hypothesis. Find an upper bound for L among the elements of A.

Solution (Hint 2). Let $\lambda_0 := \sup L$, which exists by Theorem 7.2.2; the claim is that $\lambda_0 = \inf A$.

Solution.

(a) Here $\bigcap A_0 = \{r \in \mathbb{Q} : r < 1/n \text{ for every } n \in \mathbb{Z}_{>0}\}$, the intersection taken in $\mathbb{Q}$. We claim it is $\{r \in \mathbb{Q} : r \leq 0\}$. If $r \leq 0$, then $r \leq 0 < 1/n$ for every n (Proposition 2.5.2 (a)). If $r > 0$, Corollary 2.5.4 gives n with $1/n < r$, so $r \notin D_{1/n}$ and $r \notin \bigcap A_0$. The set $\{r : r \leq 0\}$ has the largest element 0, so it fails (C3) and is not a cut. (The set A_0 is non-empty and bounded below, by D_0, for instance.)
(b) Let $L := \{\lambda \in \mathbb{R}_{\mathrm{D}} : \lambda \leq \alpha \text{ for every } \alpha \in A\}$, the set of lower bounds of A. It is

non-empty, since A is bounded below. It is bounded above: fix $\alpha_0 \in A$ ($A \neq \emptyset$); every $\lambda \in L$ has $\lambda \leq \alpha_0$. By Theorem 7.2.2, L has a supremum λ_0.
λ_0 is a lower bound of A. Let $\alpha \in A$. Every $\lambda \in L$ has $\lambda \leq \alpha$, so α is an upper bound of L, and (S2) for $\lambda_0 = \sup L$ gives $\lambda_0 \leq \alpha$.
λ_0 is the greatest lower bound. We just showed $\lambda_0 \in L$. If λ is any lower bound of A, then $\lambda \in L$, so $\lambda \leq \lambda_0$ by (S1) for λ_0. Hence $\lambda_0 = \inf A$.

(c) $\inf A_0 = D_0$. It is a lower bound: $D_0 \subseteq D_{1/n}$ since $0 < 1/n$. Let λ be any lower bound. Then $\lambda \subseteq \bigcap A_0 = \{r : r \leq 0\}$ by (a). And $0 \notin \lambda$: otherwise (C3) for λ gives $p' \in \lambda$ with $p' > 0$, but λ contains no positive rational. So every element of λ is negative, $\lambda \subseteq D_0$, and $\lambda \leq D_0$. So the infimum is the intersection with its largest element removed.

Marking scheme. (10 points) - 2: (a) computes $\bigcap A_0 = \{r : r \leq 0\}$, using the Archimedean property for the positive rationals, and names (C3) as the failing condition. - 2: (b) shows that the set of lower bounds is non-empty and bounded above by a member of A. - 2: (b) shows that its supremum is a lower bound of A, through (S2). - 2: (b) shows that it is the greatest lower bound, through (S1). - 2: (c) identifies D_0, with both inequalities. Fatal errors (0 for the step): asserting that $\bigcap A$ is the infimum without checking (C3); using a supremum of L without checking that L is non-empty and bounded above.

Exercise. C2: Find the error

Here is a proof. Its conclusion is the first half of Theorem 7.2.2.

> **Claim.** Let $A \subseteq \mathbb{R}_{\mathrm{D}}$ be non-empty and bounded above. Then $\gamma := \bigcup A$ is a cut.
>
> (1) Since A is non-empty, pick $\alpha_0 \in A$; since $\alpha_0 \neq \emptyset$, pick $p_0 \in \alpha_0$. Then $p_0 \in \gamma$, so $\gamma \neq \emptyset$. (2) Each $\alpha \in A$ is a cut, so $\alpha \neq \mathbb{Q}$. Hence $\gamma \neq \mathbb{Q}$. (3) Let $p \in \gamma$ and $q < p$. Then $p \in \alpha$ for some $\alpha \in A$, so $q \in \alpha$ by (C2), and $q \in \gamma$. (4) Let $p \in \gamma$. Then $p \in \alpha$ for some $\alpha \in A$; by (C3) there is $p' \in \alpha$ with $p' > p$, and $p' \in \gamma$. (5) So γ satisfies (C1)–(C3).

Is the conclusion true? Does the argument prove it? If not, name the line that fails, say what is illegal about it, and give a counterexample to that step.

Solution. The conclusion is true: it is proved in Theorem 7.2.2. The argument does not prove it, because line (2) fails.

What is illegal. Line (2) passes from "for every $\alpha \in A$ there is a rational outside α" to "there is a rational outside every $\alpha \in A$", which is what $\gamma \neq \mathbb{Q}$ means. The rational outside α depends on α, and the line uses it as if one rational served every member at once. This is the move (N2) of Logic and Proof, Example (What is not allowed): a witness found inside the box of one α is treated as uniform in α. A union of proper subsets of $\mathbb{Q}$ can be all of $\mathbb{Q}$.

Counterexample to the step. Let $A' := \{D_n : n \in \mathbb{N}\}$. Each D_n is a cut, so $D_n \neq \mathbb{Q}$, and the premise of the step holds; but $\bigcup A' = \mathbb{Q}$, since every rational lies below some natural number (Theorem 2.5.3). So the step leads from a true premise to a false conclusion. (This A' is not bounded above, so it is not a counterexample to the claim, which is true.)

The tell. The hypothesis "bounded above" is never used. A correct proof must use it, since the warning after Theorem 7.2.2 shows that the claim fails without it.

Repair. Let β be an upper bound of A. Every $\alpha \in A$ satisfies $\alpha \subseteq \beta$, so $\gamma \subseteq \beta$; and $\beta \neq \mathbb{Q}$, so some $b \notin \beta$, hence $b \notin \gamma$ and $\gamma \neq \mathbb{Q}$.

Exercise. C3: Exit check: least upper bounds of cuts

Let A be a non-empty set of cuts, bounded above in $(\mathbb{R}_{\mathrm{D}}, \subseteq)$. Prove that $\bigcup A$ is a cut and that it is the supremum of A.

Solution. Let $\gamma := \bigcup A$ and let β be an upper bound of A.

(C1). Pick $\alpha_0 \in A$ and $p_0 \in \alpha_0$, possible since $A \neq \emptyset$ and $\alpha_0 \neq \emptyset$; then $p_0 \in \gamma$. Every element of γ lies in some $\alpha \in A$, and $\alpha \subseteq \beta$, so $\gamma \subseteq \beta$. Pick $b \in \mathbb{Q} \setminus \beta$ (C1 for β); then $b \notin \gamma$. So $\emptyset \neq \gamma \neq \mathbb{Q}$.

(C2). If $q < p$ with $p \in \gamma$, then $p \in \alpha$ for some $\alpha \in A$, so $q \in \alpha \subseteq \gamma$.

(C3). If $p \in \gamma$, then $p \in \alpha$ for some $\alpha \in A$, and (C3) for α gives $p' \in \alpha \subseteq \gamma$ with $p' > p$.

Upper bound. Each $\alpha \in A$ satisfies $\alpha \subseteq \gamma$.

Least. If δ is a cut with $\alpha \subseteq \delta$ for every $\alpha \in A$, then every $p \in \gamma$ lies in some $\alpha \in A$, hence in δ; so $\gamma \subseteq \delta$.

So γ is a cut, an upper bound of A, and below every upper bound: $\gamma = \sup A$.

Marking scheme. (10 points) - 2: $\gamma \neq \emptyset$ from $A \neq \emptyset$, and $\gamma \neq \mathbb{Q}$ from the upper bound $\beta \neq \mathbb{Q}$. - 2: γ closed downward. - 2: γ has no largest element. - 2: γ is an upper bound of A. - 2: γ is below every upper bound of A. Fatal errors (0 for the step): "$\gamma \neq \mathbb{Q}$ because each $\alpha \neq \mathbb{Q}$" (the error of Exercise C2); comparing cuts with "$\leq$" as if they were rationals, instead of by inclusion.

7.3 Addition and Negatives

Like every section of this chapter, this one is optional: nothing outside Chapter 7 depends on it. Sections 7.1 and 7.2 made cuts and ordered them, and the least upper bound property came almost for free, as a union (Theorem 7.2.2). Now the arithmetic. Adding two cuts is the obvious thing, and it works; the zero cut works too, once (C3) is used. Negatives are where the obvious recipe fails, on the simplest cuts of all, and the repair is a small positive margin. The proof that the repaired negative is right is the **walk up by** ε (Lemma 7.1.3) worked in full. The section ends with cuts forming an ordered group under addition.

Throughout, $\mathbb{Q}$ is the ordered field of Theorem 2.5.1, and every letter p, q, r, s, t, x stands for a rational. We use the rules of an ordered field for rationals without comment: (F1)–(F9) of Definition 2.3.1, (O1)–(O4) of Definition 2.3.2, Proposition 2.3.5, and the halves of Proposition 4.1.2 (b).

7.3.1 Adding two cuts

A cut is a set of rationals, so the first candidate for a sum is the set of all sums.

Definition 7.3.1. Sum of Cuts

For cuts α and β, the **sum** is the set

$$\alpha + \beta := \{p + q : p \in \alpha,\ q \in \beta\}.$$

In symbols, $x \in \alpha + \beta \iff \exists p \in \alpha\ \exists q \in \beta\ \ x = p + q$. This is the set sum $A + B$ of *Sets, Relations, and Cardinality*, Section 1.6 (Sets, Relations, and Cardinality, Example (Reflections, scalings and sums as images)), applied to two sets of rationals: the same notation with the same meaning. For instance, if $p < 1$ and $q < 2$, then $p + q < 3$, so $D_1 + D_2 \subseteq D_3$; the Quick check later in this section shows how the reverse inclusion goes for a similar pair.

The definition makes sense for any two sets of rationals. The work is to show that a sum of cuts is again a cut, and that this addition behaves.

Theorem 7.3.1. The Sum of Two Cuts Is a Cut

For all cuts α and β, the set $\alpha + \beta$ is a cut. Addition of cuts is commutative and associative: for all cuts α, β, γ,

$$\alpha + \beta = \beta + \alpha \qquad \text{and} \qquad (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma).$$

Idea. Each cut condition for $\alpha + \beta$ comes from the same condition for α or β. A rational outside a cut is above every element of it (Lemma 7.1.1), so the sum of a point outside α and a point outside β is outside the sum. To go down from $p + q$, go down in the first summand. To go up, go up in the first summand. The two laws are laws of $\mathbb{Q}$, applied element by element.

Proof. Let α and β be cuts. We check (C1)–(C3) of Definition 7.1.1 for $\alpha + \beta$.

(C1) By (C1) for α and β, there are $p_0 \in \alpha$ and $q_0 \in \beta$, so $p_0 + q_0 \in \alpha + \beta$, and $\alpha + \beta \neq \emptyset$. Also by (C1), there are $a \in \mathbb{Q} \setminus \alpha$ and $b \in \mathbb{Q} \setminus \beta$. Let $p \in \alpha$ and $q \in \beta$. By Lemma 7.1.1, $p < a$ and $q < b$, so $p + q < a + b$ by Proposition 2.3.5 (b). So no element of $\alpha + \beta$ equals $a + b$, and $a + b \notin \alpha + \beta$. Hence $\alpha + \beta \neq \mathbb{Q}$.

(C2) Let $x \in \alpha + \beta$ and $y \in \mathbb{Q}$ with $y < x$. Write $x = p + q$ with $p \in \alpha$, $q \in \beta$. Adding $-q$ to $y < p + q$ gives $y - q < p$ (O3), so $y - q \in \alpha$ by (C2) for α. Hence $y = (y - q) + q \in \alpha + \beta$.

(C3) Let $x = p + q \in \alpha + \beta$ with $p \in \alpha$, $q \in \beta$. By (C3) for α there is $p' \in \alpha$ with $p' > p$. Then $p' + q \in \alpha + \beta$ and $p' + q > p + q = x$ (O3).

So $\alpha + \beta$ is a cut.

Commutativity. If $x \in \alpha + \beta$, then $x = p + q$ with $p \in \alpha$, $q \in \beta$, and $x = q + p \in \beta + \alpha$ by (F1) in $\mathbb{Q}$. So $\alpha + \beta \subseteq \beta + \alpha$; exchanging the roles of α and β gives the reverse inclusion.

Associativity. Let γ be a cut. ($\subseteq$) If $x \in (\alpha + \beta) + \gamma$, then $x = y + s$ with $y \in \alpha + \beta$ and $s \in \gamma$, and $y = p + q$ with $p \in \alpha$, $q \in \beta$. By (F2) in $\mathbb{Q}$, $x = p + (q + s)$, and $q + s \in \beta + \gamma$, so $x \in \alpha + (\beta + \gamma)$. ($\supseteq$) If $x = p + z$ with $p \in \alpha$ and $z = q + s \in \beta + \gamma$, then

$x = (p + q) + s \in (\alpha + \beta) + \gamma$, by (F2) again. This proves the theorem.

■

Only (C3) for α was needed in the third step; (C3) for β would serve equally. The two laws used nothing about cuts at all: they hold for the set sum of arbitrary sets of rationals (commutativity for any two, associativity for any three), and Exercise B2 asks you to check associativity in this generality.

7.3.2 The zero cut

The natural zero is $D_0 = \{r \in \mathbb{Q} : r < 0\}$, the rational cut of 0 (Definition 7.1.2). Adding an element of D_0 pushes a rational down, so $\alpha + D_0$ cannot leave α. The other inclusion needs a point of α strictly above the given one.

Proposition 7.3.2. The Zero Cut

For every cut α, $\alpha + D_0 = \alpha$.

Proof. Let α be a cut.

($\subseteq$) Let $x \in \alpha + D_0$, say $x = p + q$ with $p \in \alpha$ and $q < 0$. Then $p + q < p + 0 = p$ (O3), so $x \in \alpha$ by (C2).

($\supseteq$) Let $p \in \alpha$. By (C3) there is $p' \in \alpha$ with $p' > p$; fix one [$\exists$E] (Logic and Proof, Definition 5.6.1). Then $p - p' < 0$ by Proposition 2.3.5 (a), so $p - p' \in D_0$, and

$$p = p' + (p - p') \in \alpha + D_0.$$

As $p \in \alpha$ was arbitrary, $\alpha \subseteq \alpha + D_0$ [$\forall$I] (Logic and Proof, Definition 5.4.1).

■

The second half used (C3), and it cannot do without it. The tempting shortcut "$p = p + 0$" fails because $0 \notin D_0$; Exercise C2 shows a proof built on that shortcut. A set with a largest element fails the proposition: $A := \{r \in \mathbb{Q} : r \leq 1\}$ satisfies (C1) and (C2), and $A + D_0 = D_1$, since $a \leq 1$, $q < 0$ give $a + q < 1$, and every $x < 1$ is $1 + (x - 1)$; so adding D_0 removes 1. Section 7.1's example of a set with a largest element makes a related point about cancellation (Example (Two candidates for one rational)).

7.3.3 Negatives

Before reading on. Find a cut β with $D_1 + \beta = D_0$. Then try to write down a recipe that produces such a β from any cut α, so that $\alpha + \beta = D_0$.

For D_1 the answer is $\beta = D_{-1}$. Indeed, $p < 1$ and $q < -1$ give $p + q < 0$; and for $s < 0$, put $h := -s/2 > 0$, so that $s = (1 - h) + (-1 - h)$ with $1 - h < 1$ and $-1 - h < -1$. For a general α, a natural recipe reads the picture of a cut as a downset: the negative of "the rationals below x" should be "the rationals below $-x$", and those are the negatives of the rationals **above** x, that is, of the rationals outside α. The example below tests that recipe.

Example. The Negatives of the Outside

Optional: nothing later depends on this.

For a cut α, is the set $\{-q : q \in \mathbb{Q} \setminus \alpha\}$ a cut? Test it on $\sigma = \{p \in \mathbb{Q} : p \leq 0 \text{ or } p^2 < 2\}$ (Example (The cut of the gap)) and on D_1. Then decide in general. Is the set $\{-p : p \in \alpha\}$ a cut?

Solution. For σ it works. A rational q is outside σ exactly when $q > 0$ and $q^2 \geq 2$; by Theorem 2.6.1, $q^2 \neq 2$, so $q^2 > 2$. Since $(-q)^2 = q^2$ (Proposition 2.3.2 (c)), the set is

$$\{-q : q \in \mathbb{Q} \setminus \sigma\} = \{x \in \mathbb{Q} : x < 0 \text{ and } x^2 > 2\}.$$

It is a cut. (C1): -2 is in it, 0 is not. (C2): if $y < x$ with $x < 0$ and $x^2 > 2$, then $0 < -x < -y$, so $x^2 = (-x)^2 < (-y)^2 = y^2$ by Proposition 2.3.5 (g), and $y < 0$, $y^2 > 2$. (C3): let $x < 0$ with $x^2 > 2$, put $q := -x > 0$ and

$$q' := \frac{q^2 + 2}{2q}; \quad \text{then} \quad \begin{aligned} q - q' &= \frac{q^2 - 2}{2q} > 0, \\ q'^2 - 2 &= \frac{(q^2 - 2)^2}{4q^2} > 0. \end{aligned}$$

Then $q' > 0$, and $x' := -q'$ has $x < x' < 0$ and $x'^2 > 2$: an element of the set above x.

For D_1 it fails. The outside of D_1 is $\{q : q \geq 1\}$, so the set is $\{x \in \mathbb{Q} : x \leq -1\}$. It has a largest element, -1, so (C3) fails.

In general. The set always satisfies (C1) and (C2). (C1): for $b \notin \alpha$, $-b$ is in it; for $a \in \alpha$, $-a$ is not, since $-a = -q$ forces $q = a \in \alpha$. (C2): if $y < -q$ with $q \notin \alpha$, then $-y > q$, so $-y \notin \alpha$ by Lemma 7.1.1, and $y = -(-y)$ is in the set. (C3) fails exactly when the set has a largest element $-m$, that is, when $\mathbb{Q} \setminus \alpha$ has a least element m. That happens exactly when α is a rational cut. If $\alpha = D_m$, the outside $\{q : q \geq m\}$ has least element m. Conversely, if m is least in $\mathbb{Q} \setminus \alpha$, then every $r < m$ is in α (else it would be a smaller element of the outside), and every $r \in \alpha$ has $r < m$ (Lemma 7.1.1), so $\alpha = D_m$. So the recipe gives a cut **unless** α is rational.

The negatives of the inside. For $\alpha = D_1$, $\{-p : p \in D_1\} = \{x : x > -1\}$. It contains 0 but not $-2 < 0$, so (C2) fails. In general this set is closed **upward**, the mirror image of a cut: if $p \in \alpha$ and $y > -p$, then $-y < p$, so $-y \in \alpha$ by (C2), and $y = -(-y)$ is in the set. And it is never a cut, because (C2) fails as it did for D_1. Take $a \in \alpha$ and $b \notin \alpha$ (C1). Then $a < b$ by Lemma 7.1.1, so $-b < -a$. The set contains $-a$ but not $-b$, since $-b = -p$ forces $p = b \notin \alpha$.

The recipe passes the test on σ, the first cut one tries, and fails on the rational cuts, the simplest cuts there are.

The failure is a single point: for D_1 the recipe produces D_{-1} together with the one extra point -1. The repair is to stay a positive margin away from the outside.

The negative of a cut is the set of negatives of the outside, pushed down by a positive margin.

Definition 7.3.2. Negative of a Cut

For a cut α, its **negative** is the set

$$-\alpha := \{p \in \mathbb{Q} : -p - r \notin \alpha \text{ for some rational } r > 0\}.$$

In symbols, $p \in -\alpha \iff \exists r \in \mathbb{Q}_{>0}\ \ -p - r \notin \alpha$. It is a cut (checked below). A rational r with this property is a **margin** for p.

In words: p belongs to $-\alpha$ when $-p$ is not merely outside α but outside with room to spare, since even $-p - r$, a positive distance below $-p$, is still outside. The quantifier is "some": one margin suffices, and it may depend on p. Equivalently, $p \in -\alpha$ exactly when $p < -q$ for some $q \notin \alpha$ (take $q := -p - r$, and conversely $r := -q - p$).

$-\alpha$ is a cut. Let α be a cut.

