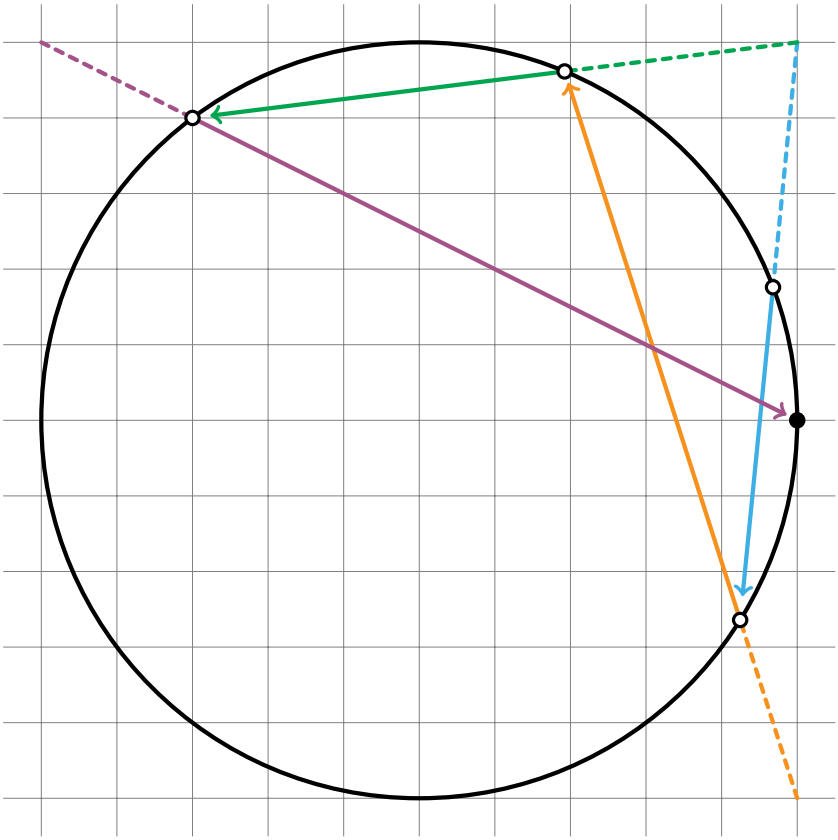
Johns Hopkins University

MATH 304

Egbert Rijke



Elementary Number Theory

Spring 2025

Cover image: Vieta jumps between the rational points on the unit circle eventually reach one of the points $(1,0)$, $(0,1)$, $(-1,0)$, or $(0,-1)$.

Contents

Contents i

Course Information v

Introduction vii

1 Mathematical Induction 1
1.1 Reasoning About Infinitely Many Natural Numbers 1
Exercises 4

2 Counting 7
2.1 Counting Bijections 7
2.2 Counting Subsets 8
2.3 The Binomial Theorem 11
2.4 The Inclusion-Exclusion Principle 12
Exercises 13

3 Euclidean Division and Representability 15
3.1 The Well-Ordering Principle of the Natural Numbers 15
3.2 Euclidean Division 16
3.3 The Representability Theorem 18
3.4 Combinatorial Applications 20
Exercises 21

4 Linear Diophantine Equations 23
4.1 Divisibility 23
4.2 Ideals of Integers 26
4.3 The Ordering by Divisibility 27
4.4 Greatest Common Divisors 28
4.5 Euclid's Algorithm 31
4.6 Linear Diophantine Equations in Multiple Variables 33
Exercises 35

5 The Fundamental Theorem of Arithmetic **37**
5.1 Prime Numbers . . . 37
5.2 The Infinitude of Primes . . . 40
5.3 The Fundamental Theorem of Arithmetic . . . 42
5.4 Legendre's Formula and Kummer's Theorem . . . 44
5.5 Bertrand's Postulate . . . 48
Exercises . . . 49

6 Polynomials **51**
6.1 Polynomials with Integer Coefficients . . . 51
6.2 Lagrange's Interpolation Theorem . . . 53
6.3 Fixed Divisors of Integer Polynomials . . . 56
Exercises . . . 58

7 Pythagorean Triples **59**
7.1 The Pythagorean Theorem . . . 59
7.2 Euclid's Parametrization of the Pythagorean Triples . . . 62
7.3 Rational Points on the Unit Circle . . . 66
7.4 The Tree of Primitive Pythagorean Triples . . . 66
Exercises . . . 67

8 Infinite Descent **69**
8.1 The Method of Infinite Descent . . . 69
8.2 The Unsolvability of $x^4 + y^4 = z^4$ . . . 71
8.3 Vieta Jumping . . . 72
Exercises . . . 73

9 Congruences **75**
9.1 The Congruence Relations . . . 75
9.2 Equivalence Relations . . . 77
9.3 Equivalence Classes and Residue Systems . . . 80
9.4 Reduced Residue Systems . . . 82
Exercises . . . 82

10 Modular Arithmetic **85**
10.1 The Integers Modulo n . . . 85
10.2 Solving Linear Congruences . . . 87
10.3 Fermat's Little Theorem . . . 89
10.4 Euler's Theorem . . . 90
10.5 Wilson's Theorem . . . 91
Exercises . . . 92

11 Systems of Linear Congruences **95**
11.1 Solving Multiple Linear Congruences Simultaneously . . . 95
11.2 The Chinese Remainder Theorem . . . 96
11.3 Linear Congruences in Multiple Variables . . . 97
11.4 Multiplicativity of Euler's Totient Function . . . 97

Exercises . . . 97

12 Polynomial Congruences **99**
12.1 Polynomial Congruences of Prime Moduli . . . 99
12.2 Polynomial Congruences of Composite Moduli . . . 101
12.3 Reduced Polynomials Modulo a Prime . . . 102
12.4 The Elementary Symmetric Polynomials . . . 102
Exercises . . . 103

13 Primitive Roots **105**
13.1 The Multiplicative Order of an Integer Modulo n . . . 105
13.2 The Infinitude of Primes Congruent to 1 Modulo Powers of 2 . . . 108
13.3 Counting Elements of a Given Order Modulo a Prime . . . 109
13.4 Primitive Roots . . . 111
Exercises . . . 114

14 Quadratic Residues **115**
14.1 Quadratic Congruences . . . 115
14.2 Quadratic Residues . . . 117
14.3 Legendre Symbols . . . 118
14.4 Euler's Criterion . . . 120
14.5 Euler's Prime-Generating Polynomial . . . 121
Exercises . . . 123

15 Quadratic Reciprocity **125**
15.1 The Quadratic Character of 2 . . . 125
15.2 The Statement of Quadratic Reciprocity . . . 128
15.3 Gauss's Lemma . . . 131
15.4 Eisenstein's Proof . . . 132
Exercises . . . 135

16 Arithmetic Functions **137**
16.1 Multiplicative Functions . . . 137
16.2 Dirichlet Convolution . . . 139
16.3 The Möbius Inversion Formula . . . 140
16.4 Dirichlet Inverses . . . 142
Exercises . . . 143

17 The Distribution of the Prime Numbers **145**
17.1 The Bachmann–Landau Notation for Asymptotic Growth . . . 145
17.2 An Elementary Estimate of the Prime Counting Function . . . 146
17.3 Chebyshev's Theorem . . . 149

Bibliography **151**

Course Information

Course Materials

We will use these course notes and *Number Theory* by Andrews [And94].

Important Dates

- First lecture: January 21st.
- First midterm: February 25th.
- Spring break: March 17-24th.
- Second midterm: April 1st.
- Final lecture: April 24th
- Final exam: May 8th.

The first midterm covers content up to February 13th. The second midterm covers content up to the spring break.

1. Midterm 1 covers Chapters 1 to 5 from these notes, and Chapters 1, 2, and 3 from Andrews *Number Theory*. Note that parts of Chapter 3 of Andrews' book do not have an equivalent in these notes, so don't forget to study Chapter 3 of Andrews.
2. Midterm 2 covers Chapters 6 and 9 to 13 from these notes, and Chapters 4, 5, 6.1, and 7 from Andrews *Number Theory*.

Grading

The course grade will be determined as follows:

- Homework: 60%
- Each exam (two midterms and final): 20% each

This adds up to a total of 120%, allowing students the opportunity to earn extra credit. The grading scale is as follows:

- $>$ 100%: A+
- 90–100%: A
- 80–90%: B
- 70–80%: C
- 60–70%: D
- $<$ 60%: F

If the median score for the class falls below a B, grades may be adjusted (curved) so that the median corresponds to a B. Any adjustments will be made consistently for all students.

Introduction

Number theory is the study of numbers, particularly the natural numbers (the numbers $0, 1, 2, \ldots$ increasing indefinitely in increments of 1), the integers (the numbers $\ldots, -2, -1, 0, 1, 2, \ldots$), and occasionally the rational numbers (fractions of integers). Its central themes include questions about divisibility, prime numbers, modular arithmetic, arithmetic functions, and finding integer solutions to equations. Additionally, number theory explores combinatorial questions, such as determining how numbers can be decomposed into sums or products of specific forms, like squares or primes. More advanced branches of number theory are occasionally also concerned with the properties of other number systems, such as the real numbers, complex numbers, or p-adic numbers. However, in *elementary number theory* these explorations are primarily motivated by how these number systems relate back to the integers.

A Short History of Number Theory

Number theory has a long history that dates back to antiquity. One of the earliest and most influential systematic treatments of mathematics is Euclid's Elements [Euc08], written around 300 BC in Alexandria. In this work, Euclid established several key results in number theory, including the infinitude of primes and the algorithm for finding the greatest common divisor, which bears his name.

Figure 1: Line engraving by S. Beyssent after Mlle. C. Reydell

Around 250 AD, the Greek mathematician Diophantus of Alexandria wrote his *Arithmetica* [Dio10], which focused on solving what are now known as Diophantine equations. This 13-volume work is often considered the first comprehensive treatment of algebra, though Diophantus relied on rhetorical descriptions rather than modern algebraic notation. Only six of the original volumes have survived, but they profoundly influenced later mathematicians.

The works of Euclid and Diophantus were preserved and studied by scholars in the Byzantine Empire, such

as Theon of Alexandria and his daughter Hypatia. They added commentaries that clarified and expanded upon the original text.

During the Islamic Golden Age, scholars like Al-Khwarizmi, Ibn al-Haytham, and Omar Khayyam preserved and expanded upon the works of Euclid and Diophantus. Al-Khwarizmi's work, especially his *Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabala* (from which the term "algebra" originates), introduced systematic methods for solving linear and quadratic equations. He used rhetorical algebra, like Diophantus, but began a transition toward symbolic representation by relying on consistent terminology for operations and equations.

After the invention of the printing press, Erhard Ratdolt, a German printer based in Venice, produced the first printed edition of the Elements in 1482. This edition included mathematical diagrams and marked an important step in the dissemination of Euclid's work. The first printed edition of Diophantus' Arithmetica was published in 1575 by Wilhelm Xylander, a German mathematician. The printing was done in Basel, Switzerland, a major hub for academic publishing at the time. By the time of Pierre de Fermat in the 17th century, both the Elements and Arithmetica had become essential texts for mathematicians across Europe, influencing the development of number theory and inspiring Fermat's groundbreaking contributions.

Pierre de Fermat (1607–1665) is often considered the founder of modern number theory. He was a contemporary of Blaise Pascal and Marin Mersenne, both of whom he corresponded with, and also of Galileo Galilei. He wrote a manuscript titled *Éléments de Géométrie*, which served as an extension and commentary on Euclid's Elements. In this work, he sought to modernize Euclid's geometry by applying the emerging methods of algebra and analytic geometry to classical geometric problems.

Figure 2: Pierre de Fermat. Engraving in Oeuvres Mathématiques Diverses.

Fermat made several groundbreaking contributions to number theory. He proved *Fermat's Little Theorem*, which is a central theorem of this course, and he formulated the *sum of two squares theorem* asserting that an odd prime p can be expressed as the sum of two squares if and only if $p \equiv 1 \pmod 4$. He also introduced the method of *infinite descent*, and formulated the *Polygonal Number Theorem*, which states that every positive integer can be expressed as the sum of at most n n-gonal numbers.

Fermat also made several famous conjectures. He conjectured that the equation $x^2 + Ny^2 = 1$ has infinitely many solutions for any non-square integer N. Euler made a significant advance to this problem by proving it for specific values of N. The general form of the solution to equations of the form $x^2 + Ny^2 = 1$ was found later by Legendre and Lagrange.

Finally, Fermat famously asserted that there could be no three nonzero integers x, y,

and z such that the equation

$$x^n + y^n = z^n$$

holds, when $n \geq 3$. This statement is now known as *Fermat's Last Theorem*. He claimed to have a "truly marvelous proof", but that the margin of his copy of *Arithmetica* was too small to contain it. His assertion, known as *Fermat's Last Theorem*, captivated mathematicians for centuries. It was finally proven in 1995 by Andrew Wiles, about 350 years after Fermat made his famous note in the margin, using advanced techniques from algebraic geometry and modular forms. Wiles's achievement is considered one of the most remarkable milestones in the history of mathematics.

While Fermat laid the groundwork for modern number theory, we should mention that the concept of *function* hasn't been crystallized until Leibniz coined the term in 1673, using it to describe quantities related to curves, such as slopes, tangents, and areas. Bernoulli played a significant role in expanding the idea of functions. He used the term explicitly in 1718, describing it as a quantity dependent on another variable. Euler provided the first modern definition of a function in his *Introductio in Analysin Infinitorum* (1748) [Eul88]. He emphasized that functions could include both algebraic (e.g., polynomials) and transcendental forms (e.g., trigonometric and exponential functions). Peter Gustav Lejeune Dirichlet was the first to define functions abstractly as a correspondence between two sets, removing the requirement for expressions or formulas. This abstraction was a key insight for the theorem that we now know as *Dirichlet's theorem*: Every arithmetic progression

$$a,\ a+b,\ a+2b,\ a+3b,\ \ldots$$

in which a and b are coprime, contains an infinitude of primes. Dirichlet proved his theorem in 1837, more than 150 years after Leibniz first coined the term of function.

Euler's Introductio in Analysin Infinitorum marked the beginning of *analytical number theory*. As we mentioned before, Euler defined functions here and established the notation $f(x)$ of a function applied to its variable, which we still use today. He derived the famous formula for complex exponentials

$$e^{ix} = \cos(x) + i\sin(x).$$

This identity connects the exponential function, trigonometric functions, and the imaginary unit. The Riemann zeta function $\zeta(s)$ makes its first appearance here in a complex variable s, and he derived the celebrated *product formula*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}.$$

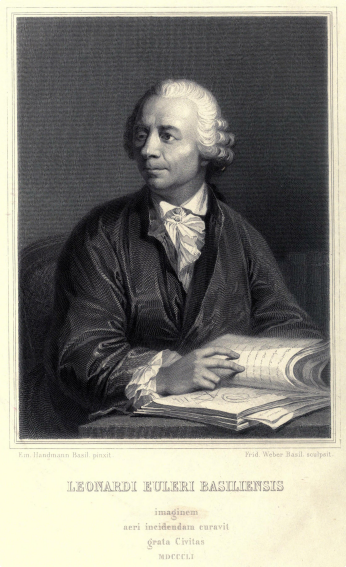


Figure 3: Leonhard Euler. Engraving by Friedrich Weber.

More generally, Euler explored infinite series and products. The *Basel summation formula*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

was first proven here, which was one of Euler's most celebrated achievements. Euler was keen to explore the connection between geometry and analysis, and in Introductio in Analysin Infinitorum, he advanced the idea that the study of functions and infinite series could provide new insights into geometric problems. He showed how techniques from analysis could be used to solve problems in geometry and vice versa, fostering the development of applied mathematics.

Building on the work of Euler, Gauss began using the properties of the zeta function and other analytic methods to investigate the distribution of the primes. His seminal work *Disquisitiones Arithmeticae* [Gau86] is often considered the starting point of the study of the distribution of primes, ultimately leading up to the *Prime Number Theorem*. Gauss conjectured that the number $\pi(n)$ of primes below a number n satisfies the asymptotic law

$$\pi(n) \sim \frac{n}{\log(n)}.$$

This conjecture was proved independently by Jacques Hadamard and Charles Jean de la Vallée Poussin. Furthermore, Gauss laid the foundation for modular arithmetic, and proved the celebrated law of quadratic reciprocity, which tells us when a quadratic equation modulo a prime number is solvable. More precisely, it states that if $q \equiv 1 \pmod 4$ then the congruence $x^2 \equiv p \pmod q$ is solvable if and only if the congruence $x^2 \equiv q \pmod p$ is solvable.

Figure 4: Carl Friedrich Gauss. Painting by Christian Albrecht Jensen.

So far, we have covered some of the most celebrated advances in humanity's understanding of numbers up to the early 19th century. Mathematics has made rapid progress since then. Bertrand's postulate, which asserts that there is always at least one prime p between n and $2n$, for any $n > 1$, was proven by Pafnuty Chebyshev in 1852, and Ben Green and Terence Tau proved in 2004 that the sequence of prime numbers contains arbitrarily long arithmetic progressions.

The 20th century saw the emergence of the Langlands program, a far-reaching research program that aims to connect number theory, representation theory, and algebraic geometry. With the contributions of mathematicians like André Weil, Robert Langlands, and many others, the program has shaped much of the direction of modern research in number theory and its applications to other branches of mathematics.

Even today fundamental concepts such as sets and functions continue to evolve. We live in an era of digital computation, with tools far beyond the reach of Fermat,

Euler, and Gauss, where essentially all recorded knowledge is readily available to almost anyone. This computational revolution has also given rise to powerful tools like proof assistants—computer programs designed to construct and formally verify mathematical proofs. Since proof assistants are programming languages, they are often based on type theory rather than set theory, slightly changing the way mathematics is done. Functions in type theory are even more general than Leibniz, Euler, and Dirichlet envisioned: unlike in set theory where functions are a specific kind of relations between sets, functions in type theory are primitive entities satisfying certain rules for evaluation, and they have an extra dependency built in, allowing types of their outputs to depend on the input. Proof assistants are being used with great success to formally verify advanced mathematical theorems. Georges Gonthier formally verified the Four Color Theorem and the Odd Order Theorem, and Tom Hales verified Kepler's conjecture about sphere packings. A current effort led by Kevin Buzzard aims to formally verify Fermat's Last Theorem.

Today, number theory continues to be one of the most active and fruitful areas of mathematical research, all of which started simply with counting on ten fingers.

Overview of the Course

An integer a is said to *divide* an integer b if there exists an integer x such that $ax = b$. In other words, the number a divides b if the equation

$$ax = b$$

has a solution in the integers. In this case we say that a is a *divisor* of b, and we write $a \mid b$. This equation is the simplest example of a Diophantine equation. Diophantine equations are equations expressed using variables, integers, and arithmetic operations. The primary goal of solving a Diophantine equation is to find integers for each of the variables for which the equation is true. The study of Diophantine equations is a cornerstone of number theory.

For example, the equation

$$x^2 - 1 = 0$$

is a *quadratic Diophantine equation* with two solutions: $x = \pm 1$. One of the most well-known quadratic Diophantine equations is the equation

$$x^2 + y^2 = z^2,$$

which is a Diophantine equation in *three variables*. Its solutions are known as *Pythagorean triples*. These triples, such as $(3, 4, 5)$ and $(5, 12, 13)$, corresponds to the side lengths of right-angled triangles and have been studied since antiquity. Perhaps the most famous Diophantine equation is Fermat's equation

$$x^n + y^n = z^n,$$

which has no solutions for nonzero integers x, y, and z when $n \geq 3$.

In this course, however, we will not go as far as proving Fermat's last theorem. The first target of this course is the Fundamental Theorem of Arithmetic, which establishes

the prime numbers as the building blocks of all natural numbers. A number n is said to be prime if it has exactly one divisor $d \mid n$ such that $d \neq n$. If this is the case, then its unique divisor that is not equal to itself is the number 1. The Fundamental Theorem of Arithmetic asserts that any nonzero natural number n can be written as a product of primes

$$n = p_1 p_2 \cdots p_k,$$

and that this decomposition of n as a product of primes is unique up to the ordering of the primes.

In order to study divisibility properties more deeply, Gauss introduced in his *Disquisitiones Arithmeticae* the congruence relations of modular arithmetic. Following Gauss, we say that a number a is congruent to b modulo n, that is,

$$a \equiv b \pmod{n}$$

if $n \mid b - a$. Gauss used this new formalism to study quadratic residues, the Chinese Remainder theorem, and a variety of other problems in number theory, a thread we will also pick up on in this course.

Some of the most important theorems in this context are Fermat's Little Theorem, Euler's Theorem, and Wilson's Theorem. Fermat's theorem asserts that given any prime number p and any number a that is not divisible by p we have

$$a^{p-1} \equiv 1 \pmod{p}.$$

Fermat's theorem is an important ingredient in primality tests: If the congruence $a^{n-1} \equiv 1 \pmod{n}$ is false then the number n cannot be prime. Testing this congruence for several values of a is a simple way of discovering that n is composite. This test is by itself, however, not conclusive because there are some numbers, the Carmichael numbers, that satisfy Fermat's congruence for all a relatively prime to n.

Euler's theorem is a sharper version of Fermat's Little Theorem, but it involves a new function: *Euler's totient function*. Euler's totient function ϕ counts the numbers less than n that are relatively prime to n, which means that they share no common divisors with n other than 1. For example, if p is a prime number then $\phi(p) = p - 1$ since every number $0 < n < p$ is relatively prime to p. Using the totient function, Euler's theorem asserts that the congruence relation

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

holds for any natural number n and any number a relatively prime to n. Euler's theorem has applications in cryptography.

Wilson's Theorem states that a number p is a prime number if and only if the congruence

$$(p-1)! \equiv -1 \pmod{p}$$

holds. Here, the exclamation mark is used for the *factorial function*: The number $n!$ is the product $1 \cdots n$ of all the numbers from 1 through n. Wilson's Theorem therefore gives another criterion for primality testing.

Euler's totient function is an example of an arithmetic function. Other such functions include the function $\tau(n)$, which returns the number of divisors of a number n, the

function $\sigma(n)$, which returns the sum of the divisors of a number n, and the *Möbius function* μ. These and other functions have important relations between them, that we will investigate next. The Möbius inversion formula, for instance, states that if f and g are two arithmetic functions, then we have

$$f(n) = \sum_{d|n} g(d) \qquad \text{if and only if} \qquad g(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right).$$

This result allows us to "invert" summation relations involving divisors.

The Möbius inversion is only one aspect of an algebraic structure that is present on the set of all arithmetic functions. Arithmetic functions can be multiplied by an operation known as *Dirichlet convolution*, which gives the set of arithmetic functions the structure of a *commutative ring*. Commutative rings are sets with operations of addition and multiplication satisfying the usual laws of arithmetic, and having such a structure on a set gives us a great opportunity to study them further. In our case, we get to study the primes more closely.

Experimentation is essential in number theory. By making lists of primes, lists of numbers that can be written as the sum of two squares, lists of square-free numbers, and so on, you will start to gather data on numbers and perhaps start seeing patterns that might otherwise feel elusive. At the end of this section, we have produced two number grids with the numbers from 1 to 1728. The great masters all have endlessly created such lists to organize and discover patterns in various kinds of numbers. Feel free to print it as many times as you like, and color them according to your own rules, or whenever something in the course piques your interest. Some exercises throughout the course will ask you to color this number grid in a certain way.

Literature

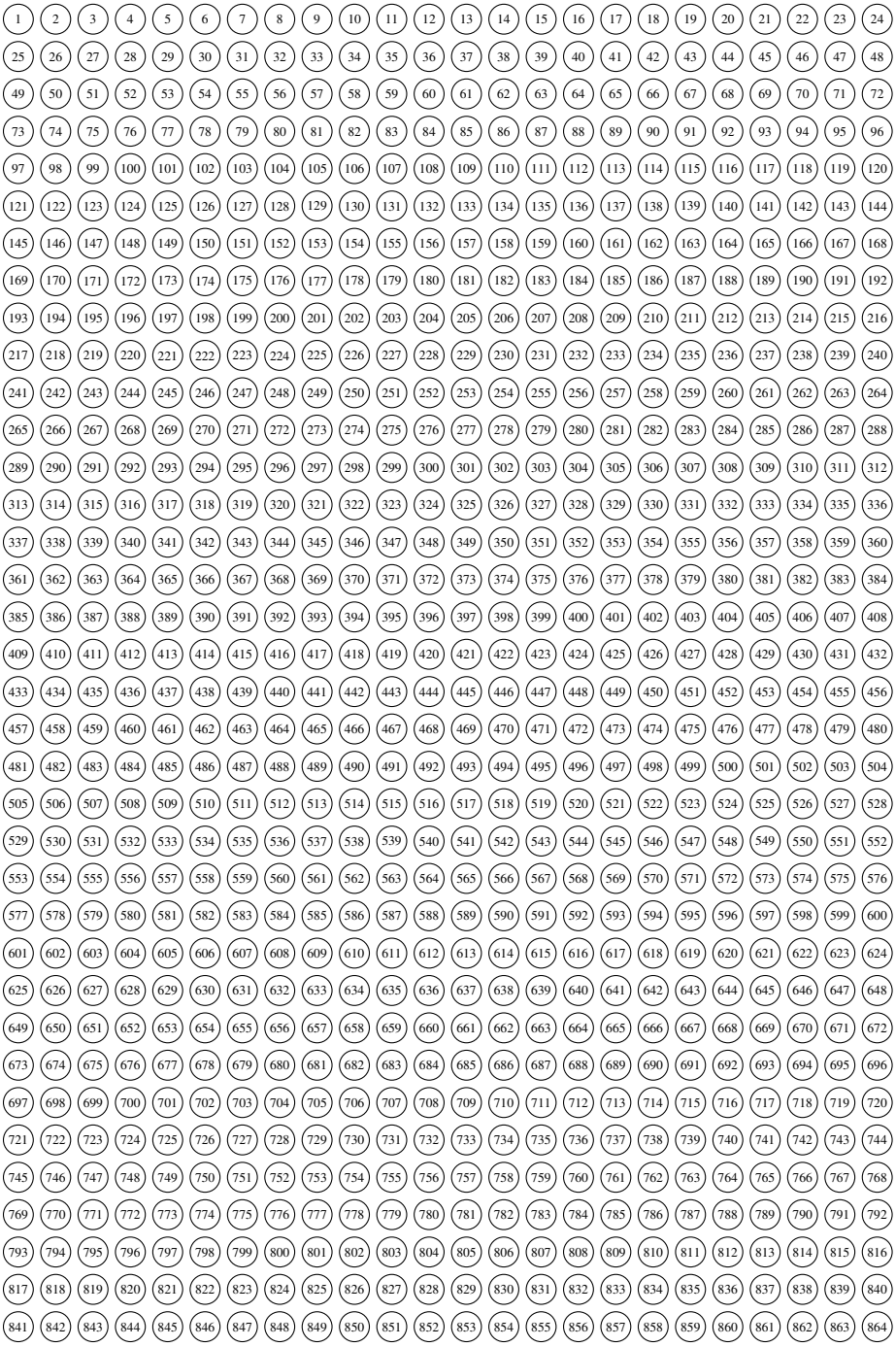
In this course we will be following [And94] fairly closely. One distinctive feature of this book is that it presents many of the most important theorems from two perspectives: a combinatorial one and an abstract one. This dual approach helps to clarify not only why these results are true but also appreciate why they are natural and inevitable within the broader framework of mathematics.

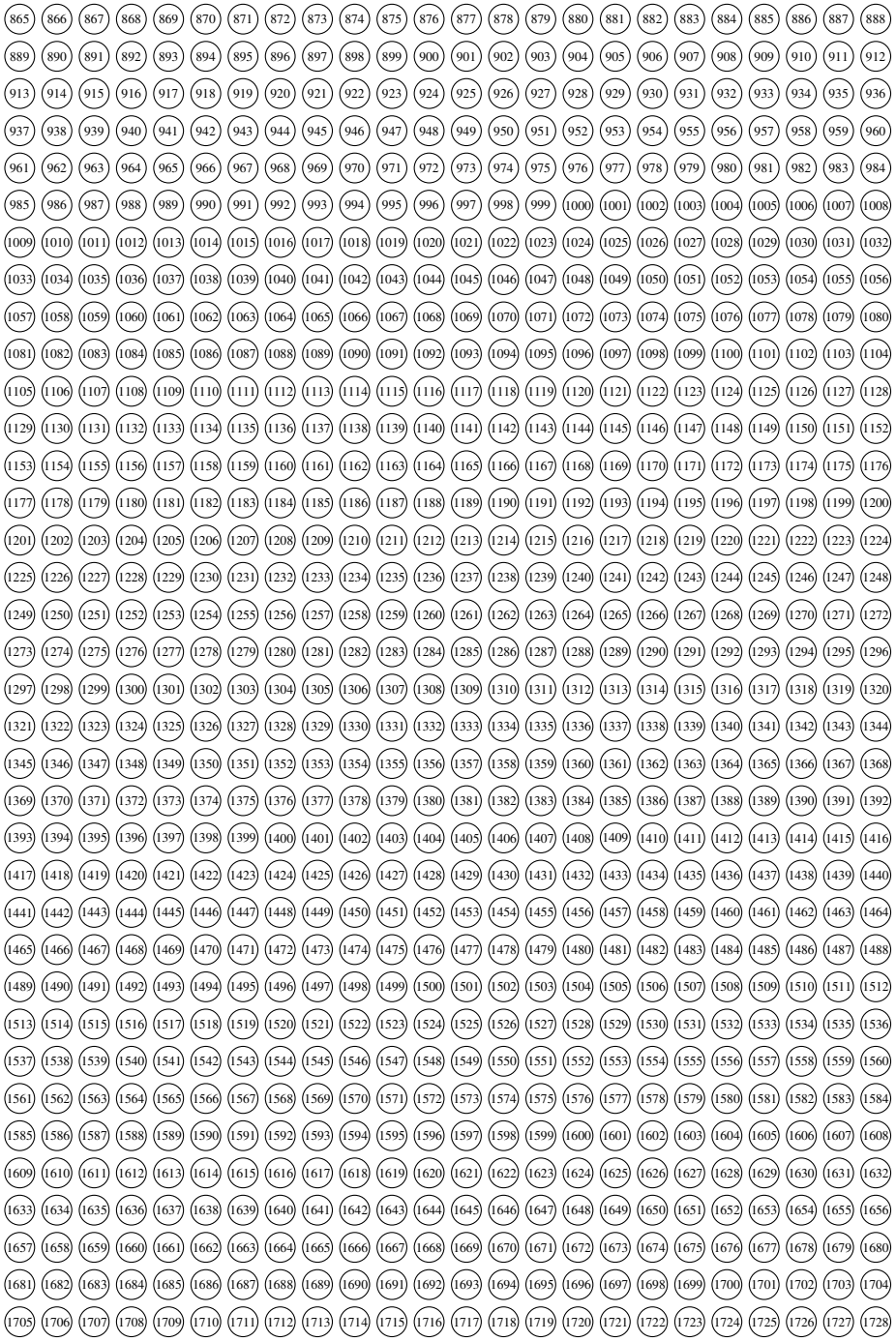
However, there are many excellent sources to learn number theory from. One of my personal favorites for its clarity and accessibility is LeVeque's *Topics in Number Theory* [LeV56a; LeV56b]. Both Andrews' and LeVeque's textbooks contain plenty of exercises, most of which are very fun.

Silverman's *Friendly Introduction to Number Theory* [Sil12] is another wonderful textbook, explaining the ideas behind many theorems and methods in number theory in a very accessible way, with practical examples and exercises. If you find some parts of Andrews or these notes hard to follow, you may well find the Friendly Introduction very helpful.

The undisputed classic textbook on number theory is Hardy and Wright's *Introduction to the Theory of Numbers* [Har+08]. This book covers all the essential topics in number theory, including elementary number theory and analytical number theory. It is more comprehensive and also provides more historical notes. The textbook of Hardy and

Wright does not provide exercises, but it contains the proofs of many important facts in number theory that are stated as exercises elsewhere.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96
97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168
169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192
193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264
265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288
289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312
313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336
337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384
385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408
409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456
457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504
505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528
529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552
553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576
577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624
625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648
649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672
673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696
697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744
745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768
769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792
793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816
817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864


865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888
889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912
913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936
937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984
985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008
1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032
1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056
1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104
1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128
1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152
1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176
1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224
1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248
1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272
1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296
1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344
1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368
1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392
1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416
1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464
1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488
1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512
1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536
1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560
1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584
1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608
1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632
1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656
1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704
1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728

Chapter 1

Mathematical Induction

1.1 Reasoning About Infinitely Many Natural Numbers

Many people are familiar with numbers through everyday experiences like counting, measuring, or comparing quantities. From a young age, we develop an intuition about how numbers behave. Our experience with numbers reveals patterns of truths and properties that seem to always be true. For example, we quickly learn that it does not matter in which order we add two numbers: Given two numbers a and b it is always true that $a + b = b + a$. Truths such as this one seem so self-evident that we might be tempted to just accept them. However, for a rigorous mathematical theory we need to be more careful. Mathematicians require proofs—conclusive arguments that establish why certain properties hold universally for all numbers, or at least for precisely defined sets of numbers. Checking a property for every number up to a hundred, a million, or even 10^{27} might confirm a pattern, but infinitely many numbers remain beyond such finite checks. To reason about all natural numbers at once, we need formal principles that extend beyond empirical observation and anecdotal evidence. In order to discover such reasoning principles, we need to find out what the natural numbers really are.

When we say that the natural numbers are the numbers 0, 1, 2, . . ., we really mean that the natural numbers are generated from 0 and the *successor function* $n \mapsto n + 1$. The collection of all natural numbers forms a set or type

$$\mathbb{N} := \{0, 1, 2, \ldots\}. \tag{1.1}$$

Mathematicians often prefer to work with sets, while computer scientists might prefer to work with types, both of which are collections of abstract objects often referred to as *elements*. The distinction is merely that of the logical framework being used: Most mathematicians rely on the Zermelo–Fraenkel set theory, while computer scientist rely on programming languages that often have a type theory in its foundation. Theorem proving in computer programs called *proof assistants* is rapidly becoming a more essential mathematical activity for students and researchers alike. However, many of the most popular proof assistants are based on type theory, in which collections of mathematical objects are called types.

The natural number zero and the successor function together are sufficient to describe all the natural numbers. The mathematician's way to express this sufficiency is by stating the principle of *mathematical induction*: Given any property $P(n)$ expressing a condition of an arbitrary natural number n, in order to prove that $P(n)$ is true for all n it suffices to prove that:

1. The property $P(0)$ is true.

2. For all n, if the property $P(n)$ holds then the property $P(n+1)$ holds.

Mathematical induction works, because every natural number is either 0 or it is the successor of a previous natural number. Proving a property for all natural numbers therefore breaks down in two cases: The *base case* in which we prove that $P(0)$ is true, and the *inductive step* where we prove that $P(n+1)$ is true provided that $P(n)$ is true.

Proofs by induction can be broken down into the following steps: First identify the property $P(n)$ that you want to prove by induction. The goal of an induction proof is to prove that $P(n)$ is true for all n. Then prove the base case. While this step is often trivial, it is a necessary and essential part of an induction proof. Finally, for the inductive step, write down the exact induction hypothesis, namely the property $P(n)$, and the exact goal of the inductive step, namely the property $P(n+1)$. Then use the induction hypothesis to prove that $P(n+1)$ then also holds. After these steps the proof that $P(n)$ holds is complete.