- (C1) By (C1) for α there is $b \notin \alpha$. For $p := -b-1$, the margin $r := 1$ gives $-p-1 = b \notin \alpha$, so $p \in -\alpha$. Also there is $a \in \alpha$, and $-a \notin -\alpha$: for every rational $r > 0$, $-(-a)-r = a-r < a$, so $a - r \in \alpha$ by (C2), and no margin exists.
- (C2) Let $p \in -\alpha$ with margin r, and $p' < p$. Then $-p' - r > -p - r$, and $-p - r \notin \alpha$, so $-p' - r \notin \alpha$, because the outside of a cut is closed upward (Lemma 7.1.1). So $p' \in -\alpha$, with the same margin.
- (C3) Let $p \in -\alpha$ with margin r. Then $p + r/2 > p$, and $-(p + r/2) - r/2 = -p - r \notin \alpha$, so $p + r/2 \in -\alpha$ with margin $r/2$.

So $-\alpha \in \mathbb{R}_{\mathrm{D}}$, and $\alpha \mapsto -\alpha$ is a function $\mathbb{R}_{\mathrm{D}} \to \mathbb{R}_{\mathrm{D}}$.

Examples. The degenerate case is $-D_0 = D_0$. A rational p is in $-D_0$ when $-p - r \geq 0$ for some $r > 0$, that is, $p \leq -r < 0$; and every $p < 0$ qualifies, with margin $r := -p$. The case that matters is $-D_1 = D_{-1}$. Here $p \in -D_1$ when $-p - r \geq 1$ for some $r > 0$, that is, $p \leq -1 - r$ for some $r > 0$. That holds for every $p < -1$, with margin $r := -1 - p$, and for no $p \geq -1$. The point -1 that spoiled the recipe has no margin. The same computation gives $-D_q = D_{-q}$ for every rational q (Exercise B1; the Quick check below does $q = 3$).

A non-example by minimal change. Allow the margin $r = 0$. The condition becomes $-p \notin \alpha$, and the set is the recipe of Example (The Negatives of the Outside): for D_1 it contains -1 and has a largest element. Exactly (C3) breaks, and the proof of (C3) above shows why: it moves up from p by half the margin, which needs $r > 0$.

Why the margin. The margin is what makes $-\alpha$ satisfy (C3) when the outside of α has a least element, and it changes nothing when the outside has none: then $-\alpha$ is exactly the recipe $\{-q : q \in \mathbb{Q} \setminus \alpha\}$ of Example (The Negatives of the Outside). For every cut, if $p \in -\alpha$, then $p < -q$ for some $q \notin \alpha$ by the equivalence above, so $-p > q$ is outside α (Lemma 7.1.1), and $p = -(-p)$ is in the recipe. Conversely, if the outside has no least element and $q \notin \alpha$, there is $q' \notin \alpha$ with $q' < q$; then $-q < -q'$, so $-q \in -\alpha$ by the equivalence above. The fold below works this out for σ.

↻ **Second reading** · *Edge case: for the cut of the square root of two, the margin changes nothing*

For σ, the definition gives $-\sigma = \{x \in \mathbb{Q} : x < 0 \text{ and } x^2 > 2\}$.

($\subseteq$) Let $p \in -\sigma$ with margin r. Then $-p - r \notin \sigma$, so $-p - r > 0$ and $(-p - r)^2 \geq 2$. Hence

$-p > -p - r > 0$, so $p < 0$, and $p^2 = (-p)^2 > (-p-r)^2 \geq 2$ by Proposition 2.3.5 (g).

($\supseteq$) Let $p < 0$ with $p^2 > 2$, and put $q := -p$. The rational $q' := (q^2+2)/(2q)$ satisfies $0 < q' < q$ and $q'^2 > 2$, by the identities $q - q' = (q^2-2)/(2q)$ and $q'^2 - 2 = (q^2-2)^2/(4q^2)$. With the margin $r := q - q' > 0$, $-p - r = q'$, which is positive with square at least 2, so $q' \notin \sigma$, and $p \in -\sigma$.

The outside of σ has no least element, and that is why no margin is needed.

Warning. The negative of a cut is **not** the set $\{-p : p \in \alpha\}$. The book's notation $-A$ for a set of numbers means that set, but for a cut $-\alpha$ always means Definition 7.3.2. The two differ even in shape: for $\alpha = D_1$, $\{-p : p \in D_1\} = \{x : x > -1\}$ contains 0 but not -2, so it fails (C2), while $-D_1 = D_{-1}$.

Quick check. Compute $-D_3$ from Definition 7.3.2, and $D_{1/2} + D_{1/3}$ from Definition 7.3.1.

Solution. $-D_3 = D_{-3}$. A rational p is in $-D_3$ when $-p - r \geq 3$ for some $r > 0$, that is, $p \leq -3 - r$. Every $p < -3$ qualifies, with margin $r := -3 - p > 0$; no $p \geq -3$ does, since then $-p - r \leq 3 - r < 3$.

$D_{1/2} + D_{1/3} = D_{5/6}$. ($\subseteq$) $p < 1/2$ and $q < 1/3$ give $p + q < 5/6$. ($\supseteq$) For $x < 5/6$, put $h := (5/6 - x)/2 > 0$. Then $x = (1/2 - h) + (1/3 - h)$, with $1/2 - h < 1/2$ and $1/3 - h < 1/3$. Section 7.5 proves $D_p + D_q = D_{p+q}$ for all rationals by the same split.

7.3.4 A cut plus its negative

Here is the theorem the margin was designed for. Its proof is a key proof of the chapter.

Theorem 7.3.3. Additive Inverses of Cuts

For every cut α, $\alpha + (-\alpha) = D_0$.

Idea. The inclusion $\subseteq$ is separation: an element of α is below every outside point, in particular below $-t - r$ for $t \in -\alpha$.

For $\supseteq$, work backwards from the goal. Given $s < 0$, we want $p \in \alpha$ and $t \in -\alpha$ with $p + t = s$. The second summand is forced, $t = s - p$, and "$t \in -\alpha$ with margin r" unwinds to $-t - r = p + (-s - r) \notin \alpha$. So we need **a point inside α and a point outside, at distance $-s - r$**, a little less than $-s$. That is what the walk up by ε supplies (Lemma 7.1.3): start inside, add ε until you leave, and the last point inside and the first point outside are exactly ε apart. Take ε and the margin r both equal to $-s/2$; then $-s - r = \varepsilon$.

Before reading on: why is $\varepsilon = -s/2$ and not $-s$?

Proof. Let α be a cut. By Definition 7.3.2 and Theorem 7.3.1, $-\alpha$ and $\alpha + (-\alpha)$ are cuts.

($\subseteq$) Let $x \in \alpha + (-\alpha)$, say $x = p + t$ with $p \in \alpha$ and $t \in -\alpha$, and let $r > 0$ be a margin for t, so $-t - r \notin \alpha$. By Lemma 7.1.1, $p < -t - r$. Adding t gives $p + t < -r < 0$. So $x \in D_0$.

($\supseteq$) Let $s \in D_0$, so $s < 0$, and let

$$\varepsilon := -s/2;$$

it depends only on s, and $\varepsilon > 0$ by Proposition 4.1.2 (b), since $-s > 0$. By (C1) there is $a \in \alpha$; fix one [$\exists$E]. By Lemma 7.1.3, applied to α, a and ε, there is $n \in \mathbb{N}$ with

$$a + n\varepsilon \in \alpha \qquad \text{and} \qquad a + (n+1)\varepsilon \notin \alpha;$$

fix one such n [$\exists$E] and put $p := a + n\varepsilon$, so that $p \in \alpha$ and $p + \varepsilon \notin \alpha$. Now let

$$t := s - p.$$

Since $s = -2\varepsilon$,

$$-t - \varepsilon = p - s - \varepsilon = p + 2\varepsilon - \varepsilon = p + \varepsilon \notin \alpha,$$

so ε is a margin for t, and $t \in -\alpha$ [$\exists$I] (Logic and Proof, Definition 5.5.1), with witness $r := \varepsilon$. Finally $p + t = s$, so $s \in \alpha + (-\alpha)$ [$\exists$I], with witnesses p and t. As $s \in D_0$ was arbitrary, $D_0 \subseteq \alpha + (-\alpha)$ [$\forall$I].

The two inclusions give $\alpha + (-\alpha) = D_0$ (Sets, Relations, and Cardinality, Proposition 1.1.1).

■

The answer to the prompt. With $\varepsilon = -s$, the walk gives $p \in \alpha$ with $p - s \notin \alpha$, and $t = s - p$ would need a margin $r > 0$ with $p - s - r \notin \alpha$. Nothing guarantees one: $p - s - r$ lies below the first outside point of the walk, and may be inside. For $\alpha = D_1$, $s = -1$ and $a = 0$, the walk stops at $p = 0$, and $p - s - r = 1 - r \in D_1$ for every $r > 0$. The margin must be strictly smaller than the gap, and halving $-s$ leaves room for both.

The proof treats rational and irrational cuts alike: it never asks whether the outside of α has a least element. The margin in Definition 7.3.2 is exactly what absorbs the rational case; without it the $\supseteq$ half would still go through, but for a rational cut α the set it lands in would not be a cut (Example (The Negatives of the Outside)).

7.3.5 An ordered group

The results so far say that cuts under $+$ behave like the rationals under $+$, with the order of Section 7.2 respected. We collect them, together with the consequences that Section 7.4 uses.

Theorem 7.3.4. Cuts Form an Ordered Group under Addition

For all cuts α, β, γ:

(a) $\alpha + \beta = \beta + \alpha$ and $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
(b) $\alpha + D_0 = \alpha$;
(c) $\alpha + (-\alpha) = D_0$;
(d) $\leq$ is a total order on $\mathbb{R}_{\mathrm{D}}$; if $\alpha \leq \beta$, then $\alpha + \gamma \leq \beta + \gamma$; and if $\alpha < \beta$, then $\alpha + \gamma < \beta + \gamma$;

(e) (cancellation and unique inverses) if $\alpha+\gamma=\beta+\gamma$, then $\alpha=\beta$; if $\alpha+\beta=D_0$, then $\beta=-\alpha$; and $-(-\alpha)=\alpha$, $-D_0=D_0$, $-(\alpha+\beta)=(-\alpha)+(-\beta)$;

(f) $\alpha<\beta$ if and only if $-\beta<-\alpha$, and $\alpha\le\beta$ if and only if $-\beta\le-\alpha$. In particular $\alpha>D_0 \iff -\alpha<D_0$, and $\alpha<D_0 \iff -\alpha>D_0$.

So $(\mathbb{R}_{\mathrm{D}},+,\le)$ is an **ordered abelian group**: (a)–(c) are (F1)–(F4) for $+$, with zero D_0, and (d) is a total order with (O3). Clauses (e) and (f) are consequences.

Proof.

(a) is Theorem 7.3.1, (b) is Proposition 7.3.2 and (c) is Theorem 7.3.3.

(b) By the paragraph after Definition 7.2.1, $\le$ is a partial order, and by Theorem 7.2.1 any two cuts are comparable, so $\le$ is a total order. If $\alpha\subseteq\beta$ and $x=p+q\in\alpha+\gamma$ with $p\in\alpha$, $q\in\gamma$, then $p\in\beta$ and $x\in\beta+\gamma$; so $\alpha+\gamma\le\beta+\gamma$. The strict form follows from the cancellation law in (e): if $\alpha<\beta$, then $\alpha+\gamma\le\beta+\gamma$, and equality would give $\alpha=\beta$.

(c) If $\alpha+\gamma=\beta+\gamma$, then by (b), (c) and (a),

$$\begin{aligned}\alpha&=\alpha+(\gamma+(-\gamma))=(\alpha+\gamma)+(-\gamma)\\&=(\beta+\gamma)+(-\gamma)=\beta+(\gamma+(-\gamma))=\beta.\end{aligned}$$

If $\alpha+\beta=D_0$, then $\beta+\alpha=D_0=(-\alpha)+\alpha$ by (a) and (c), and cancellation gives $\beta=-\alpha$. Since $(-\alpha)+\alpha=D_0$, the same applied to the cut $-\alpha$ gives $\alpha=-(-\alpha)$. Since $D_0+D_0=D_0$ by (b), it gives $D_0=-D_0$. Finally, by (a), (c) and (b),

$$(\alpha+\beta)+((-\alpha)+(-\beta))=(\alpha+(-\alpha))+(\beta+(-\beta))=D_0+D_0=D_0,$$

so $(-\alpha)+(-\beta)=-(\alpha+\beta)$.

(d) Let $\alpha<\beta$ and $\delta:=(-\alpha)+(-\beta)$. By (d), $\alpha+\delta<\beta+\delta$. By (a), (c) and (b), $\alpha+\delta=-\beta$ and $\beta+\delta=-\alpha$. So $-\beta<-\alpha$. Conversely, if $-\beta<-\alpha$, the same argument applied to the cuts $-\beta$ and $-\alpha$ gives $-(-\alpha)<-(-\beta)$, which is $\alpha<\beta$ by (e). The non-strict form adds the case of equality: $\alpha=\beta$ if and only if $-\alpha=-\beta$, by (e). For the last sentence, take α or β to be D_0 and use $-D_0=D_0$.

■

Half of the field axioms now hold for cuts. The other half needs a product, and Section 7.4 builds it by reducing to positive cuts. There (f) is what turns a negative cut into a positive one, and (e) is what lets the signs be put back.

7.3.6 Exercises

7.3.6.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 7.3.3 and the definition of $-\alpha$.

(a) Run the $\supseteq$ half of the proof for $\alpha=D_1$, $s=-1/2$ and the starting point $a=0$. What are ε, n, p, t and the margin? Check that $t\in-D_1$ and $p+t=s$.

(b) Where in the proof is Lemma 7.1.3 used, and what does that lemma need from $\mathbb{Q}$?

(c) In the check that $-\alpha$ is a cut, which condition uses $r > 0$? What goes wrong for $\alpha = D_1$ if the margin $r = 0$ is allowed?

(d) State the idea of the proof in two sentences.

Solution.

(a) $\varepsilon = -s/2 = 1/4$. The points $a + k\varepsilon = k/4$ are in D_1 exactly for $k < 4$, so $n = 3$, $p = 3/4 \in D_1$ and $p + \varepsilon = 1 \notin D_1$. Then $t = s - p = -1/2 - 3/4 = -5/4$, with margin $\varepsilon = 1/4$: $-t - 1/4 = 5/4 - 1/4 = 1 \notin D_1$, so $t \in -D_1$. And $p + t = 3/4 - 5/4 = -1/2 = s$. (Another starting point a may give another pair, for example $a = 1/8$ gives $p = 7/8$; any $p \in D_1$ with $p + 1/4 \geq 1$, that is $3/4 \leq p < 1$, works with $t = -1/2 - p$.)

(b) Once, in the $\supseteq$ half, to produce $p \in \alpha$ with $p + \varepsilon \notin \alpha$. Its proof in Section 7.1 needs the Archimedean property of $\mathbb{Q}$ (Theorem 2.5.3), to get some multiple of ε that leaves α, and the least element of a non-empty set of natural numbers (Proposition 2.3.6), to get the first one, and discreteness of the integers (Theorem 2.3.4 (c)), to step back from the first one to the one before it.

(c) (C3): from $p \in -\alpha$ with margin r we moved up to $p + r/2$, which is larger than p only when $r > 0$. With $r = 0$ allowed, the set for D_1 is $\{x : -x \notin D_1\} = \{x : x \leq -1\}$, which has the largest element -1, so it is not a cut.

(d) Inside α and outside it, there are points exactly ε apart, found by walking up from a point of α in steps of ε. With $\varepsilon = -s/2$, the last point p inside and $t = s - p$ split s, and the first point outside certifies $t \in -\alpha$ with margin ε.

7.3.6.2 B. Practice

Exercise. B1

Prove that $-D_q = D_{-q}$ for every $q \in \mathbb{Q}$.

Solution. Let $q, p \in \mathbb{Q}$. By Definition 7.3.2 and Definition 7.1.2, $p \in -D_q$ if and only if $-p - r \geq q$ for some rational $r > 0$.

($\subseteq$) Let $p \in -D_q$ with margin r. Then $-p \geq q + r > q$, so $p < -q$ (Proposition 2.3.5 (a)), and $p \in D_{-q}$.

($\supseteq$) Let $p < -q$. Then $r := -q - p > 0$, and $-p - r = q \geq q$, so $q \notin D_q$ shows that r is a margin for p. Hence $p \in -D_q$.

So $-D_q = D_{-q}$.

Exercise. B2: Associativity, element by element

Let A, B, C be any subsets of $\mathbb{Q}$, and let $A + B := \{a + b : a \in A,\ b \in B\}$. Prove that $(A + B) + C = A + (B + C)$, writing out both inclusions element by element and naming the rule of $\mathbb{Q}$ used in each. Which of the cut conditions (C1)–(C3) did you use?

Solution. ($\subseteq$) Let $x \in (A+B)+C$. By the definition, $x = y + c$ for some $y \in A+B$ and $c \in C$, and $y = a + b$ for some $a \in A$, $b \in B$. So $x = (a+b)+c = a+(b+c)$ by (F2) in $\mathbb{Q}$. Since $b + c \in B + C$, $x \in A + (B + C)$.

($\supseteq$) Let $x \in A + (B + C)$. Then $x = a + z$ with $a \in A$ and $z \in B + C$, and $z = b + c$ with $b \in B$, $c \in C$. So $x = a+(b+c) = (a+b)+c$ by (F2), and $a+b \in A+B$, so $x \in (A+B)+C$.

The two inclusions give the equality (Sets, Relations, and Cardinality, Proposition 1.1.1). None of (C1)–(C3) was used: associativity of the set sum holds for all sets of rationals, and the only fact used was (F2). For cuts, Theorem 7.3.1 adds that the sum is again a cut.

7.3.6.3 C. Going deeper

Exercise. C1: Exit check: a cut plus its negative

Prove that $\alpha + (-\alpha) = D_0$ for every cut α. Then explain, on the cut $\alpha = D_1$, why the recipe $\{-q : q \in \mathbb{Q} \setminus \alpha\}$ does not give an additive inverse in $\mathbb{R}_{\mathrm{D}}$, while Definition 7.3.2 does.

Solution (Hint 1). For $\supseteq$, given $s < 0$, you need a point of α and a point outside α that are a fixed positive distance apart. Which lemma of Section 7.1 produces such a pair?

Solution (Hint 2). Let $\varepsilon := -s/2$, and apply Lemma 7.1.3 to a point of α and this ε.

Solution. Let α be a cut; $-\alpha$ is a cut by Definition 7.3.2.

($\subseteq$) Let $p \in \alpha$ and $t \in -\alpha$, with a margin $r > 0$, so $-t - r \notin \alpha$. By Lemma 7.1.1, $p < -t - r$, so $p + t < -r < 0$, and $p + t \in D_0$.

($\supseteq$) Let $s < 0$ be arbitrary, and $\varepsilon := -s/2 > 0$. Fix $a \in \alpha$ (C1). By Lemma 7.1.3 there is $n \in \mathbb{N}$ with $p := a + n\varepsilon \in \alpha$ and $p + \varepsilon \notin \alpha$. Let $t := s - p$. Then $-t - \varepsilon = p - s - \varepsilon = p + \varepsilon \notin \alpha$, so $t \in -\alpha$ with margin ε, and $p + t = s$. So $s \in \alpha + (-\alpha)$.

The comparison on D_1. The recipe gives $\{-q : q \geq 1\} = \{x : x \leq -1\}$. This set has the largest element -1, so it fails (C3) and is not a cut; it is not an element of $\mathbb{R}_{\mathrm{D}}$ at all, whatever its sum with D_1 may be. In Definition 7.3.2, -1 would need a margin $r > 0$ with $1 - r \notin D_1$, and there is none, since $1 - r < 1$. So the margin removes exactly the point -1, and $-D_1 = \{x : x < -1\} = D_{-1}$, a cut, with $D_1 + D_{-1} = D_0$ by the first part.

Marking scheme. (10 points) - 2: $\subseteq$, using separation against the outside point $-t - r$ to get $p + t < 0$. - 1: fixes $s < 0$ arbitrary. - 3: produces $p \in \alpha$ with $p + \varepsilon \notin \alpha$ for an ε chosen from s alone, citing the step lemma (or proving it from the Archimedean property and least elements). - 2: exhibits $t \in -\alpha$ with its margin, and checks $p + t = s$. - 2: the D_1 comparison: the recipe's set has the largest element -1, and the margin removes it. Fatal errors (0 for the step): a margin equal to the whole gap ($\varepsilon = -s$ with margin $-s$); ε chosen depending on p after p was found from ε.

Exercise. C2: Find the error

Here is a proof that $\alpha + D_0 = \alpha$ for every cut α.

> *Proof.* (1) Let α be a cut. (2) ($\subseteq$) Let $x \in \alpha + D_0$, so $x = p + q$ with $p \in \alpha$ and $q \in D_0$. (3) Then $q < 0$, so $x = p + q < p$, and $x \in \alpha$ by (C2). (4) ($\supseteq$) Let $p \in \alpha$. (5) Then $p = p + 0$ with $p \in \alpha$ and $0 \in D_0$, so $p \in \alpha + D_0$. (6) Hence $\alpha + D_0 = \alpha$.