Theorem 1.1.1. *For any natural number n we have*

$$0 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. Let $S_n := 0 + \cdots + n$, and let $P(n)$ be the property that $S_n = \frac{n(n+1)}{2}$. We will prove that $P(n)$ is true for all n by induction. In the base case, we have to show that $P(0)$ holds, that is, that the identity

$$0 = \frac{0 \cdot 1}{2}$$

is true. This is indeed true, because the numerator in the fraction on the right-hand side is 0.

For the induction step, let n be a natural number and assume as our inductive hypothesis that $P(n)$ is true. Our goal is now to prove $P(n+1)$, that is, that the identity

$$S_{n+1} = \frac{(n+1)(n+2)}{2}$$

Note that $S_{n+1} = S_n + (n+1)$. By the inductive hypothesis we have that $S_n = \frac{n(n+1)}{2}$, which we may now use to rewrite

$$S_n + (n+1) = \frac{n(n+1)}{2} + (n+1).$$

Observe that $(n+1) = \frac{2(n+1)}{2}$. Therefore we can make the following computation:

$$\frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+2)(n+1)}{2}.$$

Since $(n+2)(n+1) = (n+1)(n+2)$ it follows that $S_{n+1} = \frac{(n+1)(n+2)}{2}$ as desired. □

We illustrate induction with another example, the *formula for the geometric series*.

Theorem 1.1.2. *For any real number x not equal to* 1*, the identity*

$$\sum_{k=0}^{n-1} x^k = \frac{x^n - 1}{x - 1}$$

Proof. The proof is by induction on n. In the case $n = 0$ the sum

$$\sum_{k=0}^{-1} x^k$$

is an empty sum, because it sums over all elements $0 \leq k \leq -1$ of which there are none. In empty sums, nothing is being added, so they are always 0. On the right hand side, we also see that

$$\frac{x^0 - 1}{x - 1} = \frac{1 - 1}{x - 1} = 0,$$

and hence the base case holds.

For the inductive step, assume that $\sum_{k=0}^{n-1} x^k = \frac{x^n - 1}{x-1}$. Our goal is to show that

$$\sum_{k=0}^{n} x^k = \frac{x^{n+1} - 1}{x - 1}.$$

Note that $\sum_{k=0}^{n} x^k = (\sum_{k=0}^{n-1} x^k) + x^n$. By applying the induction hypothesis we therefore find that

$$\sum_{k=0}^{n} x^k = \frac{x^n - 1}{x - 1} + x^n.$$

Note that $x^n = \frac{(x-1)x^n}{x-1}$. Therefore we can bring the two summands under one fraction as follows:

$$\frac{x^n - 1}{x - 1} + x^n = \frac{x^n - 1}{x - 1} + \frac{(x-1)x^n}{x - 1} = \frac{x^n - 1 + x^{n+1} - x^n}{x - 1} = \frac{x^{n+1} - 1}{x - 1}.$$

This completes the inductive step, and therefore the proof. □

Exercises

1.1 The *factorial function* $n \mapsto n!$ is defined recursively by

$$0! := 1$$
$$(n+1)! := n!(n+1).$$

Prove that $2^n < n!$ for all $n \geq 4$.

1.2 Prove that that the sum of the first n odd numbers is a perfect square. That is, prove that

$$\sum_{k=0}^{n-1} 2k + 1 = n^2.$$

Note: It is usual in mathematics that an empty sum $\sum_{k=0}^{-1} a_k$ is taken to be 0, because no numbers are added.

1.3 Prove that

$$\sum_{k=0}^{n} k(k+1) = \frac{n(n+1)(n+2)}{3}.$$

1.4 Prove the *formula for the square pyramidal numbers*

$$\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Can you explain why the numbers of the form $\sum_{k=0}^{n} k^2$ are called square pyramidal numbers?

1.5 Prove that

$$\sum_{k=0}^{n-1} (2k+1)^2 = \frac{n(4n^2-1)}{3}.$$

1.6 Prove *Nicomachus's Theorem*

$$\sum_{k=0}^{n} k^3 = \frac{n^2(n+1)^2}{4}.$$

1.7 Prove that

$$\sum_{k=0}^{n-1} (2k+1)^3 = n^2(2n^2-1).$$

1.8 Prove that

$$\sum_{k=0}^{n} k(k+1)(k+2) = \frac{n(n+1)(n+2)(n+3)}{4}$$

1.9 Prove the *formula for the difference of nth powers*

$$x^n - y^n = (x-y) \sum_{k=0}^{n-1} x^k y^{n-1-k}.$$

Note that this formula generalizes the *formula for the difference of squares*

$$x^2 - y^2 = (x-y)(x+y).$$

1.10 Prove that

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$$

1.11 Prove that

$$\prod_{k=2}^{n} \left(1 - \frac{1}{k}\right) = \frac{1}{n}.$$

1.12 Prove that

$$\prod_{k=1}^{n} \left(1 - \frac{1}{k^2}\right) = \frac{n+1}{2n}$$

1.13 Define the nth *Fermat number* to be

$$\Phi_n := 2^{2^n} + 1.$$

Prove that

$$\Phi_n - 2 = \prod_{k=0}^{n-1} \Phi_k.$$

Hint: Use the formula for the difference of squares.

1.14 Consider the Fibonacci numbers $F(n)$ given by

$$\begin{aligned} F_0 &:= 0 \\ F_1 &:= 1 \\ F_{n+2} &:= F_{n+1} + F_n. \end{aligned}$$

Prove that

$$\sum_{k=0}^{n} F_k = F_{n+2} - 1.$$

1.15 Prove that

$$\sum_{k=0}^{n-1} F_{2k+1} = F_{2n}.$$

1.16 Prove that

$$\sum_{k=0}^{n} F_{2k} = F_{2n+1} - 1.$$

1.17 Prove *Cassini's identity*

$$F_{n+1}^2 - F_n F_{n+2} = (-1)^n.$$

1.18 Consider two nonzero natural numbers a and b satisfying $a(a+b) < b^2$. Prove that the strict inequality

$$F_n < \left(\frac{b}{a}\right)^n$$

holds for all n. Use Cassini's identity to give some examples of natural numbers a and b that satisfy this inequality.

Chapter 2

Counting

Combinatorics, or discrete mathematics, is mathematics of finite structures. A central theme in combinatorics is counting, establishing a closed form for the number of objects of a certain kind. Sometimes, counting methods can be used to obtain results in number theory. In this lecture we will use a combinatorial argument to show that the product

$$n(n+1)\cdots(n+k-1)$$

of any k consecutive natural numbers is always divisible by $k!$. We will also establish the binomial theorem, which has a myriad of applications throughout mathematics.

2.1 Counting Bijections

Definition 2.1.1. A function $f : A \to B$ is said to be a *bijection* if for every $y \in B$, the *preimage* of b

$$f^{-1}(y) := \{x \in A \mid f(x) = y\}$$

has exactly one element. We write $A \cong B$ for the set of bijections from A to B.

A function $f : A \to B$ is a bijection precisely when it is *invertible* in the sense that there is a function $g : B \to A$ such that $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$. Here, the composite $g \circ f$ of g and f is the function $x \mapsto g(f(x))$ defined by applying g to the value $f(x)$ of f at the input variable x.

In the following proposition we prove the number of bijections $[n] \cong [n]$ on the *standard n-element set*

$$[n] := \{0, \ldots, n-1\}.$$

is the number $n!$. In other words, the factorial function counts the number of bijections on the standard n-element set.

Proposition 2.1.2. *The number of bijections* $[n] \cong [n]$ *from the standard n-element set to itself is exactly* $n!$.

Proof. The proof is by induction on n. For the base case, let $n = 0$. Then $[n]$ is an empty set, and there is exactly one function $[n] \to [n]$, the empty function. This function is a bijection, because assuming an element of an empty set is an inherent contradiction.

Now suppose that the number of bijections $[n] \cong [n]$ is $n!$. We claim that for each $y \in [n+1]$ there are exactly $n!$ bijections $f : [n+1] \cong [n+1]$ such that $f(n) = y$.

To see this, we first define the *transposition function* $s_{y,n}$ defined by

$$s_{y,n}(x) = \begin{cases} n & \text{if } x = y \\ y & \text{if } x = n \\ x & \text{otherwise.} \end{cases}$$

In other words, the function $s_{y,n}$ swaps the elements n and y, and leaves the other elements fixed. This function is a bijection, because we can verify that the preimage $s_{y,n}^{-1}(z)$ is a singleton set for every $z \in [n+1]$. The preimage $s_{y,n}^{-1}(n)$ is the singleton set $\{y\}$; the preimage $s_{y,n}^{-1}(y)$ is the singleton set $\{n\}$, and the preimage $s_{y,n}^{-1}(x)$ is the singleton set $\{x\}$ otherwise.

Now we observe that the function

$$([n+1] \cong [n+1]) \to ([n+1] \cong [n+1])$$

given by $g \mapsto s \circ g$ is itself a bijection, because it is an invertible function. Indeed, it is its own inverse, because $s \circ s \circ g = g$ for any bijection g.

Thus, if $f : [n+1] \cong [n+1]$ is a bijection such that $f(n) = y$, then $s \circ f$ is a bijection satisfying $s(f(n)) = n$. In other words, every bijection satisfying $f(n) = y$ corresponds uniquely to a bijection satisfying $f(n) = n$.

Now we observe that there are exactly $n!$ bijections $f : [n+1] \cong [n+1]$ satisfying $f(n) = n$. Indeed, such bijections are uniquely determined by their restriction to the set $[n]$, and by the induction hypothesis there are $n!$ such bijections. Since there are $n+1$ possible choices of a value y, we conclude that there are $n!(n+1)$ bijections altogether from $[n+1]$ to $[n+1]$. □

2.2 Counting Subsets

Definition 2.2.1. The *binomial coefficients* $\binom{n}{k}$ are defined recursively by

$$\binom{0}{0} := 1 \qquad \binom{0}{k+1} := 0$$
$$\binom{n+1}{0} := 1 \qquad \binom{n+1}{k+1} := \binom{n}{k} + \binom{n}{k+1}.$$

The binomial coefficients can be arranged in *Pascal's triangle*, where each entry is the sum of the two directly above it. At the top of the triangle we find the binomial coefficient $\binom{0}{0}$. This is 0th row of Pascal's triangle. In the nth row from the top, we find the binomial coefficients $\binom{n}{k}$ for $0 \leq k \leq n$.

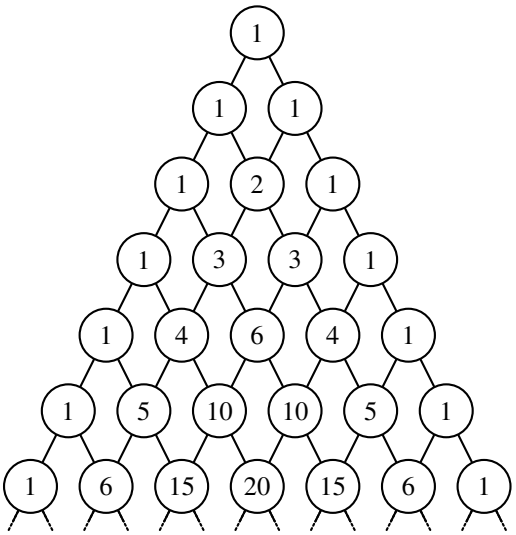


Figure 2.1: Pascal's triangle.

Theorem 2.2.2. *Consider an n-element set S, and a natural number k. Then the number of k-element subsets of S is the binomial coefficient $\binom{n}{k}$.*

Proof. Since the binomial coefficients are defined recursively, this theorem is best proven by induction on both variables.

If S has no elements and $k = 0$, then there is exactly one subset of S having k elements, the empty subset. This shows that the number of 0-element subsets of S is $\binom{0}{0}$.

If S has no elements, then there are no $(k+1)$-element subsets of S, since any $(k+1)$-element subset has at least one element, but S has no elements. This shows that the number of $(k+1)$-element subsets of S is $\binom{0}{k+1}$.

If S has $n+1$ elements, then there is exactly one subset of S with no elements, the empty subset. This shows that the number of 0-element subsets of S is $\binom{n+1}{0}$.

If S has $n+1$ elements, then S has at least one element x_0. Now there are two classes of subsets having $(k+1)$ elements: the subsets containing x_0 and the subsets not containing x_0. A $(k+1)$-element subset of S containing x_0 is equivalently described as a k-element subset of the n-element set $S \setminus \{x_0\}$, because its $(k+1)$st element is the element x_0. By the induction hypothesis, there are exactly $\binom{n}{k}$ such subsets. Furthermore, a $(k+1)$-element subset of S not containing x_0 is equivalently described as a $(k+1)$-element subset of the n-element set $S \setminus \{x_0\}$. By the induction hypothesis, there are exactly $\binom{n}{k+1}$ such subsets. Combining these two observations, we find that the number of $(k+1)$-element subsets of S is

$$\binom{n}{k+1} + \binom{n}{k} = \binom{n+1}{k+1}. \qquad \square$$

Definition 2.2.3. We define the set $P_S(k)$ of k-permutations of a set S recursively by:

1. The set $P_S(0)$ of 0-permutations of S is the set $\{*\}$ with one element.

2. The set $P_S(k+1)$ of $(k+1)$-permutations of S is the set

$$\{(s,t) \mid s \in S,\ t \in P_{S\setminus\{s\}}(r)\}$$

of pairs (s,t) consisting of an element in S and an k-permutation of the set $S \setminus \{s\}$.

Lemma 2.2.4. *The number of k-permutations of an n-element set S is the number ${}_nP_k$ defined by*

$${}_nP_k := n(n-1)\cdots(n-k+1).$$

Proof. The proof is by induction on k. There is exactly one 0-permutation of any n-element set. To see that the number of $(k+1)$-permutations of an n-element set is $n\cdots(n-k)$, note that such a $(k+1)$-permutation consists of a choice of an element of S, and a k-permutation on the remaining $n-1$-element set of elements. Thus, the number of $(k+1)$-permutations on an n-element set is n times the number of k-permutations on an $(n-1)$-element set, i.e., it is

$$n(n-1)\cdots((n-1)-k+1) = n(n-1)\cdots(n-k). \qquad \square$$

Proposition 2.2.5. *There is a bijection between the set of k-permutations of an n-element set S, and the set*

$$\{(A,f) \mid A \subseteq S, f : \{0,\dots,k-1\} \cong A\}.$$

of pairs (A,f) consisting of a subset $A \subseteq S$ and a bijection $\{0,\dots,k-1\} \cong A$.

Proof. There is exactly one 0-permutation of any n-element set S, and likewise the set of empty subsets A of S equipped with a bijection $[0] \cong A$ also contains exactly one element: the empty subset equipped with the empty bijection.

For the inductive step, consider a $(k+1)$-permutation (s,t) of S, consisting of an element s and a k-permutation t of the set $S \setminus \{s\}$. By the induction hypothesis, the k-permutation t corresponds uniquely to a subset $A \subseteq S \setminus \{s\}$ equipped with a bijection $f : [n] \cong A$. Thus, the $(k+1)$-permutation (s,t) corresponds to the subset $B := A \cup \{s\} \subseteq S$, equipped with the bijection $g : [n+1] \cong B$ given by $g(x) := f(x)$ for $x < n$ and $g(n) := s$. $\square$

Proposition 2.2.6. *The product $n(n-1)\cdots(n-k+1)$ of any k consecutive numbers is divisible by $k!$.*

Proof. By Proposition 2.2.5 there is a bijection between the set of k-permutations of the set $S := [n]$, and the set

$$\{(A,f) \mid A \subseteq S, f : \{0,\dots,k-1\} \cong A\}.$$

These two sets therefore have the same number of elements. Thus, it follows that

$$n(n-1)\cdots(n-k+1) = \binom{n}{k} \cdot k!,$$

showing that the left-hand side is divisible by $k!$. $\square$

Corollary 2.2.7. Formula for binomial coefficients. *For any two natural numbers n and k, we have*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proof. By the previous proposition we have $\binom{n}{k} \cdot k! = {}_nP_k = n!/k!$. □

2.3 The Binomial Theorem

If we expand the exponent $(x+y)^3$, we get

$$\begin{aligned}(x+y)^3 &= (x+y)^2(x+y)\\ &= (x^2+2xy+y^2)(x+y)\\ &= x^3+2x^2y+xy^2+x^2y+2xy^2+y^3\\ &= x^3+3x^2y+3xy^2+y^3.\end{aligned}$$

Going further, if we expand the exponent $(x+y)^4$, we get

$$\begin{aligned}(x+y)^4 &= (x^3+3x^2y+3xy^2+y^3)(x+y)\\ &= x^4+3x^3y+3x^2y^2+xy^3+x^3y+3x^2y^2+3xy^3+y^4\\ &= x^4+4x^3y+6x^2y^2+4xy^3+y^4.\end{aligned}$$

Now we might recognize an emerging pattern in the coefficients of these polynomials: they are all binomial coefficients! The binomial theorem generalizes this result for all exponents n.

The binomial theorem has many proofs, including combinatorial proofs and algebraic proofs. All these proofs have their own merits. We will present here an algebraic proof, using induction, since it is the most easily applicable to settings other than the integers.

Theorem 2.3.1. The binomial theorem. *In any number system, such as the natural numbers, the integers, the rational numbers, the reals, the complex numbers*[1]*, we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. We prove the statement by induction on n. In the base case, we find that both sides of the equation

$$(x+y)^0 = \sum_{k=0}^{0} \binom{n}{k} x^k y^{0-k}$$

[1] More generally, the binomial theorem applies to any semiring, conditional on the assumption that $xy = yx$.

evaluate to 1, so the equation is true. In the inductive step, we have that

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}x^{k+1}y^{n-k} + \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k+1} \\
&= x^{n+1} + \sum_{k=0}^{n-1}\binom{n}{k}x^{k+1}y^{n-k} + \sum_{k=1}^{n}\binom{n}{k}x^k y^{n-k+1} + y^{n+1} \\
&= x^{n+1} + \sum_{k=1}^{n}\binom{n}{k-1}x^k y^{n-k+1} + \sum_{k=1}^{n}\binom{n}{k}x^k y^{n-k+1} + y^{n+1} \\
&= x^{n+1} + \sum_{k=1}^{n}\left(\binom{n}{k-1} + \binom{n}{k}\right)x^k y^{n-k+1} + y^{n+1} \\
&= \sum_{k=0}^{n+1}\binom{n+1}{k}x^k y^{(n+1)-k}.
\end{aligned}$$

□

The binomial theorem has a long history. The earliest known references to aspects of the binomial theorem are found in the Chinese mathematical text *Jiuzhang Suanshu* (Nine Chapters on the Mathematical Art), from approximately the 2nd century BCE. Around the 11th century CE, the Chinese mathematician Jia Xian described a method for calculating binomial coefficients, which corresponds to what we now refer to as Pascal's triangle. This method was popularized by Yang Hui in the 13th century CE, and thus Pascal's triangle is sometimes called the *Yang Hui triangle* in China. Isaac Newton was the first to generalize the binomial theorem to non-integer number systems, in 1687 in his *Principia Mathematica*.

2.4 The Inclusion-Exclusion Principle

Consider two subsets A and B of a set X. The *union* of A and B is the set

$$A \cup B := \{x \in X \mid x \in A \text{ or } x \in B\}.$$

Given that X is a finite set, we may wish to determine the number of elements in $A \cup B$. Note, however, that in general the answer is not found by simply adding up the number of elements of A and the number of elements in B. Indeed, some elements might be contained in both subsets, and they would therefore be counted double. Since every element in the intersection $A \cap B$ is counted twice, the correct formula for the number of elements in $A \cup B$ is:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Example 2.4.1. Consider the set $X = \{1, \ldots, 100\}$, and our goal is to find the number of integers in X that are neither divisible by 2 nor by 3. This problem can be solved with the inclusion-exclusion principle.

Let A and B be the subsets of X consisting of numbers divisible by 2 and by 3, respectively. The set $A \cap B$ then consists of all the numbers divisible by both 2 and 3; that is, the numbers divisible by 6.

There are 50 numbers in X divisible by 2, there are 33 numbers in X divisible by 3, and there are 16 numbers in X divisible by 6. Thus, the total count of numbers in X not divisible by either 2 or 3 is

$$100 - 50 - 33 + 16 = 33.$$

The inclusion-exclusion principle is a generalization of the formula for the number of elements of $A \cup B$ to any finite number of subsets of a set X. To see the pattern, let us consider the case with three subsets A, B, and C of a set X. Now, if we wish to determine the number of elements in $A \cup B \cup C$, we can again start by adding the numbers $|A|$, $|B|$, and $|C|$. Every element that was in exactly one of the three subsets is correctly accounted for, but we have overcounted the elements in $A \cap B$, $A \cap C$, and $B \cap C$. If we subtract the numbers $|A \cap B|$, $|A \cap C|$, and $|B \cap C|$ from our total, then we have correctly accounted for every element that was in at most two subsets. However, now we have subtracted every element in $A \cap B \cap C$ three times from our count, so we must add them again to arrive at the correct number:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Thus, we see an alternating pattern emerging: Single sets are added, intersections of two sets are subtracted, triple intersections are added again, and so on. The inclusion-exclusion principle is stated with this alternating pattern.

Theorem 2.4.2. *Consider a set X and a finite family of subsets $A_i \subset X$ indexed by $1 \leq i \leq n$. Then the number of elements in the union of the subsets A_i is given by*

$$|A_1 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} |A_{i_1} \cap \cdots \cap A_{i_k}| \right).$$

Exercises

2.1 Prove that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

in three different ways: using the binomial theorem, by a direct proof by induction, and by a combinatorial proof.

2.2 Prove that

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k = 0.$$

Conclude that

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} = \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \binom{n}{2k+1}.$$

In this formula, the floor $\lfloor x \rfloor$ of a number x is the largest integer below or equal to x.

2.3 *The Chu–Vandermonde identity.* Use the binomial theorem at the polynomial $(x+1)^m(x+1)^n$ to show that

$$\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}.$$

2.4 Show that

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

The binomial coefficient $\binom{2n}{n}$ is also called the *central binomial coefficient*.

2.5 Show that the strict inequalities

$$\frac{4^n}{2n+1} < \binom{2n}{n} < 4^n$$

hold for all $n \geq 1$.

2.6 Show that

$$\sum_{k=0}^{n} \prod_{i=0}^{m} (k+i) = (m+1)!\binom{n+m+1}{m+2}.$$

2.7 Find the first 20 positive integers that are not divisible by any Fibonacci number other than 1. This is https://oeis.org/A147956

2.8 Define the *inclusion-exclusion sequence* a_n by $a_0 := 0$, $a_1 := 1$, and let a_{n+1} be the largest positive integer so that the number of positive integers $1 \leq x < a_{n+1}$ not divisible by any a_i for $2 \leq i \leq n$ is a_n. Determine the values for a_n for $1 \leq n \leq 10$.

Chapter 3

Euclidean Division and the Representability Theorem

3.1 The Well-Ordering Principle of the Natural Numbers

The induction principle of the natural numbers implies other important reasoning principles. One of them is the *well-ordering principle*. The well-ordering principle has many practical implications. For example, in the next section we will apply a variation of the well-ordering principle, which we will describe below, to construct division with remainder.

Theorem 3.1.1. *Any inhabited subset of the natural numbers has a least element.*

Proof. We will prove, by induction on n, that every subset $A \subseteq \mathbb{N}$ with $n \in A$ has a least element.

In the base case we have to show that every subset $A \subseteq \mathbb{N}$ containing 0 has a least element. Since 0 is the least natural number, it follows that any subset containing 0 must have a least element.

For the inductive step assume that every subset $A \subseteq \mathbb{N}$ containing n has a least element. Our goal is to show that every subset $A \subseteq \mathbb{N}$ containing $n + 1$ has a least element. There are two cases to consider: either $0 \in A$ or $0 \notin A$. In the first case, where $0 \in A$ is assumed, then A clearly has a least element. In the second case, define the subset $B \subseteq \mathbb{N}$ to be

$$\{x \in \mathbb{N} \mid x + 1 \in A\}.$$

In other words, B is the set of all natural numbers x such that $x + 1$ is in A. Since $n + 1$ is in A by assumption, it follows that $n \in B$. However, note that we are in position to apply the induction hypothesis now: The set B must have a least element m.

We claim that $m + 1$ must be the least element of A. First of all, we know that $m + 1 \in A$ because $m \in B$. Second of all, if $x \in A$ then we know that x is nonzero. This means that $x = x' + 1$ for some natural number x'. In other words, we have $x' + 1 \in A$, which means that $x' \in B$. Since m is the least element of B it follows that $m \leq x'$. This

implies that $m + 1 \leq x' + 1$, so $m + 1 \leq x$ follows. This shows that any element of A must be greater than or equal to $m + 1$. In other words, $m + 1$ is the least element of A. □

We also have some variations of the well-ordering principle that apply to subsets of the integers. For the following theorem, we say that a subset $A \subseteq \mathbb{Z}$ is *bounded from below* if there is an integer b such that $b \leq x$ for every element $x \in A$. Similarly, we say that a subset $A \subseteq \mathbb{Z}$ is *bounded from above* if there is an integer b such that $x \leq b$ for every element $x \in A$.

Theorem 3.1.2. *Let A be a subset of the integers. We make the following two claims:*

1. *If A is inhabited and bounded from below, then A has a least element.*
2. *If A is inhabited and bounded from above, then A has a largest element.*

Proof. For the first claim, assume that A is inhabited and suppose that $b \leq x$ for every element $x \in A$. Then we define the subset

$$B := \{x \in \mathbb{Z} \mid x + b \in A\}.$$

The subset B is inhabited and it is bounded from below by 0. This means that B is in fact a subset of the natural numbers, and so it has a least element by the well-ordering principle, Theorem 3.1.1. Observe that $x \in B$ if and only if $x + b \in A$, so if x is the least element of B then $x + b$ is the least element of A. This proves the first claim.

The second claim follows from the first, because if A is inhabited and bounded from above, then the set

$$B := \{x \in \mathbb{Z} \mid -x \in A\}$$

is inhabited and bounded from below. Since the set B has a least element by the previous claim, it follows that the set A has a largest element. □

3.2 Euclidean Division

The Euclidean division operation with a divisor $d > 0$ is the operation that returns for each integer a the division with remainder of a by d. More specifically, the Euclidean division operation of a by an integer d returns two integers, the *quotient* q and the *remainder* r, such that $0 \leq r < d$ and such that

$$a = q \cdot d + r$$

For example, if $a := 23$ and $d := 5$, then the Euclidean division operation returns $q := 4$ for the quotient, and $r := 3$ for the remainder. Note that we have $0 \leq 3 < 5$, and the identity

$$23 = 4 \cdot 5 + 3$$

holds. We will prove that for every $d > 0$ there is always a unique such pair (q, r) of integers satisfying the two constraints for Euclidean division. Before we do so, we first explain in more detail how to go about *unique existence* proofs.

Suppose X is a set and $P(x)$ is a property of the elements of the set X. For a concrete example, which we will work out below, take X to be the set $\mathbb{N}$ of natural numbers and $P(x)$ to be the property that $x = 0$. Now suppose someone asks you to prove that there is a unique element x that satisfies $P(x)$. Then your task splits up in two parts:

1. First, you have to prove existence. Here, your task is to find a concrete element $a \in X$ and prove that it satisfies the property $P(a)$.

2. Second, you have to prove uniqueness. This means that you have to show for any two elements x and y, if both $P(x)$ and $P(y)$ hold then we must have $x = y$. In other words, in the uniqueness part of the proof you are tasked with showing that there is *at most one* element satisfying P.

The existence part of the proof ensures that at least one solution exists, while the uniqueness part of the proof ensures that there is at most one solution. Both steps combined allow us to conclude that there exists exactly one element of x such that $P(x)$ holds. In a logical formula, it is customary to write $\exists!_{(x \in X)} P(x)$ for the unique existence property.

The example where we take $X := \mathbb{N}$ and $P(x) := x = 0$ is simple, but it captures the essence of unique existence proofs. To show that there is a unique natural number equal to 0, we first have to give an example of a natural number that is equal to 0. We simply take the number 0 itself. For the uniqueness part of the proof, we have to show that any two numbers that are equal to 0 are equal to each other. This is evident, because they are both equal to 0. Thus we conclude that

$$\exists!_{(n \in \mathbb{N})} n = 0.$$

Theorem 3.2.1. *Consider an integer a and an integer $d > 0$. Then there exists a unique pair of integers (q, r) such that $0 \leq r < d$ satisfying the identity*

$$a = q \cdot d + r.$$

Proof. Consider the set

$$A := \{x \in \mathbb{Z} \mid xd \leq a\}$$

of integers x whose product with d is at most a. We will show that A is bounded from above and that A is inhabited. Together, this will imply by Theorem 3.1.2 that A has a largest element, which will be our quotient q.

First, we show that the set A has an upper bound. There are two cases to consider: either $a \geq 0$ or $a < 0$. If $a \geq 0$, the integer a is an upper bound because for any $x \in A$ we have $x \leq \max(0, xd) \leq a$. If $a < 0$ then the integer 0 is an upper bound because $xd \leq a < 0$ implies that x must be negative.

Next, we observe that the integer $-|a|$ is always in the set A, because $-|a|d \leq -|a| \leq a$. Therefore it follows that the set A has a largest element, which we will call q.

By definition q satisfies $qd \leq a$ and q is the largest such element. By maximality of q it follows that $a < (q + 1)d = qd + d$. Now we take $r := a - qd$. Since $qd \leq a$ it follows that $0 \leq r$, and since $a < qd + d$ it follows that $r < d$. This completes the existence of q and r.

It remains to prove uniqueness. Suppose that (q, r) and (q', r') are two pairs of integers satisfying $0 \leq r < d$, $0 \leq r' < d$, and the equations

$$a = qd + r, \qquad \text{and} \qquad a = q'd + r'.$$

It follows that

$$(q - q')d = r' - r.$$

However, the integer $r' - r$ satisfies the strict inequalities $-d < r' - r < d$ and it is divisible by d. This implies that $r' - r = 0$, so it follows that $r = r'$. Now it also follows that $(q - q')d = 0$. Since d was assumed to be positive, this implies that $q - q' = 0$, from which it follows that $q = q'$. This completes the proof of uniqueness. □

3.3 The Representability Theorem

The *decimal representation* of integers is well-embedded in our culture: Any integer can be uniquely represented as a sequence of digits from 0 to 9, where the leading digit is nonzero. When we write a number such as 231, it represents the number

$$2 \cdot 10^2 + 3 \cdot 10^1 + 1 \cdot 10^0.$$

Similar representations exists in binary (using only the digits 0 and 1), and in hexadecimal notation (using the digits 0 to 9 and the letters A to F representing digits of values 10 through 15). The binary representation of the number 231 is 11100111, while the hexadecimal representation of this number is $E7$. More generally, for any *base* k there is a unique representation of all integers using powers of k.

Theorem 3.3.1. *Consider a natural number $k > 1$. Then there exists for every natural number n a unique list ℓ of digits $d_0, \ldots, d_{l-1}$ of length l, such that the inequalities $0 \leq d_i < k$ hold for each $i < l$, such that the* leading digit *$d_{l-1} \neq 0$, and satisfying the equation*

$$n = \sum_{i=0}^{l-1} d_i k^i.$$

Before proving the theorem, let's first discuss the main idea behind the proof. Since we assumed that $k > 0$, it follows from the Euclidean division theorem, Theorem 3.2.1, that there is a unique pair (q, r) of numbers such that $n = qk + r$ and $0 \leq r < k$. This gives us the first digit in the representation theorem[1]: the number $d_0 := r$.

The representability theorem would now follow if we already knew that q has a unique representation. If $d_0 = r$ and if $d_1, \ldots, d_{l-1}$ is the unique representation of q, then the list $d_0, \ldots, d_{l-1}$ is a unique representation of n. However, the usual induction principle doesn't allow us to make this step, because n is not necessarily the successor of q. Therefore we need something stronger: the *strong induction principle*.

Theorem 3.3.2. *Consider a property $P(n)$ of the natural numbers, and assume that the following two conditions hold:*

[1] It should be noted that this digit appears last in the actual representation. It is the "ones digit" of the representation.

1. *The property $P(0)$ is true.*

2. *If the property $P(k)$ is true for all $k \leq n$, then the property $P(n+1)$ is true.*

Then $P(n)$ is true for all n.

The strong induction principle can be formulated more succinctly using the *universal quantifier*. The universal quantifier $\forall$ is used to express that a property is true for all elements in a domain of discourse. In other words, when we say that a property $P(n)$ is true for all natural numbers n, we can write that as the logical formula

$$\forall_{(n \in \mathbb{N})} P(n).$$

Using universal quantification, we can state the strong induction principle as follows. Suppose that the following two conditions hold:

1. The property $P(0)$ is true.

2. If the property $\forall_{k \leq n} P(k)$ is true, then the property $P(n+1)$ is true.

Then $P(n)$ is true for all n. The universal quantifier is helpful in the proof of the strong induction principle.

Proof. Let $P(n)$ be a property of an arbitrary natural number n, and define

$$Q(n) := \forall_{(m \leq n)} P(m)$$

Note that in order to prove that $P(n)$ is true for all n, it suffices to prove that $Q(n)$ is true for all n. Indeed, if $Q(n)$ is true for all n, that just means that $P(m)$ is true for all $m \leq n$, for all n. In particular $P(n)$ is true for all n.

With that out of the way, assume that $P(0)$ is true, and that $Q(n)$ implies $P(n+1)$ as in the induction step of the strong induction principle. Our goal is to show that $P(n)$ is true for all n, and we have already shown that it suffices to show that $Q(n)$ is true for all n. We proceed by induction on n.

For the base case we need to prove that $Q(0)$ is true, meaning that $P(m)$ is true for all $m \leq 0$. Note that $m \leq 0$ implies $m = 0$, so the base case follows from the assumption that $P(0)$ is true.

For the inductive step, assume that $Q(n)$ is true. Our goal is to show that $Q(n+1)$ is true. However, we have assumed that $Q(n)$ implies $P(n+1)$. Now notice that if $Q(n)$ is true, meaning that if $P(m)$ is true for all $m \leq n$, and if $P(n+1)$ is true, then $P(m)$ is true for all $m \leq n+1$. In other words, $Q(n+1)$ is true. This completes the inductive step and therefore the proof. □

Our earlier idea to prove the representability theorem can finally be made into a proof.

Proof of Theorem 3.3.1. We prove the theorem by strong induction. For the base case, note that the empty list satisfies the requirements vacuously since there are no digits in the empty list and empty sums are set to be 0. Thus the empty list is a representation of

0. Now suppose that $\ell = (d_0, \ldots, d_{l-1})$ is a representation of 0 and ℓ is a list of nonzero length l. By the requirement that the leading digit d_{l-1} is nonzero it follows that the sum

$$\sum_{i=0}^{l-1} d_i k^i$$

is nonzero. Such a list is therefore not a representation of 0. The only representation of 0 is therefore the empty list.