(a) Is the conclusion true?

(b) Which line fails, and why?

(c) Give a counterexample to the step itself.

(d) Repair the proof.

(e) The flawed proof never uses (C3). Show that without it the conclusion can fail: compute $A + D_0$ for $A := \{r \in \mathbb{Q} : r \le 1\}$.

Solution.

(a) Yes: it is Proposition 7.3.2.

(b) Line (5). It asserts $0 \in D_0$, which is false: $D_0 = \{r : r < 0\}$, and $0 < 0$ fails by (O1). The line cites a "fact" that is neither proved nor true, error (N9) of Logic and Proof, Example (What is not allowed). Lines (1)–(4) and (6) are correct.

(c) Take $\alpha = D_1$ and $p = 1/2$. The decomposition $1/2 = 1/2 + 0$ of line (5) is not a decomposition into an element of D_1 and an element of D_0, because $0 \notin D_0$. (The conclusion $1/2 \in D_1 + D_0$ is still true, by $1/2 = 3/4 + (-1/4)$; the step does not show it.)

(d) Replace line (5) by: "By (C3) there is $p' \in \alpha$ with $p' > p$. Then $p - p' < 0$, so $p - p' \in D_0$, and $p = p' + (p - p') \in \alpha + D_0$." This is the proof of Proposition 7.3.2.

(e) Without (C3) the conclusion fails. The set $A = \{r : r \le 1\}$ satisfies (C1) and (C2) but not (C3), and $A + D_0 = D_1 \neq A$. Indeed, if $a \le 1$ and $q < 0$, then $a + q < 1$; and every $x < 1$ is $1 + (x - 1)$ with $1 \in A$ and $x - 1 < 0$. So $A + D_0 = D_1$, which does not contain $1 \in A$. An argument that never uses (C3) would prove the same equation for A, where it is false; that is the tell that some line is wrong.

7.4 Multiplication

Like every section of this chapter, this one is optional: nothing outside Chapter 7 depends on it. Section 7.3 made the cuts an ordered group under addition (Theorem 7.3.4). This section adds a product and finishes the field axioms. On the Cauchy road of Chapter 5, multiplication was termwise and cost little; on this road it is the hard part. The set of all products of elements is useless, so we define the product of two **positive** cuts first, prove the laws there, and then extend to all cuts by signs, once, through one lemma (Lemma 7.4.4). The reciprocal needs the walk up by ε again, in a multiplicative form. The section ends with the theorem that the cuts form an ordered field.

As in Section 7.3, letters p, q, r, s, t, x stand for rationals, and the rules of the ordered field $\mathbb{Q}$ (Theorem 2.5.1, Proposition 2.3.5, Proposition 2.5.2) are used for them without comment.

7.4.1 Why not the set of products

For sums, the set of all sums worked. The analogous set for products,

$$\{pq : p \in \alpha,\ q \in \beta\},$$

fails in two different ways.

It can be everything. Suppose $0 \in \alpha$ and $0 \in \beta$, as for $\alpha = \beta = D_1$. Then both cuts contain every negative rational (C2), and α contains some $a > 0$ (C3). Every rational is a product: for $x > 0$, $x = (-1)(-x)$ with $-1 \in \alpha$, $-x \in \beta$; $0 = 0 \cdot 0$; and for $x < 0$, $x = a \cdot (x/a)$ with $x/a < 0$, so $x/a \in \beta$. So the set is $\mathbb{Q}$, which fails (C1). The culprit is that negatives times negatives are unbounded above.

It can be upside down. For $\alpha = \beta = D_{-1}$, every product pq with $p, q < -1$ is $(-p)(-q) > 1$, and every $x > 1$ is such a product: $x = pq$ with $p := -(1+x)/2 < -1$ and $q := -2x/(1+x) < -1$. So the set is $\{x \in \mathbb{Q} : x > 1\}$. It contains 2 but not $0 < 2$, so it fails (C2): it is closed upward.

Both failures come from negative elements, and every cut has them. For a cut $\alpha > D_0$, the positive elements carry all the information, since the others are exactly the rationals $p \leq 0$: by Proposition 7.1.2, $\alpha > D_0$ means $0 \in \alpha$, so every $p \leq 0$ is in α by (C2), and α has a positive element by (C3). So for positive cuts we multiply positive elements only, then put back everything below.

Warning. The product of two cuts is **not** the set of products of their elements, even for positive cuts: D_1 and D_1 give $\mathbb{Q}$, as shown above. Every proof in this section works with the definitions below, never with that set.

7.4.2 Positive cuts

Definition 7.4.1. Product of Positive Cuts

Let α and β be cuts with $\alpha > D_0$ and $\beta > D_0$ (equivalently, $0 \in \alpha$ and $0 \in \beta$, by Proposition 7.1.2). Their **product** is

$$\alpha\beta := \{p \in \mathbb{Q} : p \leq rs \text{ for some } r \in \alpha \text{ and } s \in \beta \text{ with } r > 0 \text{ and } s > 0\}.$$

In words: take the products of **positive** elements, and then every rational at or below one of them. The second step is what makes $\alpha\beta$ closed downward; it also puts every $p \leq 0$ in $\alpha\beta$, since positive $r \in \alpha$ and $s \in \beta$ exist and $p \leq 0 < rs$. For instance, in $D_2 D_3$ every product rs of positive $r < 2$, $s < 3$ is below 6, so $D_2 D_3 \subseteq D_6$; Exercise B1 proves equality.

Quick check. Let $\alpha, \beta > D_0$. Suppose the definition said "$p = rs$" instead of "$p \leq rs$". Which set would it give, and which cut condition would that set fail?

Solution. It would give the set of products rs of positive elements, a set of positive rationals.

It is non-empty, but it fails (C2): it contains a positive rs and not $0 < rs$.

Proposition 7.4.1. Positive Cuts under Multiplication

Let α, β, γ be cuts with $\alpha, \beta, \gamma > D_0$. Then:

(a) $\alpha\beta$ is a cut, and $\alpha\beta > D_0$;
(b) $\alpha\beta = \beta\alpha$;
(c) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$;
(d) $\alpha D_1 = \alpha$.

Idea. Each cut condition for $\alpha\beta$ is inherited from α or β, as for sums: an outside point of each factor bounds the products, and moving up in α moves the product up. The laws (b) and (c) are (F6) and (F5) of $\mathbb{Q}$, applied to positive elements. For (d), multiplying by an element of D_1 pushes down, so one inclusion is (C2); the other writes p as a larger element of α times a factor below 1.

Proof. Throughout, α, β, γ contain 0 and a positive element (by Proposition 7.1.2 and (C3)), and every element outside one of them is positive (Lemma 7.1.1, against 0).

(a) (C1): $0 \in \alpha\beta$, as noted after the definition. By (C1) there are $a \notin \alpha$ and $b \notin \beta$; then $a, b > 0$. For positive $r \in \alpha$, $s \in \beta$, separation gives $r < a$ and $s < b$, so $rs < as < ab$ by Proposition 2.3.5 (c), since $s > 0$ and $a > 0$. So every $p \in \alpha\beta$ has $p \leq rs < ab$ for suitable r, s, and $ab \notin \alpha\beta$. (C2): if $p \leq rs$ and $q < p$, then $q \leq rs$, with the same r, s. (C3): let $p \in \alpha\beta$, with $p \leq rs$ for positive $r \in \alpha$, $s \in \beta$. By (C3) for α there is $r' \in \alpha$ with $r' > r$; then $r' > 0$, so $r's \in \alpha\beta$, and $r's > rs \geq p$ by Proposition 2.3.5 (c). So $\alpha\beta$ is a cut, and since $0 \in \alpha\beta$, $\alpha\beta > D_0$ by Proposition 7.1.2.

(b) The condition "$p \leq rs$ for some positive $r \in \alpha$, $s \in \beta$" is the same as "$p \leq sr$ for some positive $s \in \beta$, $r \in \alpha$", by (F6) in $\mathbb{Q}$. So the two sets are equal.

(c) Let T be the set of rationals p with $p \leq rst$ for some positive $r \in \alpha$, $s \in \beta$, $t \in \gamma$. We show $(\alpha\beta)\gamma = T$. ($\subseteq$) Let $p \leq ut$ with $u \in \alpha\beta$, $t \in \gamma$ positive. Then $u \leq rs$ for some positive $r \in \alpha$, $s \in \beta$, and $ut \leq (rs)t$ since $t > 0$; so $p \leq rst$ and $p \in T$. ($\supseteq$) If $p \leq (rs)t$ with r, s, t as in T, then $u := rs$ is positive and in $\alpha\beta$ (as $rs \leq rs$), so $p \in (\alpha\beta)\gamma$. The same argument, grouping $r(st)$, gives $\alpha(\beta\gamma) = T$, using (F5) in $\mathbb{Q}$.

(d) ($\subseteq$) Let $p \leq rs$ with positive $r \in \alpha$ and $s \in D_1$, so $0 < s < 1$. Then $rs < r$, so $p < r$ and $p \in \alpha$ by (C2). ($\supseteq$) Let $p \in \alpha$. If $p \leq 0$, then $p \in \alpha D_1$ as noted after the definition. If $p > 0$, take $p' \in \alpha$ with $p' > p$ by (C3). Then $0 < p/p' < 1$, so $p/p' \in D_1$ is positive, and $p = p' \cdot (p/p') \in \alpha D_1$.

■

The proof of (d) used (C3) in the same place as the proof that $\alpha + D_0 = \alpha$ (Proposition 7.3.2): to reach p from inside, one needs an element of α strictly above it.

The reciprocal will need, as the negative did, a point inside a cut and a point outside that are close. For products "close" means "close in ratio", and the walk up by ε gives that too.

Lemma 7.4.2. The Walk Up, Multiplicatively

Let α be a cut with $\alpha > D_0$, and let $\mu \in \mathbb{Q}$ with $\mu > 1$. Then there is $r \in \alpha$ with $r > 0$ and $\mu r \notin \alpha$.

Idea. Walk up from a positive point a in steps of $\varepsilon := (\mu - 1)a$. The last point r inside is at least a, so the step ε is at most $(\mu - 1)r$, and $\mu r = r + (\mu - 1)r$ is at least the first point outside.

Proof. By Proposition 7.1.2, $0 \in \alpha$, and by (C3) there is $a \in \alpha$ with $a > 0$; fix one [$\exists$E] (Logic and Proof, Definition 5.6.1). Let $\varepsilon := (\mu - 1)a$; it is positive, as a product of positives. By Lemma 7.1.3 there is $n \in \mathbb{N}$ with $r := a + n\varepsilon \in \alpha$ and $r + \varepsilon \notin \alpha$. Since $n\varepsilon \geq 0$, $r \geq a > 0$. Then

$$\mu r - (r + \varepsilon) = (\mu - 1)r - (\mu - 1)a = (\mu - 1)(r - a) \geq 0,$$

so $\mu r \geq r + \varepsilon$. As $r + \varepsilon \notin \alpha$ and the outside of a cut is closed upward (Lemma 7.1.1), $\mu r \notin \alpha$.

■

The last law for positive cuts is distributivity. The sum $\beta + \gamma$ of two positive cuts contains $0 = 0 + 0$, since $0 \in \beta$ and $0 \in \gamma$, so it is positive (both steps by Proposition 7.1.2 (a)), and $\alpha(\beta + \gamma)$ is defined.

Lemma 7.4.3. Distributivity for Positive Cuts

For all cuts $\alpha, \beta, \gamma > D_0$, $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Idea. An element of $\beta + \gamma$ is a sum $s + t$, and $r(s + t) = rs + rt$. The only snag is that s or t may be negative, and the definition of $\alpha\beta$ wants positive factors; raising them to positive elements only increases $r(s + t)$, which is harmless because both sides are closed downward. For the other inclusion, two different first factors r_1, r_2 are replaced by the larger.

Proof. Let $\alpha, \beta, \gamma > D_0$. By Theorem 7.3.1 and Proposition 7.4.1 (a), both sides are cuts.

($\subseteq$) Let $p \in \alpha(\beta + \gamma)$: $p \leq ru$ for some positive $r \in \alpha$ and positive $u \in \beta + \gamma$, and $u = s + t$ with $s \in \beta$, $t \in \gamma$. Let $\tilde{s} := s$ if $s > 0$, and otherwise let $\tilde{s}$ be a positive element of β (one exists by (C3), since $0 \in \beta$). Then $\tilde{s} \in \beta$, $\tilde{s} > 0$ and $\tilde{s} \geq s$. Choose $\tilde{t} \in \gamma$ in the same way. Since $r > 0$,

$$p \leq r(s + t) = rs + rt \leq r\tilde{s} + r\tilde{t},$$

and $r\tilde{s} \in \alpha\beta$, $r\tilde{t} \in \alpha\gamma$. So $r\tilde{s} + r\tilde{t} \in \alpha\beta + \alpha\gamma$, and p is in that cut, being equal to this element or below it (C2).

($\supseteq$) Let $x \in \alpha\beta$ and $y \in \alpha\gamma$: $x \leq r_1 s$ and $y \leq r_2 t$ with positive $r_1, r_2 \in \alpha$, $s \in \beta$, $t \in \gamma$. Let $r := \max\{r_1, r_2\}$, one of r_1, r_2, so $r \in \alpha$ and $r > 0$. Then

$$x + y \leq r_1 s + r_2 t \leq rs + rt = r(s + t),$$

and $s + t \in \beta + \gamma$ is positive. So $x + y \in \alpha(\beta + \gamma)$.

■

7.4.3 Signs

Every cut is positive, zero or negative (Theorem 7.2.1), and a negative cut has a positive negative (Theorem 7.3.4 (f)). So the product of positive cuts determines all others, as in school: multiply the sizes and fix the sign.

Definition 7.4.2. Absolute Value, Sign and Product of Cuts

For a cut α, its **absolute value** and **sign** are

$$|\alpha| := \begin{cases} \alpha & \text{if } \alpha \geq D_0, \\ -\alpha & \text{if } \alpha < D_0, \end{cases} \qquad \operatorname{sgn}\alpha := \begin{cases} 1 & \text{if } \alpha > D_0, \\ 0 & \text{if } \alpha = D_0, \\ -1 & \text{if } \alpha < D_0. \end{cases}$$

For cuts α and β, their **product** is

$$\alpha\beta := \begin{cases} D_0 & \text{if } \alpha = D_0 \text{ or } \beta = D_0, \\ |\alpha||\beta| & \text{if } \operatorname{sgn}\alpha = \operatorname{sgn}\beta \neq 0, \\ -(|\alpha||\beta|) & \text{if } \operatorname{sgn}\alpha \neq \operatorname{sgn}\beta \text{ and neither is } 0, \end{cases}$$

where $|\alpha||\beta|$ is the product of positive cuts of Definition 7.4.1.

The cases do not overlap and cover every cut and every pair, by Theorem 7.2.1. For $\alpha \neq D_0$, $|\alpha| > D_0$ (Theorem 7.3.4 (f)), so in the last two cases $|\alpha||\beta|$ is defined, and it is a cut (Proposition 7.4.1 (a)). So $\alpha\beta$ is a cut in every case: D_0, a product of positive cuts, or the negative of one (Definition 7.3.2, with the check after it). The product is a function $\mathbb{R}_D \times \mathbb{R}_D \to \mathbb{R}_D$. The new product agrees with the old one where both are defined: for $\alpha, \beta > D_0$ the signs agree, $|\alpha| = \alpha$, $|\beta| = \beta$, and the second case returns $\alpha\beta$ of Definition 7.4.1. So writing $\alpha\beta$ for both is safe. The degenerate cases: $|D_0| = D_0$, $\operatorname{sgn} D_0 = 0$, and $\alpha D_0 = D_0\alpha = D_0$ for every cut α. Two small examples. Since $-1 < 0$, $D_{-1} < D_0$ (Section 7.2, "Two translations"), so $\operatorname{sgn} D_{-1} = -1$ and $|D_{-1}| = -D_{-1} = D_1$, because $p \in -D_{-1}$ exactly when $-p - r \geq -1$ for some $r > 0$, that is, when $p < 1$; and so $D_{-1}D_{-1} = D_1D_1 = D_1$, by Proposition 7.4.1 (d).

Warning. The absolute value $|\alpha|$ is defined here before we know that $\mathbb{R}_D$ is an ordered field. Once Theorem 7.4.7 below has shown that it is, $|\alpha|$ is the absolute value of that ordered field in the sense of Definition 4.1.1, by the same formula, so the two readings never disagree. Before that theorem, no rule of Proposition 4.1.3 may be used for cuts; the rules we need are proved in the next lemma.

The next lemma is where the reduction to positive cuts pays off. That reduction is the last of the chapter's named moves.

Reduce to the positive case: define and prove for positive cuts, then extend to all cuts by signs once, through one lemma, never axiom by axiom.

The lemma is proved once, by the definition's cases, and the later sign laws (distributivity, the reciprocal of a negative cut, the field axioms in Theorem 7.4.7) cite it instead of repeating the case analysis.

Lemma 7.4.4. Sign Rules for Cuts

For all cuts α, β, γ:

(a) $\alpha = \operatorname{sgn}(\alpha)|\alpha|$, reading $1 \cdot \xi \coloneqq \xi$, $0 \cdot \xi \coloneqq D_0$ and $(-1) \cdot \xi \coloneqq -\xi$; so two cuts with the same sign and the same absolute value are equal;

(b) $|\alpha\beta| = |\alpha||\beta|$ and $\operatorname{sgn}(\alpha\beta) = \operatorname{sgn}\alpha \operatorname{sgn}\beta$;

(c) $|-\alpha| = |\alpha|$ and $\operatorname{sgn}(-\alpha) = -\operatorname{sgn}\alpha$;

(d) $(-\alpha)\beta = -(\alpha\beta) = \alpha(-\beta)$;

(e) $\alpha\beta = \beta\alpha$;

(f) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$;

(g) $\alpha D_1 = \alpha$.

In words: a cut is its sign times its size; sizes multiply as positive cuts do, and signs multiply as the integers $-1, 0, 1$ do.

Idea. Clause (a) says a cut is determined by two pieces of data, its sign and its absolute value. Clauses (b) and (c) compute both pieces for a product and a negative. After that, each of (d)–(g) is an equation between two cuts whose signs and absolute values are visibly equal: the signs by arithmetic in $\{-1, 0, 1\}$, the absolute values by Proposition 7.4.1 once a factor D_0 is set aside. No axiom is checked over the sign combinations.

Proof. We use Theorem 7.3.4 (e), (f) freely: $-(-\xi) = \xi$, $-D_0 = D_0$, and $\xi > D_0 \iff -\xi < D_0$.

(a) If $\alpha > D_0$, then $\operatorname{sgn}(\alpha)|\alpha| = 1 \cdot \alpha = \alpha$. If $\alpha = D_0$, it is $0 \cdot D_0 = D_0$. If $\alpha < D_0$, it is $(-1) \cdot (-\alpha) = -(-\alpha) = \alpha$. So two cuts with equal signs and equal absolute values are both equal to $\operatorname{sgn}(\alpha)|\alpha|$.

(b) If $\alpha = D_0$ or $\beta = D_0$, then $\alpha\beta = D_0$, so $|\alpha\beta| = D_0$ and $\operatorname{sgn}(\alpha\beta) = 0$; and $|\alpha||\beta| = D_0$, since one factor is $|D_0| = D_0$, and $\operatorname{sgn}\alpha \operatorname{sgn}\beta = 0$. Otherwise let $\pi \coloneqq |\alpha||\beta|$, a product of positive cuts, so $\pi > D_0$ by Proposition 7.4.1 (a). If the signs agree, $\alpha\beta = \pi$, with absolute value π and sign $1 = \operatorname{sgn}\alpha \operatorname{sgn}\beta$. If they differ, $\alpha\beta = -\pi < D_0$, with absolute value $-(-\pi) = \pi$ and sign $-1 = \operatorname{sgn}\alpha \operatorname{sgn}\beta$.

(c) If $\alpha > D_0$, then $-\alpha < D_0$, so $|-\alpha| = -(-\alpha) = \alpha = |\alpha|$, and the signs are -1 and 1. If $\alpha = D_0$, then $-\alpha = D_0$. If $\alpha < D_0$, then $-\alpha > D_0$, so $|-\alpha| = -\alpha = |\alpha|$, and the signs are 1 and -1.