Now suppose that every number $m \leq n$ has a unique representation. To show that the number $n + 1$ has a unique representation, first note that there is a unique pair (q, r) of numbers q and $0 \leq r < k$ such that

$$n + 1 = qk + r.$$

There are two cases to consider: either $q = 0$ or $q \neq 0$. In the first case we find that $n + 1 = r < k$. We let ℓ be the list (d_0) of length 1 with $d_0 := r$. Its leading digit is nonzero because $n + 1$ is nonzero, and clearly we have

$$\sum_{i=0}^{0} d_i k^i = d_0 = n + 1.$$

Furthermore, this representation is unique because any representation of $n + 1$ must be a list of digits of length 1, which fixes its digit d_0 to $n + 1$.

If $q \neq 0$, then we have the inequalities

$$q + 1 \leq 2q \leq qk \leq qk + r = n + 1.$$

The second inequality follows from the assumption that $k > 1$. Since $q + 1 \leq n + 1$ it follows that $q \leq n$. This puts us in position to apply the induction hypothesis of the strong induction principle: The number q has a unique representation $(e_0, \ldots, e_{l-1})$. Now we define the representation of $n + 1$ to be $(d_0, \ldots, d_l)$, where $d_0 := r$ and $d_{i+1} := e_i$ for $0 \leq i < l$.

This representation of $n + 1$ is unique, because for any representation of $n + 1$, its d_0 must be equal to r, and the list $(d_1, \ldots, d_{l-1})$ must be a representation of q, which is unique, thus completing the proof. □

3.4 Some Combinatorial Applications of Euclid's Division Theorem

Theorem 3.4.1. *Consider two positive integers $m < n$. In any set of n integers, there is a pair of distinct integers such that their difference is divisible by m.*

Proof. Consider a set A of size n of integers, and list, for every integer in A its remainder after division by m. Since the remainder is a number between 0 and $m - 1$, it follows from the pigeonhole principle that there must be at least two distinct integers x and y with the same remainder. It follows that $x - y$ is divisible by m. □

Theorem 3.4.2. *Consider two positive integers $m \leq n$. In any set of n integers, there is a nonempty subset such that the sum of its elements is divisible by m.*

Proof. Consider the n integers $a_1, \ldots, a_n$, and consider the sums

$$0,\ a_1,\ a_1 + a_2,\ \ldots,\ a_1 + \cdots + a_n.$$

By Theorem 3.4.1 it follows that at there are at least two of these sums that differ by a multiple of m. Notice that the difference of two such sums is of the form $a_k + \cdots + a_l$, so the theorem is proven. □

Theorem 3.4.3. *Consider positive integers m and n such that $m < 2^n - 1$, and consider a set of integers S of size n. Then there are two distinct nonempty subsets of S such that the difference of the sums of their elements is divisible by m.*

Proof. There are 2^k subsets of S, of which $2^k - 1$ are nonempty. By Theorem 3.4.1 it follows that there are two distinct nonempty subsets of S such that the difference of their sums is divisible by m. □

Exercises

3.1 Show that a number written as $d_{l-1} \cdots d_0$ in its decimal representation is divisible by 11 if and only if the alternating sum

$$\sum_{i=0}^{l-1} (-1)^i d_i$$

is divisible by 11. For example, the number 2541 is divisible by 11 because $-2 + 5 - 4 + 1 = 0$ is divisible by 11.

3.2 *Zeckendorf's representation theorem.* Show that for any natural number n there is a unique list $\ell = (d_0, \ldots, d_{l-1})$ of digits 0 and 1, such that its leading digit d_{l-1} is nonzero, no two consecutive digits are both 1, and satisfying the identity

$$n = \sum_{i=0}^{l-1} d_i F_{i+2},$$

where F_j is the jth Fibonacci number. Zeckendorf representations have some fun applications:

1. The Zeckendorf representation theorem can be used in the design of any document to get meaningful relative font sizes: If $d_{l-1} \cdots d_0$ is the Zeckendorf representation of a font size n, then adding an extra zero at the end gives the Zeckendorf representation $d_{l-1} \cdots d_0 0$ of a good next font size. If you are working with font size 12, what is the next sensible font size?
2. Zeckendorf representations can also be used to convert miles to kilometers with a fair degree of accuracy. Again, given a number n of miles, with Zeckendorf representation $d_{l-1} \cdots d_0$, add a 0 at the end to get the Zeckendorf representation $d_{l-1} \cdots d_0 0$ of the number of kilometers. How fast is 65 mph in kph according to this conversion?

Chapter 4

Linear Diophantine Equations

In this section we study *linear Diophantine equations*. These are equations of the form

$$a_0x_0 + \cdots + a_kx_k = b,$$

with integer coefficients $a_0, \ldots, a_k$. The objective of a linear Diophantine equation such as the one above is to find integers $x_0, \ldots, x_k$ for which the equation holds, to study conditions under which the equation has integer solutions, and ultimately to characterize all solutions.

4.1 Divisibility

The simplest linear Diophantine equation is the equation $ax = b$ in one variable x. The condition of its solvability leads to the concept of *divisibility*.

Definition 4.1.1. We say that an integer a *divides* an integer b if there exists an integer x such that the equation

$$ax = b$$

holds. When a divides b, we write $a \mid b$ and we say that a is a *divisor* of b. Sometimes we also refer to b as the *dividend*.

The integers -1, 0, and 1 have special properties with respect to divisibility:

1. The integers 1 and -1 divide any integer b. Indeed, the integer b itself is a solution to the equation $1x = b$, and the integer $-b$ is a solution to the equation $-1x = b$.

2. Every integer a divides 0. Indeed, $x = 0$ is a solution to the equation $ax = 0$.

3. If $0 \mid b$ for some integer b, then it follows that $b = 0$. Indeed, if $0x = b$ has a solution, then $b = 0$ because $0x = 0$. Furthermore, we observe that *every* integer is a solution to the equation $0x = 0$.

   On the other hand, it follows that $0 \nmid 1$, and indeed that $0 \nmid b$ for any nonzero integer b.

4. Similarly, but slightly more complicated, if $a \mid 1$ for some integer a, then it follows that $a = \pm 1$. Indeed, if $ax = 1$ has a solution, then the absolute value $|a|$ of a cannot be greater than 1, because its nonzero multiples would be of absolute value greater than 1. Therefore it follows that $|a| = 1$, which proves the claim. In the situation where $a \mid 1$, we say that a is *invertible*, or that it is a *multiplicative unit*, because the solution to the equation $ax = 1$ is a multiplicative inverse of a.

If a is a nonzero integer, then there is always at most one solution to the equation $ax = b$. In other words, if a nonzero integer a divides an integer b, then its quotient is uniquely defined. On the other hand, we just saw that every integer is a solution to the equation $0x = 0$. The quotient is therefore not uniquely defined in this case. In the following lemma we remove this ambiguity by imposing an extra bound on the solution, that $|x| \leq |b|$. In the case where a is nonzero, then this condition is automatically satisfied, while in the case where $a = 0$ it imposes the zero solution.

Lemma 4.1.2. *Suppose that $a \mid b$. Then there is a unique integer x such that $|x| \leq |b|$ satisfying the equation $ax = b$. We will write a/b for this unique integer, which is called the* quotient *of b divided by a.*

Proof. There are two cases to consider: either $a = 0$ or $a \neq 0$. In the case where $a = 0$, having a solution $ax = b$ implies that $b = 0$. In this case, any x is a solution to the equation $ax = b$. However, only one solution x satisfies $|x| \leq b$, namely $x = 0$. Thus, the quotient $0/0$ is set to be 0.

If $a \neq 0$, then the function $x \mapsto ax$ is *injective*. This means that $ax = ay$ implies $x = y$. Consequently, there can be at most one integer x such that $ax = b$. Since we assumed that there is at least one such a solution, it follows that it is unique. Moreover, if $ax = b$, then $|x| \leq |a||x| = |ax| = |b|$, so x indeed satisfies the required bound. □

Remark 4.1.3. While we have given an informal description, the definition of divisibility features the *existential quantifier* $\exists$. When we write the logical formula $\exists_{(x \in X)} P(x)$, this means that there exists an element $x \in X$ for which $P(x)$ holds. The definition of divisibility translates to the following logical formula:

$$a \mid b := \exists_{(x \in \mathbb{Z})} ax = b.$$

In order to use existential quantification in proofs, it is useful to know how to break them down systematically. There are two cases to consider: (1) How to *prove* an existence claim, and (2) how to *use* an existence claim as an assumption to prove something else.

1. The principal way to prove that there exists an element $x \in X$ such that $P(x)$ holds, is to construct an element x of X and prove that the property $P(x)$ is holds.

2. To use that there exists an $x \in X$ such that $P(x)$ holds in the proof of another property Q, you may assume to have an element x and you may assume that the property $P(x)$ holds. However, in this scenario we may not assume anything else about x unless that is otherwise warranted.

In the case of divisibility, in order to prove that $a \mid b$ holds we must find a solution to the equation $ax = b$, and in order to use that $a \mid b$ holds as an assumption, we may assume x is an integer for which the equation $ax = b$ holds.

Proposition 4.1.4. *If d divides a and b, then d divides $ax + by$ for any two integers x and y.*

Proof. Suppose that d divides both a and b. Let u and v be integers such that $du = a$ and $dv = b$. Then we have

$$ax + by = dux + dvy = d(ux + vy),$$

which shows that $ax + by$ is a multiple of d. □

Sometimes, the hardest thing is to figure out how to rigorously prove a property that looks completely self-evident. The following proposition, for instance, asserts that in any set of k consecutive integers, exactly one of them is divisible by k.

There are, as is commonly the case in mathematics, several ways to approach this problem, but the question is how to generate ideas towards a full proof. One good way of generating ideas is by looking at special cases. In our problem at hand, a special case of interest is the set $\{0, 1, \ldots, k-1\}$. There should be exactly one element in the set $\{0, 1, \ldots, k-1\}$ divisible by k.

Indeed, this is a direct consequence of the Euclidean division theorem: the only element in this set that has remainder 0 after division by k is the integer 0. This suggests using the Euclidean division theorem for the general claim as well. Furthermore, we will also use the common proof technique of assuming something *without loss of generality*.

Proposition 4.1.5. *In any set of $k > 0$ consecutive integers*

$$\{a, a+1, \ldots, a+k-1\}$$

there is exactly one element divisible by k.

Proof. First, consider the function

$$r : \{a, a+1, \ldots, a+k-1\} \to \{0, 1, \ldots, k-1\},$$

where $r(x)$ is defined to be the remainder of x after division by k. We claim that this map is injective, meaning that if $r(x) = r(y)$ for two elements x and y, then $x = y$. To see this, consider two elements x and y in the set $\{a, \ldots, a+k-1\}$. Without loss of generality we may assume that $x \leq y$. Indeed, either $x \leq y$ or $y \leq x$ holds, and these situations are completely similar, so nothing is lost if we just assume that $x \leq y$.

We have the equalities $x = q(x)k + r(x)$ and $y = q(y)k + r(y)$, where $q(x)$ and $q(y)$ are the quotients of x and y after division by k, respectively. Since we assumed that $r(x) = r(y)$, it follows that $y - x = (q(y) - q(x))k$. In other words, $y - x$ is divisible by k. Furthermore, we have $0 \leq y - x < k$ because both x and y are in the set $\{a, \ldots, a+k-1\}$. It follows from Euclidean division theorem that the only integer $0 \leq z < k$ divisible by k is 0, so we find that $y - x = 0$. We conclude that $x = y$.

It now follows that every element $y \in \{0, 1, \ldots, k-1\}$ is the value of r of exactly one element $x \in \{a, a+1, \ldots, a+k-1\}$, because both sets have k elements and the map r doesn't take any value more than once. In particular, there is exactly one element $x \in \{a, a+1, \ldots, a+k-1\}$ whose remainder after division by k is 0. In other words, exactly one element of this set is divisible by k. □

Another approach that often works well, which was suggested in class by Lucy, is a proof by contradiction. In proofs by contradiction, you assume the contrary and derive a contradiction. If the contrary is impossible, then the original claim must be true.

Proof by contradiction. Let's assume that there isn't a unique integer in the set

$$\{a, \ldots, a + k - 1\}$$

divisible by k. There are two cases to consider: Either there are no elements divisible by k, or there are at least two distinct elements divisible by k.

In the first case, by Euclid's division theorem there is a unique pair (q, r) such that $a = qk + r$, and since a is not divisible by k it must be the case that $r \neq 0$. Now it follows that $qk + k$ is an element in $\{a, \ldots, a + k - 1\}$, which is divisible by k, contrary to our assumption. Thus we conclude that it is not true that none of the elements of $\{a, \ldots, a + k - 1\}$ is divisible by k.

In the second case, let x and y be distinct elements of $\{a, \ldots, a + k - 1\}$ that are both divisible by k. Assume without loss of generality, that $x \leq y$. Then we have $0 \leq y - x < k$, and $y - x$ is divisible by k. By the Euclidean division theorem, we must have $y - x = 0$, which contradicts our assumption that x and y are distinct. □

4.2 Ideals of Integers

Ideals are a concept from *ring theory*, which studies *rings*, which are sets equipped with addition, subtraction, and multiplication, satisfying the most familiar laws of arithmetic:

$$\begin{aligned} (x + y) + z &= x + (y + z) & (xy)z &= x(yz) \\ 0 + x &= x & 1x &= x \\ x + 0 &= x & x1 &= x \\ x - x &= 0 & x(y + z) &= xy + xz \\ -x + x &= 0 & (x + y)z &= xz + yz \\ x + y &= y + x. \end{aligned}$$

Rings such as the integers satisfy additionally the law $xy = yx$ of commutativity; such rings are called *commutative rings*. Notice that the set $\mathbb{N}$ of natural numbers isn't a ring, because it lacks subtraction. Sets with the structure of a ring appeared in David Hilbert's *Zahlbericht*, but their formal definition was first given by Emmy Noether in *Idealtheorie in Ringbereichen*, who was also the first to study them systematically.

An *ideal of integers* is a subset $I \subseteq \mathbb{Z}$ that contains 0, and contains the linear combination $kx + ly$ for any $x, y \in I$, and any $k, l \in \mathbb{Z}$. The subset $\{0\} \subseteq Z$ satisfies the conditions of being an ideal in a trivial way; this ideal is called the *zero ideal*. The subset $(a) := \{ka \mid k \in \mathbb{Z}\}$ also satisfies the conditions of being an ideal; ideals of this form are called *principal ideals*, since they are generated by a single integer. In the following theorem we show that every ideal of integers is of this form.

Theorem 4.2.1. *For every ideal $I \subseteq \mathbb{Z}$ there is a unique natural number n such that $I = (n)$.*

Proof. There are two cases to consider: Either I is the zero ideal, or I contains a nonzero integer. In the first case it is clear that $I = (0)$, and that 0 is the unique natural number n such that $I = (n)$. Thus we focus on the nonzero case.

Suppose that I contains a nonzero integer x. Then I contains the integer $|x|$, because $|x| = \pm x$. In particular, I contains a positive integer. Therefore it follows that I contains a least positive integer n. To see that $I = (n)$, consider an element $x \in I$. By the Euclidean division theorem, it follows that $x = qn + r$, where $0 \leq r < n$. Notice that $r \in I$, since it is the difference $x - qn$ of two elements in I. Since r is nonnegative and n is the least positive integer in I, it follows that $r = 0$, which gives us that $x = qn$. Thus, every element in I is a multiple of n.

Now suppose that $I = (n)$ and $I = (m)$ for two nonzero natural numbers n and m. Then it follows that $m = qn$ and $n = pm$. Thus, we see that $m = (pq)m$, which implies that $pq = 1$. Furthermore, both p and q are positive integers, so we have $p = 1$ and $q = 1$. This shows that $m = 1n = n$, establishing uniqueness. □

The set of ideals of $\mathbb{Z}$ is ordered by inclusion: We write $I \subseteq J$ for two ideals I and J, if every element of I is an element of J. Note that $(b) \subseteq (a)$ holds if and only if $a \mid b$. The theory of ideals of integers is therefore closely related to the theory of divisibility of integers.

4.3 The Ordering by Divisibility

The divisibility relation equips the sets of integers and natural numbers with a useful extra structure, a *partial ordering* of their elements. The set of natural numbers equipped with the divisibility relation is a poset, while the set of integers equipped with the divisibility relation is a preorder, a slightly weaker structure.

Definition 4.3.1. A *preorder* consists of a set X and an ordering $\leq$ of the elements of X, such that

1. The ordering is *reflexive*. This means that $x \leq x$ holds for every x.
2. The ordering is *transitive*. This means that if $x \leq y$ and $y \leq z$ both hold, then $x \leq z$ also holds.

A *poset*, or *partially ordered set* in full, is a preorder $(X, \leq)$ satisfying additionally the condition:

3. The ordering is *antisymmetric*. This means that if both $x \leq y$ and $y \leq x$ hold, then $x = y$.

Theorem 4.3.2. *The set of integers equipped with the divisibility relation $\mid$ is a preorder, and the set of natural numbers equipped with the divisibility relation $\mid$ is a poset.*

Proof. We first show that the set of integers with divisibility is a preorder. To see that the divisibility relation is reflexive, note that $x = 1$ is a solution of the equation

$$ax = a.$$

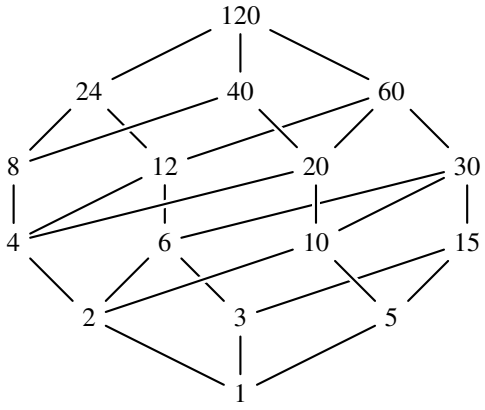
Therefore, it follows that $a \mid a$ for every integer a. To see that the divisibility relation is transitive, assume that $a \mid b$ and $b \mid c$. That is, assume that we have integers x and y such that $ax = b$ and $by = c$. Then it follows that

$$a(xy) = (ax)y = by = c.$$

In other words, c is divisible by a. This completes the proof that $\mathbb{Z}$ equipped with the divisibility relation is a preorder.

To see that the natural numbers equipped with divisibility forms a poset, it remains to show that the divisibility relation is antisymmetric on the natural numbers. Suppose that $m \mid n$ and $n \mid m$; that is, that there are natural numbers x and y so that $mx = n$ and $ny = m$. By transitivity of divisibility it follows that $m(xy) = m$. □

The diagram below is the *Hasse diagram* of the positive divisors of the number 120, a fragment of the poset of the natural numbers ordered by divisibility.



Despite lacking antisymmetry, the preorder of the integers with divisibility is essentially equivalent to the poset of natural numbers with divisibility. For every integer a there is a unique natural number n satisfying $n \mid a$ and $a \mid n$. This unique number is of course the absolute value of a. The absolute value function $x \mapsto |x|$ preserves divisibility, which means that $a \mid b$ implies $|a| \mid |b|$, and it satisfies the logical equivalence

$$(a \mid b) \wedge (b \mid a) \Leftrightarrow |a| = |b|$$

for any integer a and any natural number n. Order preserving maps $f : (P, \leq_P) \to (Q, \leq_Q)$ from a preorder $(P, \leq_P)$ to a poset $(Q, \leq_Q)$ that satisfy the logical equivalence $(x \leq_P y) \wedge (y \leq_P x) \Leftrightarrow f(x) = f(y)$ for every $x, y \in P$ are also called *poset reflections*. Every preorder has a poset refection, but we will not need to go into that aspect of order theory in this course.

4.4 Greatest Common Divisors

Some preorders $(P, \leq)$ satisfy the property that for any two elements $a, b \in P$ there is a maximal element below both of them. This property can be formulated succinctly as

follows: An element d is a maximal element below a and b if it satisfies the logical equivalence

$$(x \leq a) \wedge (x \leq b) \Leftrightarrow (x \leq d)$$

for every element $x \in P$. If d satisfies this property, we also say that d is the *meet* of a and b. It is common to write $a \wedge b$ for the meet of a and b, if P is a poset in which any two elements have a meet.

The definition of meets needs some explanation. It uses the *conjunction* ($\wedge$) and it uses *logical equivalence* ($\Leftrightarrow$). The conjunction $p \wedge q$ of two propositions p and q is the proposition that both p and q are true. The logical equivalence $p \Leftrightarrow q$ is the proposition that p is true if and only if q is true. In other words, it is the conjunction of the proposition that p implies q and the proposition that q implies p.

If d is the meet of a and b, then the logical equivalence gives two implications back and forth:

1. Any element x that is below both a and b, must be below d.

2. Any element x that is below d, must be below both a and b. In particular, d itself satisfies $d \leq d$, so it follows that $d \leq a$ and $d \leq b$. That is, d itself is below both a and b, and by the first property it is the maximal such element.

We wouldn't have this discussion if it didn't apply for us. Indeed, we will show that the preorder of integers with divisibility has meets. In the case of the integers, we will call such meets *greatest common divisors*.

Definition 4.4.1. Consider two integers a and b. An integer d is said to be a *greatest common divisor* of a and b if it satisfies the following logical equivalence

$$(x \mid a) \wedge (x \mid b) \Leftrightarrow (x \mid d).$$

for every integer x.

Thus, a greatest common divisor of a and b is an integer d that divides both a and b, and any integer that divides both a and b is a divisor of d. In other words, the signifier "greatest" refers to the divisibility ordering of the integers.

Example 4.4.2. The greatest common divisor of 0 and 0 is 0. To see this, simply note that the proposition

$$(x \mid 0) \wedge (x \mid 0) \Leftrightarrow (x \mid 0)$$

is true because any proposition is logically equivalent to the conjunction with itself. Here we see why it was relevant to note that the word "greatest" refers to the divisibility ordering: Any integer is a divisor of 0, but none of them are the greatest in the standard ordering $\leq$.

Theorem 4.4.3. *Any two integers a and b have a greatest common divisor, for which we write* $\gcd(a, b)$. *Furthermore, there are integers k and l satisfying* Bézout's identity

$$ax + by = \gcd(a, b).$$

Proof. Consider two integers a and b, and let $I := (a)$ and $J := (b)$ be the principal ideals generated by a and b. Now consider the set

$$I + J := \{x + y \mid x \in I,\ y \in J\}$$

of all sums of elements of I and elements of J. We claim that the set $I + J$ is again an ideal. It contains 0, because $0 = 0 + 0$ is a sum of elements in I and J. Furthermore, to see that $I + J$ is closed under linear combinations of its elements, consider two elements $u, v \in I + J$. If $u = x_u + y_u$ and $v = x_v + y_v$, then we can write any linear combination $ku + lv$ of u and v in the form

$$ku + lv = (kx_u + lx_v) + (ky_u + ly_v),$$

which is the sum of an element in I and an element in J. Thus, it follows that $I + J$ is an ideal.

Now let d be the unique natural number such that $I + J = (d)$, which exists by Theorem 4.2.1. We claim that d is the greatest common divisor of a and b. Notice that $a = a + 0 \in I + J$, so a must be a multiple of d. Similarly, we have $b = 0 + b \in I + J$, so b must be a multiple of d. In other words, d is a common divisor of a and b. Furthermore, if e is any common divisor of a and b, then e is a divisor of any element in $I + J$. In particular, we must have $e \mid d$, allowing us to conclude that d is the greatest common divisor.

The last claim follows, because d is an element of $I + J$, which consists elements of the form $ax + by$. □

A very important condition on pairs of integers a and b, which occurs as an assumption in many lemmas, propositions, and theorems, is the condition that $\gcd(a, b) = 1$. In other words, that any common divisor of a and b must be 1. Such numbers are called *relatively prime* or *coprime*.

Definition 4.4.4. Two integers a and b are said to be *relatively prime* or *coprime* if

$$\gcd(a, b) = 1.$$

The following proposition is a must-know fact about the integers.

Proposition 4.4.5. *Consider two relatively prime integers a and b, and an arbitrary integer c. Then we have*

$$a \mid bc \Leftrightarrow a \mid c.$$

Proof. The fact that $a \mid c$ implies $a \mid bc$ is obvious, since if c is a multiple of a, then so is bc. This establishes the converse direction of the logical equivalence.

For the forward direction, assume that $a \mid bc$. Since we have assumed that a and b are relatively prime, there is a solution to the equation

$$ax + by = 1.$$

Consequently, there is a solution to the equation $cax + cby = c$. However, both cax and cby are divisible by a, so c must be divisible by a. □

4.5 Euclid's Algorithm

The previous theorem shows that any two integers a and b have a greatest common divisor $\gcd(a, b)$, and that there are integers k and l such that

$$ka + lb = \gcd(a, b),$$

but it doesn't reveal much about how the integers k and l can be found. Euclid's algorithm is an efficient way to compute the greatest common divisor and to find integers k and l establishing the greatest common divisor as a linear combination.

Proposition 4.5.1. *Suppose a and b are two integers and $b = qa + r$. Then we have*

$$\gcd(b, a) = \gcd(a, r).$$

Proof. Suppose that d is a divisor of a. Then we claim that $d \mid b$ if and only if $d \mid r$. Indeed, since $b = qa + r$ it follows that b is a linear combination of a and r, and also that r is a linear combination of a and b. Thus, any divisor of a and b is also a divisor of r, and any divisor of a and r is also a divisor of b. In other words, we have

$$d \mid b \Leftrightarrow d \mid r.$$

It follows that d is a common divisor of a and b if and only if d is a common divisor of a and r. This implies that their greatest common divisors coincide. □

As an immediate corollary, we obtain:

Corollary 4.5.2. *If b is of the form $ka + 1$, then a and b are always relatively prime.*

Proposition 4.5.1 implies a useful way of calculating the greatest common divisor of two numbers. Suppose we want to calculate the greatest common divisor of 578 and 732. Then we write 732 in the form $q \cdot 578 + r$; that is, we write

$$732 = 1 \cdot 578 + 154.$$

By the proposition, it follows that $\gcd(732, 578) = \gcd(578, 154)$. We now proceed in the same manner, by writing

$$578 = 3 \cdot 154 + 116,$$

and we observe that $\gcd(732, 578) = \gcd(154, 116)$. We proceed with this process until we find that $\gcd(732, 578) = \gcd(d, 0)$ for some number d. Once we reach this stage, we conclude that $\gcd(732, 578) = d$. Putting words to action:

$$\begin{aligned} 154 &= 1 \cdot 116 + 38 & \gcd(732, 578) &= \gcd(116, 38) \\ 116 &= 3 \cdot 38 + 2 & \gcd(732, 578) &= \gcd(38, 2) \\ 12 &= 6 \cdot 2 + 0 & \gcd(732, 578) &= \gcd(2, 0). \end{aligned}$$

Thus, we conclude that $\gcd(732, 578) = 2$. The process outlined above is called *Euclid's Algorithm*.

By closer inspection of Euclid's algorithm, we can even find an integer solution to the equation

$$\gcd(a, b) = ax + by.$$

The best way to see how this works is by example, so let us use the previous example. Euclid's algorithm has given us that $\gcd(732, 578) = 2$. In the second to last equation, we find that 2 is a linear combination of 38 and 116; that is,

$$2 = 116 - 3 \cdot 38$$

However, by the equation before that we see that 38 itself is a linear combination of 154 and 116. This can be used to express 2 as a linear combination of 154 and 116; that is,

$$2 = 116 - 3 \cdot (154 - 116) \qquad \Rightarrow \qquad 2 = 4 \cdot 116 - 3 \cdot 154.$$

Furthermore, 116 is a linear combination of 154 and 578, giving us that

$$2 = 4 \cdot (578 - 3 \cdot 154) - 3 \cdot 154 \qquad \Rightarrow \qquad 2 = 4 \cdot 578 - 15 \cdot 154.$$

In the final step, we use that 154 is a linear combination of 732 and 578 to find

$$2 = 4 \cdot 578 - 15 \cdot (732 - 578) \qquad \Rightarrow \qquad 2 = -15 \cdot 732 + 19 \cdot 578.$$

Example 4.5.3. In class, we computed the greatest common divisor of 721 and 450. These numbers were suggested by Marwa and Dylan.

$$\begin{aligned}
721 &= 1 \cdot 450 + 271 & 1 &= 181 \cdot (721 - 450) - 109 \cdot 450 \\
&& &= 181 \cdot 721 - 290 \cdot 450 \\
450 &= 1 \cdot 271 + 179 & 1 &= 72 \cdot 271 - 109 \cdot (450 - 271) \\
&& &= 181 \cdot 271 - 109 \cdot 450 \\
271 &= 1 \cdot 179 + 92 & 1 &= 72 \cdot (271 - 179) - 37 \cdot 179 \\
&& &= 72 \cdot 271 - 109 \cdot 179 \\
179 &= 1 \cdot 92 + 87 & 1 &= 35 \cdot 92 - 37 \cdot (179 - 92) \\
&& &= 72 \cdot 92 - 37 \cdot 179 \\
92 &= 1 \cdot 87 + 5 & 1 &= 35 \cdot (92 - 87) - 2 \cdot 87 \\
&& &= 35 \cdot 92 - 37 \cdot 87 \\
87 &= 17 \cdot 5 + 2 & 1 &= 5 - 2 \cdot (87 - 17 \cdot 5) \\
&& &= 35 \cdot 5 - 2 \cdot 87 \\
5 &= 2 \cdot 2 + 1 & 1 &= 5 - 2 \cdot 2 \\
2 &= 2 \cdot 1 + 0
\end{aligned}$$

Thus we find that $\gcd(721, 450) = 1$, in other words, that 721 and 450 are relatively prime, and that

$$181 \cdot 721 - 290 \cdot 450 = 1.$$

4.6 Linear Diophantine Equations in Multiple Variables

In this section we turn our attention to linear Diophantine equations in two variables, that is, integer equations of the form

$$ax + by = c.$$

We begin by applying the theory of ideals to establish that this equation has a solution if and only if $\gcd(a, b) \mid c$.

Theorem 4.6.1. *Consider integers a, b, and c, and let $d = \gcd(a, b)$. The linear Diophantine equation*

$$ax + by = c$$

is solvable with integers x and y if and only if $d \mid c$.

Proof. The first part of the claim is equivalent to the claim that we have an equality of ideals

$$(a) + (b) = (\gcd(a, b)).$$

We recall that the ideal $I + J$ consists of sums $x + y$ of $x \in I$ and $y \in J$. Furthermore, the ideals (a) and (b) consist of multiples of a and b, respectively. Thus the ideal $(a) + (b)$ consists of all the linear combinations of a and b; in other words, all the possible integers c for which there is a solution to the equation $ax + by = c$.

In Theorem 4.4.3 we defined $\gcd(a, b)$ to be the unique natural number d such that $(a) + (b) = (d)$. Such a natural number exists by Theorem 4.2.1. Now we observe that the equality

$$(a) + (b) = (\gcd(a, b))$$

tells us exactly that the set of integers c for which there exists a solution of the equation $ax + by = c$ has the same elements as the set of multiples of $\gcd(a, b)$. In other words, a solution to $ax + by = c$ exists if and only if c is a multiple of the greatest common divisor of a and b. □

As an immediate corollary of the previous theorem, we note that:

Corollary 4.6.2. *If a and b are relatively prime integers, then the linear Diophantine equation*

$$ax + by = c$$

is solvable for any c.

Now that we have established a necessary and sufficient condition for its solvability, it remains to describe a way of finding all solutions, if there are any. Notice that we have already collected quite a bit of useful information. Euclid's algorithm allows find a solution to the equation

$$ax + by = d,$$

where $d = \gcd(a, b)$. Thus, if $c = kd$, then we can solve

$$ax + by = c$$

by first finding a solution $ax_0 + by_0 = d$ by Euclid's algorithm. Then we find that

$$a(kx_0) + b(ky_0) = kd = c.$$

In other words, we can use Euclid's algorithm to find a solution to the equation $ax+by = c$ if it has any.

Proposition 4.6.3. *Consider integers a, b, and c, and assume that a and b are relatively prime. If $ax_0 + by_0 = c$ is a solution to the Diophantine equation*

$$ax + by = c,$$

then every solution is of the form

$$x = x_0 + bk, \qquad \textit{and} \qquad y = y_0 - ak.$$

Proof. Suppose that $ax + by = c$ is another solution. Then it follows that

$$a(x - x_0) + b(y - y_0) = 0$$

In other words, we have $a(x - x_0) = -b(y - y_0)$. Thus we see that $b(y - y_0)$ is divisible by a. Since a and b are relatively prime, this implies that $a \mid y - y_0$. Similarly, $a(x - x_0)$ is divisible by b, and therefore it follows that $b \mid x - x_0$. Furthermore, if $ka = y - y_0$ and $lb = x - x_0$. Then the equation

$$lab - kab = 0$$

implies that $k = l$. In other words, $x = x_0 + bk$ and $y = y_0 - ak$. □

Theorem 4.6.4. *Suppose that $ax_0, by_0 = c$ is a solution to the Diophantine equation*

$$ax + by = c.$$

Then every solution is of the form

$$x = x_0 + \frac{b}{d}k, \qquad \textit{and} \qquad y = y_0 - \frac{a}{d}k.$$

Proof. This theorem follows from the previous proposition. If $d = \gcd(a, b)$, then we find that

$$\frac{a}{d}x + \frac{b}{d}y = \frac{c}{d}.$$

The integers $\frac{a}{d}$ and $\frac{b}{d}$ are relatively prime, so we find that $x = x_0 + \frac{b}{d}k$ and $y = y_0 - \frac{a}{d}k$ for some integer k. □

Exercises

4.1 Show that if $a \mid b$, then $d \mid b/a$ if and only if $ad \mid b$.

4.2 Suppose that $a \mid b$ and $b \mid c$. Show that if $d \mid b/a$ then $d \mid c/a$.

4.3 Show that $a \mid b$ implies $ac \mid bc$. Furthermore, show that if c is nonzero, then $ac \mid bc$ implies $a \mid b$. Conclude that $a^n \mid b^n$ if and only if $a \mid b$ for any natural number n.

4.4 Suppose that d is a common divisor of two nonzero integers a and b.

(a) Show that the greatest common divisor of a/d and b/d is $\gcd(a, b)/d$.

(b) Show that d is a greatest common divisor of a and b if and only if a/d and b/d are relatively prime.

4.5 Show that $a \mid bc$ if and only if $a/\gcd(a, b) \mid c$.

4.6 Show that if $d \mid a + b$, then $\gcd(d, a) \mid \gcd(a, b)$ and $\gcd(d, b) \mid \gcd(a, b)$. Conclude that if a and b are relatively prime, then d is relatively prime to both a and b.

4.7 Show that $a^n - 1$ is divisible by $a - 1$ for every integer a.

4.8 Show that $a^n + b^n$ is divisible by $a + b$ for any odd natural number n.

4.9 Show that $a^2 - b^2$ is divisible by 8 for any two odd integers a and b.

4.10 Show that $a^3 - a$ is divisible by 6 for every integer a.

4.11 Show that $a^5 - a$ is divisible by 30 for every integer a.

4.12 Show that any two consecutive Fibonacci numbers are relatively prime.

4.13 *Goldbach's theorem.* Show that any two distinct Fermat numbers are relatively prime.

4.14 Show that $2^m - 1$ and $2^n + 1$ are relatively prime, provided that m is odd.

4.15 Show that $n! + 1$ and $(n + 1)! + 1$ are relatively prime for all nonzero n.

4.16 (a) Prove the *Fibonacci addition formula* $F_{m+n+1} = F_{m+1}F_{n+1} + F_mF_n$.