(d) By (b) and (c), $(-\alpha)\beta$ has absolute value $|-\alpha||\beta| = |\alpha||\beta|$ and sign $-\operatorname{sgn}\alpha \operatorname{sgn}\beta$. By (c) and (b), $-(\alpha\beta)$ has absolute value $|\alpha\beta| = |\alpha||\beta|$ and sign $-\operatorname{sgn}(\alpha\beta) = -\operatorname{sgn}\alpha \operatorname{sgn}\beta$. By (a) they are equal. The same computation, with β negated instead of α, gives $\alpha(-\beta) = -(\alpha\beta)$.

(e) The signs of $\alpha\beta$ and $\beta\alpha$ are $\operatorname{sgn}\alpha \operatorname{sgn}\beta = \operatorname{sgn}\beta \operatorname{sgn}\alpha$. Their absolute values are $|\alpha||\beta|$ and $|\beta||\alpha|$, by (b): both are D_0 if α or β is D_0, and otherwise they are equal by Proposition 7.4.1 (b). By (a), $\alpha\beta = \beta\alpha$.

(f) By (b), twice, the absolute values of the two sides are $(|\alpha||\beta|)|\gamma|$ and $|\alpha|(|\beta||\gamma|)$. If one of α, β, γ is D_0, both are D_0, since a product with a factor D_0 is D_0; otherwise they are equal by Proposition 7.4.1 (c). The signs are $(\operatorname{sgn}\alpha \operatorname{sgn}\beta)\operatorname{sgn}\gamma = \operatorname{sgn}\alpha(\operatorname{sgn}\beta \operatorname{sgn}\gamma)$. By (a), the two sides are equal.

(g) $D_1 > D_0$, since $0 \in D_1$ (Proposition 7.1.2), so $|D_1| = D_1$ and $\operatorname{sgn} D_1 = 1$. By (b), αD_1 has sign $\operatorname{sgn}\alpha$ and absolute value $|\alpha|D_1$, which is $D_0 = |\alpha|$ if $\alpha = D_0$, and $|\alpha|$ by Proposition 7.4.1 (d) otherwise. By (a), $\alpha D_1 = \alpha$.

■

Quick check. Compute $D_{-2} \cdot D_3$ through the cases of Definition 7.4.2. Which case applies, and what are $|D_{-2}|$ and $\operatorname{sgn} D_{-2}$? Express the product through $D_2 D_3$.

Solution. Since $0 \notin D_{-2}$, $D_{-2} > D_0$ fails (Proposition 7.1.2), and $D_{-2} \neq D_0$ (as $-1 \in D_0 \setminus D_{-2}$); so $D_{-2} < D_0$ and $\operatorname{sgn} D_{-2} = -1$, while $\operatorname{sgn} D_3 = 1$ ($0 \in D_3$). The signs differ, so the third case applies: $D_{-2}D_3 = -(|D_{-2}|\,|D_3|)$. Here $|D_{-2}| = -D_{-2} = D_2$, since $p \in -D_{-2}$ exactly when $-p - r \geq -2$ for some $r > 0$, that is, when $p < 2$; and $|D_3| = D_3$. So $D_{-2}D_3 = -(D_2 D_3)$. Since $D_2 D_3 = D_6$ (Exercise B1) and $-D_6 = D_{-6}$ (by the same computation), the product is D_{-6}, as $(-2) \cdot 3 = -6$ suggests. Section 7.5 proves $D_p D_q = D_{pq}$ for all rationals.

Distributivity for all cuts is the one law where the reduction to positive cuts needs an idea, because $\beta + \gamma$ may have a sign different from both β and γ. The device is to move the negative summand to the other side, where it becomes positive.

Lemma 7.4.5. Distributivity for All Cuts

For all cuts α, β, γ, $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Idea. Fix $\alpha > D_0$ and write $f(\xi) \coloneqq \alpha\xi$. We know two things about f: it is additive on cuts $> D_0$ (Lemma 7.4.3), and trivially when one summand is D_0, and it commutes with negation (Lemma 7.4.4 (d)). If β and γ have the same sign, negate both if necessary. If $\beta \geq D_0 \geq \gamma$, one of β and $-\gamma$ is a sum of two cuts $\geq D_0$: $\beta = (\beta + \gamma) + (-\gamma)$ when $\beta + \gamma \geq D_0$, and $-\gamma = (-(\beta + \gamma)) + \beta$ otherwise. Apply additivity to that sum and solve for $f(\beta + \gamma)$. For $\alpha < D_0$, pull the sign out of α.

Before reading on: which case uses that $-\gamma \geq D_0$?

Proof. Throughout we use Theorem 7.3.4 and Lemma 7.4.4 (d) as in the Idea. If $\alpha = D_0$, both sides are D_0 (Definition 7.4.2, Proposition 7.3.2).

Let $\alpha > D_0$ and $f(\xi) \coloneqq \alpha\xi$ for every cut ξ. Two facts:

- (A) $f(\xi + \eta) = f(\xi) + f(\eta)$ whenever $\xi, \eta \geq D_0$. If $\xi, \eta > D_0$, this is Lemma 7.4.3. If $\xi = D_0$, both sides are $f(\eta)$, since $f(D_0) = D_0$ and $D_0 + \eta = \eta$; likewise if $\eta = D_0$.

- (N) $f(-\xi) = -f(\xi)$ for every cut ξ, by Lemma 7.4.4 (d).

By totality (Theorem 7.2.1), each of β and γ is $\geq D_0$ or $\leq D_0$, so one of the following cases applies.

Case 1: $\beta, \gamma \geq D_0$. This is (A).

Case 2: $\beta, \gamma \leq D_0$. Then $-\beta, -\gamma \geq D_0$ (Theorem 7.3.4 (f)), and $-(\beta+\gamma) = (-\beta)+(-\gamma)$ (Theorem 7.3.4 (e)). By (N), (A) and (N),

$$\begin{aligned} f(\beta+\gamma) &= -f\big((-\beta)+(-\gamma)\big) = -\big(f(-\beta)+f(-\gamma)\big) \\ &= -\big((-f(\beta))+(-f(\gamma))\big) = f(\beta)+f(\gamma), \end{aligned}$$

where the last step is Theorem 7.3.4 (e) twice.

Case 3: $\beta \geq D_0 \geq \gamma$ *and* $\beta+\gamma \geq D_0$. Then $\beta = (\beta+\gamma)+(-\gamma)$, with both summands $\geq D_0$. By (A) and (N), $f(\beta) = f(\beta+\gamma)+f(-\gamma) = f(\beta+\gamma)+(-f(\gamma))$. Adding $f(\gamma)$ to both sides gives $f(\beta)+f(\gamma) = f(\beta+\gamma)$.

Case 4: $\beta \geq D_0 \geq \gamma$ *and* $\beta+\gamma < D_0$. Then $-\gamma = \big(-(\beta+\gamma)\big)+\beta$, since $-(\beta+\gamma)+\beta = (-\beta)+(-\gamma)+\beta = -\gamma$; both summands are $\geq D_0$. By (N) and (A), $-f(\gamma) = -f(\beta+\gamma)+f(\beta)$. Adding $f(\beta+\gamma)+f(\gamma)$ to both sides gives $f(\beta+\gamma) = f(\beta)+f(\gamma)$.

Case 5: $\gamma \geq D_0 \geq \beta$. Exchange β and γ in Cases 3 and 4; both sides of the claim are symmetric in β and γ (Theorem 7.3.4 (a)).

This proves the lemma for $\alpha > D_0$. Let $\alpha < D_0$. Then $-\alpha > D_0$, and by Lemma 7.4.4 (d), $\alpha\xi = -\big((-\alpha)\xi\big)$ for every cut ξ. So, by the case just proved (applied to $-\alpha$),

$$\begin{aligned} \alpha(\beta+\gamma) &= -\big((-\alpha)(\beta+\gamma)\big) = -\big((-\alpha)\beta+(-\alpha)\gamma\big) \\ &= \big(-\big((-\alpha)\beta\big)\big)+\big(-\big((-\alpha)\gamma\big)\big) = \alpha\beta+\alpha\gamma, \end{aligned}$$

where the third equality is Theorem 7.3.4 (e).

■

The answer to the prompt. Case 3 uses $-\gamma \geq D_0$ to apply (A) to $(\beta+\gamma)+(-\gamma)$, and Case 2 uses it, with $-\beta \geq D_0$, to apply (A) to $(-\beta)+(-\gamma)$. Case 4 does not: there $-\gamma$ is the sum, and the summands are $-(\beta+\gamma)$ and β.

7.4.4 Reciprocals

For the reciprocal we repeat the idea of the negative, with division in place of subtraction. For $\alpha > D_0$, the reciprocal should consist of the rationals $p \leq 0$ together with the positive rationals below $1/q$ for some rational q outside α. The analogue of the margin is a **strict** inequality.

The reciprocal of a positive cut: the non-positive rationals, and the positive rationals strictly below the reciprocal of some point outside.

Definition 7.4.3. Reciprocal of a Cut

Let α be a cut. If $\alpha > D_0$, its **reciprocal** is the set

$$\alpha^{-1} := \{p \in \mathbb{Q} : p \leq 0\} \cup \{p \in \mathbb{Q} : p > 0 \text{ and } pq < 1 \text{ for some } q \in \mathbb{Q} \setminus \alpha\}.$$

If $\alpha < D_0$, its reciprocal is $\alpha^{-1} := -((-\alpha)^{-1})$. The cut D_0 has no reciprocal. The quantifier in the set for $\alpha > D_0$ is **some**: one point q outside α suffices, and it may depend on p.

Every $q \notin \alpha$ is positive when $\alpha > D_0$ (Lemma 7.1.1 against $0 \in \alpha$), so for $p > 0$ the condition $pq < 1$ says $p < 1/q$. For $\alpha < D_0$, the cut $-\alpha$ is positive (Theorem 7.3.4 (f)), so the definition for $\alpha < D_0$ refers to the one for $\alpha > D_0$.

Example. $D_2^{-1} = D_{1/2}$. The outside of D_2 is $\{q : q \geq 2\}$. A positive $p < 1/2$ is in D_2^{-1}, with $q = 2$: $2p < 1$. A positive $p \geq 1/2$ is not, since every $q \geq 2$ gives $pq \geq 1$. With the non-positive rationals, $D_2^{-1} = \{p : p < 1/2\}$.

A non-example by minimal change. Replace "$pq < 1$" by "$pq \leq 1$". For D_2 the set becomes $\{p : p \leq 1/2\}$, which has a largest element. This is the multiplicative mirror of Example (The Negatives of the Outside), and Exercise C1 shows that, as there, it fails exactly for the rational cuts.

Theorem 7.4.6. Reciprocals of Cuts

Let α be a cut with $\alpha \neq D_0$. Then α^{-1} is a cut, $\alpha^{-1} > D_0$ if $\alpha > D_0$, $\alpha^{-1} < D_0$ if $\alpha < D_0$, and

$$\alpha\alpha^{-1} = D_1.$$

Idea. For $\alpha > D_0$, the cut conditions are checked as for the negative; the strict inequality $pq < 1$ leaves room above p, as the margin did. The inclusion $\alpha\alpha^{-1} \subseteq D_1$ is separation: $r < q$ for $r \in \alpha$, $q \notin \alpha$, so $rs < qs < 1$.

For $\supseteq$, work backwards. Given $0 < p < 1$, we want positive $r \in \alpha$ and $s \in \alpha^{-1}$ with $rs = p$. So $s = p/r$, and "$s \in \alpha^{-1}$" asks for some $q \notin \alpha$ with $(p/r)q < 1$, that is, $q < r/p$. So we need **a point r inside and a point outside below r/p**: a point inside and a point outside whose **ratio** is less than $1/p$. The multiplicative walk (Lemma 7.4.2) supplies the ratio μ, and μ must be chosen strictly between 1 and $1/p$, for the same reason that ε was $-s/2$ and not $-s$: we take the midpoint $\mu := (1 + 1/p)/2$. For $\alpha < D_0$, the sign rules move everything to $-\alpha$.

Exercise B2 asks for the last step of the proof; the fold below has the proof in full.

↻ **Second reading** · *The proof of the reciprocal theorem (try Exercise B2 first)*

Proof. *Let $\alpha > D_0$.* Then $0 \in \alpha$ (Proposition 7.1.2), so every $q \notin \alpha$ is positive (Lemma 7.1.1), and α has a positive element (C3).

(C1) $0 \in \alpha^{-1}$. Let $a \in \alpha$ with $a > 0$. Then $1/a \notin \alpha^{-1}$: it is positive, and every $q \notin \alpha$ has $q > a$ (Lemma 7.1.1), so $(1/a)q > (1/a)a = 1$.

(C2) Let $p \in \alpha^{-1}$ and $p' < p$. If $p' \le 0$, then $p' \in \alpha^{-1}$. Otherwise $0 < p' < p$, so $p > 0$ and $pq < 1$ for some $q \notin \alpha$; as $q > 0$, $p'q < pq < 1$, and $p' \in \alpha^{-1}$ with the same q.

(C3) Let $p \in \alpha^{-1}$. If $p \le 0$, take any $q \notin \alpha$ (C1); then $1/(2q) > 0 \ge p$, and $(1/(2q))q = 1/2 < 1$, so $1/(2q) \in \alpha^{-1}$. If $p > 0$ with $pq < 1$, $q \notin \alpha$, then $p < 1/q$; the midpoint $p'' \coloneqq (p+1/q)/2$ has $p < p'' < 1/q$ (Proposition 4.1.2 (b)), so $p'' > 0$ and $p''q < 1$, and $p'' \in \alpha^{-1}$.

So α^{-1} is a cut, and $\alpha^{-1} > D_0$ because $0 \in \alpha^{-1}$ (Proposition 7.1.2). The product $\alpha\alpha^{-1}$ is therefore given by Definition 7.4.1.

($\subseteq$) Let $x \in \alpha\alpha^{-1}$: $x \le rs$ with positive $r \in \alpha$ and $s \in \alpha^{-1}$, and $sq < 1$ for some $q \notin \alpha$. By Lemma 7.1.1, $r < q$, so $rs < qs < 1$ since $s > 0$. Hence $x < 1$ and $x \in D_1$.

($\supseteq$) Let $p \in D_1$, so $p < 1$. If $p \le 0$, then $p \in \alpha\alpha^{-1}$, since positive elements of α and of α^{-1} exist and their product is above p. Let $0 < p < 1$, and put

$$\mu \coloneqq \frac{1}{2}\Big(1 + \frac{1}{p}\Big);$$

it depends only on p. Since $1/p > 1$ (Proposition 2.5.2 (b)), μ is the midpoint of $1 < 1/p$, so $1 < \mu < 1/p$. By Lemma 7.4.2 there is $r \in \alpha$ with $r > 0$ and $\mu r \notin \alpha$; fix one [$\exists$E]. Put $s \coloneqq p/r > 0$. Then

$$s \cdot (\mu r) = p\mu < p \cdot \frac{1}{p} = 1,$$

and $\mu r \notin \alpha$, so $s \in \alpha^{-1}$ [$\exists$I] (Logic and Proof, Definition 5.5.1), with witness $q \coloneqq \mu r$. Since $p = rs \le rs$, $p \in \alpha\alpha^{-1}$ [$\exists$I]. As p was arbitrary, $D_1 \subseteq \alpha\alpha^{-1}$ [$\forall$I] (Logic and Proof, Definition 5.4.1).

Let $\alpha < D_0$. Then $-\alpha > D_0$, so by the first part $\beta \coloneqq (-\alpha)^{-1}$ is a cut with $\beta > D_0$ and $(-\alpha)\beta = D_1$. Then $\alpha^{-1} = -\beta$ is a cut (Definition 7.3.2), and $-\beta < D_0$ (Theorem 7.3.4 (f)). By Lemma 7.4.4 (d), used twice,

$$\alpha\alpha^{-1} = \alpha(-\beta) = -(\alpha\beta) = (-\alpha)\beta = D_1.$$

■

The walk up by ε has now produced both inverses: the additive one through Lemma 7.1.3 directly, the multiplicative one through its multiplicative form. No other law of this section constructs a point by walking up.

7.4.5 The cuts form an ordered field

Theorem 7.4.7. The Cuts Form an Ordered Field

With the addition of Definition 7.3.1, the multiplication of Definition 7.4.2 and the order of Definition 7.2.1, $\mathbb{R}_{\mathrm{D}}$ is an ordered field (Definition 2.3.2), with zero D_0 and one D_1.

Proof. Let α, β, γ be cuts. Both operations map $\mathbb{R}_{\mathrm{D}} \times \mathbb{R}_{\mathrm{D}}$ into $\mathbb{R}_{\mathrm{D}}$: sums by Theorem 7.3.1, products as noted after Definition 7.4.2. The table names, for each clause of Definition 2.3.1

and Definition 2.3.2, the line that proves it.

Clause	Statement for cuts	Proved by
(F1), (F2)	$+$ commutative, associative	Theorem 7.3.1
(F3)	$\alpha + D_0 = \alpha$	Proposition 7.3.2
(F4)	$\alpha + (-\alpha) = D_0$	Theorem 7.3.3
(F5)	$(\alpha\beta)\gamma = \alpha(\beta\gamma)$	Lemma 7.4.4 (f)
(F6)	$\alpha\beta = \beta\alpha$	Lemma 7.4.4 (e)
(F7)	$\alpha D_1 = \alpha$	Lemma 7.4.4 (g)
(F8)	$\alpha\alpha^{-1} = D_1$ for $\alpha \neq D_0$	Theorem 7.4.6
(F9)	$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$	Lemma 7.4.5
$1 \neq 0$	$D_1 \neq D_0$	$0 \in D_1 \setminus D_0$
(O1)	exactly one of $\alpha < \beta$, $\alpha = \beta$, $\beta < \alpha$	Theorem 7.2.1
(O2)	$\alpha < \beta < \gamma \Rightarrow \alpha < \gamma$	below
(O3)	$\alpha < \beta \Rightarrow \alpha + \gamma < \beta + \gamma$	Theorem 7.3.4 (d)
(O4)	$D_0 < \alpha, \beta \Rightarrow D_0 < \alpha\beta$	Proposition 7.4.1 (a)

For (O2): if $\alpha \subsetneq \beta \subsetneq \gamma$, then $\alpha \subseteq \gamma$ (Sets, Relations, and Cardinality, Proposition 1.1.2), and $\alpha = \gamma$ would give $\beta \subseteq \gamma = \alpha \subseteq \beta$, so $\alpha = \beta$ (Sets, Relations, and Cardinality, Proposition 1.1.1), against $\alpha < \beta$. For (O4): for $\alpha, \beta > D_0$ the product of Definition 7.4.2 is that of Definition 7.4.1, as noted after the definition. Every clause is covered, so $\mathbb{R}_{\mathrm{D}}$ is an ordered field.

■

Nothing in the table is left to an exercise. Two rows are proved by walking up: (F4), through Lemma 7.1.3, and (F8), through Lemma 7.4.2. Several others, such as (O3), (F5)–(F7) and (F9), lean on (F4) through the cancellation law and $-(-\alpha) = \alpha$ of Theorem 7.3.4 (e). The order rows cost almost nothing, because the order is inclusion; on the Cauchy road of Chapter 5 the order took a section of its own. Section 7.5 puts $\mathbb{Q}$ inside $\mathbb{R}_{\mathrm{D}}$ and adds completeness, and Section 7.6 compares the result with Chapter 5's field.

7.4.6 Exercises

7.4.6.1 A. Check your understanding

Exercise. A1

(a) Why is the set $\{pq : p \in \alpha,\ q \in \beta\}$ useless as a product for $\alpha = \beta = D_1$?

(b) In the table of the proof of Theorem 7.4.7, which rows are proved by a lemma that walks up through a cut, Lemma 7.1.3 or Lemma 7.4.2? Which further rows depend on such a row?

(c) Compute $|D_{-1}|$, $\operatorname{sgn} D_{-1}$, $|D_{1/2}|$, $\operatorname{sgn} D_{1/2}$ and $D_{-1}D_{1/2}$ from the definitions.

(d) True or false: if $\alpha > D_0$, then α^{-1} contains every negative rational. Give a reason.

Solution.

(a) It is all of $\mathbb{Q}$, so it fails (C1): $x > 0$ is $(-1)(-x)$, $0 = 0 \cdot 0$, and $x < 0$ is $(1/2)(2x)$, with every factor in D_1.

(b) (F4), the additive inverse: Theorem 7.3.3 uses Lemma 7.1.3. (F8), the multiplicative inverse: Theorem 7.4.6 uses Lemma 7.4.2, whose proof uses Lemma 7.1.3. Further rows depend on (F4): the strict form of (O3) uses cancellation, and (F5)–(F7) and (F9) use Lemma 7.4.4 and Lemma 7.4.5, which use $-(-\alpha) = \alpha$, cancellation, or both (Theorem 7.3.4 (e)), all proved from $\alpha + (-\alpha) = D_0$. The rows (F1)–(F3), $1 \neq 0$, (O1), (O2) and (O4) use neither lemma.

(c) $D_{-1} < D_0$: $D_{-1} > D_0$ fails since $0 \notin D_{-1}$, and $D_{-1} \neq D_0$ since $-1/2 \in D_0 \setminus D_{-1}$. So $\operatorname{sgn} D_{-1} = -1$ and $|D_{-1}| = -D_{-1} = D_1$ (Exercise (B1) with $q = -1$). Since $0 \in D_{1/2}$, $D_{1/2} > D_0$, so $\operatorname{sgn} D_{1/2} = 1$ and $|D_{1/2}| = D_{1/2}$. The signs differ, so

$$D_{-1}D_{1/2} = -(D_1 D_{1/2}) = -(D_{1/2}D_1) = -D_{1/2} = D_{-1/2},$$

by Proposition 7.4.1 (b) and (d), and Exercise (B1) with $q = 1/2$.

(d) True: by Definition 7.4.3, α^{-1} contains every $p \leq 0$, with no condition.

7.4.6.2 B. Practice

Exercise. B1

Prove from Definition 7.4.1 that $D_2 D_3 = D_6$.

Hint: for $0 < x < 6$, scale 2 and 3 down by suitable factors.

Solution. $D_2, D_3 > D_0$, since $0 \in D_2$ and $0 \in D_3$.

($\subseteq$) Let $p \leq rs$ with $0 < r < 2$, $0 < s < 3$. Then $rs < 2s < 6$ (Proposition 2.3.5 (c), twice), so $p < 6$.

($\supseteq$) Let $x < 6$. If $x \leq 0$, then $x \leq 1 \cdot 1$ with $1 \in D_2$ and $1 \in D_3$. Let $0 < x < 6$, and put

$$\lambda := \frac{1}{2}\left(1 + \frac{x}{6}\right), \qquad r := 2\lambda, \qquad s := \frac{x}{2\lambda}.$$

Since $0 < x/6 < 1$, λ is the midpoint of $x/6 < 1$, so $x/6 < \lambda < 1$. Hence $0 < r < 2$, so $r \in D_2$. Also $s > 0$, and $s < 3$ because $x < 6\lambda$; so $s \in D_3$. Finally $rs = x$, so $x \in D_2 D_3$. Hence $D_2 D_3 = D_6$.

Exercise. B2: The last step of the reciprocal

Let α be a cut with $\alpha > D_0$, and let $p \in \mathbb{Q}$ with $0 < p < 1$. We show that $p \in \alpha\alpha^{-1}$.

(1) Let $\mu := (1 + 1/p)/2$. Since $1/p > 1$, μ is the midpoint of 1 and $1/p$, so $1 < \mu < 1/p$.

(2) By Lemma 7.4.2 there is $r \in \alpha$ with $r > 0$ and $\mu r \notin \alpha$.

(3) Hence $p \in \alpha\alpha^{-1}$.

Name the move, then complete it: write out step (3).

Solution. The move is the **walk up by** ε, in its multiplicative form: step (2) walks up through α until the next point is outside, which gives a point inside and a point outside whose ratio is μ, less than $1/p$.

(3) Put $s := p/r$; it is positive. Then

$$s \cdot (\mu r) = p\mu < p \cdot \frac{1}{p} = 1,$$

using $\mu < 1/p$ and $p > 0$. Since $\mu r \notin \alpha$, the rational $q := \mu r$ witnesses $s \in \alpha^{-1}$ (Definition 7.4.3). Now $r \in \alpha$ and $s \in \alpha^{-1}$ are positive and $p = rs \leq rs$, so $p \in \alpha\alpha^{-1}$ by Definition 7.4.1.

Why μ and not $1/p$ itself: with $\mu = 1/p$, step (2) gives r with $r/p \notin \alpha$, and $s = p/r$ would need some $q \notin \alpha$ with $(p/r)q < 1$, that is, $q < r/p$. The point r/p itself does not qualify, and there may be no smaller one outside. The strict room between μ and $1/p$ plays the part of the margin.

Exercise. B3

Prove that $(-\alpha)(-\beta) = \alpha\beta$ for all cuts α and β.

Solution. By Lemma 7.4.4 (d), applied first with $-\alpha$ in front and then to $\alpha(-\beta)$,

$$(-\alpha)(-\beta) = -\big(\alpha(-\beta)\big) = -\big(-(\alpha\beta)\big) = \alpha\beta,$$

where the last step is $-(-\xi) = \xi$ (Theorem 7.3.4 (e)) with $\xi = \alpha\beta$.

7.4.6.3 C. Going deeper

Exercise. C1: The reciprocal without strictness

For a cut $\alpha > D_0$, let

$$\begin{aligned}\rho(\alpha) := {} & \{p \in \mathbb{Q} : p \leq 0\} \\ & \cup \{p \in \mathbb{Q} : p > 0 \text{ and } pq \leq 1 \text{ for some } q \in \mathbb{Q} \setminus \alpha\}.\end{aligned}$$

Prove that $\rho(\alpha)$ is a cut if and only if α is not a rational cut, that is, $\alpha \neq D_a$ for every $a \in \mathbb{Q}$.

Solution (Hint 1). Compute $\rho(D_a)$ first. Which element of the outside of D_a does all the work?

Solution (Hint 2). If α is not rational, the outside $\mathbb{Q} \setminus \alpha$ has no least element: a least element m would force $\alpha = D_m$.

Solution. Let $\alpha > D_0$. Then $0 \in \alpha$, every $q \notin \alpha$ is positive (Lemma 7.1.1), and α has a positive element.

If $\alpha = D_a$, then $\rho(\alpha)$ is not a cut. Here $a > 0$, since $0 \in D_a$, and the outside is $\{q : q \geq a\}$. For $p > 0$: if $pq \leq 1$ with $q \geq a$, then $pa \leq pq \leq 1$; conversely $pa \leq 1$ gives the witness $q = a$. So $\rho(D_a) = \{p : p \leq 0\} \cup \{p > 0 : p \leq 1/a\} = \{p : p \leq 1/a\}$. It has the largest element $1/a$, so (C3) fails.

If α is not rational, then $\rho(\alpha)$ is a cut. First, $\mathbb{Q} \setminus \alpha$ has no least element. Suppose m were one. Every $r < m$ is then in α (otherwise r is a smaller element of the outside), and every $r \in \alpha$ has $r < m$ (Lemma 7.1.1). So $\alpha = D_m$, against the hypothesis.

(C1) $0 \in \rho(\alpha)$. For a positive $a \in \alpha$, $1/a \notin \rho(\alpha)$: every $q \notin \alpha$ has $q > a$, so $(1/a)q > 1$.

(C2) As for α^{-1}: if $0 < p' < p$ and $pq \leq 1$ with $q > 0$, then $p'q < pq \leq 1$; and every $p' \leq 0$ is in $\rho(\alpha)$.

(C3) If $p \leq 0$, then for any $q \notin \alpha$, $1/(2q)$ is in $\rho(\alpha)$ and larger. Let $p > 0$ with $pq \leq 1$, $q \notin \alpha$. Since q is not least in the outside, there is $q' \notin \alpha$ with $q' < q$; then $q' > 0$ and $pq' < pq \leq 1$, so $p < 1/q'$. The midpoint $p'' := (p + 1/q')/2$ has $p < p'' < 1/q'$, so $p''q' < 1$, and $p'' \in \rho(\alpha)$ is larger than p.

So $\rho(\alpha)$ is a cut exactly when α is not rational. For such α, $\rho(\alpha) = \alpha^{-1}$: the argument for (C3) shows that $pq \leq 1$ with $q \notin \alpha$ gives $pq' < 1$ with $q' \notin \alpha$, and the converse is immediate. As with the negatives of the outside in Section 7.3, the non-strict recipe works for the cut σ of Example (The cut of the gap), which is not a rational cut (Exercise C1 of Section 7.1), and fails on the simplest cuts; the strict inequality in Definition 7.4.3 is the multiplicative margin.

Marking scheme. (10 points) - 3: shows that $\rho(D_a)$ has the largest element $1/a$ ($1/a \in \rho(D_a)$ and every element is $\leq 1/a$; the full equality $\rho(D_a) = \{p : p \leq 1/a\}$ also earns this), and names (C3) as the failing condition. - 2: proves that the outside of a non-rational cut has no least element. - 1: (C1) and (C2) for $\rho(\alpha)$. - 3: (C3) for $\rho(\alpha)$, using a smaller outside point q' to make the inequality strict. - 1: states the equivalence with both directions. Second route: showing $\rho(\alpha) = \alpha^{-1}$ for non-rational α ($\subseteq$ by a smaller outside point q' making $pq' < 1$, $\supseteq$ since $<$ implies $\leq$) and citing Theorem 7.4.6 earns the 1 point for (C1) and (C2) and the 3 points for (C3). The 2 points for the no-least-element step still require that step. Fatal errors (0 for the step): assuming $q > 0$ without separation against $0 \in \alpha$; "a smaller outside point exists" asserted for every cut.

7.5 The Rationals Inside the Cuts, and a Complete Ordered Field

This section is part of the optional Chapter 7, and nothing outside the chapter depends on it. Section 7.4 finished the arithmetic: by Theorem 7.4.7 the cuts form an ordered field $\mathbb{R}_D$, with zero D_0 and one D_1. Three questions are left before $\mathbb{R}_D$ can be called a model of the real numbers. Do the rational cuts D_q form a faithful copy of $\mathbb{Q}$, with the arithmetic of $\mathbb{Q}$? Is the new field complete? And does the cut σ of Example (The cut of the gap), built to sit at the gap of $\mathbb{Q}$, really square to D_2? This section answers all three with yes. On the Cauchy road, Section 5.4 asked the first question about the map κ (Theorem 5.4.1), and Sections 5.5 and 5.6

together answered the second; here the second answer is already in hand, from Section 7.2.

7.5.1 A copy of the rationals

Every ordered field F contains a copy of $\mathbb{Q}$: by Theorem 4.2.2 there is exactly one map $\iota_F\colon \mathbb{Q} \to F$ that sends 1 to 1_F and respects sums and products, and it is strictly increasing. For $F = \mathbb{R}_{\mathrm{D}}$ we have a candidate in hand, the map $q \mapsto D_q$ of Definition 7.1.2. It sends 1 to D_1, the one of the field. The theorem below checks the rest: sums and positive products by computing with sets, negatives and the other signs by the field rules; and the uniqueness clause then says the candidate is $\iota_{\mathbb{R}_{\mathrm{D}}}$.

Theorem 7.5.1. The Rationals Inside the Cuts

The map $\mathbb{Q} \to \mathbb{R}_{\mathrm{D}}$, $q \mapsto D_q$, is injective, and for all $p, q \in \mathbb{Q}$:

(a) $p < q$ if and only if $D_p < D_q$;
(b) $D_p + D_q = D_{p+q}$;
(c) $-D_q = D_{-q}$;
(d) $D_p D_q = D_{pq}$;
(e) D_1 is the one of the ordered field $\mathbb{R}_{\mathrm{D}}$ of Theorem 7.4.7.

Consequently $D_q = \iota_{\mathbb{R}_{\mathrm{D}}}(q)$ for every $q \in \mathbb{Q}$, where $\iota_{\mathbb{R}_{\mathrm{D}}}$ is the map of Theorem 4.2.2.

Idea. The identity (b), and (d) for positive p, q, are equalities of sets of rationals, so each has two halves. The $\subseteq$ halves are closure: if $r < p$ and $s < q$, then $r + s < p + q$. The $\supseteq$ halves must **split** a given rational $x < p + q$ as a sum of something below p and something below q, and the obvious split $x = p + (x - p)$ fails, because $p \notin D_p$. So we take a little off both summands: half the gap $p + q - x$ from each. For products we do the same with a factor instead of a summand. Signs other than "both positive" are reduced to that case, the chapter's named move of reducing to the positive case (Section 7.4); since $\mathbb{R}_{\mathrm{D}}$ is already known to be a field, the ring rules for signs do the reduction.

Proof. Throughout, $\mathbb{R}_{\mathrm{D}}$ is an ordered field (Theorem 7.4.7), so the rules of Proposition 2.3.5 hold in it as well as in $\mathbb{Q}$.

(a) If $p < q$, then $p \in D_q$, so $D_p \subsetneq D_q$ by Proposition 7.1.2 (a) applied to the cut D_q; that is, $D_p < D_q$ (Definition 7.2.1). Conversely, if $D_p \subsetneq D_q$, then Proposition 7.1.2 (a), applied to the cut D_q and the rational p, gives $p \in D_q$, that is, $p < q$.

Injectivity. If $p \neq q$, then $p < q$ or $q < p$ by trichotomy (O1) in $\mathbb{Q}$, and (a) gives $D_p \neq D_q$ in either case.

(b) ($\subseteq$) An element of $D_p + D_q$ is $r + s$ with $r < p$ and $s < q$ (Definition 7.3.1), and $r + s < p + q$ by Proposition 2.3.5 (b). ($\supseteq$) Let $x \in \mathbb{Q}$ with $x < p + q$, and put

$$\delta := \frac{p + q - x}{2} > 0.$$

Then $p - \delta < p$ and $q - \delta < q$, so $p - \delta \in D_p$ and $q - \delta \in D_q$, and $(p - \delta) + (q - \delta) = p + q - 2\delta = x$. Hence $x \in D_p + D_q$.

(c) By (b), $D_q + D_{-q} = D_{q+(-q)} = D_0$. By Theorem 7.3.3, also $D_q + (-D_q) = D_0$. The negative of an element of the commutative ring $\mathbb{R}_{\mathrm{D}}$ is unique (Proposition 2.3.2 (a)), so $-D_q = D_{-q}$.

(d), both positive. Let $p > 0$ and $q > 0$. By (a), $D_p > D_0$ and $D_q > D_0$, so the product is that of Definition 7.4.1:

$$D_p D_q = \{x \in \mathbb{Q} : x \le rs \text{ for some } r \in D_p,\ s \in D_q \text{ with } r > 0,\ s > 0\}.$$

($\subseteq$) If $0 < r < p$ and $0 < s < q$, then $rs < ps < pq$ by Proposition 2.3.5 (c), used twice. So $x \le rs$ gives $x < pq$.

($\supseteq$) Let $x < pq$. If $x \le 0$, take $r := p/2$ and $s := q/2$: they are positive, lie in D_p and D_q, and $x \le 0 < rs$. If $0 < x < pq$, put

$$\lambda := \frac{1}{2}\Big(1 + \frac{x}{pq}\Big), \qquad r := \lambda p, \qquad s := \frac{x}{\lambda p}.$$

Since $0 < x/(pq) < 1$, the midpoint λ satisfies $x/(pq) < \lambda < 1$. Hence $0 < r < p$; and $s > 0$, with $s < q$ because this is equivalent, after multiplying by $\lambda p > 0$, to $x < \lambda pq$, that is, to $x/(pq) < \lambda$. So $r \in D_p$, $s \in D_q$, both are positive, and $rs = x$. Hence $x \in D_p D_q$.

(d), the other signs. If $p = 0$ or $q = 0$, then one factor is D_0, so $D_p D_q = D_0$ by Definition 7.4.2, and $D_{pq} = D_0$. For negative entries, (c) gives $D_p = -D_{-p}$ for every p, since $-(-\alpha) = \alpha$ (Proposition 2.3.2 (a)). If $p < 0 < q$, then by the sign rule $(-a)b = -(ab)$ of Proposition 2.3.2 (c) in $\mathbb{R}_{\mathrm{D}}$ (in the cut model it is part of Lemma 7.4.4) and the positive case,

$$D_p D_q = (-D_{-p})D_q = -(D_{-p}D_q) = -D_{-pq} = D_{pq},$$

the last step by (c). The case $q < 0 < p$ follows by commutativity of the product. If $p < 0$ and $q < 0$, then the rule $(-a)(-b) = ab$ of Proposition 2.3.2 (c) and the positive case give

$$D_p D_q = (-D_{-p})(-D_{-q}) = D_{-p}D_{-q} = D_{(-p)(-q)} = D_{pq}.$$

(e) This is part of Theorem 7.4.7.

The consequence. By (e), (b) and (d), the map $q \mapsto D_q$ sends 1 to the one of $\mathbb{R}_{\mathrm{D}}$ and respects sums and products. By the uniqueness clause of Theorem 4.2.2, applied to the ordered field $F = \mathbb{R}_{\mathrm{D}}$, it is $\iota_{\mathbb{R}_{\mathrm{D}}}$. This proves the theorem.

■

So the rationals live inside $\mathbb{R}_{\mathrm{D}}$ as the rational cuts, with their order and arithmetic unchanged. From now on, a computation with rational cuts may be done in $\mathbb{Q}$ and carried over: $D_{1/2}+D_{1/3} = D_{5/6}$, $D_{-2}D_3 = D_{-6}$, $D_2^{-1} = D_{1/2}$. The last one holds because $D_2 D_{1/2} = D_1$ by (d), and in a field the inverse is the only element with that property (Lemma 4.1.1 (a)).

Quick check. In the ($\supseteq$) half of (b), why not write $x = p + (x - p)$? It has $x - p < q$.

Solution. Because $p \notin D_p$: the cut D_p consists of the rationals **strictly** below p. The split must take something from p as well, and $\delta > 0$ is that something. The flawed proof of

Exercise C2 of Section 7.3 falls into the same trap, with $0 \notin D_0$.

On the Cauchy road the same theorem was Theorem 5.4.1, and its proof was a computation with constant sequences. There the sum of two constant sequences is the constant sequence of the sum, termwise. Here the sum of two rational cuts is not obviously a rational cut, and the $\supseteq$ half needs the room that δ provides.

Proposition 7.5.2. A Cut Is the Union of the Rational Cuts Below It

Optional: nothing later depends on this.

For every cut α,

$$\alpha = \bigcup\{D_q : q \in \alpha\} = \sup\{D_q : q \in \alpha\}.$$

Proof. ($\supseteq$) If $q \in \alpha$ and $r < q$, then $r \in \alpha$ by (C2) of Definition 7.1.1; so $D_q \subseteq \alpha$. ($\subseteq$) If $p \in \alpha$, then (C3) gives $p' \in \alpha$ with $p' > p$, and $p \in D_{p'}$. This proves the first equality. The set $\{D_q : q \in \alpha\}$ is non-empty, since $\alpha \neq \emptyset$, and bounded above by α, by the $\supseteq$ half; so by Theorem 7.2.2 its supremum is its union. ∎

The proposition is the cut model's way of saying that every real number is approached from below by rationals. Nothing later in the chapter uses it.

7.5.2 Between two cuts there is a rational cut

In $\mathbb{Q}$ between two rationals lies a third, their midpoint. In an Archimedean field, Theorem 4.2.5 puts a rational between any two elements. For $\mathbb{R}_{\mathrm{D}}$ we prove the corresponding statement directly from the cut conditions. We do not yet know that $\mathbb{R}_{\mathrm{D}}$ is Archimedean, and the direct proof is shorter anyway.

Theorem 7.5.3. The Rational Cuts Are Dense

Let α and β be cuts with $\alpha < \beta$. Then there is $q \in \mathbb{Q}$ with $\alpha < D_q < \beta$.

Try it first (ten minutes): given cuts $\alpha < \beta$, which rational would you put between them, and why is one point of $\beta \setminus \alpha$ not quite enough?

Idea. Since $\alpha \subsetneq \beta$, there is a rational p in β but not in α, and the natural guess is $q = p$. By Proposition 7.1.2, $p \notin \alpha$ gives $\alpha \subseteq D_p$ (part (b)), and $p \in \beta$ gives $D_p \subsetneq \beta$ (part (a)). So $\alpha \leq D_p < \beta$. The first inequality need not be strict: for $\alpha = D_1$, $\beta = D_2$ and $p = 1$, the guess gives $D_p = \alpha$. The cure is to move up once more **inside** β: condition (C3) of β gives $p' \in \beta$ above p, and $D_{p'}$ is strictly above D_p, so strictly above α, and still strictly below β.

Proof. Since $\alpha \subsetneq \beta$ (Definition 7.2.1), there is $p \in \beta$ with $p \notin \alpha$. By (C3) of Definition 7.1.1 for β, there is $p' \in \beta$ with $p' > p$. Then:

- $\alpha \subseteq D_p$, by Proposition 7.1.2 (b), since $p \notin \alpha$;
- $D_p \subsetneq D_{p'}$, by 1 (a), since $p < p'$;
- $D_{p'} \subsetneq \beta$, by Proposition 7.1.2 (a), since $p' \in \beta$.