(b) For any integer d and any two natural numbers m and n, show that if two of the three following conditions hold, then so does the third:

1. $d \mid F_m$.
2. $d \mid F_n$.
3. $d \mid F_{n+m}$.

(c) Show that the Fibonacci sequence preserves and reflects divisibility: we have $m \mid n$ if and only if $F_m \mid F_n$.

4.17 Consider three integers a, b, and c. Show that the following are equivalent:

1. The integers a and b are both relatively prime to c.
2. The integer ab is relatively prime to c.

4.18 Prove that if a and b are relatively prime integers, and n is an arbitrary natural number, then in the arithmetical progression

$$a + kb, \qquad k = 0, 1, 2, \ldots$$

there are infinitely many numbers relatively prime to n.

Chapter 5

The Fundamental Theorem of Arithmetic

5.1 Prime Numbers

Definition 5.1.1. An integer a is said to be *prime* if it has exactly one positive proper divisor. Any integer $a \neq \pm 1$ that is not prime is said to be a *composite number*.

The numbers 0 and 1 are not prime. To see that 0 isn't prime, simply note that any positive integer is a proper divisor of 0. To see that 1 isn't prime, note that 1 doesn't have any proper divisors. Indeed, it has exactly one positive divisor, namely 1 itself, but this divisor isn't proper. We also note that if $a > 1$ is prime, then the number 1 is always a positive proper divisor. Thus we see that $a > 1$ is prime if and only if the number 1 is its unique positive proper divisor.

To see that the number 2 is prime, note that its positive divisors are a subset of the set $\{1, 2\}$ consisting of all the positive integers below 2. Its proper divisors are therefore a subset of $\{1\}$. The integer 1 is indeed a positive divisor, so we see that 2 has exactly one positive proper divisor.

To see that the number 3 is prime, note that its positive divisors are a subset of the set $\{1, 2, 3\}$ consisting of all positive integers below 3. The number 1 is a proper divisor, the number 2 doesn't divide 3, and 3 itself is a divisor but it isn't proper. Therefore we see that the number 3 has exactly one positive proper divisor.

The number 4 isn't prime, because the numbers 1 and 2 are two distinct proper divisors of 4. This process of finding all primes up to a desired bound is formalized in the *sieve of Eratosthenes*, which we will now describe. Figure 5.1 displays the sieve of Eratosthenes up to 1350.

The sieve of Eratosthenes is an iterative process that generates at stage n a set P_n of numbers known to be prime at stage n, and a set Q_n of prime candidates. In other words, the sets P_n form an increasing sequence

$$P_0 \subseteq P_1 \subseteq P_2 \subseteq \cdots$$

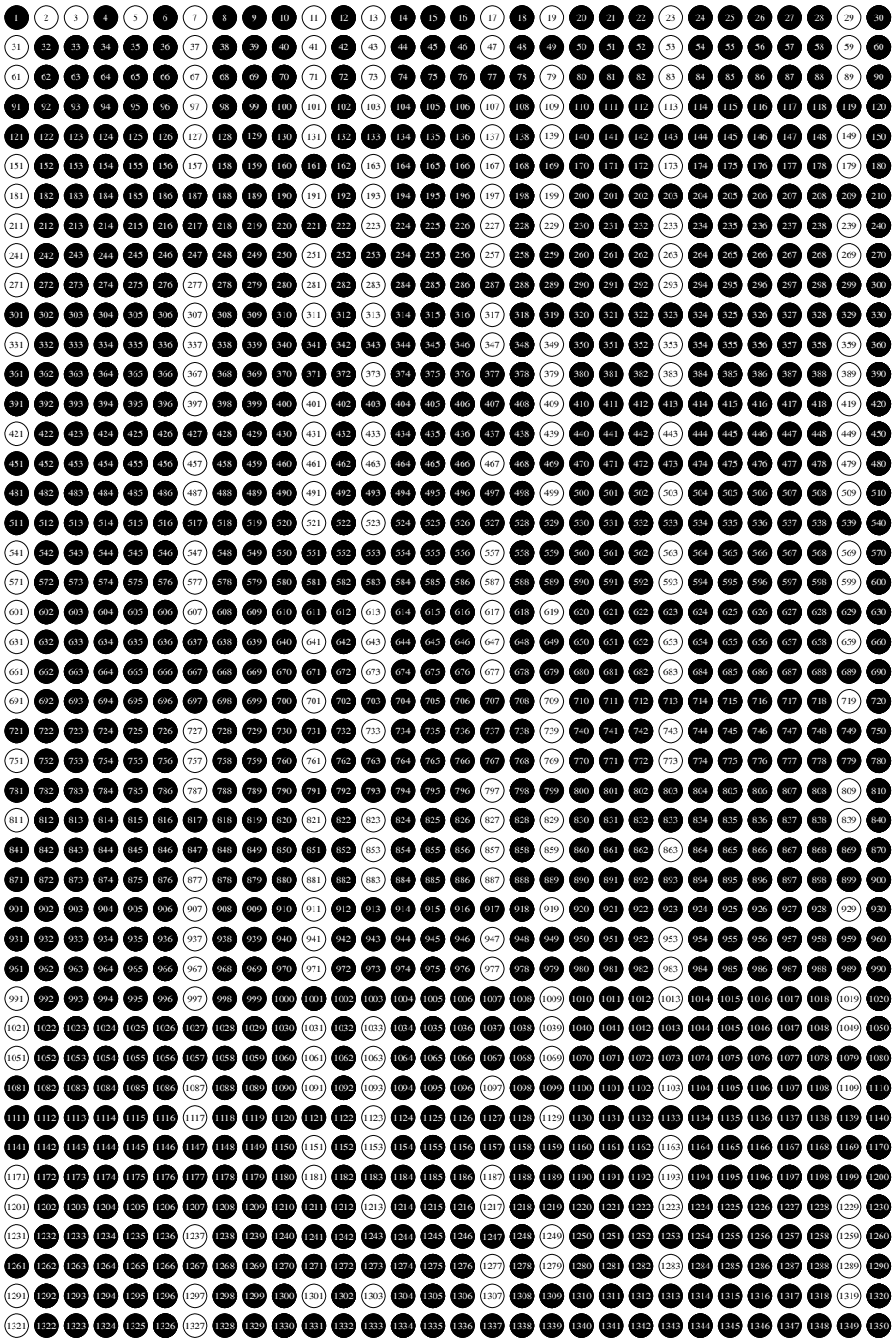


Figure 5.1: The sieve of Eratosthenes up to 1350: The number 1 and every composite number are marked in black, leaving just the primes in white circles.

of sets of numbers known to be prime, where at each higher stage the set of known primes becomes larger, and the sets Q_n form a decreasing sequence

$$Q_0 \supseteq Q_1 \supseteq Q_2 \supseteq \cdots$$

of candidates, from which we pick our next prime. We will define the sets P_n and Q_n by a recursive process. Initially, the set P_0 is empty and $Q_0 = \{n \in \mathbb{N} \mid n \geq 2\}$ is the set of all numbers ≥ 2. Now we define P_{n+1} to be $P_n \cup \{p(n)\}$, where $p(n) := \min(Q_n)$, and we define

$$Q_{n+1} := Q_n \setminus \{k\,p(n) \mid k \geq 1\}.$$

In other words, the nth prime $p(n)$ is the minimal element of $Q(n)$, and the set Q_{n+1} of candidates is the set Q_n minus all the multiples of $p(n)$. In the first five stages, these sets look as follows:

n	P_n	Q_n
0	$\emptyset$	$\{2, 3, 4, 5, 6, 7, 8, \ldots\}$
1	$\{2\}$	$\{3, 5, 7, 9, 11, 13, 15, \ldots\}$
2	$\{2, 3\}$	$\{5, 7, 11, 13, 17, 19, 23, \ldots\}$
3	$\{2, 3, 5\}$	$\{7, 11, 13, 17, 19, 23, 29, \ldots\}$
4	$\{2, 3, 5, 7\}$	$\{11, 13, 17, 19, 23, 29, 31, \ldots\}$
5	$\{2, 3, 5, 7, 11\}$	$\{13, 17, 19, 23, 29, 31, 37, \ldots\}$

Notice that not only the minimal element of Q_n is prime, but every element $q \in Q_n$ such that $q < (\max(P_n))^2$ is prime. This is because if $q < (\max(P_n))^2$ is composite, then its lowest factor must be a prime number $p < \max(P_n)$, i.e., it must be a prime number in P_n. However, all the multiples of the primes in P_n have been removed from Q_n, so there are no such composite numbers.

We can use this observation to obtain the list of all primes below 100, at stage 5 in the sieve of Eratosthenes. The list of primes below 100 is

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.$$

The sieve of Eratosthenes either terminates, i.e., it reaches a stage for which Q_n is empty, or there are infinitely many primes. We will now show that there are infinitely many primes. In the following lemma we use the concept of *nontrivial divisors*, which are divisors d of a natural number n such that $d \neq 1$. The concept of nontrivial divisors should not be confused with the concept of proper divisors, which are divisors d of a natural number n such that $d \neq n$.

Theorem 5.1.2. *Any natural number $n > 1$ has a least nontrivial divisor, and this divisor is prime.*

Proof. Suppose that $n > 1$. Then the set of nontrivial divisors $d \neq 1$ of n contains the number n, so it has a least element p by the well-ordering principle of the natural numbers. Then the set of divisors of p itself is the set $\{1, p\}$, because any divisor of p also divides n, and p is assumed to be the least nontrivial divisor. This shows that p has exactly one proper divisor, and therefore p is prime. □

5.2 The Infinitude of Primes

The theorem asserting the infinitude of primes asserts that no finite set of primes is the set of all primes. Many proofs of the infinitude of primes require the following lemma:

Lemma 5.2.1. *If a and b are relatively prime integers, then any prime divisors of a are distinct from any prime divisors of b.*

Proof. Suppose that a and b are relatively prime, and that p and q are prime divisors of a and b, respectively. By the assumption that a and b are relatively prime, it follows that $p \nmid b$ and that $q \nmid a$. Therefore it follows that p and q cannot be the same. □

Theorem 5.2.2. *There are infinitely many prime numbers.*

Euclid's proof of Theorem 5.2.2. We will prove that there are infinitely many prime numbers by showing that for any finite set of prime numbers, there is a prime number not belonging to the finite set.

Consider a finite set $\{p_0, \ldots, p_k\}$ of primes, and define

$$n := 1 + \prod_{i=0}^{k} p_i$$

Then there exists, for each $0 \leq i \leq k$, a natural number q_i such that $n = q_i p_i + 1$. Indeed, we can simply define

$$q_i := \prod_{j=0, j\neq i}^{k} p_j.$$

Since n can be written in this way, it follows that n is relatively prime to each p_i. Consequently, the least prime divisor of n is relatively prime to each p_i. In other words, the least prime divisor of n is not in the set of primes $\{p_0, \ldots, p_k\}$. □

Saidak's Proof

There are many proofs of the infinitude of primes. The following is due to Filip Saidak [Sai06], who published his proof in 2006, based on the idea that if $n = ab$ is the product of two numbers that are relatively prime and both at least 2, then n must have two prime factors.

Saidak's proof of Theorem 5.2.2. Let $N_0 := 1$, and define

$$N_{n+1} := N_n(N_n + 1).$$

We claim that N_n has at least n prime factors. In the base case, we note that it is indeed true that 1 has at least zero prime factors. For the inductive step, suppose that N_n has at least n prime factors. The number $N_n + 1$ is relatively prime to N_n, so its prime factors are all distinct from the prime factors of N_n. Furthermore, we have $1 < N_n + 1$, so $N_n + 1$ indeed has at least one prime factor. Thus it follows that N_{n+1} has at least $n + 1$ prime factors, and therefore there must be infinitely many primes. □

Goldbach's Proof

Our next proof of the infinitude of primes is due to Goldbach, from 1730. In the 18th century, it was customary among mathematicians to share their findings private correspondences. In July of 1730, Goldbach wrote a letter to Euler, in which he established that all distinct Fermat numbers are relatively prime, and that hence there must be infinitely many primes.

Recall that the Fermat numbers Φ_n are defined by

$$\Phi_n := 2^{2^n} + 1.$$

Lemma 5.2.3. *Any two distinct Fermat numbers are relatively prime.*

Proof. Consider two distinct natural numbers m and n, and assume that $m < n$. Recall from Exercise 1.13 that

$$\Phi_n - 2 = \prod_{k<n} \Phi_k.$$

Thus we see that $\Phi_m \mid \Phi_n - 2$. It follows that any common divisor d of Φ_m and Φ_n is also a divisor of 2, i.e., $d = 1$ or $d = 2$. However, all Fermat numbers are clearly odd, and this implies that any common divisor of Φ_m and Φ_n is 1. In other words, Φ_m and Φ_n are relatively prime. □

Goldbach's proof of Theorem 5.2.2. Since any two distinct Fermat numbers are relatively prime, and since any number $n > 1$ has a prime divisor, it follows that if we pick a prime divisor of each Fermat number Φ_n, then we picked infinitely many distinct prime numbers. □

The Furstenberg proof, following Cass–Wildenberg

The following proof of the infinitude of primes is, in its essence, due to Hillel Furstenberg [Fur55]. In its original formulation, this proof makes clever use of a topology on $\mathbb{Z}$. Although topologies are not hard to define, motivating them is beyond the scope of this course. We will therefore present a revised version of Furstenberg's proof, following Daniel Cass and Gerald Wildenberg [CW03].

The Cass–Wildenberg approach to prove the infinitude of primes makes use of the concept of a periodic subset of $\mathbb{Z}$. A subset $A \subseteq \mathbb{Z}$ is said to be *periodic with period n*, or simply *periodic* if we merely assume such a natural number n to exist, if we have

$$x \in A \qquad \Leftrightarrow \qquad x + n \in \mathbb{Z}.$$

Thus, if A is a periodic subset of $\mathbb{Z}$ with period n, and $x \in A$, then

$$x + kn \in A$$

for all $k \in \mathbb{Z}$. In particular, all inhabited periodic subsets of $\mathbb{Z}$ are infinite. We note two basic facts about periodic subsets of $\mathbb{Z}$, which we leave to the reader to verify:

1. If $A_1, \ldots, A_m$ is a finite family of periodic subsets of $\mathbb{Z}$, each with their own periods, then the union
$$A_1 \cup \cdots \cup A_m$$
is again a periodic subset of k. The period of the union is the least common multiple of the periods of the subsets A_i.

2. If A is a periodic subset of $\mathbb{Z}$, then its complement $A^{\mathrm{c}} := \{x \in \mathbb{Z} \mid x \notin A\}$ is periodic with the same period.

The Furstenberg proof of Theorem 5.2.2, following Cass–Wildenberg. Consider the subsets
$$S_p := \{kp \mid k \in \mathbb{Z}\},$$
where p is a prime number, and define
$$S := \bigcup_{p \text{ prime}} S_p.$$
Since every integer besides ± 1 is divisible by a prime, it follows that S consists of all the integers besides ± 1. The complement of S is therefore the set $\{-1, 1\}$. Note that the subset $\{-1, 1\}$ of $\mathbb{Z}$ cannot be periodic, since is inhabited and finite. This allows us to conclude that the subset S of $\mathbb{Z}$ is not periodic. However, if there were only finitely many primes, then S would be periodic because it would be a finite union of periodic subsets of $\mathbb{Z}$. Thus we conclude that there must be infinitely many primes. □

5.3 The Fundamental Theorem of Arithmetic

The following proposition is due to Euclid, who included it as proposition 30 of book VII of the *Elements*.

Lemma 5.3.1. *Suppose that p is prime, and that a is an integer. Then either $p \mid a$ or $\gcd(p, a) = 1$.*

Proof. The greatest common divisor of p and a is in particular a divisor of p, so it is either 1 or p. If $\gcd(p, a) = p$, then we have $p \mid a$. □

Proposition 5.3.2. *Consider a positive integer a. The following are equivalent:*

1. *The integer a is prime.*

2. *We have $a > 1$, and for any two integers b and c such that $a \mid bc$, we have that $a \mid b$ or $a \mid c$.*

Proof. Suppose first that a is prime. Since 0 and 1 aren't prime, it follows that $a > 1$. Now consider two integers b and c such that $a \mid bc$. There are two cases to consider: either $a \mid b$, in which case we are done immediately, or $a \nmid b$. In the case where $a \nmid b$, it follows from Lemma 5.3.1 that a and b are relatively prime. This allows us to apply Proposition 4.4.5, by which we conclude that $a \mid c$.

For the converse, suppose that for any two integers b and c such that $a \mid bc$, we have $a \mid b$ or $a \mid c$, and let d be a proper divisor of a. Now consider any positive proper divisor d of a, and write $a = dk$. Then it follows in particular that $a \mid dk$, so by assumption we have either $a \mid d$ or $a \mid k$. The first condition can be ruled out, because d is assumed to be a proper divisor and therefore $a \nmid d$. Thus we find ourselves in the second case, where $a \mid k$ and $k \mid a$. The integer k is also positive, since a and d are positive, so it follows that $a = k$. This implies that $d = 1$, and hence we conclude that a is prime. □

Theorem 5.3.3. *For any natural number $n > 0$, there is a unique list $\ell = (p_0, \ldots, p_{l-1})$ of length l, consisting of primes $p_0 \leq p_1 \leq \cdots \leq p_{l-1}$ such that*

$$n = \prod_{i=0}^{l-1} p_i.$$

This unique list of primes is called the prime decomposition of *n.*

Proof. We apply strong induction on n, with base case 1. In the base case, we let ℓ be the empty list. The empty list satisfies the increasing primes condition vacuously, and empty products are 1 by definition, so the empty list is indeed a prime decomposition of 1.

For the inductive step, recall that any natural number $n + 1 > 1$ has a least prime divisor. If p is the least prime divisor of $n + 1$, then it follows that $(n + 1)/p \leq n$. By the strong induction hypothesis we have a unique prime decomposition $(q_0, \ldots, q_{l-1})$ of $(n + 1)/p$. Now define the list $\ell := (p_0, p_1, \ldots, p_l)$ by $p_0 := p$ and $p_{i+1} := q_i$. Then each p_i is a prime divisor of $n + 1$, and since p_0 is the least prime divisor of $n + 1$ we have the inequalities

$$p_0 \leq p_1 \leq \cdots \leq p_l.$$

Furthermore, we have

$$n + 1 = p \cdot ((n + 1)/p) = p_0 \cdot \prod_{i=1}^{l} p_i = \prod_{i=0}^{l} p_i.$$

This proves that the prime decomposition of $n + 1$ exists.

To show that the prime decomposition of $n + 1$ is unique, let $(r_0, \cdots, r_k)$ be a prime decomposition of $n + 1$. The smallest prime divisor p of $n + 1$ then divides the product

$$p \mid \prod_{i=0}^{k} r_i.$$

This implies that $p \mid r_i$ for some $0 \leq i \leq k$, i.e., that $p = r_i$. However, p is assumed to be the smallest prime factor of $n + 1$, so it follows that $p = r_0$. Now it follows that $(r_1, \ldots, r_k)$ is a prime decomposition of $(n + 1)/p$, which is unique, and therefore we conclude that $n + 1$ has a unique prime factorization. □

5.4 Legendre's Formula and Kummer's Theorem

In some cases we can give precise expressions of the prime factorization of a number. We will describe here two such cases: The prime factorization of the factorial $n!$, which is due to Adrien-Marie Legendre, and the prime factorization of the binomial coefficients $\binom{n}{m}$, which is due to Ernst Kummer. These explicit formulas will be useful in the study of the distribution of the primes.

Definition 5.4.1. The *p-adic valuation* of a positive integer n is the largest exponent m such that $p^m \mid n$. That is, it is the unique natural number m such that

$$p^k \mid n \qquad \Leftrightarrow \qquad k \leq m$$

for every natural number k. We will write $\nu_p(n)$ for the p-adic valuation of n.

For example, the 2-adic valuation of 24 is $\nu_2(24) = 3$ because the exponent of 2 in $24 = 2^3 \cdot 3$ is 3, and its 3-adic valuation is $\nu_3(24) = 1$ because the exponent of 3 in $2^3 \cdot 3$ is 1. Another useful fact is that the p-adic valuation of the greatest common divisor of two numbers is the minimum of their respective p-adic valuations. More generally, if the greatest common divisor is taken of multiple natural numbers we obtain

$$\nu_p(\gcd(n_1, \ldots, n_k)) = \min(\nu_p(n_1), \ldots, \nu_p(n_k)).$$

Likewise, the p-adic valuation of the least common multiple of $n_1, \ldots, n_k$ is

$$\nu_p(\operatorname{lcm}(n_1, \ldots, n_k)) = \max(\nu_p(n_1), \ldots, \nu_p(n_k)).$$

For example the greatest common divisor and the least common multiple of $60 = 2^2 \cdot 3 \cdot 5$ and $525 = 3 \cdot 5^2 \cdot 7$ are

$$\gcd(60, 525) = 3 \cdot 5 = 15 \qquad \text{and} \qquad \operatorname{lcm}(60, 525) = 2^2 \cdot 3 \cdot 5^2 \cdot 7 = 2100.$$

Theorem 5.4.2 (Legendre's formula)**.** *For any natural number n, we have*

$$n! = \prod_{p \leq n} p^{\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \cdots}.$$

Proof. Another way of stating Legendre's formula for $n!$ is that the p-adic valuation of $n!$ is given by

$$\nu_p(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

To see this, note that each integer $1 \leq x \leq n$ divisible by p contributes a factor p to $n!$. Each integer $1 \leq x \leq n$ divisible by p^2 contributes a second factor of p to $n!$, and more generally, each integer $1 \leq x \leq n$ divisible by p^k contributes a kth factor of p.

Since there are $\left\lfloor \frac{n}{p^k} \right\rfloor$ integers from 1 to n divisible by p^k, it follows that the number of factors of p in $n!$ is

$$\sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor. \qquad \square$$

Corollary 5.4.3 (de Polignac's formula)**.** *Consider a natural number $n \geq 1$ written in its base p representation*

$$\sum_{i=0}^{l} a_i p^i,$$

where we have $0 \leq a_i < p$ for each i. Then p-adic valuation of $n!$ is

$$\nu_p(n!) = \frac{n - \sum_{i=0}^{l} a_i}{p-1}.$$

Proof. Given the base-p representation of n, the base-p representation of $\lfloor \frac{n}{p^k} \rfloor$ is given by dropping the digits first k digits in the base-p representation of n:

$$\left\lfloor \frac{n}{p^k} \right\rfloor = a_k + a_{k+1} p + \cdots + a_l p^{l-k}.$$

Using this observation, we obtain:

$$\sum_{k=1}^{l} \left\lfloor \frac{n}{p^k} \right\rfloor = \sum_{k=1}^{l} \sum_{j=k}^{l} a_j p^{j-k} = \sum_{j=1}^{l} a_j \sum_{k=1}^{j} p^{j-k} = \sum_{j=1}^{l} a_j \frac{p^j - 1}{p-1}.$$

On the other hand, we have

$$n - \sum_{j=0}^{l} a_j = \sum_{j=0}^{l} a_j p^j - a_j = \sum_{j=0}^{l} a_j (p^j - 1).$$

This gives

$$\frac{n - \sum_{j=0}^{l} a_j}{p-1} = \sum_{j=0}^{l} a_j \frac{p^j - 1}{p-1} = \sum_{k=1}^{l} \left\lfloor \frac{n}{p^k} \right\rfloor. \qquad \square$$

Example 5.4.4. We can use Legendre's formula and de Polignac's formula to compute the prime factorizations of binomial coefficients. For example, it follows immediately from Legendre's formula that all the prime factors of the central binomial coefficient

$$\binom{64}{32}$$

are bounded from above by 64. Moreover, every prime $32 < p \leq 64$ occurs with p-adic valuation 1. Furthermore, the central binomial coefficient $\binom{64}{32}$ is not divisible by any prime $\frac{2}{3}32 < p \leq 32$, showing that the 23-, 29-, and 31-adic valuations of $\binom{64}{32}$ are 0. To compute the p-adic valuations for the lower primes, we use de Polignac's formula, for which it is necessary to write 32 and 64 in base p. Using letters A, B, C, and so forth for digits representing values of 10 and higher, we gather all the calculations for de Polignac's formula in the following table:

p	32 base p	64 base p	$\frac{2s_p(32)-s_p(64)}{p-1}$
2	$(100000)_2$	$(1000000)_2$	1
3	$(1012)_3$	$(2101)_3$	2
5	$(112)_5$	$(224)_5$	0
7	$(44)_7$	$(121)_7$	2
11	$(2A)_{11}$	$(59)_{11}$	1
13	$(26)_{13}$	$(4C)_{13}$	0
17	$(1F)_{17}$	$(3D)_{17}$	1
19	$(1D)_{19}$	$(37)_{19}$	1

Thus, we see that

$$\binom{64}{32} = 2 \cdot 3^2 \cdot 7^2 \cdot 11 \cdot 17 \cdot 19 \cdot 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59 \cdot 61.$$

Even though the binomial coefficient $\binom{64}{32}$ is a fairly large number with 19 decimal digits, we didn't need to evaluate it in order to obtain its prime factorization. In fact, without fully evaluating it, we can now determine the last decimal digit of $\binom{64}{32}$ by the following procedure: Start with $d_0 := 1$ and let d_{i+1} be the remainder of the product $d_i p_i^{m_i}$ after division by 10, where $p_i^{m_i}$ is the ith prime power in the prime factorization of $\binom{64}{32}$. This gives:

$$d_1 = 2,\ d_2 = 8,\ d_3 = 2,\ d_4 = 2,\ \ldots,\ d_{13} = 4.$$

The last decimal digit of $\binom{64}{32}$ is therefore a 4.

In the following theorem, which is due to Ernst Kummer [Kum52], we will compute the p-adic valuation of the binomial coefficients in full generality. Before we do this, we need to recall addition of two natural numbers written in base p. Suppose that

$$m = \sum_{i=0}^{l_m} a_i p^i \qquad \text{and} \qquad n = \sum_{i=0}^{l_n} b_i p^i.$$

Then their sum $m + n$ has base-p representation

$$m + n = \sum_{i=0}^{l_n} c_i p^i,$$

where c_i is obtained by the following recursive procedure, simultaneously with a number $u(m,n)_i \in \{0,1\}$ recording whether a 1 needs to be carried over to the next digit: Set $u(m,n)_0 = 0$ and define

$$(c_i, u(m,n)_{i+1}) := \begin{cases} (a_i + b_i + u(m,n)_i, 0) & \text{if } a_i + b_i + u(m,n)_i < p \\ (a_i + b_i + u(m,n)_i - p, 1) & \text{if } a_i + b_i + u(m,n)_i \geq p. \end{cases}$$

Theorem 5.4.5 (Kummer's Theorem)**.** *Consider two natural numbers m and n. The p-adic valuation of the binomial coefficient $\binom{m+n}{m}$ is the number of carries in the addition*

of m and n in base p: If m and n have base p-representations

$$m = \sum_{i=0}^{l_m} a_i p^i \qquad \text{and} \qquad n = \sum_{i=0}^{l_n} b_i p^i,$$

then

$$\nu_p\binom{m+n}{m} = \#\{0 \leq i \leq l_n \mid u(m,n)_i = 1\}.$$

Proof. Write $s_p(x)$ for the sum of the digits of the p-adic representation of natural number x. By de Polignac's formula for the p-adic valuation of $(m+n)!$ we find that

$$\nu_p\binom{m+n}{m} = \nu_p((m+n)!) - \nu_p(m!) - \nu_p(n!) = \frac{s_p(m) + s_p(n) - s_p(m+n)}{p-1}.$$

The p-adic representation of $m+n$ is given by

$$m+n = \sum_{i=0}^{l_{m+n}} c_i p^i,$$

where

$$c_i = \begin{cases} a_i + b_i + u(m,n)_i & \text{if } a_i + b_i + u(m,n)_i < p \\ a_i + b_i + u(m,n)_i - p & \text{otherwise.} \end{cases}$$

We therefore find that the numerator in de Polignac's expression for $\nu_p\binom{m+n}{m}$ contains a term $p - u(m,n)_{i+1}$ for every time $u(m,n)_{i+1} = 1$. That is, the numerator is of the form $C(p-1)$ where C is the number of carries. This proves the theorem. □

Kummer's theorem can be used to show that the binomial coefficient $\binom{n}{k}$ always divides the least common multiple of $1, 2, \ldots, n$. In order to prove this result, note that the largest exponent m such that $p^m \leq n$ is the number

$$m = \lfloor \log_p(n) \rfloor.$$

In other words, if m is the largest integer not exceeding $\log_p(n)$, then the leading digit in the base-p representation of n belongs to p^m.

Corollary 5.4.6. *For any $0 \leq k \leq n$ we have*

$$\binom{n}{k} \;\Big|\; \operatorname{lcm}(1, \ldots, n).$$

Proof. Suppose that the base-p representation of n is

$$n = \sum_{i=0}^{l} a_i p^i.$$

Then there are clearly at most l carries when we add k to $n-k$. Since $p^l \leq n$, it follows that

$$p^{\nu_p\binom{n}{k}} \mid \operatorname{lcm}(1, \ldots, n).$$

Since this is true for every prime $p \leq n$, and $\binom{n}{k}$ contains only prime divisors up to n, the claim follows. □

5.5 Bertrand's Postulate

In 1845, Joseph Bertrand observed that for any $1 \leq n \leq 3,000,000$ there is always a prime between n and $2n$, and conjectured that this must be true for all n [Ber45]. Nowadays on a computer, it is really easy to generate lists of primes p_i so that the prime p_{i+1} is strictly below $2p_i$ for each i. Taking each time the largest possible prime satisfying this requirement, we obtain the list

$$2,\ 3,\ 5,\ 7,\ 13,\ 23,\ 43,\ 83,\ 163,\ 317,\ 631,\ 1259,$$
$$2503,\ 5003,\ 9973,\ 19937,\ 39869,\ 79699,\ 159389,$$
$$318751,\ 637499,\ 1274989,\ 2549951,\ 5099893,\ \ldots.$$

The primes in this list are sometimes called the *Bertrand primes*, and they are listed as A006992 in the On-Line Encyclopedia of Integer Sequences [Fou25]. Bertrand's conjecture was proven five years later by Pafnuty Chebyshev, who contributed many more results about the distribution of the primes using analytical methods [Che50]. We will give a proof by Paul Erdős [Erd32], which can be obtained through the methods of the previous section.

Lemma 5.5.1. *For any positive integer n, we have*

$$\prod_{p \leq n} p < 4^n.$$

Proof. The proof is by induction on n. For $n = 1$, there are no primes $p \leq 1$, so the product over all the primes $p \leq 1$ is just 1, which is indeed strictly below 4. Likewise, the theorem holds for $n = 2$, because $2 < 4^2$.

For the inductive step, assume that $\prod_{p \leq n} p < 4^n$. In order to show that

$$\prod_{p \leq n+1} p < 4^{n+1},$$

there are two cases to consider: Either $n + 1$ is even or $n + 1$ is odd. The even case is immediate: We have

$$\prod_{p \leq n+1} p = \prod_{p \leq n} p < 4^n < 4^{n+1}.$$

In the odd case, there is an m such that $2m + 1 = n + 1$. Since every prime $m + 2 \leq p \leq 2m + 1$ divides the binomial coefficient $\binom{2m+1}{m}$, it follows that

$$\prod_{p \leq 2m+1} p \leq \binom{2m+1}{m} \cdot \prod_{p \leq m+1} p < \binom{2m+1}{m} 4^{m+1}.$$

To finish the proof, it suffices to show that $\binom{2m+1}{m} \leq 4^m$. This follows, since the binomial coefficients $\binom{2m+1}{m}$ and $\binom{2m+1}{m+1}$ occur separately in the expansion of $(1+1)^{2m+1}$. However, they are equal, so neither of them can exceed $\frac{1}{2}2^{2m+1} = 4^m$, which proves the lemma. □

Theorem 5.5.2. *For any positive integer n, there is a prime $n < p \leq 2n$.*

Proof. Consider an integer $n \geq 2$ for which there is no prime $n < p < 2n$, and let p be a prime divisor of the central binomial coefficient $\binom{2n}{n}$. Such a prime divisor must then be less than or equal to $\frac{2}{3}n$, since $q^3 \nmid (2n)!$ for any prime $\frac{2}{3}n < q \leq n$, while $q^2 \mid (n!)^2$.

Now, if m is the largest exponent such that $p^m \mid \binom{2n}{n}$, then it follows from Corollary 5.4.6 that

$$p^m \leq 2n.$$

This allows us to estimate the number of primes for which $p^2 \mid \binom{2n}{n}$. Indeed, if $m \geq 2$, then it follows that

$$p \leq \sqrt[m]{2n} \leq \sqrt{2n}.$$

This shows that there are at most $\sqrt{2n}$ primes in the prime factorization of $\binom{2n}{n}$, of which the exponent is larger than 1. It follows that

$$\frac{4^n}{2n+1} < \binom{2n}{n} \leq (2n)^{\sqrt{2n}} \prod_{p \leq 2n/3} p < (2n)^{\sqrt{2n}} 4^{\frac{2n}{3}},$$

where the first inequality was established in Exercise 2.5, and the last inequality follows from Lemma 5.5.1. Rearranging this strict inequality, we obtain

$$4^{\frac{n}{3}} < (2n+1)(2n)^{\sqrt{2n}} < (2n)^{\sqrt{2n}+1}.$$

Taking logarithms base 2 on both sides, we find that

$$\frac{n}{2} < \frac{2n}{3} < (\sqrt{2n}+1)\log_2(2n)$$

We see that this strict inequality fails for $n \geq 2^9 = 512$, since

$$2^8 = 512 \not< 330 = (2^5+1) \cdot 10.$$

We conclude that Bertrand's postulate must be true for all $n \geq 512$. Using the sieve of Eratosthenes up to 1350, as displayed in Figure 5.1, we see that the Bertrand primes are indeed primes up to 631, which is sufficient and thus the theorem is proven. □

Exercises

5.1 In the number grid provided at the end of the introduction, shade the numbered cells in two colors:

1. Leave all the prime numbers unshaded.
2. Shade the number 1 and all square-free composite numbers in one color.
3. Shade all non-square-free numbers in the other color.

Explain your method.

5.2 Show that there are no primes p such that $p+8$ and $p+22$ are also prime.

5.3 Show that

$$p \,\Big|\, \binom{p}{k}$$

for any $0 < k < p$, and any prime number p.

5.4 Show that

$$4^{\frac{p-1}{2}} \equiv 1 \pmod{p}$$

for any odd prime number p.

5.5 Define the nth *Mersenne number* M_n to be

$$M_n := 2^n - 1.$$

Show that if n is composite, then so is M_n. Conclude that if M_n is prime, then so is n. Prime numbers of the form M_p for some prime number p are called *Mersenne primes*.

5.6 A natural number n is said to be *perfect* if it is the sum of its proper divisors. Show that an even number is perfect if and only if it is of the form $2^{p-1}M_p$, where M_p is a Mersenne prime.

5.7 Consider a fixed set $\{p_1, \ldots, p_n\}$ of n distinct primes. Given a natural number k, how many numbers can be written as a product

$$p_1^{m_1} \cdots p_n^{m_n}$$

such that $m_1 + \cdots + m_n = k$?