Hence $\alpha \subsetneq D_{p'} \subsetneq \beta$, that is, $\alpha < D_{p'} < \beta$. Take $q := p'$.

■

Each hypothesis has a job. "$\alpha < \beta$" supplies p, and the strictness is needed: for $\alpha = \beta$ there is nothing strictly between. Condition (C3), used for β, supplies the second step up; without it, D_1 and the set $\{r \in \mathbb{Q} : r \leq 1\}$ would be two "cuts", the first strictly inside the second, with no D_q strictly between them: $D_1 \subsetneq D_q$ forces $q > 1$, and then D_q contains $(1+q)/2 > 1$, which the second set lacks.

Warning. "Dense" does not mean that every cut is rational. The rational cuts are dense, and still σ of Example (The cut of the gap) is not one of them, as Proposition 7.5.5 below proves. Density says only that between two cuts there is a rational cut, never that a cut **is** one.

Quick check. Name a rational q with $\sigma < D_q < D_{3/2}$, and check both inequalities.

Solution. Take $q = 29/20$. Then $q > 0$ and $q^2 = 841/400 > 800/400 = 2$, so $q \notin \sigma$, and Proposition 7.1.2 (b) gives $\sigma \subseteq D_q$. The inclusion is strict: $10/7 > 0$ and $(10/7)^2 = 100/49 > 98/49 = 2$, so $10/7 \notin \sigma$, while $10/7 < 29/20$ ($200/140 < 203/140$), so $10/7 \in D_q$. Hence $\sigma < D_q$. And $29/20 < 30/20 = 3/2$, so $D_q < D_{3/2}$ by 1 (a). The rational $q = 10/7$ works the same way, with $17/12$ as the witness of strictness: $(17/12)^2 = 289/144 > 2$ and $17/12 < 10/7$ ($119/84 < 120/84$).

7.5.3 Archimedes in the cuts

An ordered field is Archimedean (Definition 4.2.2) if no element lies above every $n \cdot 1_F$, $n \in \mathbb{N}$. In $\mathbb{R}_{\mathrm{D}}$, $n \cdot 1 = \iota_{\mathbb{R}_{\mathrm{D}}}(n) = D_n$, by Theorem 4.2.2 and 1. A cut is bounded inside $\mathbb{Q}$ by any rational it misses, and $\mathbb{Q}$ is Archimedean, so the property comes almost for free.

Proposition 7.5.4. The Cuts Are Archimedean

Optional: nothing later depends on this.

For every cut α there is $n \in \mathbb{N}$ with $\alpha < D_n$. Hence $\mathbb{R}_{\mathrm{D}}$ is an Archimedean ordered field.

Proof. By (C1) of Definition 7.1.1 there is $b \in \mathbb{Q}$ with $b \notin \alpha$, and Proposition 7.1.2 (b) gives $\alpha \subseteq D_b$. By Theorem 2.5.3 there is $n \in \mathbb{N}$ with $b < n$, so $D_b < D_n$ by 1 (a). Hence $\alpha < D_n = n \cdot 1_{\mathbb{R}_{\mathrm{D}}}$, which is Definition 4.2.2 for $\mathbb{R}_{\mathrm{D}}$.

■

7.5.4 The square root of two

The cut $\sigma := \{p \in \mathbb{Q} : p \le 0 \text{ or } p^2 < 2\}$ of Example (The cut of the gap) was introduced as "the rationals below $\sqrt{2}$". Now that cuts can be multiplied, that description becomes a theorem: $\sigma\sigma = D_2$. The proof is the chapter's multiplicative form of **walk up by** ε, Lemma 7.4.2: step up by a factor $\mu > 1$ until you leave σ.

Proposition 7.5.5. The Cut of the Square Root of Two

The cut σ of Example (The cut of the gap) satisfies $\sigma > D_0$ and $\sigma\sigma = D_2$. Consequently $\sigma \neq D_q$ for every $q \in \mathbb{Q}$.

Idea. The $\subseteq$ half is a bound: two positive elements of σ have squares below 2, so their product has square below 4. The $\supseteq$ half must write a given $x < 2$ as a product rs of positive elements of σ. Small x are easy: $x \le 1 \cdot 1$. For $1 \le x < 2$ we want $r \in \sigma$ with $r^2 > x$, since then $s := x/r$ is smaller than r and so lies in σ. Such an r is found by walking up by a factor: if $r \in \sigma$ and $\mu r \notin \sigma$, then $r^2 \ge 2/\mu^2$, and this exceeds x as soon as $\mu^2 < 2/x$. So we need a rational $\mu > 1$ with $\mu^2 < 2/x$, written down without square roots.

Proof. $\sigma > D_0$. Every $p < 0$ has $p \le 0$, so $D_0 \subseteq \sigma$; and $0 \in \sigma \setminus D_0$. Hence $D_0 \subsetneq \sigma$. So $\sigma\sigma$ is the product of Definition 7.4.1.

($\subseteq$) Let $x \in \sigma\sigma$, so $x \le rs$ with $r, s \in \sigma$, $r > 0$, $s > 0$. Since $r > 0$ and $r \in \sigma$, $r^2 < 2$; likewise $s^2 < 2$. By Proposition 2.3.5 (c), used twice, $(rs)^2 = r^2 s^2 < 2s^2 < 4 = 2^2$. Since $rs \ge 0$ and $2 \ge 0$, Proposition 2.3.5 (g) gives $rs < 2$. So $x < 2$ and $x \in D_2$.

($\supseteq$), *small* x. Let $x < 1$. Then $1 \in \sigma$, since $1^2 = 1 < 2$, and $x \le 1 \cdot 1$ with $r = s = 1$ positive. So $x \in \sigma\sigma$. This covers every $x \le 0$ as well.

($\supseteq$), $1 \le x < 2$. Put

$$t := \frac{2}{x} - 1, \qquad \mu := 1 + \frac{t}{3}.$$

Since $1 \le x < 2$, we have $1 < 2/x \le 2$, so $0 < t \le 1$ and $\mu > 1$. Moreover $t^2/9 < t/3$, because this is equivalent to $t < 3$. Hence

$$\mu^2 = 1 + \frac{2t}{3} + \frac{t^2}{9} < 1 + \frac{2t}{3} + \frac{t}{3} = 1 + t = \frac{2}{x}.$$

Since $\sigma > D_0$ and $\mu > 1$, Lemma 7.4.2 gives $r \in \sigma$ with $r > 0$ and $\mu r \notin \sigma$. As $\mu r > 0$ and $\mu r \notin \sigma$, the definition of σ gives $(\mu r)^2 \ge 2$, so

$$r^2 \ge \frac{2}{\mu^2} > x,$$

the second inequality because $\mu^2 < 2/x$. Put $s := x/r$. Then $s > 0$, and $s < r$ because $x < r^2$. Since $r \in \sigma$, (C2) of Definition 7.1.1 gives $s \in \sigma$. Finally $rs = x$, so $x \in \sigma\sigma$.

So $\sigma\sigma = D_2$.

The consequence. Suppose $\sigma = D_q$ for some $q \in \mathbb{Q}$. Then $D_0 < D_q$, so $q > 0$ by 1 (a). By

1 (d), $D_{q^2} = D_q D_q = \sigma\sigma = D_2$, and by injectivity $q^2 = 2$. This contradicts Theorem 2.6.1. Hence $\sigma \neq D_q$ for every $q \in \mathbb{Q}$.

■

The constant μ depends on x only, and it is chosen before r: the lemma is applied to the fixed factor μ, and it returns r. For $x = 3/2$, for instance, $t = 1/3$ and $\mu = 10/9$, with $\mu^2 = 100/81 < 108/81 = 4/3 = 2/x$. Exercise C1 of Section 7.1 proved $\sigma \neq D_q$ directly from the definition; the proof above gets it as a by-product of $\sigma\sigma = D_2$ and needs no step below a rational q with $q^2 > 2$.

Quick check. Why does the case $1 \leq x < 2$ need Lemma 7.4.2, while the case $x < 1$ does not? Where in the $(\supseteq)$ half is the hypothesis $x < 2$ used?

Solution. For $x < 1$ the fixed element $1 \in \sigma$ already gives $x \leq 1 \cdot 1$. For $x \geq 1$ no fixed pair works for all x: we need $r \in \sigma$ with $r^2 > x$, and how close to the gap r must be depends on how close x is to 2. The hypothesis $x < 2$ is used to get $2/x > 1$, so $t > 0$ and $\mu > 1$, which the lemma requires.

7.5.5 A complete ordered field

The last question has the shortest answer. Completeness in the sense of Definition 5.6.1 asks that every non-empty subset bounded above have a supremum for the order $\leq$ of the field. In $\mathbb{R}_{\mathrm{D}}$, that order is inclusion (Definition 7.2.1), and Theorem 7.2.2 found the supremum: the union.

Theorem 7.5.6. The Cuts Form a Complete Ordered Field

$\mathbb{R}_{\mathrm{D}}$, with the sum of Definition 7.3.1, the product of Definition 7.4.2 and the order of Definition 7.2.1, is a complete ordered field (Definition 5.6.1).

Proof. By Theorem 7.4.7, $\mathbb{R}_{\mathrm{D}}$ is an ordered field. Its relation $\leq$ ("$<$ or $=$", Definition 2.3.2) is "$\subsetneq$ or $=$", that is, $\subseteq$, the order of Definition 7.2.1. Let $A \subseteq \mathbb{R}_{\mathrm{D}}$ be non-empty and bounded above for this order. By Theorem 7.2.2, $\bigcup A$ is a cut and is the supremum of A in $(\mathbb{R}_{\mathrm{D}}, \leq)$. So every non-empty subset bounded above has a supremum, and $\mathbb{R}_{\mathrm{D}}$ is complete.

■

That is the whole proof, and it is the point of the chapter's comparison. On the Cauchy road, Section 5.5 proved that every Cauchy sequence of real numbers converges, and Section 5.6 then needed bisection and the Archimedean property to reach suprema (Theorem 5.6.2). On the Dedekind road the supremum is a union, and all the work went into the arithmetic of Sections 7.3 and 7.4.

With this theorem, Theorem 6.1.2 gives a second proof of Proposition 7.5.4: every complete ordered field is Archimedean. By Theorem 6.3.1, $\mathbb{R}_{\mathrm{D}}$ is therefore a copy of the field $\mathbb{R}$ of Chapter 5, by exactly one order-preserving field isomorphism. Section 7.6 writes that isomorphism down.

7.5.6 Exercises

7.5.6.1 A. Check your understanding

Exercise. A1

(a) Which result shows that the map $q \mapsto D_q$ is the map $\iota_{\mathbb{R}_D}$ of Theorem 4.2.2, and which property of $\iota_{\mathbb{R}_D}$ does that argument use?

(b) True or false, with a reason from the cut conditions: every cut is a union of rational cuts.

(c) In the proof of Theorem 7.5.3, the rational $p \in \beta \setminus \alpha$ was not used directly. Give cuts $\alpha < \beta$ and a rational $p \in \beta \setminus \alpha$ with $D_p = \alpha$.

(d) True or false: if α is a cut and $\alpha \neq D_q$ for every $q \in \mathbb{Q}$, then $\alpha\alpha \neq D_r$ for every $r \in \mathbb{Q}$.

Solution.

(a) (a) Parts (b), (d) and (e) say that $q \mapsto D_q$ respects sums and products and sends 1 to the one of $\mathbb{R}_D$; the **uniqueness** clause of Theorem 4.2.2 says only one map $\mathbb{Q} \to \mathbb{R}_D$ does that, namely $\iota_{\mathbb{R}_D}$.

(b) True. Let α be a cut. For $q \in \alpha$, (C2) gives $D_q \subseteq \alpha$, so the union of the D_q with $q \in \alpha$ is contained in α. Conversely, for $q \in \alpha$, (C3) gives $q' \in \alpha$ with $q' > q$, and then $q \in D_{q'}$. So α is the union of the rational cuts D_q, $q \in \alpha$. (This is the optional proposition on unions of rational cuts in this section; the argument above does not need it.)

(c) $\alpha = D_1$, $\beta = D_2$, $p = 1$: then $1 \in D_2$, $1 \notin D_1$, and $D_p = D_1 = \alpha$. So D_p is not strictly above α, and the proof moves up to $p' \in \beta$ with $p' > p$.

(d) False. σ is not rational (Proposition 7.5.5), and $\sigma\sigma = D_2$ is.

7.5.6.2 B. Practice

Exercise. B1

Let $p, q \in \mathbb{Q}$ with $p < 0 < q$. Prove that $D_p D_q = D_{pq}$ directly from the case distinction of Definition 7.4.2, using the positive case of 1 (d) and part (c), but not the ring rules for signs (Proposition 2.3.2 (c), Lemma 7.4.4).

Solution. By 1 (a), $D_p < D_0 < D_q$. So neither factor is D_0, the signs differ, and Definition 7.4.2 gives

$$D_p D_q = -(|D_p||D_q|).$$

Since $D_q \geq D_0$, $|D_q| = D_q$. Since $D_p < D_0$, $|D_p| = -D_p$, and $-D_p = D_{-p}$ by 1 (c). Now $-p > 0$ and $q > 0$, so the positive case of (d) gives $D_{-p}D_q = D_{(-p)q} = D_{-pq}$. Hence $D_p D_q = -D_{-pq} = D_{pq}$, the last step by (c) again.

Exercise. B2

Let $\rho := \{p \in \mathbb{Q} : p \le 0 \text{ or } p^2 < 3\}$. Exercise B2 of Section 7.1 shows that ρ is a cut (there the set is written with $p < 0$; since $0^2 < 3$, it is the same set). Prove that $\rho > D_0$ and $\rho\rho = D_3$.

Hint: follow Proposition 7.5.5, and check that the same μ still works when $2/x$ is replaced by $3/x$.

Solution. $\rho > D_0$. $D_0 \subseteq \rho$, since $p < 0$ gives $p \le 0$, and $0 \in \rho \setminus D_0$.

($\subseteq$) If $x \le rs$ with $r, s \in \rho$ positive, then $r^2 < 3$ and $s^2 < 3$, so $(rs)^2 = r^2 s^2 < 3s^2 < 9 = 3^2$ by Proposition 2.3.5 (c), and $rs < 3$ by Proposition 2.3.5 (g), since $rs, 3 \ge 0$. So $x < 3$.

($\supseteq$) Let $x < 3$. If $x < 1$, then $x \le 1 \cdot 1$ with $1 \in \rho$. Let $1 \le x < 3$, and put $t := 3/x - 1$ and $\mu := 1 + t/3$. Then $1 < 3/x \le 3$, so $0 < t \le 2$ and $\mu > 1$. Since $t < 3$, $t^2/9 < t/3$, and

$$\mu^2 = 1 + \frac{2t}{3} + \frac{t^2}{9} < 1 + t = \frac{3}{x}.$$

By Lemma 7.4.2 ($\rho > D_0$, $\mu > 1$) there is $r \in \rho$, $r > 0$, with $\mu r \notin \rho$. Then $(\mu r)^2 \ge 3$, so $r^2 \ge 3/\mu^2 > x$. Put $s := x/r$: then $0 < s < r$, so $s \in \rho$ by (C2), and $rs = x$. So $x \in \rho\rho$.

Hence $\rho\rho = D_3$.

The only change from σ is the range of t: now t can be as large as 2, and the inequality $t^2/9 < t/3$ still holds, because it needs only $t < 3$.

7.5.6.3 C. Going deeper

Exercise. C1: An irrational cut in every gap

Prove that between any two cuts $\alpha < \beta$ lies a cut that is not rational: there is a cut γ with $\alpha < \gamma < \beta$ and $\gamma \ne D_r$ for every $r \in \mathbb{Q}$.

Solution (Hint 1). Build an irrational cut between D_0 and D_1 from σ, then move it with rational coefficients.

Solution (Hint 2). By Theorem 7.5.3, used twice, there are rationals q, q' with $\alpha < D_q < D_{q'} < \beta$.

Solution. Two rationals. By Theorem 7.5.3 there is $q \in \mathbb{Q}$ with $\alpha < D_q < \beta$; applied again to $D_q < \beta$, it gives q' with $D_q < D_{q'} < \beta$. By 1 (a), $q < q'$.

An irrational cut between D_0 and D_1. Since $1^2 < 2$, $1 \in \sigma$, so $D_1 < \sigma$ (Proposition 7.1.2 (a)). Since $3/2 > 0$ and $(3/2)^2 = 9/4 > 2$, $3/2 \notin \sigma$, so $\sigma \subseteq D_{3/2}$ (Proposition 7.1.2 (b)), and $D_{3/2} < D_2$ (1 (a)). Hence $D_1 < \sigma < D_2$. Adding $-D_1 = D_{-1}$ (translation in the ordered field $\mathbb{R}_{\mathrm{D}}$, Theorem 7.4.7, and 1 (b), (c)) gives $D_0 < \sigma - D_1 < D_1$.

The cut. Put $\gamma := D_q + D_{q'-q}(\sigma - D_1)$. Since $q' > q$, $D_{q'-q} > D_0$ by 1 (a), and multiplying

by it preserves strict inequalities (Proposition 2.3.5 (c) in $\mathbb{R}_D$), so

$$D_0 < D_{q'-q}(\sigma - D_1) < D_{q'-q}D_1 = D_{q'-q}.$$

Adding D_q gives $D_q < \gamma < D_q + D_{q'-q} = D_{q'}$. Hence $\alpha < D_q < \gamma < D_{q'} < \beta$.

γ is not rational. Suppose $\gamma = D_r$. In the field $\mathbb{R}_D$, $D_{q'-q} \neq D_0$ has the inverse $D_{1/(q'-q)}$ (by (d), $D_{q'-q}D_{1/(q'-q)} = D_1$). Solving for σ,

$$\sigma = D_1 + D_{1/(q'-q)}(D_r - D_q) = D_{1+(r-q)/(q'-q)},$$

by 1 (b), (c), (d). This contradicts Proposition 7.5.5. So γ is not rational.

Marking scheme. (10 points) - 2: two rationals $q < q'$ with $\alpha < D_q < D_{q'} < \beta$, from density used twice. - 3: a non-rational cut strictly between D_0 and D_1 (or between any two rational cuts), with both inequalities proved. - 3: the affine image lies strictly between D_q and $D_{q'}$, using the order rules of the ordered field $\mathbb{R}_D$ with a positive factor. - 2: the image is not rational, because rational cuts are closed under the field operations (1). Fatal errors (0 for the step): treating σ as a number "$\sqrt{2}$" with decimal digits; claiming $\sigma - D_1 < D_1$ without a rational bound such as $3/2 \notin \sigma$.

Exercise. C2: A cube root of two

Let $\tau := \{p \in \mathbb{Q} : p \leq 0 \text{ or } p^3 < 2\}$. We show that τ is a cut with $\tau > D_0$ and that $\tau\tau\tau = D_2$. Everything below is proved in full except the last case, which is left to you.

Monotone cubing. For rationals $0 \leq a < b$, $b^3 - a^3 = (b-a)(b^2 + ab + a^2) > 0$, since $b - a > 0$, $b^2 > 0$ (Proposition 2.3.5 (e)), $ab \geq 0$, $a^2 \geq 0$, and a product of positives is positive (Proposition 2.3.5 (c)). Only non-negative a, b are needed, because of the clause "$p \leq 0$ or".

τ is a cut and $\tau > D_0$. (C1) $0 \in \tau$, and $2 \notin \tau$ since $2^3 = 8 \geq 2$. (C2) Let $q < p \in \tau$. If $q \leq 0$, then $q \in \tau$. If $0 < q < p$, then $p > 0$, so $p^3 < 2$, and monotone cubing gives $q^3 < p^3 < 2$. (C3) If $p \leq 0$, then $1 \in \tau$ is larger. If $p > 0$ and $p^3 < 2$, put $d := 2 - p^3$, so $0 < d \leq 2$, and

$$h := \frac{d}{3p^2 + 3p + 2}.$$

Then $0 < h < 1$, because $d \leq 2 < 3p^2 + 3p + 2$. So $h^2 < h$ and $h^3 < h$, and since $p > 0$,

$$\begin{aligned}(p+h)^3 &= p^3 + 3p^2h + 3ph^2 + h^3 \\ &< p^3 + h(3p^2 + 3p + 1) \\ &< p^3 + h(3p^2 + 3p + 2) = p^3 + d = 2,\end{aligned}$$

so $p + h \in \tau$ and $p + h > p$. Finally $D_0 \subseteq \tau$ and $1 \in \tau \setminus D_0$, so $\tau > D_0$.