5.8 Suppose that $n \geq 3$ and p is a prime such that $\frac{2}{3}n < p \leq n$. Prove that

$$p \nmid \binom{2n}{n}.$$

5.9 Prove the following identity due to Bakir Farhi [Far09]:

$$\operatorname{lcm}\left(\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}\right) = \frac{\operatorname{lcm}(1, \ldots, n+1)}{n+1}.$$

5.10 Determine the number of trailing zeros in the decimal representations of 100! and 2^{10}!.

5.11 (a) Show that $\binom{n}{m}$ is odd if and only if there are no carries occur when adding m and $n - m$ in their binary representations.

(b) Show that $\binom{n}{m}$ is odd for every $0 \leq m \leq n$ if and only if n is of the form $2^k - 1$.

Chapter 6

Polynomials

6.1 Polynomials with Integer Coefficients

Definition 6.1.1. An *polynomial with integer coefficients* in one variable x is an expression of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

in which we require that if $n \neq 0$ then the *leading coefficient* a_n is nonzero. The integers a_k are called the *coefficients* of $f(x)$. The *zero polynomial* is the constant polynomial 0 and the *unit polynomial* is the constant polynomial 1. The *degree* of a polynomial is the highest power of x with a nonzero coefficient.

There is a subtle, yet important distinction between polynomials and the functions they describe, which has to do with equality of polynomials. Two polynomials

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 \qquad \text{and} \qquad b_n x^n + b_{n-1} x^{n-1} + \cdots + b_0$$

are said to be equal if their coefficients are equal, meaning that $a_i = b_i$ for all $0 \leq i \leq n$, while functions are considered equal if their values are equal. This is an important distinction, because for some number systems such as the integers modulo n, polynomials aren't necessarily equal as polynomials when they are equal as functions. For instance, by Fermat's Little Theorem we have that

$$x^p \equiv x \pmod{p},$$

while the polynomials x^p and x clearly can be distinguished by their coefficients. Nevertheless, we will show in Corollary 6.2.7 that two polynomials with integer coefficients have equal coefficients if and only if they define equal functions.

The viewpoint that polynomials are formal expressions, and not functions, was emphasized by Leopold Kronecker, in the 19th century. With this distinction in mind, we make the following definition.

Definition 6.1.2. A function $f : \mathbb{Z} \to \mathbb{Z}$ is said to be *polynomial in the integers* if there is a polynomial with integer coefficients such as the above, such that

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

for all $x \in \mathbb{Z}$.

Remark 6.1.3. Polynomials can be added, subtracted, and multiplied in the same way as integers. To add two polynomials, we add their corresponding coefficients, taking missing coefficients as zero when necessary. Polynomials are multiplied by the formula

$$\begin{aligned}\left(\sum_{i=0}^{n} a_i x^i\right)\left(\sum_{j=0}^{m} b_j x^j\right) &= \sum_{i=0}^{n}\sum_{j=0}^{m} a_i b_j x^{i+j} \\ &= \sum_{k=0}^{n+m}\sum_{i=0}^{n} a_i b_{k-i} x^k.\end{aligned}$$

These operations of addition, subtraction, and multiplication, polynomials satisfy the laws of arithmetic. More precisely, the polynomials with integer coefficients form a commutative ring.

Definition 6.1.4. A *root* of a polynomial $f(x)$ is an element r such that

$$f(r) = 0.$$

Sometimes we will write $N(f)$ for the set of roots of f.

Theorem 6.1.5 (Factor Theorem)**.** *Consider a nonzero polynomial $f(x)$ of degree n. An integer r is a root of $f(x)$ if and only if there is a polynomial $g(x)$ of degree $n-1$ such that $f(x) = (x-r)g(x)$.*

Proof. We will prove the claim by induction on n. In the base case $f(x)$ is a nonzero constant polynomial, so it has no roots and there is no polynomial $g(x)$ of degree -1 such that $f(x) = (x-r)g(x)$.

For the inductive step, it is clear that if there is a polynomial $g(x)$ such that $f(x) = (x-r)g(x)$, then r is a root of the polynomial $f(x)$. For the converse, suppose that r is a root of $f(x)$. Then

$$f(x) = f(x) - f(r) = a_n(x^n - r^n) + \cdots + a_1(x-r).$$

By the formula for the difference of nth powers, we can factor $x^n - r^n$ as

$$x^k - r^k = (x-r)q_k(x),$$

for some polynomial $q_k(x)$ of degree $k-1$. Now let

$$g(x) := \sum_{k=1}^{n} a_k q_k(x).$$

Then we have $f(x) = (x-r)q_k(x)$ by rearranging the terms. □

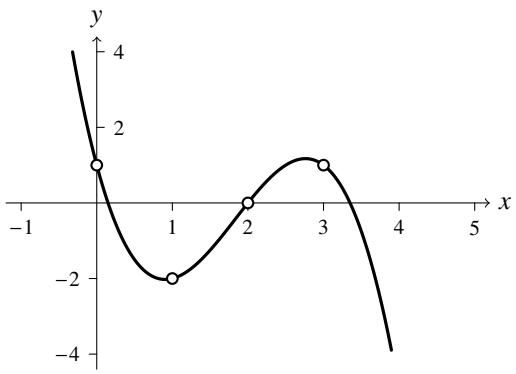


Figure 6.1: A polynomial of degree 3 obtained by Lagrange interpolation.

The factor theorem has some useful corollaries. The first is that a polynomial of degree has at most n roots, and the second is a factor theorem for modular arithmetic, which we will establish in the next section.

Corollary 6.1.6. *A polynomial of degree n with nonzero leading coefficient has at most n roots.*

Proof. The proof is by induction on the degree n. If $f(x)$ is a constant polynomial of degree 0 and its leading coefficient is nonzero, then $f(x)$ is always nonzero and so $f(x)$ has no roots.

For the inductive step, assume that all polynomials of degree n have at most n roots, and consider a polynomial $f(x)$ of degree $n+1$. Then either $f(x)$ has no roots, in which case we are done immediately, or $f(x)$ has a root r, in which case $f(x)$ factors as

$$f(x) = (x - r)g(x)$$

by the Factor Theorem, where $g(x)$ is a polynomial of degree n. Then the product $(x - r)g(x) = 0$ if and only if $x - r = 0$ or $g(x) = 0$. Thus we see that x is a root of $f(x)$ if and only if $x = r$ or x is a root of $g(x)$, so there are at most $n + 1$ roots of the polynomial $f(x)$. □

6.2 Lagrange's Interpolation Theorem

Lagrange's Interpolation Theorem establishes that for any set of data points, there is always a way of defining a polynomial that exactly matches the data. Moreover, it gives the unique way of defining such a polynomial in the lowest degree. Lagrange's Interpolation Theorem has the important consequence that any polynomial of degree n is uniquely determined by any $n+1$ of its values. We will also use Lagrange's Interpolation Theorem in the next section to derive a divisibility property of polynomials due to Hensel.

The idea underlying Lagrange's Interpolation Theorem is to break the problem into several subproblems: For each data point (x_i, y_i), we create a polynomial that has roots

at x_j for all the other data points (x_j, y_j). This is very easy to do: we just take the product

$$g_i(x) = (x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_j).$$

Then we have that $g_i(x_j) = 0$ for all $j \neq i$ and $g_i(x_i) \neq 0$. Thus there is some constant c_i such that $c_i g_i(x_i) = y_i$. Adding them all up gives the Lagrange Interpolation Theorem.

Example 6.2.1. In Figure 6.1, we started with the set of points

$$(0, 1),\ (1, -2),\ (2, 0),\ (3, 1).$$

Then we define the polynomials

$$\begin{aligned} g_0 &:= \tfrac{1}{6}(x - 1)(x - 2)(x - 3) \\ g_1 &:= -x(x - 2)(x - 3) \\ g_2 &:= 0 \\ g_3 &:= \tfrac{1}{6}x(x - 1)(x - 2), \end{aligned}$$

which we have already rescaled so that $g_0(0) = 1$, $g_1(1) = -2$, $g_2(2) = 0$, and $g_3(3) = 1$. Adding them up gives the polynomial

$$f(x) := \frac{1}{6}(x - 1)(x - 2)(x - 3) - x(x - 2)(x - 3) + \frac{1}{6}x(x - 1)(x - 2),$$

which we have plotted.

Theorem 6.2.2 (Lagrange's Interpolation Theorem)**.** *For any choice of $n + 1$ points (x_i, y_i) such that the x_i are pairwise distinct, there is a unique polynomial $f(x)$ of degree at most n such that*

$$f(x_i) = y_i$$

for all $0 \leq i \leq n$.

Proof. For any $0 \leq i \leq n$, define the *basis polynomial* for Lagrange interpolation $g_i(x)$ by

$$g_i(x) := \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}.$$

Then $g_i(x)$ is a polynomial of degree n, whose n roots are the points x_j for $j \neq i$. Furthermore, we have defined g_i in such a way that $g_i(x_i) = 1$. Now we define the polynomial $f(x)$ of degree n by

$$f(x) = \sum_{i=0}^{n} y_i g_i(x).$$

Then $f(x_i) = y_i$ since $g_j(x_i) = 0$ for $j \neq i$ and $g_i(x_i) = 1$. Thus f satisfies the criteria of the theorem.

To see that the polynomial $f(x)$ is the unique polynomial of degree at most n that satisfies $f(x_i) = y_i$, consider a second polynomial $g(x)$ of degree at most n that satisfies the same requirement. Then the polynomial

$$f(x) - g(x)$$

is a polynomial of degree at most n with $n + 1$ roots. By virtue of Corollary 6.1.6 this is only possible if $f(x) - g(x)$ is the zero polynomial, that is, if $f(x) = g(x)$. □

Corollary 6.2.3. *Any polynomial $f(x)$ of degree n is uniquely determined by any $n + 1$ of its values. In particular, it is uniquely determined by the values*

$$f(0), \ldots, f(n).$$

Even though the basis polynomials of Lagrange's Interpolation Theorem don't always have integer coefficients, we can use Lagrange's Interpolation Theorem to show that for any polynomial f with real coefficients, if

$$f(0), \ldots, f(n)$$

are integers, then $f(x)$ is an integer for every integer x.

Definition 6.2.4. A polynomial $f(x)$ with real coefficients is said to be *integer valued* if $f(x)$ is an integer for every integer x.

A simple example of an integer valued polynomial that doesn't have integer coefficients is the polynomial

$$\frac{x(x+1)}{2},$$

which returns the triangular numbers.

Lagrange's Interpolation Theorem establishes a polynomial $f(x)$ as a linear combination of the basis polynomials, thus obtaining

$$\sum_{k=0}^{n} f(k) g_i(x).$$

However, we can turn this perspective around. Each value $f(x)$ can be written as a linear combination of the values $f(k)$ for $0 \leq k \leq n$. This observation is important enough that we establish it in its own lemma, which is a direct corollary of Lagrange's Interpolation Theorem.

Lemma 6.2.5. *Consider a polynomial $f(x)$. Then each value $f(x)$ can be written as a linear combination of the values $f(0), \ldots, f(n)$, i.e., there are coefficients $a_i(x)$ for each $0 \leq i \leq n$ depending on x such that*

$$f(x) = \sum_{i=0}^{n} a_i(x) f(i).$$

Proof. The coefficients $a_k(x)$ are given by the basis polynomials $g_k(x)$. □

Equipped with this change of perspective, we can prove that a polynomial of degree n is integer values as soon as its values $f(0), \ldots, f(n)$ are integers.

Theorem 6.2.6. *Consider a polynomial $f(x)$ of degree n with real coefficients. If the values*

$$f(0), \ldots, f(n)$$

are integers, then $f(x)$ is an integer for every integer x.

Proof. The basis polynomials of f are

$$g_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x-j}{i-j}.$$

We can express $g_i(x)$ using binomial coefficients as

$$g_i(x) = \binom{x}{i}\binom{x-i-1}{n-i},$$

which is an integer for every integer x. Therefore it follows by Lagrange's Interpolation Theorem that every value $f(x)$ can be written as an integer linear combination of the integers $f(0), \ldots, f(n)$, which proves the theorem. □

Corollary 6.2.7. *Two polynomials $f(x)$ and $g(x)$ with integer coefficients have equal coefficients if and only if their values are equal.*

Proof. The forward direction, which asserts that if $f(x)$ and $g(x)$ have equal coefficients then they define the same function, is immediate. For the converse, assume that $f(x) = g(x)$ for every x. Then the polynomial $f(x) - g(x)$ must be the 0 polynomial, since any other polynomial has a degree and a polynomial of degree n has at most n roots by Corollary 6.1.6.

On the other hand, the polynomial $f(x) - g(x)$ is defined as

$$f(x) - g(x) = (a_n - b_n)x^n + \cdots (a_0 - b_0),$$

where a_i and b_i are the coefficients of f and g, respectively, or taken to be 0 if there is no such coefficient. Thus it follows that $a_i - b_i = 0$ for all i, which proves that $a_i = b_i$. □

6.3 Fixed Divisors of Integer Polynomials

In the previous chapter, we have done a few exercises which asked us to show that integers in certain polynomial forms satisfy a divisibility property. For example, we showed that $a^3 - a$ is always divisible by 6, and that $a^5 - a$ is always divisible by 30. We also showed that the product of any n consecutive integers is always divisible by $n!$. In this section we will establish a vast generalization of these facts.

Definition 6.3.1. The *fixed divisor* of a function $f : X \to \mathbb{Z}$ into the integers is the greatest common divisor of all values of f, i.e.,

$$\gcd(\{f(x) \mid x \in X\}).$$

Lagrange's Interpolation Theorem can be used to determine the greatest common divisor of all the values of a polynomial, by examining only one more value than its degree. This observation is due to Kurt Hensel, who was a student of Leopold Kronecker.

Theorem 6.3.2 (Hensel's Fixed Divisor Theorem [Hen96])**.** *Consider a polynomial $f(x)$ of degree n, with integer coefficients. Then the fixed divisor of f is the greatest common divisor*

$$\gcd(f(a), \ldots, f(a+n))$$

of any $n+1$ values of consecutive inputs of f.

Proof. First, we note that by translating the polynomial along the x-axis, which we can do by considering the polynomial $f(a)$, it suffices to show that for any polynomial $f(x)$ of degree n with integer coefficients, the fixed divisor of f is the greatest common divisor of the integers

$$f(0), \ldots, f(n).$$

Now recall from Theorem 6.2.6 that the values $f(x)$ are integer linear combinations of the values $f(0), \ldots, f(n)$:

$$f(x) = \sum_{i=0}^{n} f(i) g_i(x).$$

It follows that

$$\gcd(f(0), \ldots, f(n)) \mid f(x).$$

Since the fixed divisor of f divides this greatest common divisor, this establishes the theorem. □

Example 6.3.3. Consider the polynomial

$$f(x) = x^n - x.$$

When n is even, then f is an *even function*, which means that $f(-x) = f(x)$, and when n is odd, then f is an *odd function* which means that $f(-x) = -f(x)$. In either case, this implies that the fixed divisor of $x^n - x$ is equal to the greatest common divisor of the elements

$$f(0), \ldots, f(m),$$

where m is the largest integer such that $2m \leq n+1$. In other words, to compute the fixed divisor of $x^n - x$, we only have to inspect about half the amount of values of Hensel's Fixed Divisor Theorem. Furthermore, the polynomial $f(x) = x^n - x$ always satisfies $f(0) = 0$ and $f(1) = 0$, reducing the amount of cases even further. Thus, the fixed divisor of $x^n - x$ is

$$\gcd(\{f(k) \mid 2 \leq k \leq (n+1)/2\}).$$

1. For $n = 3$, the only number $2 \leq k \leq (n+1)/2$ is the number 2. The polynomial $f(x) = x^3 - x$ from Exercise 4.10 has the value

   $$f(2) = 6,$$

   so the fixed divisor of $x^3 - x$ is 6, which is $2 \cdot 3$.

2. For $n = 5$, the only two numbers $2 \leq k \leq (n+1)/2$ are the numbers 2 and 3. The polynomial $f(x) = x^5 - x$ from Exercise 4.11 has the values

$$f(2) = 30 \qquad \text{and} \qquad f(3) = 240.$$

Since 240 is divisible by 30 it follows that the fixed divisor of $x^5 - x$ is 30, which is $2 \cdot 3 \cdot 5$.

3. For $n = 7$, the only three numbers $2 \leq k \leq (n+1)/2$ are the numbers 2, 3, and 4. The polynomial $f(x) = x^7 - x$ has the values

$$f(2) = 126 \qquad \text{and} \qquad f(3) = 2184.$$

Furthermore, using that $4^7 = (2^7)^2$ we see that $f(4)$ factors as $(2^7 - 2)(2^7 + 2)$ by the formula for the difference of squares. Therefore it follows that the fixed divisor of $f(x)$ is the greatest common divisor of $f(2)$, $f(3)$, and $f(4)$, is just the greatest common divisor of $f(2)$ and $f(3)$, which is

$$\gcd(126, 2184) = 42 = 2 \cdot 3 \cdot 7.$$

We will generalize these examples in Theorem 13.4.6.

Exercises

6.1 Show that $x^{13} - x$ has fixed divisor

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 13.$$

Chapter 7

Pythagorean Triples

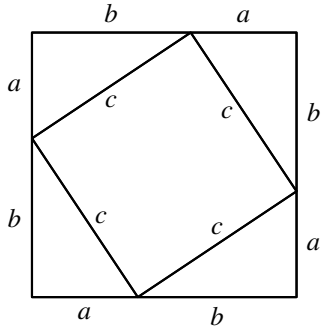
7.1 The Pythagorean Theorem

Recall that a right triangle is a triangle in which one of the angles is a right angle, i.e., it is exactly 90°. The Pythagorean theorem relates the lengths of the sides of a right triangle to the length of its hypotenuse. There are many ways to prove this theorem. We will present the *Chinese proof*, *Euclid's proof*, and the trigonometric proof of Ne'Kiya Jackson and Calcea Johnson, who made headlines in 2023 as high school students with their remarkable discovery of a trigonometric proof [Gua23].

Theorem 7.1.1. *Consider a right triangle with side lengths a and b, and a hypotenuse of length c. Then*

$$a^2 + b^2 = c^2.$$

The Chinese proof. Consider a square with side lengths $a + b$, and include four right triangles with side lengths a and b in the square, so that their right angles coincide with the four right angles of the square.



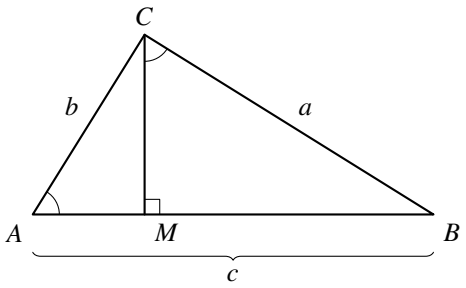
The area of the square is $(a + b)^2$. On the other hand, the area of each triangle is $ab/2$. Combined with the area of the inner square, whose sides are the hypotenuses of the four

triangles, we obtain that

$$(a+b)^2 = c^2 + 4\frac{ab}{2} = c^2 + 2ab.$$

Since $(a+b)^2 = a^2 + 2ab + b^2$, we find that $a^2 + b^2 = c^2$. □

Euclid's proof. Consider a right triangle ACB with hypotenuse AB, and draw a line from C through AB at a right angle. The intersection point is called M.



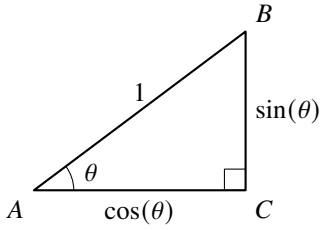
Let x be the length of the line segment AM, and let y be the length of the line segment BM, so that $x + y = c$. The line CM divides the triangle ACB into two triangles CMA and CMB, both of which are similar to the original triangle ACB. Thus, the ratio $a : c$ is the same as the ratio $y : a$, and the ratio $b : c$ is the same as the ratio $x : b$. This gives us the equalities

$$a^2 = cy \qquad \text{and} \qquad b^2 = cx.$$

Since $x + y = c$, we find that $a^2 + b^2 = c^2$. □

Our third proof and final proof in this section, which is due to Ne'Kiya Jackson and Calcea Johnson, is remarkable because it makes essential use of trigonometric functions sine and cosine and the addition law, which can be derived trigonometrically, rather than on calculations involving areas of shapes. Besides its use of trigonometry, it is remarkable because Ne'Kiya Jackson and Calcea Johnson were still in high school when they found their proof, and it stands out for its inventive creativity.

Recall that the sine and cosine functions are defined to be the lengths of the vertical and horizontal sides of a right triangle, as indicated in the diagram on the side.



This means that if ABC is a right triangle with side BC of length a, side AC of length b, hypotenuse AB of length c, and angle $\theta := \angle CAB$, then we have

$$\sin(\theta) = \frac{a}{c} \qquad \text{and} \qquad \cos(\theta) = \frac{b}{c}.$$

In the proof of the Pythagorean theorem, we will make use of the addition formulas for sine and cosine:

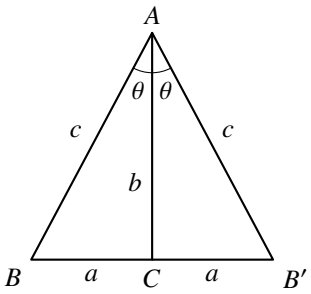
$$\sin(\theta + \phi) = \sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi)$$
$$\cos(\theta + \phi) = \sin(\theta)\sin(\phi) - \cos(\theta)\cos(\phi).$$

In particular, $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$, and $\cos(2\theta) = \sin^2(\theta) - \cos^2(\theta)$. These laws can be derived geometrically if $\theta + \phi$ is an acute angle, i.e., if it is strictly less than $90°$, see Exercise 7.2.

The Jackson–Johnson proof [JJ24]. Consider a right triangle ABC with the right angle at C, and label the sides $a = BC$, $b = AC$, and $c = AB$ so c is the hypotenuse.

In the special case where ABC is isosceles, i.e., if $a = b$, the hypotenuse is the diagonal of an a-by-a square, which gives us that $c^2 = 2a^2$. Hence, we may assume that ABC is nonisosceles.

Without loss of generality, we will assume that $a < b$, i.e., that BC is the shorter side. We reflect the triangle along the longer side AC, as indicated in the diagram below:



Observe that the angle $\theta := \angle BAC$ is strictly less than $45°$, so that the angle $2\theta = \angle BAB'$ is still an acute angle.

Using this combined triangle, we draw a line from B perpendicular to the line AB, and we extend the line AB' until they meet in an intersection point, which we call D. This construction yields a new right triangle ABD with the right angle at B, as in Figure 7.1.

We will now subdivide the triangle BDB' by first dropping a line down from B' at a right angle from the line BB', creating a new right triangle BEB'. Since this triangle also has an angle θ, it is similar to the original triangle ABC. Its sides have lengths $2a$ and $2a^2/b$.

We continue subdividing the triangle BDB' into ever smaller right triangles, as indicated in the figure. The successive triangles in this sequence are all right triangles with an angle θ, so they are all similar. Corresponding sides in these successive triangles have a ratio of a/b. The hypotenuses of these triangles therefore have length

$$2c\left(\frac{a}{b}\right)^n$$

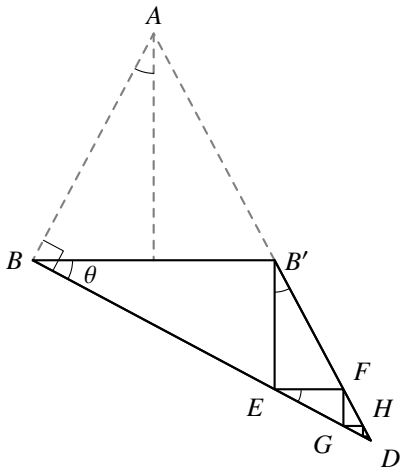


Figure 7.1: The recursive subdivision of the triangle BDB'.

This allows us to calculate the length of the side $B'D$ via the formula for the geometric series, which we proved in Theorem 1.1.2:

$$|B'D| = \sum_{k=0}^{\infty} 2c\left(\frac{a}{b}\right)^{2k+2} = 2c\left(\frac{a}{b}\right)^2 \sum_{k=0}^{\infty}\left(\frac{a}{b}\right)^{2k} = \frac{2c\left(\frac{a}{b}\right)^2}{1-\left(\frac{a}{b}\right)^2} = \frac{2ca^2}{b^2-a^2}.$$

Now we can find two expressions for the length of the line AD. First, it consists of the lines AB' and $B'D$, so it is of length $c + |B'D|$, which we calculated above. On the other hand, by trigonometry we have that $c/|AD| = \cos(2\theta)$, which gives us that $|AD| = c/\cos(2\theta)$. Since both expressions give the length of the line AD, both involving a factor of c which we can factor out, we obtain the following equation:

$$1 + \frac{2a^2}{b^2-a^2} = \frac{1}{\cos(2\theta)}.$$

Furthermore, the addition formula for cosines gives us that $\cos(2\theta) = \sin^2(\theta) - \cos^2(\theta)$. Since $\sin(\theta) = \frac{b}{c}$ and $\cos(\theta) = \frac{a}{c}$, we find that

$$1 + \frac{2a^2}{b^2-a^2} = \frac{c^2}{b^2-a^2}.$$

Rearranging this equation gives us that $(b^2 - a^2) + 2a^2 = c^2$, which yields

$$a^2 + b^2 = c^2.$$ □

7.2 Euclid's Parametrization of the Pythagorean Triples

Our goal in this section is to describe all the solutions of the quadratic Diophantine equation

$$x^2 + y^2 = z^2,$$

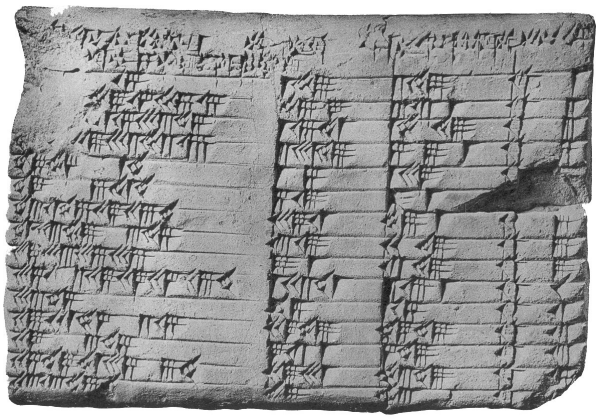
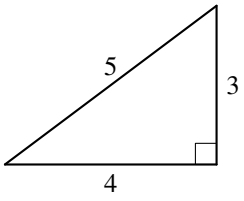


Figure 7.2: The Babylonian tablet Plimpton 322, listing Pythagorean triples.

where z is a nonzero integer. By the Pythagorean Theorem, any such triple (x, y, z) describes the lengths of the sides and hypotenuse of a right triangle. Such a triple is therefore called a *Pythagorean triple*. Notice that if (x, y, z) is a Pythagorean triple, then so is (kx, ky, kz), and their corresponding triangles are of the same shape. Such Pythagorean triples are therefore called *similar*, and we are interested in the Pythagorean triples up to similarity.

The trivial solutions are $(0, 1, 1)$ and $(1, 0, 1)$. The simplest nontrivial solution is the famous equation

$$3^2 + 4^2 = 5^2.$$



The triangle with side lengths 3 and 4 and a hypotenuse of length 5 is sometimes called the *Egyptian triangle*, since it is believed that the ancient Egyptians used this triangle for practical applications. The oldest surviving table of Pythagorean triples was created by the Babylonians, in what is now known as the tablet *Plimpton 322*. This clay tablet, which is estimated to date from 1800 BCE, contains a mathematical table written in cuneiform script, where each row describes a Pythagorean triple in sexagesimal notation, i.e., base 60.

Some further Pythagorean triples include

$$(5, 12, 13), \qquad (21, 20, 29), \qquad \text{and} \qquad (15, 8, 17).$$

Fibonacci devised in his book *Liber Quadratorum* (The Book of Squares) [Fib25] a clever way to quickly show that there are infinitely many Pythagorean triples. If an odd square a^2 is the nth odd number, i.e., if $a^2 = 2n + 1$, then there are positive integers b and c such that

$$b^2 = \sum_{i=0}^{n-1} (2i + 1), \qquad \text{and} \qquad c^2 = \sum_{i=0}^{n} (2i + 1)$$

by Exercise 1.2. For this choice of a, b, and c, we have

$$a^2 + b^2 = c^2.$$

Proposition 7.2.1. *There are infinitely many Pythagorean triples (a, b, c) such that $c = b + 1$.*

Note that while Fibonacci used the sum of odd numbers to find a triple (a, b, c) of integers such that $a^2 + b^2 = c^2$, these Fibonacci triples follow the pattern

$$(2b + 1) + b^2 = (b + 1)^2.$$

Here we immediately see that if $2b + 1$ is a square number a^2, and indeed any odd square is of this form, then we obtain a Pythagorean triple of the form $(a, b, b + 1)$. Rewriting $a = 2k + 1$, we see that Fibonacci's Pythagorean triples are of the form

$$(2k + 1,\ 2k^2 + 2k,\ 2k^2 + 2k + 1).$$

The Pythagorean triples (a, b, c) generated by Fibonacci's method always satisfy $c = b + 1$. However, we have already seen that $(21, 20, 29)$ is a Pythagorean triple that is not of this form. Therefore we see that, even though Fibonacci's method can be used to see that there are infinitely many Pythagorean triples, it does not generate all of them.

In the remainder of this section we will derive the classic formula that describes all the Pythagorean triples. The method leading to this formula was described by Euclid in Book X, Proposition 29 and Book 6, Proposition 21 of the Elements.

Definition 7.2.2. A *primitive Pythagorean triple* is a triple (x, y, z) satisfying

$$x^2 + y^2 = z^2$$

such that $\gcd(x, y) = 1$.

Example 7.2.3. All the Pythagorean triples generated by Fibonacci's method are primitive, since $\gcd(2b + 1, b) = 1$.

Lemma 7.2.4. *For any Pythagorean triple (x, y, z) there is a unique primitive Pythagorean triple (x', y', z') consisting of positive integers for which there is an integer k such that $(kx', ky', kz') = (x, y, z)$.*

Proof. Given a Pythagorean triple (x, y, z), let $k := \gcd(x, y)$. Since $k \mid x$ and $k \mid y$, it follows from the equation $x^2 + y^2 = z^2$ that $k \mid z$. Thus, we define

$$x' := x/k, \qquad y' := y/k, \qquad \text{and} \qquad z' := z/k.$$

Then x' and y' are relatively prime, and we have

$$x'^2 + y'^2 = \frac{x^2 + y^2}{k^2} = \frac{z^2}{k^2} = z'^2.$$

The triple (x', y', z') is therefore a primitive Pythagorean triple such that

$$(kx', ky', kz') = (x, y, z).$$

For uniqueness, suppose that (x'', y'', z'') is an other such primitive Pythagorean triple, satisfying $(lx'', ly'', lz'') = (x, y, z)$. Since $\gcd(x'', y'') = 1$, it follows that

$$k = \gcd(x, y) = \gcd(lx'', ly'') = l$$

Furthermore, since at least one of x or y is nonzero, it follows both k and l are nonzero. Thus we get $x' = x''$ from the equation $kx' = kx''$, and we get $y' = y''$ from the equation $ky' = ky''$. We conclude that $(x', y', z') = (x'', y'', z'')$. □

By the previous lemma we can restrict our attention to primitive Pythagorean triples. Next, we do a parity analysis of the components of a primitive Pythagorean triple.

Lemma 7.2.5. *Consider a primitive Pythagorean triple (x, y, z). Then exactly one of x and y is even, the other is odd, and z is always odd.*

Proof. Since (x, y, z) is a primitive Pythagorean triple, it is clearly impossible for both x and y to be even. It is also impossible for both of them to be odd, since in that case we would have

$$z^2 = x^2 + y^2 \equiv 2 \pmod 4 .$$

However, every square is either congruent to 0 or to 1 modulo 4, so this is impossible.

Thus, one of x and y is even and the other is odd. This gives us that z^2 is odd, which is only possible if z is odd. □

By the previous lemma we may assume, without loss of generality, that in a primitive Pythagorean triple (x, y, z), the integer x is odd and the integer y is even.

Next, we note that $x^2 = z^2 - y^2$, which factors by the formula for the difference of squares as

$$x^2 = (z - y)(z + y).$$

In the following lemma we will show that for any primitive Pythagorean triple (x, y, z), the factors $z - y$ and $z + y$ of x^2 are relatively prime. This immediately implies that they are both squares, which is the key to Euclid's parametrization of the Pythagorean triples.

Lemma 7.2.6. *Suppose that (x, y, z) is a primitive Pythagorean triple in which y is even. Then the integers $z - y$ and $z + y$ are relatively prime, and they are both squares.*

Proof. Consider a common divisor d of $z - y$ and $z + y$. Then we have

$$d \mid (z + y) - (z - y) = 2y, \qquad \text{and} \qquad d \mid (z + y) + (z - y) = 2z.$$

The integer d is therefore a common divisor of $2y$ and $2z$. Since y and z are relatively prime, it follows that $d \mid 2$.

Furthermore, we note that $d \mid (z - y)(z + y) = z^2 - y^2 = x^2$. Since x is odd, it follows that d must be odd, and therefore we conclude that $d = 1$. This shows that $z + y$ and $z - y$ are relatively prime.

Now we note again that $(z - y)(z + y) = x^2$. Since $z - y$ and $z + y$ are both factors of x^2 and they are relatively prime, it follows that they must both be squares. □

Theorem 7.2.7 (The Pythagorean Triples Theorem)**.** *For every primitive Pythagorean triple (x, y, z) in which y is even there is a unique pair (s, t) of relatively prime odd integers with $1 \leq t < s$, such that*

$$x = st, \qquad y = \frac{s^2 - t^2}{2}, \qquad \textit{and} \qquad c = \frac{s^2 + t^2}{2}.$$

Proof. First, we show that for any two odd integers $0 \leq t < s$ that are relatively prime, the integers

$$x = st, \qquad y = \frac{s^2 - t^2}{2}, \qquad \text{and} \qquad z = \frac{s^2 + t^2}{2}.$$

form a primitive Pythagorean triple. It is easy to see that they form a Pythagorean triple. We simply calculate

$$s^2t^2 + \frac{(s^2 - t^2)^2}{4} = \frac{s^4 + 2s^2t^2 + t^4}{4} = \frac{(s^2 + t^2)^2}{4}.$$

To see that st and $(s^2 - t^2)/2$ are relatively prime, consider a common prime divisor p of st and $(s^2 - t^2)/2$. Since $p \mid st$, it follows that either $p \mid s$ or $p \mid t$. Since $p \mid s^2 - t^2$ and p divides one of the integers s and t, it follows that p also divides the other one. However, s and t have no common prime factors, so we have reached a contradiction. Thus we conclude that there are no common prime factors of st and $(s^2 - t^2)/2$, and hence they must be relatively prime. This proves the claim that (x, y, z) is a primitive Pythagorean triple.

Now consider a primitive Pythagorean triple (x, y, z) in which y is even. By Lemma 7.2.6 it follows that $z - y$ and $z + y$ are relatively prime, and are both squares. Thus there are nonnegative integers s and t with $t < s$ such that

$$s^2 = z + y, \qquad \text{and} \qquad t^2 = z - y.$$

Since z is odd and y is even, it follows that both s and t are odd. Furthermore, since $z + y$ and $z - y$ are relatively prime, it follows that s and t are relatively prime.