The $\subseteq$ half of $\tau\tau\tau = D_2$. Since $\tau > D_0$, the proof of Proposition 7.4.1 (c), with $\alpha = \beta = \gamma = \tau$, shows that $(\tau\tau)\tau = \tau(\tau\tau)$ (so we write $\tau\tau\tau$) is the set of $x \in \mathbb{Q}$ with $x \leq rst$ for some positive $r, s, t \in \tau$. For such r, s, t, let m be the largest. Then $rst \leq m^3 < 2$, by Proposition 2.3.5 (c) and monotone cubing. So $\tau\tau\tau \subseteq D_2$.

The $\supseteq$ half, easy cases. Let $x < 1$. Then $1 \in \tau$ ($1^3 < 2$) and $x \leq 1 \cdot 1 \cdot 1$, so $x \in \tau\tau\tau$. This covers $x \leq 0$ and $0 < x < 1$ together.

The $\supseteq$ half, the case $1 \le x < 2$. Let $1 \le x < 2$. Find $r \in \tau$, $r > 0$, with $r^3 > x$, and then write x as a product of three positive elements of τ.

Name the move, then complete it.

Solution. The move is **walk up by** ε in its multiplicative form, Lemma 7.4.2: step up by a fixed factor $\mu > 1$ until you leave τ, and choose μ so small that the last step inside is already beyond what we need.

The factor. We need $\mu > 1$ with $\mu^3 < 2/x$. Put $t := 2/x - 1$, so $0 < t \le 1$, and $\mu := 1 + t/7$. Since $0 < t \le 1$, $t^2 \le t$ and $t^3 \le t$, so

$$\begin{aligned}\mu^3 &= 1 + \frac{3t}{7} + \frac{3t^2}{49} + \frac{t^3}{343} \\ &\le 1 + t\Big(\frac{147 + 21 + 1}{343}\Big) = 1 + \frac{169}{343}t < 1 + t = \frac{2}{x}.\end{aligned}$$

The step. Since $\tau > D_0$ and $\mu > 1$, Lemma 7.4.2 gives $r \in \tau$ with $r > 0$ and $\mu r \notin \tau$. As $\mu r > 0$, this means $(\mu r)^3 \ge 2$, so $r^3 \ge 2/\mu^3 > x$.

The product. Put $u := x/r^2$. Then $u > 0$, and $u < r$ because $x < r^3$; so $u \in \tau$ by (C2). Now $x = r \cdot r \cdot u$ with r, r, u positive elements of τ, so $x \in \tau\tau\tau$.

Hence $\tau\tau\tau = D_2$: the cut τ is a cube root of D_2 in $\mathbb{R}_{\mathrm{D}}$. The factor $1/7$ is one convenient choice; any $\mu = 1 + ct$ with $3c + 3c^2 + c^3 < 1$ works by the same estimate, and $c = 1/3$, which served for σ, does not: for $x = 1$ it gives $\mu = 4/3$ and $\mu^3 = 64/27 > 2$.

7.6 The Two Roads Give the Same Real Numbers

This section is part of the optional Chapter 7, and nothing outside the chapter depends on it. We now have two complete ordered fields. Chapter 5 built one from classes of Cauchy sequences of rationals (Theorem 5.6.3), and Section 7.5 built another from cuts (Theorem 7.5.6). Their elements have nothing in common: a class of sequences is not a set of rationals. Chapter 6 proved that any two complete ordered fields are linked by exactly one order-preserving field isomorphism (Theorem 6.3.1). This section applies that theorem to the two models, computes the isomorphism explicitly, and compares what each road had to pay for.

7.6.1 Two models on one page

With both models in view, one letter $\mathbb{R}$ is not enough. In this section only, $\mathbb{R}_{\mathrm{C}}$ denotes the field $\mathbb{R}$ of Chapter 5, the quotient set $\mathcal{C}/\!\sim$ of Definition 5.2.1, and $\mathbb{R}_{\mathrm{D}}$ the field of cuts of Definition 7.1.1. Each contains its own copy of $\mathbb{Q}$, given by the map of Theorem 4.2.2. In $\mathbb{R}_{\mathrm{D}}$ that copy is $q \mapsto D_q$ (1). In $\mathbb{R}_{\mathrm{C}}$ we write

$$\iota := \iota_{\mathbb{R}_{\mathrm{C}}} : \mathbb{Q} \to \mathbb{R}_{\mathrm{C}},$$

which is the map $\kappa(q) = [(q)_n]$ of Definition 5.4.1, by Theorem 5.4.1 (d). From the identification in Section 5.4 on, Chapter 5 wrote q for $\kappa(q)$. Here we keep ι visible, since a bare q could now mean an element of either field.

Recall from Section 6.1 the rationals below an element x of an ordered field F (Definition 6.1.2):

$$\mathbb{Q}_F(x) \coloneqq \{q \in \mathbb{Q} : \iota_F(q) < x\}.$$

Section 6.1 remarked that this set is closed downward. In $\mathbb{R}_{\mathrm{D}}$, the rationals below a cut α are $\{q : D_q < \alpha\}$, and Proposition 7.1.2 (a) says that this set is α itself. So in the cut model **every element is its own set of rationals below**. The theorem below says what the isomorphism does with this.

7.6.2 The isomorphism

The uniqueness theorem gives the existence and uniqueness of an isomorphism, but says nothing about what it does to a given element. For these two models, what it does can be written down.

Theorem 7.6.1. The Two Constructions Give the Same Reals

Let $\mathbb{R}_{\mathrm{C}}$ be the field of Chapter 5, $\mathbb{R}_{\mathrm{D}}$ the field of cuts, and $\iota = \iota_{\mathbb{R}_{\mathrm{C}}}$.

(a) There is **exactly one** order-preserving field isomorphism $\Psi\colon \mathbb{R}_{\mathrm{C}} \to \mathbb{R}_{\mathrm{D}}$ (Definition 6.1.1).

(b) $\Psi(\iota(q)) = D_q$ for every $q \in \mathbb{Q}$.

(c) For every $x \in \mathbb{R}_{\mathrm{C}}$,

$$\Psi(x) = \mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x) = \{q \in \mathbb{Q} : \iota(q) < x\}.$$

Idea. Part (a) is Chapter 6's theorem, once both fields are known to be complete; that work is already done. Part (b) says that Ψ carries each copy of $\mathbb{Q}$ to the other. Chapter 6 showed that every map respecting sums, products and 1 fixes the rationals (Proposition 6.1.1). Here "fixes" means $\Psi \circ \iota_{\mathbb{R}_{\mathrm{C}}} = \iota_{\mathbb{R}_{\mathrm{D}}}$, and we know $\iota_{\mathbb{R}_{\mathrm{D}}}$. Part (c) is then a chain of three equivalences. A rational q lies in the cut $\Psi(x)$ exactly when D_q lies below it. By (b), D_q is $\Psi(\iota(q))$. And Ψ compares elements the way $\mathbb{R}_{\mathrm{C}}$ does.

Before reading on: where is the uniqueness in Theorem 6.3.1 used, and where only existence?

Proof. *(a)* $\mathbb{R}_{\mathrm{C}}$ is a complete ordered field by Theorem 5.6.3, and $\mathbb{R}_{\mathrm{D}}$ is one by Theorem 7.5.6. By Theorem 6.3.1, with $F = \mathbb{R}_{\mathrm{C}}$ and $L = \mathbb{R}_{\mathrm{D}}$, there is exactly one order-preserving field isomorphism $\Psi\colon \mathbb{R}_{\mathrm{C}} \to \mathbb{R}_{\mathrm{D}}$.

(b) Ψ respects sums and products and sends 1 to 1 (Definition 6.1.1). By Proposition 6.1.1, $\Psi \circ \iota_{\mathbb{R}_{\mathrm{C}}} = \iota_{\mathbb{R}_{\mathrm{D}}}$. By 1, $\iota_{\mathbb{R}_{\mathrm{D}}}(q) = D_q$ for every $q \in \mathbb{Q}$. Hence $\Psi(\iota(q)) = D_q$.

(c) First, Ψ reflects order: if $\Psi(u) < \Psi(v)$, then $u < v$. Indeed, $u = v$ would give $\Psi(u) = \Psi(v)$, and $v < u$ would give $\Psi(v) < \Psi(u)$, and trichotomy (O1) in $\mathbb{R}_{\mathrm{D}}$ excludes both; so $u < v$ by (O1) in $\mathbb{R}_{\mathrm{C}}$. Together with strict increase, $u < v \iff \Psi(u) < \Psi(v)$.

Now let $x \in \mathbb{R}_{\mathrm{C}}$ and $q \in \mathbb{Q}$. The value $\Psi(x)$ is a cut, so

$$q \in \Psi(x) \iff D_q < \Psi(x) \iff \Psi(\iota(q)) < \Psi(x) \iff \iota(q) < x.$$

The first equivalence is Proposition 7.1.2 (a), $q \in \alpha \iff D_q \subsetneq \alpha$, read with $\subsetneq$ as $<$ (Definition 7.2.1). The second is (b). The third holds because Ψ preserves and reflects $<$. The last condition says $q \in \mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x)$ (Definition 6.1.2). Since $\Psi(x)$ and $\mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x)$ are subsets of

$\mathbb{Q}$ with the same elements, they are equal. This proves the theorem.

■

The answer to the prompt: uniqueness is used only for the words "exactly one" in (a). Parts (b) and (c) use nothing about Ψ except that it respects sums, products and 1, and is strictly increasing. So any order-preserving field map $\mathbb{R}_{\mathrm{C}} \to \mathbb{R}_{\mathrm{D}}$ would satisfy (c), and (c) determines the map. That gives a second proof of the uniqueness in (a), by computation. The existence in (a) is the part that computation does not give cheaply. A direct proof would show that $x \mapsto \mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x)$ lands in the cuts, is onto, is strictly increasing (hence injective), sends 1 to D_1, and respects sums and products. The first two are Exercises C2(a) and C2(b), each a few lines. Strict increase is Lemma 6.1.3, since $\mathbb{R}_{\mathrm{C}}$ is Archimedean (Theorem 6.1.2), and $\mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(\iota(1)) = D_1$ holds by the definitions. Sums are Lemma 6.2.1 read in $\mathbb{R}_{\mathrm{C}}$ (Exercise B1). The product, with its sign cases, is the work Chapter 6 did once for all complete ordered fields.

Part (c) can also be read off Chapter 6's construction. By property $(\star)$ of Theorem 6.2.3, $\mathbb{Q}_{\mathbb{R}_{\mathrm{D}}}(\Psi(x)) = \mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x)$, and $\mathbb{Q}_{\mathbb{R}_{\mathrm{D}}}(\alpha) = \alpha$ for every cut, as we saw above. The proof given here avoids $(\star)$ and uses only that Ψ fixes the rationals.

So part (c) says: *the cut attached to a Cauchy real is the set of rationals below it.* Dedekind's definition is what the Cauchy model sees when it looks at one of its elements through the rationals. The map also runs the other way. Its inverse $\Psi^{-1}\colon \mathbb{R}_{\mathrm{D}} \to \mathbb{R}_{\mathrm{C}}$ is again an order-preserving field isomorphism (the second-reading fold after Corollary 6.3.2 checks this for any isomorphism). So by the uniqueness clause of Theorem 6.2.3, applied with $F = \mathbb{R}_{\mathrm{D}}$, which is Archimedean by Theorem 6.1.2, and $L = \mathbb{R}_{\mathrm{C}}$, it is the map of that theorem:

$$\begin{aligned}\Psi^{-1}(\alpha) &= \sup_{\mathbb{R}_{\mathrm{C}}}\{\iota(q) : q \in \mathbb{Q}_{\mathbb{R}_{\mathrm{D}}}(\alpha)\} \\ &= \sup_{\mathbb{R}_{\mathrm{C}}}\{\iota(q) : q \in \alpha\}.\end{aligned}$$

A cut is sent to the least upper bound, in the Cauchy model, of its own rationals. Exercise C2 below proves the same correspondence directly in any complete ordered field F: there $x \mapsto \mathbb{Q}_F(x)$ is onto the cuts, with inverse $\alpha \mapsto \sup_F\{\iota_F(q) : q \in \alpha\}$. For an Archimedean F, C2 also shows that this map is onto exactly when F is complete.

Quick check. What is $\Psi(\iota(-3/4))$? Is Ψ determined by its values on the elements $\iota(q)$, $q \in \mathbb{Q}$?

Solution. By (b), $\Psi(\iota(-3/4)) = D_{-3/4} = \{r \in \mathbb{Q} : r < -3/4\}$; by (c) this is also $\{q : \iota(q) < \iota(-3/4)\}$, the same set, since ι preserves and reflects order. Yes: by (c), $\Psi(x)$ is the set of q with $\iota(q) < x$, which is determined by how x compares with the elements $\iota(q)$. The order of $\mathbb{R}_{\mathrm{C}}$, together with the values $\Psi(\iota(q)) = D_q$, decides everything.

Warning. "The same real numbers" means **isomorphic as ordered fields, by a unique isomorphism**. It does not mean that the two sets have the same elements. The cut σ contains the rational 0; the element of $\mathbb{R}_{\mathrm{C}}$ that corresponds to it is a class of sequences, and its elements are functions $\mathbb{N} \to \mathbb{Q}$, not rationals. Informally, the isomorphism carries every statement about sums, products and order in one field to the same statement in the other, so no such statement can tell the two fields apart; Section 6.5 explains this pattern and why it is not stated as a theorem.

7.6.3 The square root of two, twice

Chapter 5 found a square root of 2 as the class of the Newton sequence (Example (The Class of the Newton Sequence)). Chapter 7 found one as the cut σ (Proposition 7.5.5). The isomorphism should match them.

Example. The Two Square Roots of Two

Optional: nothing later depends on this.

Let $x \in \mathbb{R}_C$ with $x > 0$ and $x^2 = \iota(2)$; for instance the class r of the Newton sequence (Example (The Class of the Newton Sequence)), which has $r > 0$ and $r^2 = \kappa(2) = \iota(2)$ (Theorem 5.4.1 (d)). Show that $\Psi(x) = \sigma$.

Solution. Since Ψ is multiplicative, $\Psi(x)^2 = \Psi(x^2) = \Psi(\iota(2)) = D_2$, by Theorem 7.6.1 (b). By Proposition 7.5.5, $\sigma\sigma = D_2$ as well. Both are positive: $\Psi(x) > \Psi(\iota(0)) = D_0$, since $x > 0 = \iota(0)$ and Ψ is strictly increasing, and $\sigma > D_0$ by Proposition 7.5.5.

Positive square roots are unique in an ordered field. If $u, v > 0$ and $u \neq v$, say $u < v$, then $u^2 < v^2$ by Proposition 2.3.5 (g). Applied in $\mathbb{R}_D$ to $u = \Psi(x)$ and $v = \sigma$, which have the same square D_2, this gives $\Psi(x) = \sigma$.

For the Newton class there is a second route through (c): $\Psi(r) = \mathbb{Q}_{\mathbb{R}_C}(r)$, and Example (The Rationals below the Square Root of Two) computed this set as the rationals $q < 0$ together with those $q \geq 0$ with $q^2 < 2$. That is σ, written differently.

Exercise C1 below runs the correspondence backwards: from any cut α it builds a Cauchy sequence of rationals whose class Ψ sends to α. Applied to $\alpha = \sigma$, it produces a positive element of $\mathbb{R}_C$ with square $\iota(2)$. As a **source** for such an element, though, that route would assume what the example illustrates, so the example takes x from Chapter 5.

7.6.4 What each road pays for

Both constructions start from $\mathbb{Q}$ and end at a complete ordered field, and by Theorem 7.6.1 they end at the same one. They spend their effort in different places.

	Cauchy sequences (Chapter 5)	Dedekind cuts (Chapter 7)
an element is	a class of Cauchy sequences of rationals (Definition 5.2.1)	a set of rationals with (C1)–(C3) (Definition 7.1.1)
equality	equality of classes; every operation and the order must be shown independent of representatives (Theorem 5.2.1, Proposition 5.3.1)	equality of sets; nothing to check

	Cauchy sequences (Chapter 5)	Dedekind cuts (Chapter 7)
sum and product	termwise on representatives (Definition 5.2.2)	the sum is the set sum (Definition 7.3.1); the negative needs a margin (Definition 7.3.2); the product is defined for positive cuts, then by signs (Definition 7.4.2)
inverses	through the sign theorem for Cauchy sequences (Theorem 5.2.3)	a margin again, by "walk up" (Definition 7.4.3, Theorem 7.4.6)
order	positive classes, eventually above a fixed rational margin (Definition 5.3.1)	inclusion (Definition 7.2.1), total by Theorem 7.2.1
completeness	every Cauchy sequence converges (Section 5.5), then bisection (Theorem 5.6.2)	a union (Theorem 7.2.2)
generalizes to	the completion of a metric space	the completion of an ordered set

Read down the columns. The Cauchy road gets its arithmetic cheaply, termwise, and pays at both ends: every definition must be checked against a change of representative, the order needs a margin, and completeness takes two sections. The Dedekind road gets order and completeness almost free, because a cut is a set and inclusion is already an order, and pays in the middle: negatives need a margin, and multiplication needs a detour through positive cuts and signs.

The last row looks beyond this volume. The Cauchy construction works for any metric space, with sequences whose terms get close: *Metric and Topological Spaces*, Chapter 2 builds the completion of a metric space that way. The Dedekind construction has an analogue for any ordered set, in which a cut is a set of elements that equals the set of lower bounds of its upper bounds; the result is called the Dedekind–MacNeille completion, and no volume of this series develops it. Each road leads somewhere the other does not.

↻ **Second reading** · *History*

Richard Dedekind published the construction by cuts in 1872, in *Stetigkeit und irrationale Zahlen*; he says there that the idea came to him in 1858, while preparing lectures on the differential calculus. Charles Méray had built the irrational numbers from convergent sequences of rationals in 1869, and Georg Cantor and Eduard Heine gave constructions from fundamental sequences, our Cauchy sequences, in 1872. Karl Weierstrass had presented his own construction of the real numbers in his Berlin lectures of the 1860s. The two roads of this volume were thus laid out within a few years of each other.

7.6.5 Exercises

7.6.5.1 A. Check your understanding

Exercise. A1

These questions are about the proof of Theorem 7.6.1.

(a) Name the earlier results the proof uses, and say for which of (a), (b), (c) each is used.

(b) Run part (c) for $x = \iota(1/2)$: which rationals lie in $\Psi(x)$? Check your answer against (b).

(c) State the idea of the proof in two sentences.

(d) Which step of (c) uses that Ψ reflects order, and which uses that $\Psi(x)$ is a cut?

Solution.

(a) For (a): Theorem 5.6.3 and Theorem 7.5.6 (both fields are complete), and Theorem 6.3.1. For (b): Proposition 6.1.1, and 1 for $\iota_{\mathbb{R}_D}(q) = D_q$. For (c): Proposition 7.1.2 (a), part (b) of the theorem, and trichotomy (O1) of Definition 2.3.2 in both fields, to see that Ψ reflects order.

(b) $q \in \Psi(\iota(1/2)) \iff \iota(q) < \iota(1/2) \iff q < 1/2$, since ι is strictly increasing and reflects order (Theorem 4.2.2). So $\Psi(\iota(1/2)) = \{q \in \mathbb{Q} : q < 1/2\} = D_{1/2}$, which is what (b) says.

(c) Both fields are complete, so Chapter 6 supplies exactly one isomorphism, and like every map that respects sums, products and 1, it carries the copy of $\mathbb{Q}$ in one field to the copy in the other. Since a cut consists of the rationals q with D_q below it, the isomorphism sends x to the set of rationals below x.

(d) The third equivalence, $\Psi(\iota(q)) < \Psi(x) \Rightarrow \iota(q) < x$, uses that Ψ reflects order; the reverse implication uses strict increase. The first equivalence, $q \in \Psi(x) \iff D_q < \Psi(x)$, is Proposition 7.1.2, which holds only for cuts.

7.6.5.2 B. Practice

Exercise. B1

Let $x, y \in \mathbb{R}_C$.

(a) Deduce from Lemma 6.2.1, without using Ψ, that $\mathbb{Q}_{\mathbb{R}_C}(x + y) = \mathbb{Q}_{\mathbb{R}_C}(x) + \mathbb{Q}_{\mathbb{R}_C}(y)$, where the right side is the set sum of Definition 7.3.1. Say why the lemma applies to $\mathbb{R}_C$.

(b) Explain how (a), read through Theorem 7.6.1 (c), is the statement $\Psi(x + y) = \Psi(x) + \Psi(y)$.

Solution. Write $\mathbb{Q}(z)$ for $\mathbb{Q}_{\mathbb{R}_C}(z)$.