Then we see by a direct calculation that

$$x = st, \qquad y = \frac{s^2 - t^2}{2}, \qquad \text{and} \qquad z = \frac{s^2 + t^2}{2}. \qquad \square$$

7.3 Rational Points on the Unit Circle

7.4 The Tree of Primitive Pythagorean Triples

The set of all primitive Pythagorean triples has the structure of a rooted ternary tree. This structure was first discovered by Berggren in 1934 [Ber34].

$$A = \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & 2 \\ 2 & -2 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix}, \qquad \text{and} \qquad C = \begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & 2 \\ -2 & 2 & 3 \end{pmatrix}$$

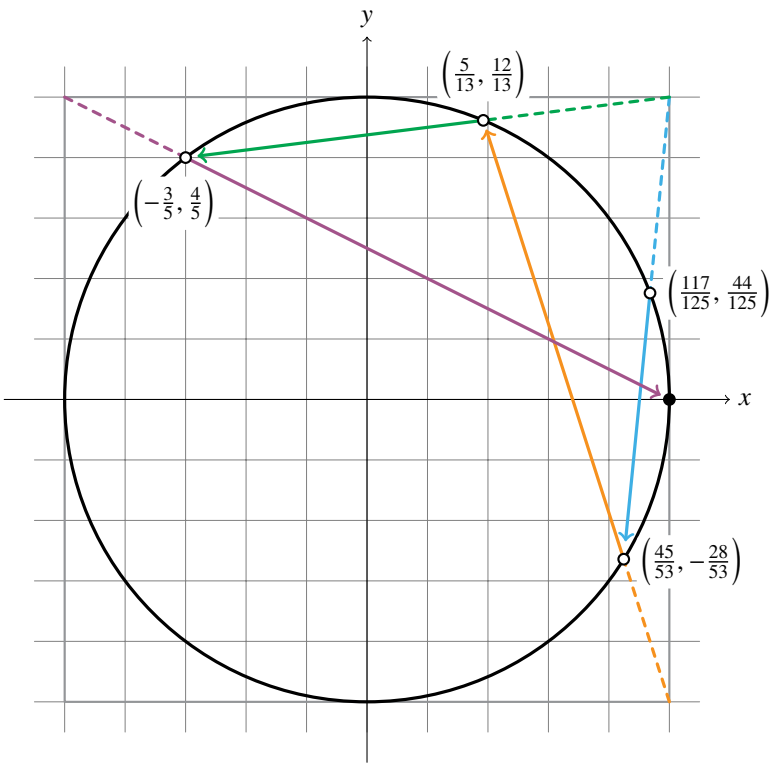
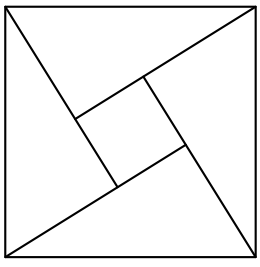


Figure 7.3: For any rational point (x, y) on the unit circle apart from $(0, \pm 1)$ and $(\pm 1, 0)$, draw a straight line through it from the outer corner of its quadrant to find a new rational point (x', y') on the unit circle. The common denominator of (x', y') is strictly smaller than that of (x, y), so that this process eventually terminates at $(0, \pm 1)$ or $(\pm 1, 0)$.

Exercises

7.1 *Bhaskara's proof of the Pythagorean Theorem.* Use the following diagram to give a proof of the Pythagorean Theorem:



7.2

7.3 *The Brahmagupta–Fibonacci identity.* Show that

$$(a^2 + b^2)(c^2 + d^2) = (ac \pm bd)^2 + (ad \mp bc)^2.$$

Conclude that if m and n are two integers that can be written as the sum of two squares, then their product mn can also be written as the sum of two squares.

Chapter 8

Infinite Descent

8.1 The Method of Infinite Descent

The method of infinite descent was invented by Fermat. He used this method effectively to solve various problems, including that there are no solutions to $x^4 + y^4 = z^4$ in the positive integers, and that the area of a right triangle is not a perfect square. The method of infinite descent is used to show that certain properties or identities are impossible, by showing that if they do hold then an infinitely descending sequence of natural numbers can be constructed. The following theorem shows that there is no such sequence.

Proposition 8.1.1. *There is no infinite descending sequence*

$$a_0 > a_1 > a_2 > \cdots$$

of natural numbers.

Proof. Suppose, by way of contradiction, that there is such a sequence

$$a_0 > a_1 > a_2 > \cdots .$$

Then the set

$$A := \{a_0, a_1, a_2, \ldots\}$$

is a nonempty subset of the natural numbers, and by the well-ordering principle of the natural numbers it follows that this set has a least element. However, all the elements in A are of the form a_n for some natural number $\mathbb{N}$, and none of these elements are minimal, since we have assumed that $a_n > a_{n+1}$. This contradiction shows that our assumption that there exists an infinite descending sequence must be false, and thus there is no such sequence. □

The previous proposition leads to Fermat's method of infinite descent. Fermat observed that if an assumption H leads to an infinite descending sequence, then the assumption H must be false. We will illustrate Fermat's method with the following theorem:

Theorem 8.1.2. *If a natural number n is not a perfect square, then its square root $\sqrt{n}$ is irrational.*

Proof. We will prove the theorem by deriving a contradiction under the assumption that there is a rational number $\frac{a}{b}$ such that

$$\frac{a^2}{b^2} = n.$$

This equation can be written in the form $a^2 = nb^2$, where n is not a perfect square. In other words, our goal is to show that the set

$$A := \{(a, b) \mid a^2 = nb^2\}$$

is empty. We will do so by constructing a function $(a, b) \mapsto (a', b')$ on the set A, so that $a > a'$. If the set A was nonempty, then we would obtain an infinite descending sequence of natural numbers by iterating this function, thus reaching a contradiction by Proposition 8.1.1.

Consider a and b such that $a^2 = nb^2$. Since n is not a perfect square, it has a prime divisor p such that

$$n = p^{2k+1}m$$

for an odd number $2k + 1$ and a number m not divisible by p. It follows that

$$p^{2k+1} \mid a^2.$$

By the pigeonhole principle it follows that $p^{k+1} \mid a$, which also implies that a is divisible by p. However, this implies the slightly stronger divisibility relation

$$p^{2(k+1)} \mid nb^2.$$

However, n is not divisible by $p^{2(k+1)}$ so it follows that b is divisible by p. Thus, we have shown that both a and b are divisible by p. We may therefore define

$$a' := \frac{a}{p} \qquad \text{and} \qquad b' := \frac{b}{p}.$$

Since multiplication by p^2 is injective we find that $a'^2 = nb'^2$, completing the proof. □

In the previous proof we introduced a set A of which the elements were pairs (a, b) of natural numbers satisfying $a^2 = nb^2$, and we defined a function $f : A \to A$ in order to show that A is empty. In the following theorem we state the method of infinite descent using a generalization of this setup.

Theorem 8.1.3 (Method of Infinite Descent)**.** *Consider a set A equipped with a function $f : A \to A$ and a function $h : A \to \mathbb{N}$. If $h(f(a)) < h(a)$ for all a, then A is empty.*

Proof. Suppose, by way of contradiction, that the set A is inhabited by an element $a \in A$. By iterating f, we define a sequence of elements a_i by

$$a_0 := a$$
$$a_{i+1} := f(a_i).$$

Then we have $h(a_0) > h(a_i) > h(a_2) > \cdots$, which is an infinite descending sequence of natural numbers. Thus, the assumption that A is inhabited was false. □

8.2 The Unsolvability of $x^4 + y^4 = z^4$

Theorem 8.2.1. *There are no three positive integers x, y, and z for which the equation*

$$x^4 + y^4 = z^2$$

holds.

Proof. Consider the set

$$A := \{(x, y, z) \in \mathbb{Z}_{>0} \mid x^4 + y^4 = z^2\}$$

with $h(x, y, z) := z$. We will construct a function $f : A \to A$ such that $h(f(x, y, z)) < h(x, y, z)$ for every $(x, y, z) \in A$.

First, we notice that if (x, y, z) is a triple of positive integers satisfying the equation $x^4 + y^4 = z^2$, then (x^2, y^2, z) is a Pythagorean triple. We may assume that x and y are relatively prime, since we can always divide out any common factors. Furthermore, we may assume that x is even, so that

$$\begin{aligned} x^2 &= 2st \\ y^2 &= s^2 - t^2 \\ z &= s^2 + t^2, \end{aligned}$$

where s and t are relatively prime.

The key observation is now in the second equation: We obtain a new Pythagorean triple $y^2 + t^2 = s^2$, allowing us to apply the parametrization of Pythagorean triples once more. Since y was assumed to be odd it follows that t is even, so we find relatively prime integers u and v such that

$$\begin{aligned} t &= 2uv \\ y &= u^2 - v^2 \\ s &= u^2 + v^2. \end{aligned}$$

Now it follows that $x^2 = 4uv(u^2 + v^2)$. We claim that the numbers uv and $u^2 + v^2$ are relatively prime. To see this, note that any prime divisor p of uv must be a prime divisor of u or v. On the other hand, given that p is a prime divisor of one of u and v, if p were also a prime divisor of $u^2 + v^2$, then it would follow that p is a prime divisor of both u and v. This contradicts the assumption that u and v are relatively prime.

Since uv and $u^2 + v^2$ are relatively prime, the fact that $4uv(u^2 + v^2)$ is a square implies that both uv and $u^2 + v^2$ are squares. Given that u and v are relatively prime, this implies that u and v are squares, i.e., that there are integers a and b such that $u = a^2$ and $v = b^2$. Thus, it follows that

$$u^2 + v^2 = a^4 + b^4$$

is a square, say c^2. We have therefore shown that any triple of positive integers x, y, and z such that $x^4 + y^4 = z^2$ we can construct positive integers a, b, and c such that

$$a^4 + b^4 = c^2$$

This finishes the construction of the function $f : A \to A$. Furthermore, the strict inequality $c^2 = s < s^2 + t^2 = z^2$ shows that $h(f(x, y, z)) < h(x, y, z)$, so we conclude by the method of infinite descent that the set A must be empty. □

Corollary 8.2.2. *There is no solution to the equation*

$$x^{4n} + y^{4n} = z^{4n},$$

where x, y, and z are positive integers.

Proof. Given such a solution, set $a = x^n$, $b = y^n$, and $c = z^{2n}$. Then we have

$$a^4 + b^4 = c^2,$$

which is impossible. □

8.3 Vieta Jumping

Consider a polynomial with roots a_1, a_2, and a_3, i.e., the polynomial

$$(x - a_1)(x - a_2)(x - a_3).$$

If we expand this polynomial, it takes the form

$$x^3 - (a_1 + a_2 + a_3)x^2 + (a_1a_2 + a_2a_3 + a_3a_1)x - a_1a_2a_3.$$

François Viète recognized in the 16th century that this yields a relation between the coefficients and the roots of a monic polynomial, i.e., a polynomial with leading coefficient 1. In the case of a monic polynomial $x^3 + c_2x^2 + c_1x + c_0$ with roots a_1, a_2, and a_3, we get

$$\begin{aligned} c_2 &= -(a_1 + a_2 + a_3) \\ c_1 &= a_1a_2 + a_1a_3 + a_2a_3 \\ c_0 &= -a_1a_2a_3. \end{aligned}$$

More generally, the coefficients of the polynomial

$$f(x) = (x - a_1) \cdots (x - a_n)$$

are of the form

$$e_k(a_1, \ldots, a_n) := \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} a_{i_1} a_{i_2} \cdots a_{i_k}.$$

Notice that these coefficients are themselves polynomials in the variables $a_1, \ldots, a_n$. These polynomials are called the *elementary symmetric polynomials*. For example, for $n = 3$ we have

$$f(x) = x^3 - (a_1 + a_2 + a_3)x^2 + (a_1a_2 + a_1a_3 + a_2a_3)x - a_1a_2a_3.$$

Vieta jumping is a modern problem-solving technique that became popular in international contest problems.

Exercises

8.1 Use the method of infinite descent to show that if a and b are relatively prime and the product ab is a square, then both a and b are squares.

8.2 Let x and y be positive integers satisfying the equation

$$x^2 - xy + y^2 = 1.$$

Prove that $x = y = 1$.

8.3 Show that if x and y are positive integers such that xy divides $x^2 + y^2 + 1$, then

$$\frac{x^2 + y^2 + 1}{xy} = 3.$$

8.4 Let x, y, and z be positive integers such that $xy + yz + zx$ divides $x^2 + y^2 + z^2$. Prove that

$$\frac{x^2 + y^2 + z^2}{xy + yz + zx}$$

is a perfect square.

Chapter 9

Congruences

9.1 The Congruence Relations

Carl Friedrich Gauss introduced the congruence relations in his monumental work, *Disquisitiones Arithmeticae* [Gau86]. The congruence relations allowed him to systematically simplify the study of the integers. Prior to Gauss's invention, number theory was pursued with a variety of ad hoc techniques that often made proofs less transparent than they could be. His key insight was that many properties of the integers depend not on their exact size, but rather on their remainder when divided by a fixed number.

Definition 9.1.1. We say that a is *congruent to b modulo c*, written

$$a \equiv b \pmod{c},$$

if $c \mid a - b$. The number c is called the *modulus* of the congruence.

Example 9.1.2. We have

$$17 \equiv 2 \pmod{3}$$

because $17 - 2 = 15$ and 15 is divisible by 3. Similarly, we have

$$28 \equiv 3 \pmod{5}$$

because $28 - 3 = 25$ and 25 is divisible by 5.

Example 9.1.3. The 24-hour clock is widely used in Europe, Asia, and Africa, while in North and South America, the 12-hour clock is more common. For instance, what is written as 19:05 in the 24-hour format corresponds to 7:05pm in the 12-hour format. The numbers 19 and 7 are congruent modulo 12 because

$$19 \equiv 7 \pmod{12}.$$

This follows from the fact that $19 - 7 = 12$ is a multiple of 12.

Remark 9.1.4. Congruence modulo 0 is equality, because $0 \mid b - a$ holds if and only if $b - a = 0$. Congruence modulo 1, on the other hand, is true for any two integers a and b, because $1 \mid b - a$ is always true.

Also, we note that $a \equiv b \pmod{c}$ holds if and only if $a \equiv b \pmod{-c}$ holds. For this reason, the modulus is usually taken to be a natural number n.

In the following lemma we establish that a and b are congruent modulo n if and only if they have the same remainder after division by n.

Lemma 9.1.5. *Consider integers a and b, and a positive natural number n, and suppose that $a = qn + r$ and $b = pn + s$, where $0 \leq r, s < n$. Then the congruence*

$$a \equiv b \pmod{n}$$

holds if and only if $r = s$.

Proof. By definition, the congruence $a \equiv b \pmod{n}$ holds if and only if $n \mid a - b$. Since $a = qn + r$ and $b = pn + s$, where $0 \leq r, s < n$ we see that $n \mid a - b$ holds if and only if $n \mid (qn + r) - (pn + s)$. Rewriting the expressing $(qn + r) - (pn + s)$, we see that

$$n \mid (qn + r) - (pn + s) \qquad \text{if and only if} \qquad n \mid (q - p)n + (r - s).$$

Now we note that the number $(q-p)n$ is divisible by n, so we find that $n \mid (q-p)n+(r-s)$ holds if and only if $n \mid r - s$. However, the number $r - s$ satisfies the strict inequalities

$$-n < r - s < n,$$

and the only integer in this range that is divisible by n is the integer 0. Thus we see that $n \mid r - s$ holds if and only if $r = s$, completing the chain of logical equivalences. □

Example 9.1.6. Recall from Theorem 3.3.1 that any number n can be written as

$$n = \sum_{i=0}^{l-1} d_i b^i$$

in base $b > 1$. Given that n is written in this way, we have $n \equiv d_0 \pmod{b}$. Indeed, the digit d_0 was constructed using the Euclidean Division Theorem as the remainder of n after division by b. For example, the congruence $37 \equiv 7 \pmod{10}$ holds.

An important property of the congruence relations, which makes them so useful in the study of the integers, is that they are compatible with the arithmetic operations of addition and multiplication.

Proposition 9.1.7. *Suppose that $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$. Then the following congruences hold:*

$$a + b \equiv a' + b' \pmod{n},$$
$$ab \equiv a'b' \pmod{n}.$$

Consequently, for any integers x and y,

$$ax + by \equiv a'x + b'y \pmod{n}.$$

Proof. By assumption we have that $n \mid a - a'$ and $n \mid b - b'$. Thus it follows that n divides the sum

$$n \mid (a - a') + (b - b') = (a + b) - (a' + b'),$$

which shows that $a + b \equiv a' + b' \pmod{n}$. Moreover, n divides any linear combination of $a - a'$ and $b - b'$, so in particular we have

$$n \mid (a - a')b + a'(b - b') = ab - a'b + a'b - a'b' = ab - a'b'.$$

This shows that $ab \equiv a'b' \pmod{n}$. □

One way of thinking about the following corollary, is that every polynomial is periodic modulo n, i.e., every polynomial repeats itself modulo n after every n steps.

Corollary 9.1.8. *If $x \equiv y \pmod{n}$, then we have*

$$\sum_{k=0}^{l-1} a_k x^k \equiv \sum_{k=0}^{l-1} a_k y^k \pmod{n}.$$

The power of modular arithmetic is effectively demonstrated with divisibility tests. Consider a number n written in base 10 as

$$n = \sum_{k=0}^{l-1} d_k 10^k.$$

Then n is divisible by 9 if and only if the sum of its digits is divisible by 9. Indeed, since $10 \equiv 1 \pmod{9}$, it follows that

$$\sum_{k=0}^{l-1} d_k 10^k \equiv \sum_{k=0}^{l-1} d_k 1^k \pmod{9},$$

and we recognize that the right-hand side is just the sum of the digits of n. For example, the number 34, 524 is divisible by 9 because $3 + 4 + 5 + 2 + 4 = 18$ and the number 18 is divisible by 9 because $1 + 8 = 9$.

Remark 9.1.9. One might wonder whether congruence relations are also preserved by operations like the greatest common divisor. However, this is not the case. For a counter example, consider the congruences $6 \equiv 16 \pmod{10}$ and $15 \equiv 25 \pmod{10}$. The greatest common divisor of 6 and 15 is 3, while the greatest common divisor of 16 and 25 is 1. Thus we see that, even though $6 \equiv 16$ and $15 \equiv 25$ modulo 10, we have

$$\gcd(6, 15) \not\equiv \gcd(16, 25) \pmod{10}.$$

9.2 Equivalence Relations

The congruence relations are examples of equivalence relations, which we will now define.

Definition 9.2.1. An equivalence relation on a set A is a binary relation $\sim$ satisfying the following three conditions:

1. *Reflexivity.* For any element $a \in A$, we have
$$a \sim a.$$
2. *Symmetry.* For any two elements $a, b \in A$, we have
$$(a \sim b) \Rightarrow (b \sim a).$$
3. *Transitivity.* For any three elements $a, b, c \in A$, we have
$$(a \sim b) \wedge (b \sim c) \Rightarrow (a \sim c).$$

Thus, where the ordering relation of a poset is reflexive, *antisymmetric* and transitive, an equivalence relation is reflexive, *symmetric* and transitive. The standard example of an equivalence relation is equality itself, which is indeed reflexive, symmetric and transitive.

Proposition 9.2.2. *For any natural number n, the congruence relation* $\equiv$ *modulo n is an equivalence relation.*

Proof. To see that the congruence relation modulo n is reflexive, note that $a - a = 0$ and any number divides 0. Thus the congruence $a \equiv a \pmod{n}$ always holds. To see that the congruence relation modulo n is symmetric, note that if $n \mid a - b$ then $n \mid -(a - b) = b - a$. This shows that $a \equiv b \pmod{n}$ implies $b \equiv a \pmod{n}$. To see that the congruence relation modulo n is transitive, assume that $n \mid a - b$ and that $n \mid b - c$. Then it follows that
$$n \mid (a - b) + (b - c) = a - c,$$
showing that $a \equiv c \pmod{n}$. □

Equivalence relations come about in many forms, and one can get quite creative in defining them. We illustrate the generality of equivalence relations with some examples.

Example 9.2.3. Consider a prime number p. Every nonzero integer a can be written in the form $p^k u$ where u is not divisible by p. The number u is called the *p-core* or the *p-free part* of a. The relation $a \sim b$ defined by the condition that a and b have the same p-core is an equivalence relation.

The equivalence relation of the previous example is an instance of a general class of equivalence relations. Given a function $f : A \to B$, there is an equivalence relation $\sim_f$ on the set A, given by
$$x \sim_f y \qquad \text{if and only if} \qquad f(x) = f(y).$$
This relation is also called the *kernel* of f. In the previous example, the function $f : \mathbb{Z}_{\neq 0} \to \mathbb{Z}_{\neq 0}$ is the function that sends a to the unique integer u not divisible by p for which there is a *valuation* k such that $p^k u = a$.

Proposition 9.2.4. *For any function $f : A \to B$, the relation $\sim_f$ is an equivalence relation.*

Proof. To see that any relation of the form $\sim_f$ is an equivalence relation, we need to check reflexivity, symmetry, and transitivity. Note that we have reflexivity because $f(x) = f(x)$ for any $x \in A$. The relation $\sim_f$ is symmetric because $f(y) = f(x)$ holds whenever $f(x) = f(y)$ holds. Finally, the relation $\sim_f$ is transitive because $f(x) = f(z)$ holds whenever both $f(x) = f(y)$ and $f(y) = f(z)$ hold. □

Example 9.2.5. Recall from Exercise 3.2 that Zeckendorf's representation theorem established that any natural number n can be written uniquely as a sum

$$n = \sum_{k=0}^{l_n - 1} d_{n,k} F_{k+2}$$

of nonconsecutive Fibonacci numbers with index at least 2. Now we define an equivalence relation $\sim_F$ by declaring that two natural numbers m and n are equivalent, if one can be obtained from the other by shifting their Zeckendorf representation. More precisely, we define $m \sim_F n$ to hold if and only if there exist natural numbers s and t such that

$$\sum_{k=0}^{l_m - 1} d_{m,k} F_{k+s+2} = \sum_{k=0}^{l_n - 1} d_{n,k} F_{k+t+2}.$$

For example, the number 17 is written $(100101)_F$ in the Zeckendorf notation, and its shifts are the numbers $(1001010)_F = 28$, $(10010100)_F = 45$, and so on. Thus we have

$$17 \sim_F 28 \sim_F 45 \sim_F \cdots.$$

The equivalence relation of the previous example is an instance of another general way of constructing equivalence relations. Given an operation $T : A \to A$, define the equivalence relation $\sim_T$ by declaring that

$$x \sim_T y \qquad \text{if and only if} \qquad \exists_{(s,t \in \mathbb{N})} T^s(x) = T^t(y).$$

In technical terminology, the set A equipped with the operation T form a *dynamical system*, and when $x \sim_T y$ holds we say that x and y have the same *eventual orbit*. The operation T in the equivalence relation $\sim_F$ of the previous example is the shift operation on the Zeckendorf representation, which simply appends a digit 0.

Proposition 9.2.6. *For any operation T on a set A, the relation $\sim_T$ is an equivalence relation.*

Proof. To see that the relation $\sim_T$ is indeed an equivalence relation, we need to verify reflexivity, symmetry, and transitivity.

The relation $\sim_T$ is reflexive, because $T^0(x) = x = T^0(x)$, and the relation $\sim_T$ is symmetric because the condition that $\exists_{(s,t \in \mathbb{N})}(T^s(x) = T^t(y))$ is symmetric in x and y. To see that $\sim_T$ is transitive, consider s, t, u, v such that $T^s(x) = T^t(y)$ and $T^u(y) = T^v(z)$. Then we have

$$T^{u+s}(x) = T^u(T^s(x)) = T^u(T^t(y)) = T^t(T^u(y)) = T^t(T^v(z)) = T^{t+v}(z),$$

showing that $x \sim_T z$. □

Perhaps one of the most famous dynamical systems on the set of positive integers is due to Collatz, whose function T is defined by

$$T(n) := \begin{cases} n/2 & \text{if } n \text{ is even} \\ 3n+1 & \text{otherwise.} \end{cases}$$

For example, starting at the number 3, repeated application of the Collatz function produces the sequence:

$$3,\ 10,\ 5,\ 16,\ 8,\ 4,\ 2,\ 1,\ 4, \cdots .$$

Once the sequence reaches 1, it enters the cycle

$$1 \mapsto 4 \mapsto 2 \mapsto 1 \mapsto \cdots$$

The famous and elusive *Collatz conjecture* states that for every positive integer n, the sequence

$$n,\ T(n),\ T^2(n),\ \ldots$$

eventually reaches the value 1. In our notation, the Collatz conjecture asserts that $n \sim_T 1$ for every positive natural number n.

9.3 Equivalence Classes and Residue Systems

Definition 9.3.1. Consider an equivalence relation $\sim$ on a set A. Then the *equivalence class* of an element $a \in A$ is the set

$$[a] := \{x \in A \mid a \sim x\}.$$

Example 9.3.2. In the case where the equivalence relation under consideration is a congruence relation with modulus n, then the equivalence classes are called *congruence classes*. For any positive integer n, there are exactly n distinct congruence classes modulo n, corresponding to the n possible residues $0 \leq r < n$ after division by n with remainder.

The congruence class modulo 2 of any even integer consists precisely of all the even integers, and the congruence class of any odd integer consists precisely of all the odd integers.

To see this, suppose that a is an even integer. Then $a \equiv x \pmod 2$ holds if and only if a and x have the same remainder after division by 2. Since the remainder of a divided by 2 is 0, it follows that a is congruent to x precisely when x is divisible by 2, i.e., when x is even. This shows that for an even integer a, the congruence class

$$[a] := \{x \in A \mid a \equiv x \pmod 2\}$$

is the set of even numbers.

Similarly, if a is an odd integer, then $a \equiv x \pmod 2$ holds if and only if x is also odd, because $a \equiv x \pmod 2$ holds if and only if x divided by 2 has remainder 1, which means that x is odd.

The equivalence classes of an equivalence relation $\sim$ on a set A form a set

$$A/\sim := \{U \subseteq A \mid \exists_{(a \in A)} U = [a]\}$$

In other words, the set of equivalence classes of $\sim$ is the set of subsets $U \subseteq A$ such that $U = [a]$ for some $a \in A$. The set $A/\sim$ is called the *quotient* of A by the equivalence relation $\sim$. Furthermore, the function

$$a \mapsto [a] : A \to A/\sim$$

sending an element to its equivalence class is called the *quotient map*. We often write q for the quotient map.

Proposition 9.3.3. *For any equivalence relation $\sim$ on a set A, the quotient map*

$$a \mapsto [a] : A \to A/\sim$$

satisfies the following two conditions:

1. *The quotient map is* surjective*. This means that for any equivalence class $U \subseteq A$ there is an element $a \in A$ such that $U = [a]$.*
2. *The quotient map is* effective*. This means that for any two elements $a, b \in A$ we have*
$$[a] = [b] \qquad \textit{if and only if} \qquad a \sim b.$$

Proof. The first claim is true by definition: the set of equivalence classes is defined as

$$\{U \subseteq A \mid \exists_{(a \in A)} U = [a]\},$$

i.e., it is defined as the set of subsets of A of the form $[a]$ for some $a \in A$.

For the second claim, let $a, b \in A$. If we have the equality $[a] = [b]$, then it follows that $b \in [a]$, so that $a \sim b$. For the converse, if $a \sim b$ holds, and $x \in A$, then we have $a \sim x$ if and only if $b \sim x$ by symmetry and transitivity of the relation $\sim$. This shows that $x \in [a]$ if and only if $x \in [b]$. Thus, the subsets $[a]$ and $[b]$ contain the same elements, so they must be the same. □

Definition 9.3.4. A *complete residue system* modulo n is a choice of exactly one element from each congruence class modulo n. In other words, a complete residue system is a set

$$\{r_1, \ldots, r_n\}$$

satisfying the condition that for every integer a there is a exactly one index $1 \leq i \leq n$ such that

$$a \equiv r_i \pmod{n}.$$

Example 9.3.5. For every nonnegative integer n, the set

$$\{0, \ldots, n-1\}$$

forms a complete residue system modulo n. Similarly, any set

$$\{a, \ldots, a+n-1\}$$

of n consecutive integers is a complete residue system modulo n.

Remark 9.3.6. Since complete residue systems contain a unique element in every congruence class, there is for every complete residue system $\{r_0, \ldots, r_{n-1}\}$ a bijection

$$\mathbb{N}/(\equiv \pmod{n}) \cong \{r_0, \ldots, r_{n-1}\}.$$

9.4 Reduced Residue Systems

Before we introduce reduced residue systems, let us prove a lemma.

Lemma 9.4.1. *Suppose that* $a \equiv b \pmod{n}$. *Then* $\gcd(a, n) = \gcd(b, n)$. *Consequently, we have* $\gcd(a, n) = 1$ *if and only if* $\gcd(b, n) = 1$.

Proof. By the assumption that a and b are congruent modulo n, it follows that they have the same remainder after division by n. Let r be this remainder. Since $\gcd(a, n) = \gcd(r, n)$ and $\gcd(b, n) = \gcd(r, n)$ we see that

$$\gcd(a, n) = \gcd(b, n).$$

This equality implies that the left hand side equals 1 if and only if the right hand side equals 1. □

By the previous lemma, we can test whether an integer is relatively prime to n by testing any element in its congruence class. This allows us to make the following definition.

Definition 9.4.2. A *reduced residue system* modulo n is a set

$$\{r_1, \ldots, r_k\}$$

of integers satisfying the condition that for every integer a relatively prime to n, there is exactly one index $1 \leq i \leq k$ such that

$$a \equiv r_i \pmod{n}.$$

Corollary 9.4.3. *Any two reduced residue systems modulo n have the same size, for which we write* $\phi(n)$. *The function* ϕ *is called* Euler's totient function.

Proof. Reduced residue systems are choices of representatives of the congruence classes in which all elements are relatively prime to n. Thus, the size of a reduced residue system is always equal to the number of such congruence classes. □

Exercises

9.1 Describe a complete residue system modulo 7 consisting entirely of multiples of 3.

9.2 List all the numbers $0 \leq a < 6$ such that $a \equiv x^2 \pmod{6}$ for some x, and all the numbers $0 \leq a < 8$ such that $a \equiv x^2 \pmod{8}$ for some x.

9.3 Compute $\phi(n)$ for $n = 1, \ldots, 12$.

9.4 (a) Find integers k, a, and b such that

$$ka \equiv kb \pmod 4, \qquad \text{but} \qquad a \not\equiv b \pmod 4.$$

(b) Prove that if $\gcd(k, n) = d$, then we have

$$ka \equiv kb \pmod n \qquad \text{if and only if} \qquad a \equiv b \pmod{\frac{n}{d}}.$$

9.5 Show that the following are equivalent for any $n > 0$ and any integer a:

1. $\gcd(a, n) = 1$.
2. There exists an integer b such that $ab \equiv 1 \pmod n$.

9.6 Consider a prime number p and an integer a. Show that

$$a^2 \equiv 1 \pmod p \qquad \text{if and only if} \qquad a \equiv \pm 1 \pmod p.$$

9.7 Prove that for any $n > 1$ and any k, we have

$$(n-1)^2 \mid n^k - 1 \qquad \text{if and only if} \qquad n - 1 \mid k.$$

9.8 A pair of *twin primes* is a pair of primes p and q such that $q - p = 2$. Show that if p and q are twin primes strictly greater than 3, then $pq \equiv 8 \pmod 9$.

9.9 Given a prime $p > 3$, show that $p^2 \equiv 1 \pmod{24}$.

Chapter 10

Modular Arithmetic

10.1 The Integers Modulo n

We have seen that the congruence relations modulo a natural number n form equivalence relations that are compatible with the arithmetic operations of addition and multiplication. We have also seen that equivalence relations give rise to equivalence classes, and that the set of all equivalence classes for a given equivalence relation is called the *quotient*. Quotients are important mathematical constructions, which are used to simplify mathematical problems and constructions by treating similar or equivalent elements of a set as the same.

Definition 10.1.1. We define the set $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n as the set of congruence classes modulo n.

The set of integers modulo n can be equipped with addition and multiplication in the following way:

$$[a] + [b] := [a + b]$$
$$[a][b] := [ab]$$

Notice that the result of these operations depends only on the equivalence classes, not on the particular integers a and b chosen to represent the equivalence classes $[a]$ and $[b]$. Recall from Proposition 9.3.3 that we have $[a] = [a']$ if and only if $a \equiv a' \pmod{n}$, and similarly we have $[b] = [b']$ if and only if $b \equiv b' \pmod{n}$. Thus, for any $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$, it follows from Proposition 9.1.7 that $a + b \equiv a' + b' \pmod{n}$ and $ab \equiv a'b' \pmod{n}$. In other words, if $[a] = [a']$ and $[b] = [b']$, then we have

$$[a] + [b] = [a'] + [b']$$
$$[a][b] = [a'][b'].$$

This ensures that the definitions of addition and multiplication on the integers modulo n is *well-defined* on the congruence classes modulo n.

Remark 10.1.2. The definitions of addition and multiplication on the integers modulo n follow a general pattern of defining functions $f : A/\sim \to B$ out of quotient sets. Given a function $g : A \to B$, a function $f : A/\sim \to B$ is said to be *defined by descent* from g if f is defined by

$$f([a]) := g(a).$$

For this definition to make sense, the values of f must depend only on the equivalence classes. In other words, if a and a' represent the same equivalence class, i.e., if $a \sim a'$, then we must have that $f([a]) = f([a'])$, i.e., we must have that

$$g(a) = g(a').$$

By verifying that $a \sim a'$ implies $f([a]) = f([a'])$, we ensure that f is *well-defined*. The well-definedness condition ensures that f returns the same output for all elements of the same equivalence class.

Example 10.1.3. The well-definedness condition is important to verify, because not all functions are well-defined with respect to the congruence relations modulo n.

Take, for example, the absolute value function $a \mapsto |a|$ on the integers. Then the operation

$$[a] \mapsto |a|$$

is not well-defined on the integers modulo 3, because $2 \equiv -1 \pmod 3$ while

$$|2| \not\equiv |-1| \pmod 3.$$

Typical functions that *are* well defined on congruence classes modulo n, include functions that are entirely built up from arithmetic operations, such as polynomials.

Theorem 10.1.4. *The set $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n form a commutative ring, i.e., it satisfies the following laws of arithmetic:*

$$\begin{aligned}
(x+y)+z &= x+(y+z) & (xy)z &= x(yz)\\
0+x &= x & 1x &= x\\
x+0 &= x & x1 &= x\\
x-x &= 0 & x(y+z) &= xy+xz\\
-x+x &= 0 & (x+y)z &= xz+yz\\
x+y &= y+x & xy &= yx.
\end{aligned}$$

Proof. To see that $(x + y) + z = x + (y + z)$, we first note that the variables x, y, and z represent congruence classes of integers modulo n. Since for every congruence class there is an integer representing it, it suffices to prove that

$$([a] + [b]) + [c] = [a] + ([b] + [c]).$$

By unfolding the definitions, this equation reduces to

$$[(a + b) + c] = [a + (b + c)].$$

However, the integers $(a + b) + c$ and $a + (b + c)$ are equal, so their equivalence classes are the same.

The proofs of the other properties follow a very similar pattern. We will give the proofs of commutativity of addition and distributivity of multiplication over addition from the left, but leave the other proofs to the reader.

To see that $x + y = y + x$, it suffices to show that

$$[a] + [b] = [b] + [a].$$

By unfolding the definitions, this equation reduces to

$$[a + b] = [b + a],$$

and indeed those equivalence classes are the same, because $a + b = b + a$ holds for any two integers a and b.

To see that $x(y + z) = xy + xz$, it suffices to show that

$$[a]([b] + [c]) = [a][b] + [a][c].$$

By unfolding the definitions, this equation reduces to

$$[a(b + c)] = [ab + ac].$$

Again, this identification of equivalence classes holds, because we have the equality $a(b + c) = ab + ac$. □

10.2 Solving Linear Congruences

The simplest linear congruence is the linear congruence

$$ax \equiv b \pmod{n}.$$

To solve this linear congruence, we are tasked with finding an integer for which the congruence $ax \equiv b \pmod{n}$ holds. However, we may immediately observe that if $x \equiv x' \mod n$, then the congruence $ax \equiv b \pmod{n}$ holds if and only if the congruence $ax' \equiv b \pmod{n}$ holds. Thus, we are really interested in solving such congruences in $\mathbb{Z}/n\mathbb{Z}$, i.e., in finding a congruence class $[x]$ modulo n for which the congruence $ax \equiv b \pmod{n}$ holds.

A complete solution to the linear congruence $ax \equiv b \pmod{n}$ is a description of the set of all congruence classes $[x]$ modulo n that satisfy the linear congruence. We typically list these solutions by listing the numbers $0 \leq x < n$ for which the congruence holds. For example, the linear congruence

$$4x \equiv 2 \pmod{6}$$

has two incongruent solutions modulo 6, the congruence classes of 2 and 5 modulo 6.

Definition 10.2.1. We say that *a divides b* modulo n if there exists an integer x for which the congruence

$$ax \equiv b \pmod{n}.$$

By the definition of congruence relations, the congruence $ax \equiv b \pmod{n}$ holds if and only if $n \mid ax - b$, which is equivalent to the problem of finding an integer y such that the equation

$$ax + ny = b$$

holds. By Theorem 4.6.4 we have an exact description of the set of solutions of this equation: Given a solution $ax_0 + ny_0 = b$, every solution is of the form

$$x = x_0 + k\frac{n}{d}, \qquad \text{and} \qquad y = y_0 - k\frac{a}{d},$$

where $d = \gcd(a, n)$. Furthermore, a solution can be found if and only if $d \mid b$, and in this case we can find x_0 and y_0 through Euclid's algorithm for finding the greatest common divisor.

Now, observe that there are exactly d incongruent solutions of the equation $ax+ny = b$ modulo n. Thus, we can reformulate Theorem 4.6.4 in modular arithmetic as follows.

Theorem 10.2.2. *Consider two integers a and b, and a natural number n with $d = \gcd(a, n)$. The linear congruence*

$$ax \equiv b \pmod{n}$$

is solvable if and only if $d \mid b$, and in this case the number of incongruent solutions is exactly d. Furthermore, if we have one solution $ax_0 \equiv b \pmod{n}$, then all the solutions are of the form

$$x \equiv x_0 + k\frac{n}{d}$$

for $0 \leq k < d$.

Example 10.2.3. Consider the linear congruence

$$6x \equiv 15 \pmod{21}.$$

The greatest common divisor of 6 and 21 is $\gcd(6, 21) = 3$. Since 15 is divisible by 3, we expect to find exactly 3 incongruent solutions modulo 21.

To find one solution, we first express 3 as a linear combination of 6 and 21. Normally, we would use Euclid's algorithm to do this, but the numbers here are small enough to immediately see that

$$3 = 4 \cdot 6 - 21.$$

From this expression we find that

$$15 = 20 \cdot 6 - 5 \cdot 21.$$

This gives us the solution $6 \cdot 20 \equiv 15 \pmod{21}$. The remaining solutions are now of the form

$$x \equiv 20 + k \cdot 7$$

for $0 \leq k < 3$. In other words, the full set of incongruent solutions of the equation $6x \equiv 15 \pmod{21}$ is

$$\{6, 13, 20\}.$$

Corollary 10.2.4. *The linear congruence* $ax \equiv 1 \pmod{n}$ *has at most one solution, and it is solvable if and only if* $\gcd(a, n) = 1$.

The previous corollary connects integers relatively prime to n to the integers that are *invertible* modulo n. Indeed, if $ax \equiv 1 \pmod{n}$ has a solution, then its solution x also guarantees that the equality

$$[a][x] = [1]$$

holds in $\mathbb{Z}/n\mathbb{Z}$. In other words, the integer a represents an invertible congruence class modulo n.

Example 10.2.5. The invertible congruence classes modulo 5 are the congruence classes of 1, 2, 3, and 4. Indeed, one can verify that

$$1 \cdot 1 \equiv 1 \pmod{5} \qquad 2 \cdot 3 \equiv 1 \pmod{5}$$
$$3 \cdot 2 \equiv 1 \pmod{5} \qquad 4 \cdot 4 \equiv 1 \pmod{5}.$$

Corollary 10.2.6. *Consider an integer* a *relatively prime to* n. *Then the operation* $x \mapsto [a]x$ *is a bijection on* $\mathbb{Z}/n\mathbb{Z}$.

10.3 Fermat's Little Theorem

Theorem 10.3.1 (Fermat's Little Theorem)**.** *Consider a prime number* p *and an integer* a. *Then the congruence*

$$a^p \equiv a \pmod{p}$$

holds.

Fermat's proof. Note that the congruence $(-a)^p \equiv -a^p \pmod{p}$ holds for all primes p, so it suffices to prove the claim for the nonnegative integers. We proceed by induction on a. The base case $0^p \equiv 0 \pmod{p}$ clearly holds.

For the inductive step, assume that $a^p \equiv a \pmod{p}$. The Binomial Theorem then gives us that

$$(a+1)^p = \sum_{k=0}^{p} \binom{p}{k} a^k$$

However, note that $p \mid \binom{p}{k}$ for any $0 < k < p$. We therefore find that

$$(a+1)^p \equiv a^p + 1 \pmod{p}.$$

Since $a^p \equiv a \pmod{p}$ holds by the induction hypothesis, the claim follows. □

Corollary 10.3.2. *Consider a prime number* p *and an integer* a *not divisible by* p. *Then the congruence*

$$a^{p-1} \equiv 1 \pmod{p}$$

holds.

Proof. Since a is not divisible by p, it follows that a is relatively prime to p, and therefore the claim follows from the cancellation law

$$ax \equiv ay \pmod{p} \qquad \Rightarrow \qquad x \equiv y \pmod{p}.$$ □

10.4 Euler's Theorem

Euler figured out a way to generalize Fermat's Little Theorem to the composite moduli. He made use of his totient function ϕ, which counts for each n the number of $0 \leq m < n$ relatively prime to n.

Theorem 10.4.1 (Euler's Theorem)**.** *Consider a natural number n and an integer a such that* $\gcd(a, n) = 1$. *The congruence*

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

holds

We present two proofs of Euler's Theorem. The first is the standard proof, and the other is inspired by methods of group theory.

First proof of Euler's Theorem. Consider an integer a relatively prime to n, and consider a reduced residue system $r_1, \ldots, r_{\phi(n)}$ modulo n. We claim that

$$ar_1, \ldots, ar_{\phi(n)}$$

is again a reduced residue system modulo n. Indeed, since multiplying by a induces a bijection on $\mathbb{Z}/n\mathbb{Z}$, we see that all integers $ar_1, \ldots, ar_{\phi(n)}$ are incongruent. Furthermore, the product of any two integers relatively prime to n is again relatively prime to n, so all the integers ar_i are relatively prime to n. Since any set of $\phi(n)$ incongruent integers relatively prime to n form a reduced residue system, the claim follows.

Using that the integers $ar_1, \ldots, ar_{\phi(n)}$ form a reduced residue system, it follows that

$$\prod_{i=1}^{\phi(n)} r_i \equiv \prod_{i=1}^{\phi(n)} ar_i = a^{\phi(n)} \prod_{i=1}^{\phi(n)} r_i \pmod{n}$$

Since $\prod_{i=1}^{\phi(n)} r_i$ is relatively prime to n, it follows that $a^{\phi(n)} \equiv 1 \pmod{n}$. □

Second proof of Euler's Theorem. Consider the least positive integer k such that $a^k \equiv 1 \pmod{n}$, and consider the relation $\sim$ on the set of integers $0 \leq b < n$ relatively prime to n, where

$$b \sim c \qquad \text{if and only if} \qquad \exists_{0 \leq i < k} ba^i \equiv c \pmod{n}.$$

This relation is reflexive, because $ba^0 \equiv b \pmod{n}$. It is symmetric, because if $ba^i \equiv c \pmod{n}$ for some nonzero i, then $ca^{k-i} \equiv ba^i a^{k-i} \equiv b \pmod{n}$. Furthermore, it is transitive, because if $ba^i \equiv c$ and $ca^j \equiv d$, then $ba^{i+j} \equiv d$. If $k \leq i + j$, then we find that also $ba^{i+j-k} \equiv d$. Thus, the relation $\sim$ is an equivalence relation.

We claim that each equivalence class has size k. Indeed, for each b, the set

$$\{b, ba, ba^2, \ldots, ba^{k-1}\}$$

consists of k distinct elements, because multiplying with b is a bijection. Thus we have partitioned the set of all numbers $0 \leq b < n$ relatively prime to n into equivalence classes of size k. On the other hand, the total number of elements $0 \leq b < n$ relatively prime to n is the number $\phi(n)$. This shows that $k \mid \phi(n)$, so it follows that $a^{\phi(n)} \equiv 1 \pmod{n}$. □

Remark 10.4.2. The second proof of Euler's theorem can be simplified using the language of group theory. A *group* is a set G equipped with a binary operation $x, y \mapsto xy$, a unary operation $x \mapsto x^{-1}$, and a unit element 1, satisfying the axioms

$$\begin{aligned}(xy)z &= x(yz)\\ x1 &= x\\ 1x &= x\\ xx^{-1} &= 1\\ x^{-1}x &= 1.\end{aligned}$$

One can show that the set of integers modulo n relatively prime to n forms a group, where the binary operation is multiplication, the unit element is the integer 1, and the inverse x^{-1} of an integer x modulo n is the unique solution to the equation $xy \equiv 1 \pmod{n}$. The number of elements in this group is $\phi(n)$. For any integer a modulo n relatively prime to n such that $a^k \equiv 1 \pmod{n}$, the set $\{1, a, a^2, \ldots, a^{k-1}\}$ is a subgroup, which means that it is closed under the group operations. The argument where we introduced an equivalence relation $\sim$ to partition the set of all integers modulo n relatively prime to n into equivalence classes of size k is an instance of a much more general result, Lagrange's Theorem, which asserts that the size of any subgroup of a finite group divides the size of the entire group.

Since $\phi(p) = p - 1$, it follows that Fermat's Little Theorem is also a direct corollary of Euler's Theorem.

Some questions about Euler's totient function remain unresolved. While we have seen that $\phi(p) = p - 1$ for any prime number p, Lehmer asked whether there is any composite number n such that

$$\phi(n) \mid n - 1.$$

This open problem is now known as *Lehmer's problem* [Leh33].

10.5 Wilson's Theorem

Theorem 10.5.1 (Wilson's Theorem)**.** *A natural number n is prime if and only if*

$$(n-1)! \equiv -1 \pmod{n}.$$

Proof. First, we observe that the claim holds for $n = 2$, since $1 \equiv -1 \pmod{2}$.

Suppose that n is prime strictly greater than 2. Then every integer $0 < a < n$ has a multiplicative inverse modulo n, i.e., for every integer a there is an integer a^{-1} such that

$$aa^{-1} \equiv 1.$$

Now we group the integers $1 \leq a < n$ into pairs with their own inverses. By Exercise 9.6 and the assumption that $n \neq 2$, there are exactly two integers that would pair up with themselves because they are their own inverses: The integers 1 and -1 are the only two integers a modulo n for which the congruence $a^2 \equiv 1 \pmod{n}$ holds. The other integers come in proper pairs (a, a^{-1}).

Each pair (a, a^{-1}) contributes 1 to the product

$$(n-1)! := \prod_{a=1}^{n-1} a.$$

Therefore, there is exactly one factor that contributes a something to this product: the factor -1. Thus we see that $(n-1)! \equiv -1 \pmod{n}$.

Conversely, suppose that $(n-1)! \equiv -1 \pmod{n}$. Then it follows that

$$-(n-1)! \equiv 1 \pmod{n}.$$

For any proper divisor a of n, we therefore find that a divides 1 modulo n, i.e., that there is a solution to the linear congruence

$$ax \equiv 1 \pmod{n}.$$

This implies that $\gcd(a, n) = 1$. It follows that the only proper divisor of n is 1, i.e., that n is prime. □

Exercises

10.1 Find the multiplicative inverses of 1, 2, 3, 4, 5, and 6 modulo 7.

10.2 Find the multiplicative inverses of 1, 3, 7, 9, 11, 13, 17, and 19 modulo 20.

10.3 For how many congruence classes $b \pmod{n}$ is the linear congruence

$$ax \equiv b \pmod{n}$$

solvable?

10.4 Consider two integers a and b, and a prime number p. Show that

$$(a+b)^p \equiv a^p + b^p \pmod{p}.$$

10.5 Show that

$$2^{p-2} + 3^{p-2} + 6^{p-2} \equiv 1 \pmod{p}$$

for any prime number $p \neq 3$.

10.6 (a) Prove that

$$5^{5^5} \equiv 5^5 \pmod{31}.$$

(b) Prove that

$$7^{7^7} \equiv 7^7 \pmod{43}.$$

(c) Find a prime $p > 3$ such that

$$3^{3^3} \equiv 3^3 \pmod{p}.$$

10.7 Prove that if

$$a^n \equiv a \pmod{n}$$

for every integer a, then $n = 2$ or n is odd.

10.8 Find all the solutions to the equation

$$p^q - q^p = p + q,$$

where p and q are prime numbers.

10.9 For any integer a and any prime number p, show that

$$a^{p^k} \equiv a^{p^{k-1}} \pmod{p^k}.$$

10.10 Consider two distinct primes p and q, and an integer $a \geq 0$. Prove that

$$a^{pq} + a \equiv a^p + a^q \pmod{pq}$$

Chapter 11

Systems of Linear Congruences

11.1 Solving Multiple Linear Congruences Simultaneously

Consider the linear congruences

$$\begin{aligned} 10x &\equiv 4 \pmod{12} \\ 15x &\equiv 6 \pmod{21}, \end{aligned}$$

and suppose our goal is to find a single solution x that simultaneously solves both of them. To start off, we observe that if we can find such a solution, then each individual congruence must be solvable in its own right. By Theorem 10.2.2, this is the case if and only if $\gcd(10, 12) \mid 4$ and $\gcd(15, 21) \mid 6$. Indeed, $\gcd(10, 12) = 2$ and $\gcd(15, 21) = 3$, so the divisibility requirements are satisfied.

By dividing through with $\gcd(10, 12)$, we see that the linear congruence $10x \equiv 4 \pmod{12}$ has the same set of solutions as the linear congruence $5x \equiv 2 \pmod{6}$, which we can work out to be $x \equiv 4 \pmod{6}$. Similarly the linear congruence $15x \equiv 6 \pmod{21}$ has the same set of solutions as the linear congruence $5x \equiv 2 \pmod{7}$, which we can work out to be $x \equiv 6 \pmod{7}$. Thus, the original system of linear congruences has the same set of solutions as the system of linear congruences

$$\begin{aligned} x &\equiv 4 \pmod{6} \\ x &\equiv 6 \pmod{7}. \end{aligned}$$

In other words, x is simultaneously of the form $6y + 4$ and of the form $7z + 6$. We can find such y and z by solving the linear Diophantine equation

$$6y + 4 = 7z + 6,$$

which reduces to $6y - 7z = 2$. By Theorem 4.6.1, it follows that this equation solves if and only if $\gcd(6, 7) \mid 2$, which is indeed the case since 6 and 7 are relatively prime. Using the extended Euclid's algorithm, we find that $y = 5$ and $z = 4$ gives a solution. In

this case, we have $x = 34$. Note that any integer in the congruence class of 34 modulo 42 is also a solution, so we see that

$$x \equiv 34 \pmod{42}$$

is a solution to our original system of linear congruences. Indeed, we can check that $10 \cdot 34 = 28 \cdot 12 + 4$ and $15 \cdot 34 = 24 \cdot 21 + 6$.

To summarize the method by which we found this solution, we first reduced each individual linear congruence by dividing through by the greatest common divisor of the scalar and the modulus. This way we obtained a system of linear congruences that were individually uniquely solvable. Their solutions gave two expressions for the variable x, which combined into a single linear Diophantine equation, which we solved by Theorem 4.6.1 and Euclid's algorithm. As the saying goes, any good method in mathematics begs to become a theorem. The essential theorem to solve systems of linear congruences is the Chinese Remainder Theorem, which we will state and prove in the next section.

We also note that the methods we used to solve two linear congruences simultaneously can be used to solve three or more linear congruences simultaneously. Suppose our original problem included a third linear congruence

$$\begin{aligned} 10x &\equiv 4 \pmod{12} \\ 15x &\equiv 6 \pmod{21} \\ 6x &\equiv 3 \pmod{15}\,. \end{aligned}$$

Then we'd simply solve the first two as before, and continue solving the system of two linear congruences

$$\begin{aligned} x &\equiv 34 \pmod{42} \\ x &\equiv 3 \pmod{5}\,. \end{aligned}$$

This system of linear congruences has solution $x \equiv 118 \pmod{210}$.

11.2 The Chinese Remainder Theorem

An essential step in the previous example was the reduction of a system of linear congruences, to a system of congruences of the form

$$\begin{aligned} x &\equiv a \pmod{m} \\ x &\equiv b \pmod{n}\,. \end{aligned}$$

In the following theorem we will prove that this system of congruences is uniquely solvable modulo mn, provided that m and n are relatively prime.

Theorem 11.2.1 (Chinese Remainder Theorem)**.** *Consider two linear congruences*

$$\begin{aligned} x &\equiv a \pmod{m} \\ x &\equiv b \pmod{n}\,, \end{aligned}$$

where m and n are relatively prime. Then there is a unique solution x modulo mn that solves both linear congruences simultaneously.

Proof. By the first linear congruence, we find that $x = my + a$ for some integer y. Substituting this into the second linear congruence, we obtain that

$$my \equiv b - a \pmod{n}.$$

This linear congruence has a unique solution modulo n, since $\gcd(m, n) = 1$. Therefore, there is exactly one integer $0 \leq y_0 < n$ such that $x = my_0 + a$. Consequently, there is exactly one $0 \leq x < mn$ that solves both linear congruences simultaneously. □

11.3 Linear Congruences in Multiple Variables

11.4 Multiplicativity of Euler's Totient Function

Exercises

11.1 Consider the congruence

$$xy + 7x + 4y \equiv 10 \pmod{13}.$$

Compute the congruence class of y modulo 13 for $x \equiv 2, 5, 8 \pmod{13}$.

11.2 Three musicians perform a piece with interwoven, cyclic rhythms. One plays on the first of every four beats, the second plays on the third of every five beats, and the third plays on the second of every third beat. The performance concludes when all three musicians play simultaneously for the first time. How many beats does the composition last?

11.3 Show that $m \mid n$ implies $\phi(m) \mid \phi(n)$.

11.4 Prove that if $\gcd(a, b) = d$, then

$$\phi(ab) = \frac{d\phi(a)\phi(b)}{\phi(d)}$$

11.5 For any $n > 1$, prove that

$$\sum_{a<n \gcd(a,n)=1} a = \frac{n\phi(n)}{2}$$

11.6 How many fractions $\frac{a}{b}$ are there where $\gcd(a, b) = 1$, and $0 \leq a \leq b \leq n$. Such fractions are called *Farey fractions* of order n.

Chapter 12

Polynomial Congruences

12.1 Polynomial Congruences of Prime Moduli

Even though a polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

has at most n roots in the integers, the same is not true in modular arithmetic. A fairly easy example, the quadratic congruence

$$x^2 \equiv 1 \pmod{8}$$

has four solutions, the numbers 1, 3, 5, and 7. On the other hand, we will prove in this section that the number of solutions of a polynomial congruence modulo a prime is, just as in the integer case, at most the degree of the polynomial.

Theorem 12.1.1 (Factor Theorem in Modular Arithmetic)**.** *An integer r is a solution to the polynomial congruence*

$$f(x) \equiv 0 \pmod{n},$$

if and only if

$$f(x) \equiv (x - r)g(x) \pmod{n}$$

for some polynomial $g(x)$.

Proof. The proof is very similar to the proof of Theorem 6.1.5. First, we note that the converse direction is trivial, so we focus on the forward direction. Suppose that

$$f(r) \equiv 0 \pmod{n}$$

and consider the polynomial $f(x) - f(r)$. Then we have

$$f(x) \equiv f(x) - f(r) \pmod{n},$$

and the factorization of $f(x) - f(r)$ as $(x - r)g(x)$ follows in the same way as for Theorem 6.1.5. □

Theorem 12.1.2 (Lagrange's Theorem)**.** *Consider a polynomial*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0,$$

and a prime p such that $p \nmid a_n$. Then the polynomial congruence

$$f(x) \equiv 0 \pmod{p}$$

has at most n solutions.

Proof. We prove the claim by induction on the degree n of the polynomial f. If f has degree zero, then $f(x) = a_0$, and since the leading term of f is assumed to be not divisible by p, it follows that $a_0 \not\equiv 0 \pmod{p}$. Thus, the polynomial f has no solutions modulo p.

Now suppose that any polynomial congruence modulo p of degree n has at most n solutions, and consider a polynomial congruence $f(x) \equiv 0 \pmod{p}$ of degree $n+1$. If this congruence has no roots, then the number of solutions is certainly below $n+1$.

On the other hand, if $f(r) \equiv 0 \pmod{p}$ is a solution, then by the Factor Theorem for Modular Arithmetic we can find a polynomial $g(x)$ of degree n such that

$$f(x) \equiv (x-r)g(x) \pmod{p}.$$

It follows that the polynomial congruences $f(x) \equiv 0 \pmod{p}$ and

$$(x-r)g(x) \equiv 0 \pmod{p}$$

have the same sets of solutions. Since p is prime, we have by Proposition 5.3.2 that $ab \equiv 0 \pmod{p}$ if and only if $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$. Any solution x of the congruence $(x-r)g(x) \equiv 0 \pmod{p}$ must therefore satisfy $x - r \equiv 0 \pmod{p}$ or $g(x) \equiv 0 \pmod{p}$. There are at most d solutions of the polynomial congruence $g(x) \equiv 0 \pmod{p}$, so we conclude that there are at most $d+1$ solutions to the polynomial congruence $f(x) \equiv 0 \pmod{p}$. □

The previous theorem is due to Joseph-Louis Lagrange, published in 1770 in *Réflexions sur la résolution algébrique des équations* [Lag70]. Lagrange demonstrated that the roots of polynomials can be studied by their permutations. His ideas would eventually lead to the emergence of group theory (the study of symmetry), and Galois theory.

While Lagrange's theorem establishes an upper bound for the number of incongruent solutions of a polynomial congruence modulo a prime, we can use Fermat's Little Theorem to find examples of polynomial congruences that have the maximum allowable number of solutions.

Theorem 12.1.3. *Consider a prime p and a positive integer n such that $n \mid p-1$. Then the polynomial congruence*

$$x^n \equiv 1 \pmod{p}$$

has exactly n incongruent solutions.

Proof. Note that the polynomial congruence $x^{p-1} - 1 \equiv 0 \pmod{p}$ has exactly $p-1$ solutions. By the formula for the difference of powers, we have

$$(x^{p-1} - 1) \equiv (x^n - 1)q(x) \pmod{p}$$

for some polynomial q of degree $p - 1 - n$. Now we observe that the polynomial congruences

$$x^n - 1 \equiv 0 \pmod{p} \qquad \text{and} \qquad q(x) \equiv 0 \pmod{p}$$

have at most n and $p - 1 - n$ solutions, respectively. Since their product has $p - 1$ solutions, it follows that the polynomial congruence $x^n - 1 \equiv 0 \pmod{p}$ must have exactly n solutions. □

12.2 Polynomial Congruences of Composite Moduli

In the following theorem we generalize the Chinese Remainder Theorem to polynomial congruences.

Theorem 12.2.1. *Consider an integer polynomial f and consider two relatively prime integers m and n. Then any solution of the system*

$$f(x) \equiv 0 \pmod{m}$$
$$f(x) \equiv 0 \pmod{n}$$

of polynomial congruences corresponds uniquely to a solution of the polynomial congruence

$$f(x) \equiv 0 \pmod{mn}.$$

Proof. First note that if $f(x) \equiv 0 \pmod{mn}$, then it follows that $f(x) \equiv 0 \pmod{m}$ and $f(x) \equiv 0 \pmod{n}$, simply because divisibility is transitive. Consequently, if y is the unique nonnegative integer strictly below m such that $x \equiv y \pmod{m}$, then $f(y) \equiv \pmod{m}$ is also a solution, and similar for the modulus n. Thus if we write $N_{mn}(f)$ for the set of roots of f modulo mn, and $N_m(f)$ and $N_n(f)$ for the sets of roots of f modulo m and n, respectively, then we define a map

$$\pi : N_{mn}(f) \to N_m(f) \times N_n(f)$$

by $x \mapsto (x \pmod{m}, x \pmod{n})$. We claim that the map π is a bijection.

To see that the map π is a bijection, consider $0 \leq y < m$ and $0 \leq z < n$ such that

$$f(y) \equiv 0 \pmod{m}$$
$$f(z) \equiv 0 \pmod{n}.$$

By the assumption that $\gcd(m, n) = 1$, we obtain from the Chinese Remainder Theorem there a unique $0 \leq x < mn$ such that

$$x \equiv y \pmod{m}$$
$$x \equiv z \pmod{n}.$$

Consequently, we have that $f(x) \equiv 0 \pmod{m}$ and $f(x) \equiv 0 \pmod{n}$. Applying the Chinese Remainder Theorem once more to the number $f(x)$, using that 0 is the only integer u modulo mn such that $u \equiv 0 \pmod{m}$ and $u \equiv 0 \pmod{n}$, we find that $f(x) \equiv 0 \pmod{mn}$. □

Corollary 12.2.2. *Consider $n = p_1^{k_1} \cdots p_m^{k_m}$, where all the primes p_i are distinct. A polynomial congruence*

$$f(x) \equiv 0 \pmod{n}$$

is solvable if and only if the polynomial congruence

$$f(x) \equiv 0 \left(\mathrm{mod}\ p_i^{k_i}\right)$$

is solvable for each i.

Proof. By induction on the number of prime factors of n. □

12.3 Reduced Polynomials Modulo a Prime

Definition 12.3.1. Two polynomials $f(x)$ and $g(x)$ are said to be *congruent* modulo a prime p if their coefficients are congruent modulo p.

Definition 12.3.2. Two polynomials $f(x)$ and $g(x)$ are said to be *equivalent* modulo a prime p if for each integer x the polynomial congruence

$$f(x) \equiv g(x) \pmod{p}$$

holds.

Definition 12.3.3. A polynomial $f(x)$ is said to be *reduced* modulo p if it is of degree less than p.

Theorem 12.3.4. *Any two equivalent reduced polynomials modulo p are congruent.*

Proof. Consider two equivalent reduced polynomials $f(x)$ and $g(x)$. Then $f(x) - g(x)$ is a reduced polynomial equivalent to 0. However, if $f(x) - g(x)$ were nonzero then its number of roots would be strictly less than p. Since the number of roots is p, it follows that $f(x) - g(x)$ is congruent to the zero polynomial. □

Theorem 12.3.5. *For every polynomial $f(x)$ there is a unique reduced polynomial $g(x)$ equivalent to $f(x)$.*

12.4 The Elementary Symmetric Polynomials

Using the fact that equivalent reduced polynomials must be congruent, we obtain the following factorization of the polynomial $x^{p-1} - 1$ in $\mathbb{Z}/p\mathbb{Z}$, which has roots at the numbers $1, \ldots, p-1$.

Theorem 12.4.1. *Consider a prime p. Then we have*

$$x^{p-1} - 1 \equiv (x-1)\cdots(x-p+1) \pmod{p}$$

for every x. Hence their coefficients are congruent.

Proof. Both polynomials are equivalent and reduced, so they must be congruent. □

The previous theorem provides another way of proving Wilson's Theorem. This proof is due to Lagrange, who was the first to publish a proof of Wilson's theorem.

Proof of Wilson's Theorem. The integer $(p-1)!$ is the constant coefficient of the polynomial $(x-1)\cdots(x-p+1)$, which is congruent to the polynomial $x^{p-1}-1$ with constant coefficient -1. □

Using the elementary symmetric polynomials, we find another way of stating Theorem 12.4.1:

Theorem 12.4.2. *For each prime p and each $1 \leq k < p-1$ we have*

$$e_k(1,\ldots,p-1) \equiv 0 \pmod{p}.$$

Exercises

12.1 Show that there are only two polynomials f such that any two values $f(x)$ and $f(y)$ are relatively prime for distinct inputs x and y.

12.2 Consider a prime number p. Construct for every integer $0 \leq a < p$ a polynomial $f(x)$ of degree $< p$ with coefficients in $\mathbb{Z}/p\mathbb{Z}$ such that

$$f(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise.} \end{cases}$$

Use these polynomials similarly to the basis polynomials in Lagrange's Interpolation Theorem to show that every function $\mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ is polynomial.

12.3 Define the kth *power sum*

$$p_k(a_1,\ldots,a_n) := a_1^k + \cdots + a_n^k.$$

We will simply write p_k for $p_k(a_1,\ldots,a_n)$ and e_k for $e_k(a_1,\ldots,a_n)$.

(a) *The Newton-Girard identities.* Show that for all k we have

$$p_k = e_1 p_{k-1} - e_2 p_{k-2} + \cdots + (-1)^{k-1} e_{k-1} p_1 + (-1)^{k-1} k e_k.$$

Hint: This looks more complicated than it is.

(b) Prove that the congruence

$$1^k + \cdots + (p-1)^k \equiv 0 \pmod{p}$$

holds for every $0 < k < p-1$, and every prime number p.

Chapter 13

Primitive Roots

13.1 The Multiplicative Order of an Integer Modulo n

This section is about the function $n \mapsto a^n - 1$. We will begin by showing that this function preserves divisibility.

Proposition 13.1.1. *Suppose that $m \mid n$. Then we have*

$$a^m - 1 \mid a^n - 1.$$

Proof. Suppose that $qm = n$. Then we can write

$$a^n - 1 = (a^m)^q - 1^q = (a^m - 1) \sum_{k=0}^{q-1} a^{km}$$

by the formula for the difference of qth powers (Exercise 1.9). □

We will now define the multiplicative order of an element a modulo n. If $\gcd(a, n) = 1$, then we can simply define the multiplicative order of a to be the least positive integer k such that

$$a^k \equiv 1 \pmod{n}.$$

However, if $\gcd(a, n) \neq 1$ then there will not be such a positive integer. In this case we set the multiplicative order of a modulo n to be 0. At first glance, this may look quite arbitrary. However, there is a natural way of looking at it.

Definition 13.1.2. Consider an integer a and a natural number n. We define he *multiplicative order* of a modulo n to be the unique natural number m such that the ideal

$$I_{a,n} := \{k \in \mathbb{Z} \mid a^{|k|} \equiv 1 \pmod{n}\}$$

is the ideal (m) generated by m.

Using this definition, we see that if $\gcd(a, n) = 1$, then indeed the order $\mathrm{ord}_n(a)$ of a modulo n is the least positive integer k such that $a^k \equiv 1 \pmod{n}$. If $\gcd(a, n) \neq 1$, then the ideal $I_{a,n}$ is the zero ideal (0), which implies that $\mathrm{ord}_n(a) = 0$.

We saw the multiplicative order featuring in the second proof of Euler's Theorem. Thus, using the second proof of Euler's theorem we immediately observe that

$$\mathrm{ord}_n(a) \mid \phi(n),$$

for any a and n such that $\gcd(a, n) = 1$.

In the following theorem, which is also known as the *Order Theorem*, we will see that by setting the value of the order of a modulo n to be 0 when $\gcd(a, n) \neq 1$, we prove a powerful duality principle between multiples of $\mathrm{ord}_n(a)$ and numbers dividing $a^k - 1$.

Theorem 13.1.3 (Order Theorem)**.** *The multiplicative order of an integer a modulo n satisfies the logical equivalence*

$$\mathrm{ord}_n(a) \mid k \Leftrightarrow n \mid a^k - 1$$

Proof. We have two cases to consider: either $\gcd(a, n) = 1$ or $\gcd(a, n) \neq 1$. In the first case, $\mathrm{ord}_n(a)$ is the least positive integer m such that

$$a^m \equiv 1 \pmod{n}.$$

It is immediate from this condition that if $m \mid k$, then it follows that $a^k \equiv 1 \pmod{n}$. Conversely, if $a^k \equiv 1 \pmod{n}$, then we can write $k = qm + r$ by the Euclidean division theorem, where $0 \leq r < m$. It follows that

$$a^r \equiv a^{qm} a^r \equiv a^{qm+r} \equiv a^k \equiv 1 \pmod{n}.$$

Since r is strictly less than m and m is the least positive integer such that $a^m \equiv 1 \pmod{n}$, it follows that $r = 0$. Thus we conclude that $\mathrm{ord}_n(a) \mid k$. This proves the claim in the first case.

In the second case we have $\mathrm{ord}_n(a) = 0$. Then the condition that $\mathrm{ord}_n(a) \mid k$ is equivalent to the condition that $k = 0$. Thus, our task is to show that $n \mid a^k - 1$ if and only if $k = 0$. Consider a common divisor $d > 1$ of a and n. Then we have $d \mid a^k - 1$, so it follows that $d \mid a^k$ if and only if $d \mid 1$. Since we have assumed that $d > 1$, it therefore follows that $d \nmid a^k$. However, since d divides a, the only way in which this is possible is that $k = 0$.

$$\mathrm{ord}_n(a) \mid k.$$ □

The previous theorem has strong implications about the function $k \mapsto a^k - 1$, which we will present as a corollary of the following definition and theorem.

Definition 13.1.4. A function $f : \mathbb{N} \to \mathbb{N}$ is said to be a *divisibility sequence* if it preserves divisibility, i.e., if it satisfies the condition

$$m \mid n \Rightarrow f(m) \mid f(n)$$

for all m and n. A divisibility sequence is said to be a *strong divisibility sequence* if $f(0) = 0$ and f preserves greatest common divisors, i.e., if

$$f(\gcd(m, n)) = \gcd(f(m), f(n))$$

for all m and n.