(a) The field $\mathbb{R}_C$ is Archimedean (Theorem 5.4.2), so Lemma 6.2.1 applies with $F = \mathbb{R}_C$,

$a = x$ and $b = y$: for every $q \in \mathbb{Q}$,

$$q \in \mathbb{Q}(x+y) \iff \exists r \in \mathbb{Q}\ (r \in \mathbb{Q}(x) \wedge q - r \in \mathbb{Q}(y)).$$

($\subseteq$) If $q \in \mathbb{Q}(x+y)$, take r as on the right; then $q = r + (q-r)$ with $r \in \mathbb{Q}(x)$ and $q - r \in \mathbb{Q}(y)$, so $q \in \mathbb{Q}(x) + \mathbb{Q}(y)$.
($\supseteq$) If $q = r + s$ with $r \in \mathbb{Q}(x)$ and $s \in \mathbb{Q}(y)$, then $q - r = s \in \mathbb{Q}(y)$, so the right side holds with this r, and $q \in \mathbb{Q}(x+y)$.

(b) By Theorem 7.6.1 (c), $\mathbb{Q}(x+y) = \Psi(x+y)$, $\mathbb{Q}(x) = \Psi(x)$ and $\mathbb{Q}(y) = \Psi(y)$, and the set sum of two cuts is their sum in $\mathbb{R}_{\mathrm{D}}$ (Definition 7.3.1). So (a) says $\Psi(x+y) = \Psi(x)+\Psi(y)$. We knew this already, since Ψ is additive; (a) is Lemma 6.2.1 read in $\mathbb{R}_{\mathrm{C}}$ as a statement about cuts.

7.6.5.3 C. Going deeper

Exercise. C1: From a cut to a Cauchy sequence

Let α be a cut. For $n \in \mathbb{N}$ let $2^{-n} := ((2 \cdot 1_{\mathbb{Q}})^n)^{-1}$ (Definition 4.2.1). Fix $a \in \alpha$, and for each n let k_n be the **least** $k \in \mathbb{N}$ with $a + k\,2^{-n} \notin \alpha$, and $a_n := a + (k_n - 1)\,2^{-n}$.

(a) Show that k_n exists and $k_n \geq 1$, and that $a_n \in \alpha$ and $a_n + 2^{-n} \notin \alpha$.

(b) Show that (a_n) is a Cauchy sequence of rationals (Definition 5.1.1).

(c) Show that $\Psi([(a_n)]) = \alpha$.

Solution (Hint 1). For (b), compare a_n with $a_m + 2^{-m}$: one is in α and the other is not. For (c), use Theorem 7.6.1 (c), so that you only need to decide when $\iota(q) < [(a_n)]$.

Solution (Hint 2). For (c), if $q \in \alpha$, take $q' \in \alpha$ with $q' > q$, and show that $a_n > (q + q')/2$ for all large n.

Solution. Write $\varepsilon_n := 2^{-n}$. By Lemma 5.6.1 with $F = \mathbb{Q}$, $(2 \cdot 1_{\mathbb{Q}})^n \geq n + 1 > 0$, so $\varepsilon_n > 0$.

A threshold for ε_n. Let $\varepsilon \in \mathbb{Q}$, $\varepsilon > 0$. By Theorem 2.5.3 there is $N \in \mathbb{N}$ with $N > 1/\varepsilon$, that is, $N\varepsilon > 1$. For $n \geq N$, Lemma 5.6.1 gives $(2 \cdot 1_{\mathbb{Q}})^n \varepsilon \geq (n+1)\varepsilon > N\varepsilon > 1$, so $\varepsilon_n < \varepsilon$. Call this (†): for every rational $\varepsilon > 0$ there is N with $\varepsilon_n < \varepsilon$ for all $n \geq N$.

(a) The set $S_n := \{k \in \mathbb{N} : a + k\varepsilon_n \notin \alpha\}$ is non-empty. Indeed, take $b \notin \alpha$ (C1); by Theorem 2.5.3 there is $m \in \mathbb{N}$ with $m > (b-a)/\varepsilon_n$, so $a + m\varepsilon_n > b$, and $a + m\varepsilon_n \notin \alpha$ because the complement of a cut is closed upward (Lemma 7.1.1). This is the first step of the proof of Lemma 7.1.3. So S_n is a non-empty set of integers ($\mathbb{N} \subseteq \mathbb{Z}$, Theorem 2.3.4 (b)), and 0 is a lower bound of it; by Proposition 2.3.6 it has a least element k_n. Since $a + 0 \cdot \varepsilon_n = a \in \alpha$, $k_n \neq 0$, so $k_n \geq 1$ (Theorem 2.3.4 (c)) and $k_n - 1 \in \mathbb{N}$. By minimality $k_n - 1 \notin S_n$, so $a_n \in \alpha$; and $a_n + \varepsilon_n = a + k_n\varepsilon_n \notin \alpha$. Each a_n is defined by a rule from the fixed a, so no choice principle is involved.

(b) For any $m, n \in \mathbb{N}$: $a_n \in \alpha$ and $a_m + \varepsilon_m \notin \alpha$, so $a_n < a_m + \varepsilon_m$ by Lemma 7.1.1;

symmetrically, $a_m < a_n + \varepsilon_n$. Given a rational $\varepsilon > 0$, take N from (†). For $m, n \geq N$,

$$-\varepsilon < -\varepsilon_m < a_m - a_n < \varepsilon_n < \varepsilon,$$

so $|a_m - a_n| < \varepsilon$ (Proposition 4.1.3 (d)). Hence $(a_n) \in \mathcal{C}$.

(c) Let $x := [(a_n)] \in \mathbb{R}_{\mathrm{C}}$. By Theorem 7.6.1 (c), $\Psi(x) = \{q \in \mathbb{Q} : \iota(q) < x\}$, and $\iota = \kappa$ (Theorem 5.4.1 (d)). So it suffices to show $\kappa(q) < x \iff q \in \alpha$ for every $q \in \mathbb{Q}$.

If $q \in \alpha$. By (C3) there is $q' \in \alpha$ with $q' > q$. For every n, $q' \in \alpha$ and $a_n + \varepsilon_n \notin \alpha$ give $q' < a_n + \varepsilon_n$ (Lemma 7.1.1). Apply (†) with $\varepsilon := (q' - q)/2$: for $n \geq N$,

$$a_n > q' - \varepsilon_n > q' - \frac{q' - q}{2} = \frac{q + q'}{2}.$$

By Lemma 5.3.4, applied to the constant sequence with value $(q + q')/2$ and to (a_n), $\kappa((q + q')/2) \leq x$. Since $q < (q + q')/2$, Theorem 5.4.1 (b) gives $\kappa(q) < \kappa((q + q')/2) \leq x$.

If $q \notin \alpha$. For every n, $a_n \in \alpha$ gives $a_n < q$ (Lemma 7.1.1). By Lemma 5.3.4, $x \leq \kappa(q)$, so $\kappa(q) < x$ is false by (O1).

Hence $\Psi(x) = \alpha$. The argument used neither monotonicity of (a_n) nor any law of exponents in $\mathbb{Q}$: only the inequality of Lemma 5.6.1.

Marking scheme. (10 points) - 2: (a), with the least-element principle and the reason $k_n \geq 1$. - 3: (b), from the two separation inequalities and a threshold N for 2^{-n} with a stated reason. - 3: (c) for $q \in \alpha$: the strict gap $q' > q$ from (C3), then a correct passage to $\kappa(q) < x$, either by Lemma 5.3.4 (which gives only $\leq$) against the midpoint $(q + q')/2$ followed by Theorem 5.4.1 (b), or directly by showing that $(a_n - q)_n$, a representative of $x - \kappa(q)$, is eventually $\geq$ the fixed margin $(q' - q)/2$ (Definition 5.3.1). Either route earns all 3 points. - 2: (c) for $q \notin \alpha$. Fatal errors (0 for the step): concluding $\kappa(q) < x$ from $q < a_n$ for all n (termwise $<$ gives only $\leq$ between classes); choosing N after m, n.

Exercise. C2: Exit check: the rationals below, as a cut

Let F be an Archimedean ordered field (Definition 4.2.2). Prove:

(a) $\mathbb{Q}_F(x)$ is a cut for every $x \in F$;

(b) the map $\mathbb{Q}_F \colon F \to \mathbb{R}_{\mathrm{D}}$, $x \mapsto \mathbb{Q}_F(x)$, is onto if and only if F is complete (Definition 5.6.1).

You may use Lemma 6.1.3: $x \leq y \iff \mathbb{Q}_F(x) \subseteq \mathbb{Q}_F(y)$, so the map is injective.

Solution. Throughout, ι_F is strictly increasing and reflects order (Theorem 4.2.2).

(a) *(C1).* By Definition 4.2.2 there is $n \in \mathbb{N}$ with $x < n \cdot 1_F = \iota_F(n)$; then $\iota_F(n) < x$ fails by (O1), so $n \notin \mathbb{Q}_F(x)$, and $\mathbb{Q}_F(x) \neq \mathbb{Q}$. Applied to $-x$, the same property gives $m \in \mathbb{N}$ with $-x < \iota_F(m)$, so $\iota_F(-m) = -\iota_F(m) < x$ (Proposition 2.3.5 (a)), and $-m \in \mathbb{Q}_F(x) \neq \emptyset$.
(C2). If $q \in \mathbb{Q}_F(x)$ and $p < q$, then $\iota_F(p) < \iota_F(q) < x$, so $p \in \mathbb{Q}_F(x)$.
(C3). If $q \in \mathbb{Q}_F(x)$, then $\iota_F(q) < x$, and Theorem 4.2.5 gives $q' \in \mathbb{Q}$ with $\iota_F(q) < \iota_F(q') < x$. Then $q' > q$, since ι_F reflects order, and $q' \in \mathbb{Q}_F(x)$.

(b) ($\Leftarrow$) Let F be complete and α a cut. Put $S := \{\iota_F(q) : q \in \alpha\}$. It is non-empty, since $\alpha \neq \emptyset$. Take $b \notin \alpha$ (C1); every $q \in \alpha$ has $q < b$ (Lemma 7.1.1), so $\iota_F(b)$ is an upper bound of S. By completeness $s := \sup S$ exists. We show $\mathbb{Q}_F(s) = \alpha$.

If $q \in \alpha$, (C3) gives $q' \in \alpha$ with $q' > q$, and $\iota_F(q) < \iota_F(q') \leq s$. So $q \in \mathbb{Q}_F(s)$.

If $q \notin \alpha$, every $p \in \alpha$ has $p < q$ (Lemma 7.1.1), so $\iota_F(q)$ is an upper bound of S, and $s \leq \iota_F(q)$ because s is the least one. So $\iota_F(q) < s$ fails, and $q \notin \mathbb{Q}_F(s)$.

Hence $\mathbb{Q}_F(s) = \alpha$, and the map is onto.

($\Rightarrow$) Let the map be onto. By (a) and Lemma 6.1.3, it is a bijection $F \to \mathbb{R}_{\mathrm{D}}$ with $x \leq y \iff \mathbb{Q}_F(x) \subseteq \mathbb{Q}_F(y)$. Let $A \subseteq F$ be non-empty and bounded above by u. The set $\{\mathbb{Q}_F(a) : a \in A\}$ of cuts is non-empty and bounded above by $\mathbb{Q}_F(u)$, so it has a supremum γ in $(\mathbb{R}_{\mathrm{D}}, \subseteq)$ (Theorem 7.2.2). Since the map is onto, $\gamma = \mathbb{Q}_F(s)$ for some $s \in F$. For $a \in A$, $\mathbb{Q}_F(a) \subseteq \mathbb{Q}_F(s)$, so $a \leq s$: s is an upper bound of A. If v is any upper bound of A, then $\mathbb{Q}_F(v)$ is an upper bound of the set of cuts, so $\mathbb{Q}_F(s) = \gamma \subseteq \mathbb{Q}_F(v)$, and $s \leq v$. Hence $s = \sup A$, and F is complete.

For $F = \mathbb{R}_{\mathrm{C}}$, (b) gives a second proof that $x \mapsto \mathbb{Q}_{\mathbb{R}_{\mathrm{C}}}(x)$, the map Ψ of Theorem 7.6.1 (c), is onto.

Marking scheme. (10 points) - 2: (a) (C1), with the Archimedean property used for **both** halves (a rational outside, and a rational inside). - 1: (a) (C2) by transitivity, and (C3) by density. - 1: (b) ($\Leftarrow$): states the candidate $s = \sup \iota_F(\alpha)$ and why it exists (non-empty, bounded above by $\iota_F(b)$, $b \notin \alpha$). - 2: $\alpha \subseteq \mathbb{Q}_F(s)$, using (C3) of α to get the strict inequality. - 2: $\mathbb{Q}_F(s) \subseteq \alpha$, from "$\iota_F(q)$ is an upper bound". - 2: ($\Rightarrow$) through the order isomorphism onto $(\mathbb{R}_{\mathrm{D}}, \subseteq)$ and Theorem 7.2.2. Fatal errors (0 for the step): using $\iota_F(q) \leq x$ in place of $\iota_F(q) < x$ (the set then has a largest element when $x = \iota_F(r)$); concluding $q \in \mathbb{Q}_F(s)$ from $\iota_F(q) \leq s$.

7.6.5.4 R. Review

Exercise. R1

Let $f : \mathbb{R}_{\mathrm{D}} \to \mathbb{R}_{\mathrm{D}}$ satisfy $\alpha \leq \beta \Rightarrow f(\alpha) \leq f(\beta)$ for all cuts α, β, and $f(D_q) = D_q$ for every $q \in \mathbb{Q}$. Prove that $f(\alpha) = \alpha$ for every cut α.

Solution. Let α be a cut, and suppose $f(\alpha) \neq \alpha$. By Theorem 7.2.1, $\alpha < f(\alpha)$ or $f(\alpha) < \alpha$.

If $\alpha < f(\alpha)$, Theorem 7.5.3 gives $q \in \mathbb{Q}$ with $\alpha < D_q < f(\alpha)$. Since $\alpha \leq D_q$, the hypothesis gives $f(\alpha) \leq f(D_q) = D_q$, so $f(\alpha) \leq D_q < f(\alpha)$, which is impossible. If $f(\alpha) < \alpha$, a rational cut with $f(\alpha) < D_q < \alpha$ gives $D_q = f(D_q) \leq f(\alpha) < D_q$ in the same way. Hence $f(\alpha) = \alpha$.

This is the method of Chapter 6: elements of an Archimedean field are separated by rationals (Lemma 6.1.3), so a map that fixes the rationals and respects order can move nothing. Neither sums nor products were needed.

Exercise. R2

Let F be an ordered field, $x \in F$, and (q_n) a sequence of rationals with $\iota_F(q_n) \to x$ in F.

Show that (q_n) is a Cauchy sequence in $\mathbb{Q}$, and that for every $q \in \mathbb{Q}$ with $\iota_F(q) < x$ there is $N \in \mathbb{N}$ with $q_n > q$ for all $n \geq N$.

Solution. Cauchy in $\mathbb{Q}$. The sequence $(\iota_F(q_n))$ converges in F, so it is a Cauchy sequence in F (Proposition 4.5.1). Let $\varepsilon \in \mathbb{Q}$, $\varepsilon > 0$. Then $\iota_F(\varepsilon) > \iota_F(0) = 0$, so Definition 4.5.1 with the tolerance $\iota_F(\varepsilon)$ gives N with $|\iota_F(q_m) - \iota_F(q_n)| < \iota_F(\varepsilon)$ for all $m, n \geq N$. By Proposition 4.1.3 (d) and additivity of ι_F,

$$\iota_F(-\varepsilon) < \iota_F(q_m - q_n) < \iota_F(\varepsilon),$$

and since ι_F reflects order (Theorem 4.2.2), $-\varepsilon < q_m - q_n < \varepsilon$, that is, $|q_m - q_n| < \varepsilon$. So (q_n) is Cauchy in $\mathbb{Q}$ (Definition 4.5.1 with $F = \mathbb{Q}$).

Eventually above q. Let $\iota_F(q) < x$, and put $\delta := x - \iota_F(q) > 0$. By Definition 4.3.2 there is N with $|\iota_F(q_n) - x| < \delta$ for $n \geq N$, so $\iota_F(q_n) > x - \delta = \iota_F(q)$, and $q_n > q$ because ι_F reflects order.

Both parts are from Chapter 4. The second says that every rational q with $\iota_F(q) < x$, that is, every element of the set $\mathbb{Q}_F(x)$ of Exercise C2, lies below all but finitely many q_n. No Archimedean hypothesis was needed.

Exercise. R3

Prove that $|\mathbb{R}_{\mathrm{D}}| \leq |\mathcal{P}(\mathbb{N})|$, and that the set $\{D_q : q \in \mathbb{Q}\}$ of rational cuts is countable and is not all of $\mathbb{R}_{\mathrm{D}}$.

Solution. $\mathbb{Q}$ *is countable.* This is Exercise (R2): $\mathbb{Z}$ is the image of the countable set $\mathbb{N} \times \mathbb{N}$ under $a \mapsto [a]$, and $\mathbb{Q}$ is the image of the countable set $\mathbb{Z} \times \mathbb{Z}_{>0}$ under $(p, q) \mapsto [(p, q)]$, by Sets, Relations, and Cardinality, Theorem 3.2.6, Sets, Relations, and Cardinality, Proposition 3.2.3, Sets, Relations, and Cardinality, Corollary 3.2.7 (a) and Sets, Relations, and Cardinality, Proposition 3.2.4, read on $\mathbb{N} = \omega$ through Section 1.7. (It is the argument of Sets, Relations, and Cardinality, Theorem 3.2.8, run for the constructed $\mathbb{Q}$; that theorem itself is about another $\mathbb{Q}$ and is not used.)

An injection into $\mathcal{P}(\mathbb{N})$. By Sets, Relations, and Cardinality, Theorem 3.2.2, (i) $\Rightarrow$ (ii), there is an injection $i \colon \mathbb{Q} \to \mathbb{N}$. Every cut is a subset of $\mathbb{Q}$, so $\alpha \mapsto i(\alpha) = \{i(q) : q \in \alpha\}$ maps $\mathbb{R}_{\mathrm{D}}$ into $\mathcal{P}(\mathbb{N})$. It is injective: if $i(\alpha) = i(\beta)$ and $q \in \alpha$, then $i(q) = i(q')$ for some $q' \in \beta$, so $q = q' \in \beta$ since i is injective; thus $\alpha \subseteq \beta$, and symmetrically $\beta \subseteq \alpha$. By Sets, Relations, and Cardinality, Definition 3.5.1, $|\mathbb{R}_{\mathrm{D}}| \leq |\mathcal{P}(\mathbb{N})|$.

The rational cuts. $\{D_q : q \in \mathbb{Q}\}$ is the image of the countable set $\mathbb{Q}$ under $q \mapsto D_q$, so it is countable (Sets, Relations, and Cardinality, Proposition 3.2.4). It is not all of $\mathbb{R}_{\mathrm{D}}$: σ is not a rational cut (Proposition 7.5.5).

Much more is true: $\mathbb{R}_{\mathrm{D}}$ is in bijection with $\mathbb{R}_{\mathrm{C}}$ (Theorem 7.6.1), which Exercise (R3) showed to be uncountable. So $\mathbb{R}_{\mathrm{D}}$ is uncountable, while its rational cuts are countable.

Exercise. R4

Is the set $\{\alpha \in \mathbb{R}_{\mathrm{D}} : \alpha \neq D_q \text{ for every } q \in \mathbb{Q}\}$ of irrational cuts closed under addition?

Justify your answer.

Solution. No. The tool is the additive inverse of a cut (Theorem 7.3.3) together with the arithmetic of rational cuts (1).

The cut σ is irrational (Proposition 7.5.5). So is $-\sigma$: if $-\sigma = D_q$, then $\sigma = -(-\sigma) = -D_q = D_{-q}$ (Proposition 2.3.2 (a) in the field $\mathbb{R}_{\mathrm{D}}$, and 1 (c)), which is impossible. But $\sigma + (-\sigma) = D_0$ (Theorem 7.3.3), a rational cut. So two irrational cuts can have a rational sum.

The school slogan "irrational plus irrational is irrational" is false, and its opposite is no rule either: $\sigma + \sigma$ is irrational. Indeed $\sigma + \sigma = (D_1 + D_1)\sigma = D_2\sigma$ by distributivity, and if $D_2\sigma = D_r$, then $\sigma = D_{1/2}D_r = D_{r/2}$ by 1 (d), which is impossible. Whether a sum of irrational cuts is rational depends on the summands, not on a rule; a proof has to look at them.