Theorem 13.1.5. *Suppose that $f : \mathbb{N} \to \mathbb{N}$ is a divisibility sequence. Then the following are equivalent:*

1. *There is a function $g : \mathbb{N} \to \mathbb{N}$ satisfying the logical equivalence*

$$g(m) \mid n \Leftrightarrow m \mid f(n).$$

2. *The divisibility sequence f is a strong divisibility sequence.*

Proof. Suppose first that a function g exists satisfying the logical equivalence

$$g(m) \mid n \Leftrightarrow m \mid f(n).$$

Then it follows that

$$\begin{aligned} d \mid f(\gcd(m, n)) &\Leftrightarrow g(d) \mid \gcd(m, n) \\ &\Leftrightarrow g(d) \mid m \wedge g(d) \mid n \\ &\Leftrightarrow d \mid f(m) \wedge d \mid f(n) \\ &\Leftrightarrow d \mid \gcd(f(m), f(n)). \end{aligned}$$

Since $f(\gcd(m, n))$ and $\gcd(f(m), f(n))$ have the exact same set of divisors, they must be equal.

For the converse, suppose f preserves greatest common divisors. Then consider the ideal

$$I_m := \left\{ k_1 n_1 + \cdots + k_r n_r \;\middle|\; m \mid f(n_i) \text{ for all } i \right\}$$

and define the function g by letting $g(m)$ be the unique natural number such that $I_m = (g(m))$. In other words, $g(m)$ is the least natural number that can be written as a linear combination of numbers n_i, such that $m \mid f(n_i)$.

To see that g is a divisibility sequence, suppose that $m \mid m'$, i.e., suppose that $qm = m'$. Then any linear combination of natural numbers n such that $m' \mid f(n)$ is also a linear combination of natural numbers n such that $m \mid f(n)$. In particular $g(m')$ is in the ideal $(g(m))$, which means that $g(m) \mid g(m')$.

We claim that the function g satisfies the desired logical equivalence. Suppose first that $g(m) \mid n$, say $qg(m) = n$. Note that $g(m)$ can be written in the form $k_1 n_1 + \cdots + k_r n_r$, so we can proceed by induction on r. In the base case, we have that $g(m) = 0$. Then it follows that $n = 0$, and it also follows that every linear combination of natural numbers n such that $m \mid f(n)$ must be 0. Since $f(0) = 0$ by the assumption that f is a strong divisibility sequence, this shows that $m \mid f(n)$.

For the inductive step, suppose that for every m such that $g(m)$ can be written in the form $k_1 n_1 + \cdots + k_r n_r$, we have $g(m) \mid n$ implies $m \mid f(n)$, and consider m such that $g(m)$ can be written in the form

$$k_1 n_1 + \cdots + k_r n_r + k_{r+1} n_{r+1}.$$ □

Corollary 13.1.6. *Consider an integer a and two natural numbers m and n relatively prime to a. Then*

$$\gcd(a^m - 1, a^n - 1) = a^{\gcd(m,n)} - 1.$$

The logical equivalence in the definition of the order of an element can be used effectively to compute the order of a power of an element in terms of the order of that element.

Proposition 13.1.7. *For any integer a of order $k = \mathrm{ord}_n(a)$, we have*

$$\mathrm{ord}_n(a^m) = \frac{k}{\gcd(m,k)}$$

In particular, we have $\mathrm{ord}_n(a^m) = k$ if and only if m and k are relatively prime.

Proof. We have the following chain of logical equivalences:

$$\mathrm{ord}_n(a^m) \mid l \iff n \mid (a^m)^l - 1 \iff \mathrm{ord}_n(a) \mid ml \iff \frac{\mathrm{ord}_n(a)}{\gcd(m, \mathrm{ord}_n(a))} \mid l.$$

Thus, we see that $\mathrm{ord}_n(a^m)$ and $k/\gcd(m,k)$ divide the same numbers, so they must be equal. □

Remark 13.1.8. More generally, if $(P, \leq_P)$ and $(Q, \leq_Q)$ are two posets, a *Galois connection* between P and Q consists of order preserving maps $f : P \to Q$ and $g : Q \to P$ satisfying the logical equivalence

$$g(x) \leq_P y \Leftrightarrow x \leq_Q f(x)$$

for any $x \in Q$ and $y \in P$. In this case, the map f is said to be the *upper adjoint*, and the map g is said to be the *lower adjoint* of the Galois connection. Thus we have proven in this section that the function $n \mapsto a^n - 1$ is the upper adjoint of a Galois connection, and we have used this fact to show that f preserves greatest common divisors. Theorem 13.1.5 generalizes to the setting of arbitrary posets, and indeed, some readers may recognize it as a special case of the *Adjoint Functor Theorem*.

Another famous Galois connection on the natural numbers ordered by divisibility is the Fibonacci sequence. We will give the full statement of this fact in the exercises.

13.2 The Infinitude of Primes Congruent to 1 Modulo Powers of 2

One way to show that a prime p is congruent to 1 modulo n is by finding an integer a such that

$$\mathrm{ord}_p(a) = n.$$

By the Order Theorem and Fermat's Little Theorem, this implies that $n \mid p - 1$. We will use this idea to show that there are infinitely many primes congruent to 1 modulo any power of 2. In particular, there are infinitely many primes congruent to 1 modulo 4, modulo 8, modulo 16, and so on.

Theorem 13.2.1. *For any positive natural number n, there are infinitely many primes congruent to 1 modulo 2^n.*

Proof. Since $p \equiv 1 \pmod{2^{n+1}}$ implies that $p \equiv 1 \pmod{2^n}$, it suffices to show that for any finite list $p_1, \ldots, p_k$ of primes congruent to 1 modulo 2^{n+1}, there is a prime $q \equiv 1 \pmod{2^{n+1}}$ which is not among the primes p_i.

Given such a list of primes p_i, let

$$a := 2p_1 \cdots p_k,$$

and let q be a prime divisor of the integer $b := a^{2^n} + 1$. Note that q is necessarily an odd prime, since a is even and hence b is odd. Since $p_i \nmid b$ for any i, it follows that q is not among the primes p_i. Furthermore, we observe that

$$a^{2^n} \equiv -1 \pmod{q}.$$

This implies that $\mathrm{ord}_q(a) \mid 2^{n+1}$, while clearly $\mathrm{ord}_q(a) \nmid 2^n$. Thus, we must have $\mathrm{ord}_q(a) = 2^{n+1}$, and this allows us to conclude that $q \equiv 1 \pmod{2^{n+1}}$. □

The previous theorem is a special case of a deep theorem in analytic number theory: *Dirichlet's Theorem*. Dirichlet's Theorem shows that for any integer a relatively prime to n, there are infinitely many primes

$$p \equiv a \pmod{n}.$$

The techniques of proving Dirichlet's Theorem are beyond the scope of this course. However, having taken this course, you are well equipped to read LeVeque's excellent *Topics in Number Theory* [LeV56a; LeV56b], a two-volume work that contains well-explained proofs of Dirichlet's Theorem as well as the Prime Number Theorem.

13.3 Counting Elements of a Given Order Modulo a Prime

Euler's Theorem tells us that

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

for every integer a such that $\gcd(a, n) = 1$. By the Order Theorem this implies that

$$\mathrm{ord}_n(a) \mid \phi(n).$$

Euler's theorem therefore gives an upper bound for the order of any element modulo n. For example, since 13 is a prime number, we have $\phi(13) = 12$, so all integers relatively prime to 13 have an order dividing the number 12. Computing the powers of 2, for instance, we see that $\mathrm{ord}_{13}(2) = 12$:

m	0	1	2	3	4	5	6	7	8	9	10	11
$2^m \pmod{13}$	1	2	4	8	3	6	12	11	9	5	10	7.

We can use this table to compute the orders of all the nonzero integers modulo 13. For example, since $3 \equiv 2^4 \pmod{13}$, it follows that the order of 3 modulo 13 is 3. Similarly, since $5 \equiv 2^9 \pmod{13}$, it follows that

$$5^4 \equiv (2^9)^4 \equiv (2^{12})^3 \equiv 1 \pmod{13}.$$

Computing the orders of all the elements from 1 to 12 modulo 13 this way, we obtain:

a	1	2	3	4	5	6	7	8	9	10	11	12
$\mathrm{ord}_{13}(a)$	1	12	3	6	4	12	12	4	3	6	12	2.

While the orders of the integers modulo 13 might at first glance appear to be somewhat unwieldy, there are some important patterns to uncover. First of all, we notice that there are four integers that have the maximal possible order of 12, and the number of integers $1 \leq x \leq 12$ relatively prime to 12 is also $\phi(12) = 4$. This is no coincidence, because every integer relatively prime to 13 is congruent to one of the integers

$$2^0,\ 2^1,\ 2^2,\ 2^3,\ 2^4,\ 2^5,\ 2^6,\ 2^7,\ 2^8,\ 2^9,\ 2^{10},\ \text{and } 2^{11},$$

and by Proposition 13.1.7 it follows that

$$\mathrm{ord}_{13}(2^m) = 12 \qquad \Leftrightarrow \qquad \gcd(m, 12) = 1.$$

However, we can go further: There are two integers that have order 6 modulo 13, and indeed $\phi(6) = 2$ as well. Similar for the we find that there are $\phi(1)$, $\phi(2)$, $\phi(3)$, and $\phi(4)$ integers of order 1, 2, 3, and 4, respectively. The following proposition tells us that modulo a prime p, the number of integers of order k modulo p is either 0 or $\phi(k)$. This suggests that for any prime number p and any $k \mid p - 1$, it should be true that the number of elements of order k is exactly $\phi(k)$. To prove this, we will need the following important fact about Euler's totient function:

Theorem 13.3.1. *For every positive integer n, we have*

$$\sum_{d \mid n} \phi(d) = n,$$

where the sum ranges over the positive integers d dividing n.

Proof. The sum suggests that there is a way of partitioning the set $\{1, \ldots, n\}$ into sets A_d of size $\phi(d)$, and this is indeed how this theorem is proven. For each positive divisor d of n, define the map

$$f_d : \{1 \leq x \leq d \mid \gcd(x, d) = 1\} \to \{1, \ldots, n\}$$

by $x \mapsto \frac{n}{d}x$, and define the set A_d to be the image of f_d. Since $\frac{n}{d}$ is a positive integer, it follows that the map f_d is injective, and therefore each set A_d has exactly $\phi(d)$ elements. Now notice that

$$y \in A_d \qquad \Leftrightarrow \qquad \gcd(y, n) = \frac{n}{d},$$

since if $x = y/\gcd(y, n)$, then $\gcd(x, d) = \gcd(y/\gcd(y, n), n/\gcd(y, n)) = 1$. Therefore, it follows that for every $1 \leq y \leq n$ there is exactly one $d \mid n$ such that $y \in A_d$, so it follows that the sets A_d partition the set $\{1, \ldots, n\}$. The sum of the numbers of elements of each A_d is therefore n. □

Theorem 13.3.2. *Consider a prime number p. Then there are for every $k \mid p-1$ exactly $\phi(k)$ integers of order k modulo p.*

Proof. In this proof, let us write $\psi(k)$ for the number of integers of order k modulo p. It is clear that $\psi(k) = 0$ if $k \nmid p-1$. On the other hand, we claim that for any $k \mid p-1$, we have either $\psi(k) = 0$ or $\psi(k) = \phi(k)$. In other words, we will first show that if there is an element a of order k modulo p, then there are $\phi(k)$ elements of order k.

First recall that every integer of order k modulo p is a solution to the polynomial congruence

$$x^k \equiv 1 \pmod{p}.$$

Since $\operatorname{ord}_p(a) \mid p-1$ holds for any element a, this polynomial congruence as exactly k solutions by Theorem 12.1.3. However, since we are given an element a of order k, we can describe the set of solutions to the congruence $x^k \equiv 1 \pmod{p}$: it is the set consisting of $1, a, a^2, \ldots, a^{k-1}$. In other words, every solution is congruent to a power of a. The powers of a that are also of order k are characterized by Proposition 13.1.7:

$$\operatorname{ord}_p(a^m) = k \qquad \Leftrightarrow \qquad \gcd(m,k) = 1.$$

Thus we see that there are exactly $\phi(k)$ integers of order k.

We have thus established that $\psi(k) \leq \phi(k)$ for every $k \mid p-1$. Now we see that, since every element $1 \leq x < p$ has an order modulo p, we have

$$n = \sum_{k \mid p-1} \psi(k) \leq \sum_{k \mid p-1} \phi(k) = n.$$

This inequality is therefore an equality, and since $\psi(k) \leq \phi(k)$ for every $k \mid p-1$ it follows that $\psi(k) = \phi(k)$ for every $k \mid p-1$. □

13.4 Primitive Roots

We saw in the previous section that there are four integers modulo 13 that have order 12, which is the maximal possible order by Euler's theorem. We also saw that there are exactly $\phi(k)$ elements of order k modulo p, for every $k \mid p-1$. It follows that there are

$$\phi(p-1)$$

elements of order $p-1$. In other words, for every prime number p, there is an integer a such that the integers

$$1, a, a^2, \ldots, a^{p-2}$$

form a reduced residue system modulo p. That is, there is an integer a such that every integer x relatively prime to p is congruent to exactly one power a^m, where $0 \leq m < p-1$. Such integers are called *primitive roots*. In the previous section we saw that 2 is a primitive root of the prime 13.

Definition 13.4.1. A *primitive root* modulo a natural number n is an integer a such that

$$\operatorname{ord}_n(a) = \phi(n).$$

Combining the previous remarks, we obtain:

Theorem 13.4.2. *The number of primitive roots of a prime number p is $\phi(p-1)$.*

In fact, it is possible to precisely classify all the natural numbers that have primitive roots. They are the numbers

$$2,\ 4,\ p^k,\ \text{and}\ 2p^k.$$

where p is an odd prime. For example, 3 is a primitive root of 4, and 5 is a primitive root of 6. The number 8, on the other hand, does not have primitive roots because $\phi(8) = 4$ while the nonunital odd integers 3, 5, and 7 all have order 2.

It is useful to have access to primitive roots, if they exist, because if a is a primitive root modulo n, then the integers

$$1, a, a^2, \ldots, a^{\phi(n)-1}$$

form a reduced residue system modulo n. This gives a convenient way of representing ell the elements of a reduced residue system.

Definition 13.4.3. Suppose a is a primitive root modulo n, and let x be an integer relatively prime to n. Then we define the *index*

$$\text{ind}_a(n)$$

of x base a to be the unique natural number such that $a^{\text{ind}_a(n)} = x$.

Theorem 13.4.4. *Consider a natural number n with a primitive root a.*

1. *For any two integers x and y relatively prime to n, we have*

$$x \equiv y \pmod{n} \qquad \Leftrightarrow \qquad \text{ind}_a(x) \equiv \text{ind}_a(y) \pmod{\phi(n)}.$$

2. *Furthermore, we have*

$$\text{ind}_a(xy) \equiv \text{ind}_a(x) + \text{ind}_a(y) \pmod{\phi(n)}$$
$$\text{ind}_a(x^m) \equiv m\,\text{ind}_a(x) \pmod{\phi(n)}.$$

Example 13.4.5. Solving linear congruences becomes quite a straightforward task when a primitive root is known. For example, the congruence

$$14x \equiv 6 \pmod{26}$$

is solvable if and only if $7x \equiv 3 \pmod{13}$ is. By the previous theorem, this linear congruence is equivalent to the linear congruence

$$\text{ind}_2(7) + \text{ind}_2(x) \equiv \text{ind}_2(3) \pmod{12}.$$

Since $\text{ind}_2(7) = 11$ and $\text{ind}_2(3) = 4$, we find that $\text{ind}_2(x) \equiv 5 \pmod{12}$. Thus, we conclude that

$$x \equiv 6 \pmod{13}$$

is a solution of the original linear congruence.

We now return to an old problem, the investigation of which started in Exercises 4.10 and 4.11, which asked to compute common divisors of all the integers of the form $x^3 - x$ and common divisors of all the integers of the form $x^5 - x$. In Section 6.3 we showed that the fixed divisor of $f(x) = x^n - x$ can be determined by computing the greatest common divisor of the values

$$f(2), \ldots, f(\lfloor \tfrac{n+1}{2} \rfloor).$$

Recall from Definition 6.3.1 that the fixed divisor of a polynomial $f(x)$ is the greatest common divisor of all the values of $f(x)$. Using primitive roots we can now compute the fixed divisors of all the polynomials of the form $x^n - x$.

Theorem 13.4.6. *Consider the polynomial $x^n - x$, for some natural number $n > 1$. Its fixed divisor is*

$$\prod_{\substack{p \text{ prime} \\ p-1 \mid n-1}} p.$$

Proof. First, we claim that a prime number p appears as a factor of the fixed divisor of $x^n - x$ if and only if

$$p - 1 \mid n - 1.$$

To see this, suppose that p divides every integer of the form $x^n - x$. Since $x^n - x = x(x^{n-1} - 1)$ and p is prime, it follows that $p \mid x$ or $p \mid x^{n-1} - 1$ for every x not divisible by p. From this, it follows that

$$\mathrm{ord}_p(x) \mid n - 1$$

for every x not divisible by p. In particular, if x is a primitive root modulo p, we find that

$$p - 1 \mid n - 1.$$

Thus we see that if p divides $x^n - x$ for every integer x, then we must have $p - 1 \mid n - 1$.

Conversely, if $p - 1 \mid n - 1$, then it follows by Fermat's Little Theorem that

$$x^{n-1} \equiv 1 \pmod{p}$$

for every x not divisible by p. This implies that $x(x^{n-1} - 1)$ is divisible by p for every x, completing the proof of the first claim.

To finish the proof, it remains to show that the fixed divisor of $x^n - x$ is square-free, i.e., that for each prime number p there is an integer x such that $p^2 \nmid x^n - x$. We choose $x = p$, because the integer $p^n - p = p(p^{n-1} - 1)$ is indeed not divisible by p^2. □

We conclude with an observation of Sophie Germain.

Corollary 13.4.7. *Consider a natural number n, and write $q = \prod_{p-1 \mid n-1} p$ for the product of all primes p such that $p - 1 \mid n - 1$. Then*

$$a^n \equiv a \pmod{q}$$

for every integer a. Consequently, if x, y, and z are integers such that $x^n + y^n = z^n$, then we have

$$x + y \equiv z \pmod{q}.$$

Exercises

13.1 (a) Find the least primitive root g modulo 23, and give a table of the congruence classes of its powers.
(b) Find all primitive roots modulo 23.
(c) For each $k \mid \phi(23)$, find all elements of order k modulo 23.
(d) For each integer $1, \ldots, 22$, find its index base g modulo 23.
(e) Solve the linear congruence

$$8x \equiv 9 \pmod{23}$$

using indices.

13.2 Show that if $ab \equiv 1 \pmod{n}$, then

$$\operatorname{ord}_n(a) = \operatorname{ord}_n(b).$$

13.3 Consider an integer a. Show that if

$$a^m - 1 \mid a^n - 1,$$

then $m \mid n$.

13.4 Show that $\operatorname{ord}_{a^k-1}(a) = k$.

13.5 Suppose that there is an integer a of even order modulo p. Prove that there is a unique integer of order 2 modulo p.

13.6 Prove that for any prime p and any integer a not divisible by p, there is a primitive root g modulo p such that

$$g^{\frac{p-1}{\operatorname{ord}_p(a)}} \equiv a \pmod{p}.$$

13.7 (a) Show that the function $(x, y) \mapsto (y, x + y)$ is a bijection on the set of pairs of elements $x, y \in \mathbb{Z}/n\mathbb{Z}$.
(b) Show that for any m, the Fibonacci sequence is periodic modulo m. Its period is called the *Pisano period*.
(c) Show that for any natural number m, there is a least positive number n such that

$$m \mid F_n.$$

We will call this number G_m. By virtue of its duality to the Fibonacci sequence, we will call the sequence G the *cofibonacci sequence*.
(d) Show that $G(m)$ satisfies the following logical equivalence

$$G_m \mid n \Leftrightarrow m \mid F_n.$$

Conclude that the Fibonacci sequence is a strong division sequence.

13.8 Use methods similar to the previous exercise to show that the Pell numbers P_n, which are defined by

$$P_0 := 0, \qquad P_1 := 1 \qquad \text{and} \qquad P_{n+2} = 2P_{n+1} + P_n,$$

form a strong division sequence.

Chapter 14

Quadratic Residues

14.1 Quadratic Congruences

A *quadratic congruence* is a congruence of the form

$$ax^2 + bx + c \equiv 0 \pmod{n},$$

where a is assumed to be relatively prime to n. To build intuition for such quadratic congruences, let's first recall how to solve a quadratic equation

$$ax^2 + bx + c = 0,$$

where $a \neq 0$. Since we have assumed that a is nonzero, we may divide through by a to obtain the equation

$$x^2 + \frac{b}{a}x = -\frac{c}{a}.$$

Completing the square, gives us the equation

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a^2} - \frac{c}{a} = \frac{b^2 - 4ac}{4a^2}.$$

This equation has the familiar solutions

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The number $\Delta := b^2 - 4ac$ is called the *discriminant* of the quadratic equation $ax^2 + bx + c = 0$. For example, the quadratic equation $x^2 - x - 1$ has discriminant 5, and therefore its solutions are

$$x = \frac{1 \pm \sqrt{5}}{2}.$$

Since the square root $\sqrt{k}$ of an integer k is an integer if and only if k is a perfect square, we obtain the following theorem:

Theorem 14.1.1. *Consider the quadratic equation with integer coefficients*

$$ax^2 + bx + c = 0,$$

where $a \neq 0$. This equation has a rational solution if and only if the discriminant $\Delta = b^2 - 4ac$ is a perfect square.

Now consider an odd prime p. Recall from Lagrange's Theorem that a polynomial congruence

$$f(x) \equiv 0 \pmod{p}$$

has at most n solutions, where n is the degree of f. It follows that a quadratic congruence modulo a prime has at most two solutions. In the following theorem, which is analogous to the previous theorem, we show that a quadratic congruence modulo an odd prime is solvable if and only if its discriminant is a perfect square.

The situation is slightly different for the prime $p = 2$, since the quadratic congruence

$$x^2 + x + 1 \equiv 0 \pmod{2}$$

has no solutions, even though the discriminant is a perfect square. Indeed, as we will see in the proof below, we will rely on the invertibility of 2 modulo p to solve general quadratic congruences modulo p, which is not possible if $p = 2$.

Theorem 14.1.2. *Consider a quadratic congruence*

$$ax^2 + bx + c \equiv 0 \pmod{p}$$

modulo an odd prime p, where $a \not\equiv 0 \pmod{p}$. This quadratic congruence is solvable if and only if the discriminant $\Delta = b^2 - 4ac$ is congruent to a square modulo p.

Proof. Given that p is an odd prime and that $a \not\equiv 0 \pmod{p}$, it follows that $4a$ is invertible modulo p. The quadratic congruence in the statement of the theorem is therefore solvable if and only if the quadratic congruence

$$4a^2x^2 + 4abx + 4ac \equiv 0 \pmod{p}$$

is solvable. Given a solution of this quadratic congruence, we calculate

$$\begin{aligned}(2ax + b)^2 &= 4a^2x^2 + 4abx + b^2 \\ &= 4a^2x^2 + 4abx + 4ac + b^2 - 4ac \\ &\equiv b^2 - 4ac\end{aligned}$$

modulo p. This shows that the discriminant is congruent to a square modulo p.

On the other hand, if the discriminant is congruent to a square modulo p, as in

$$y^2 \equiv b^2 - 4ac \pmod{p},$$

then we find, entirely analogous to the case of quadratic equations, that

$$x = \frac{-b \pm y}{2a}$$

solves the quadratic congruence modulo p. We leave the verification that such x is indeed a solution to the reader. □

14.2 Quadratic Residues

We have seen in the previous section that in order to solve a quadratic congruence

$$ax^2 + bx + c \equiv 0 \pmod{p}$$

modulo an odd prime p, we need to determine whether the discriminant $\Delta = b^2 - 4ac$ is congruent to a square modulo p. In other words, we need to determine whether the discriminant is a quadratic residue modulo p, which is defined as follows:

Definition 14.2.1. An integer a is said to be a *quadratic residue* modulo n if there is an integer x such that the congruence

$$x^2 \equiv a \pmod{n}$$

holds. If no such x exists, we say that a is a *quadratic nonresidue* modulo n.

For example, 0 and 1 are always quadratic residues modulo any natural number n. The number 0 is called the *trivial quadratic residue*, and we will mostly be interested in the nontrivial quadratic residues.

Modulo 3, the only quadratic residues are 0 and 1, since the quadratic congruence $x^2 \equiv 2 \pmod{3}$ does not have solutions. In other words, every square is either divisible by 3 or congruent to 1 modulo 3.

Similarly, 0 and 1 are the only quadratic residues modulo 4; that is, no square number can be congruent to 2 or 3 modulo 4. Indeed, the square of an even number is divisible by 4 and the square of an odd number is congruent to 1 modulo 4.

Modulo 5 we find that 0, 1, and 4 are quadratic residues, and we can similarly list all the quadratic. The following table lists the nontrivial quadratic residues modulo n from 4 to 10, where 0 is included as a quadratic residue if it is the square of a nonzero integer modulo n:

n	4	5	6	7	8	9	10
quadratic residues	0, 1	1, 4	1, 3, 4	1, 2, 4	0, 1, 4	0, 1, 4, 7	1, 4, 5, 6, 9

If we have access to a primitive root g of a prime p, then it is easy to characterize the nontrivial quadratic residues of p as the even powers of g. This has the immediate corollary that exactly half of the nonzero integers modulo p are quadratic residues.

Proposition 14.2.2. *Consider a prime p with a primitive root g. An integer $a \not\equiv 0 \pmod{p}$ is a nontrivial quadratic residue modulo a p if and only if it is an even power of g.*

Proof. Consider a primitive root g modulo p. If $a \equiv g^{2k} \pmod{p}$, then a is clearly a quadratic residue modulo p. On the other hand, if $x^2 \equiv a \pmod{p}$, and $x \equiv g^k \pmod{p}$, then we have $a \equiv g^{2k} \pmod{p}$. □

Corollary 14.2.3. *There are exactly $(p-1)/2$ nontrivial quadratic residues modulo p in the set $\{1, \ldots, p-1\}$, and exactly $(p-1)/2$ quadratic nonresidues.*

14.3 Legendre Symbols

The fact that the integers $1, 2, \ldots, p-1$ can be split into the sets of quadratic residues and quadratic nonresidues, which are both of size $(p-1)/2$, suggests there is an interesting dichotomy between them. We can explore this further using primitive roots. If g is a primitive root modulo p, then the quadratic residues are precisely those integers modulo p that can be written as an even power of g, and the quadratic nonresidues are precisely those integers modulo p that can be written as an odd power of g. Now recall that

$$g^m g^k = g^{m+k},$$

and that the parity of the sum of two integers can be expressed in terms of the parity of the summands: The summands m and k have the same parity if and only if $m + k$ is even, and likewise they have distinct parity if and only if $m + k$ is odd. This leads to the following multiplication rule of quadratic residues:

Proposition 14.3.1. *Consider two integers a and b modulo a prime p, both not divisible by p. Then we have*

1. *If both a and b are quadratic residues, or both are quadratic nonresidues, then ab is a quadratic residue.*
2. *If one of a and b is a quadratic residue and the other is a quadratic nonresidue, then ab is a quadratic nonresidue.*

The previous proposition was also observed by Adrien-Marie Legendre, who published his findings in *Essai sur la théorie des nombres* in 1798 [Leg98]. In this work, he also formulated the quadratic reciprocity theorem, and attempted to prove it. His approach, however, relied on unproven assumptions about certain functions, and was therefore incomplete.

Definition 14.3.2. The *Legendre symbol* of an integer a modulo a prime p, such that $p \nmid a$, is defined by

$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic nonresidue modulo } p. \end{cases}$$

An immediate consequence of Proposition 14.2.2 is that

$$\left(\frac{a}{p}\right) = (-1)^{\operatorname{ind}_g(a)}$$

for any primitive root g modulo p. Using Legendre symbols, we can reformulate Proposition 14.3.1 succinctly as the *multiplicative law of Legendre symbols*: For any two integers a and b, both not multiples of p, the equality

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$

holds. In fact, the Legendre symbol is the unique surjective function from the integers $\{1, \ldots, p-1\}$ to $\{1, -1\}$ with this property.

Theorem 14.3.3. *Consider an odd prime p. Then the function $a \mapsto \left(\frac{a}{p}\right)$ is the unique surjective function $f : \{1, \ldots, p-1\} \to \{1, -1\}$ preserving multiplication, in the sense that*

$$f(ab) = f(a)f(b)$$

for every two nonzero integers a and b modulo p[1].

Proof. We have already seen that the Legendre symbol is a surjective function preserving multiplication. Thus, our task is now to show that if $f : \{1, \ldots, p-1\} \to \{1, -1\}$ is any surjective function preserving multiplication, then f is the Legendre symbol. Notice that for any primitive root g modulo p, we have

$$f(g^m) = f(g)^m.$$

since every element is of the form g^m, this implies that $f(g) = -1$. Then it follows that $f(g^m) = 1$ if and only if m is even, i.e., $f(a) = 1$ if and only if a is a quadratic residue modulo p. □

Example 14.3.4. Recall that 2 is a primitive root of the prime 13, and is therefore a quadratic nonresidue. Using that the square of a Legendre symbol is always 1, we can compute the Legendre symbol $\left(\frac{8}{13}\right)$ as follows:

$$\left(\frac{8}{13}\right) = \left(\frac{2}{13}\right)\left(\frac{2}{13}\right)\left(\frac{2}{13}\right) = \left(\frac{2}{13}\right) = -1.$$

Thus, 8 is a quadratic nonresidue modulo 13.

Similarly, using that $5 \equiv 18 \pmod{13}$ we can compute the Legendre symbol $\left(\frac{5}{13}\right)$ as follows:

$$\left(\frac{5}{13}\right) = \left(\frac{18}{13}\right) = \left(\frac{2}{13}\right)\left(\frac{3}{13}\right)\left(\frac{3}{13}\right) = \left(\frac{2}{13}\right) = -1.$$

We see that 5 is also a quadratic nonresidue modulo 13.

On the other hand, again using that the square of a Legendre symbol is always 1, we can compute the Legendre symbol $\left(\frac{3}{13}\right)$ as follows:

$$\left(\frac{3}{13}\right) = \left(\frac{3}{13}\right)\left(\frac{3}{13}\right)\left(\frac{3}{13}\right) = \left(\frac{27}{13}\right) = \left(\frac{1}{13}\right) = 1.$$

Thus we see that 3 is a quadratic residue modulo 13. In fact, we can readily verify that $3 \equiv 4^2 \pmod{13}$. The set of all nontrivial quadratic residues modulo 13 is

$$\{1, 3, 4, 9, 10, 12\}.$$

[1] In the language of group theory: The Legendre symbol is the unique surjective group homomorphism from the group $(\mathbb{Z}/p\mathbb{Z})^\times$ of invertible elements modulo p to the 2-element group S_2.

14.4 Euler's Criterion

Euler had a deep understanding of numbers that are congruent to a square modulo a prime. In 1748 he had published a result now known as *Euler's criterion*, which characterized precisely the squares modulo a prime.

Theorem 14.4.1 (Euler's Criterion)**.** *Consider an integer a and an odd prime p. Then we have*

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}$$

Proof. First note that Fermat's Little Theorem gives that if $b := a^{\frac{p-1}{2}}$, then

$$b^2 \equiv 1 \pmod{p}.$$

This implies that $b \equiv \pm 1 \pmod{p}$.

To prove the congruence in Euler's Criterion, consider a primitive root g modulo p, and suppose that $g^m \equiv a \pmod{p}$. Then a is a quadratic residue if and only if m is even. Also, we note that

$$\frac{m(p-1)}{2}$$

is a multiple of $p-1$ if and only if m is even. Thus we see that

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

if and only if a is a quadratic residue, and it is congruent to -1 otherwise. □

Theorem 14.4.2. *For any odd prime p, the integer -1 is a quadratic residue modulo p if and only if $p \equiv 1 \pmod{4}$.*

Proof. By Euler's Criterion, it follows that -1 is a quadratic residue modulo p if and only if

$$(-1)^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Since $(-1)^m \equiv 1 \pmod{p}$ if and only if m is even, it follows that -1 is a quadratic residue modulo p if and only if $\frac{p-1}{2}$ is even, which holds if and only if

$$p \equiv 1 \pmod{4}.$$ □

Since primitive roots are always quadratic nonresidues, we obtain the following corollary. Artin conjectured that there are infinitely many primes p for which 2 is a primitive root. More generally, he conjectured that for any integer a that is neither -1 nor a perfect square, there are infinitely many primes for which a is a primitive root.

Corollary 14.4.3. *If 2 is a primitive root modulo p, then $p \equiv \pm 3 \pmod{8}$.*

We can use Theorem 14.4.2 to prove that there are infinitely many primes congruent to 1 modulo 4. The first fifteen primes congruent to 1 modulo 4 are:

$$5,\ 13,\ 17,\ 29,\ 37,\ 41,\ 53,\ 61,\ 73,\ 89,\ 97,\ 101,\ 109,\ 113,\ \text{and } 137.$$

We have already seen in Theorem 13.2.1 that there are infinitely many primes congruent to 1 modulo 2^n, for any positive natural number n. Here we give an alternative proof that there are infinitely many primes congruent to 1 modulo 4, using our earlier observation that -1 is a quadratic residue modulo a prime p if and only if $p \equiv 1 \pmod 4$. Similar techniques can also be used to show that there are infinitely many primes congruent to 1 modulo 3 or 5.

Theorem 14.4.4. *There are infinitely many primes congruent to* 1 *modulo* 4.

Proof. Consider a finite set of primes $p_1, \ldots, p_k$ that are all congruent to 1 modulo 4, and define

$$n = (2p_1 \cdots p_k)^2 + 1.$$

Notice that $n \equiv 1 \pmod 4$ is greater than 1, and $p_i \nmid n$ for any $1 \leq i \leq k$. It follows that for any prime divisor q of n, we have

$$(2p_1 \cdots p_k)^2 \equiv -1 \pmod q,$$

showing that -1 is a quadratic residue modulo q. Since -1 is a quadratic residue modulo q if and only if $q \equiv 1 \pmod 4$, it follows that any prime divisor q of n is congruent to 1 modulo 4. Since none of the prime divisors of n are among the primes p_i we started out with, the primes p_i could not have listed all the primes congruent to 1 modulo 4. □

14.5 Euler's Prime-Generating Polynomial

In the early 1770s, Leonhard Euler devised his famous prime-generating polynomial

$$n^2 + n + 41.$$

This polynomial produces prime numbers for any $0 \leq n \leq 39$. Note that

$$40^2 + 40 + 41 = 40(40 + 1) + 41 = 41^2$$

isn't prime anymore. We could easily imagine Euler pondering one day over the question of creating a quadratic polynomial that returns a surprising amount of primes, and it would be natural to wonder how he might have approached the problem.

Of course, there is no polynomial $f(x)$ other than a constant polynomial such that $f(x)$ is a prime number for every integer x.

Proposition 14.5.1. *If every value of a polynomial $f(x)$ is prime, then $f(x)$ is constant.*

Proof. Suppose $f(x)$ is a polynomial, all of whose values are prime. Then in particular $p := f(0)$ is prime. Then $p \mid f(kp)$ for every integer k. Since each value of f is prime, it follows that $f(kp) = p$ for every integer k. This implies that $f(x) - p$ has infinitely many roots, i.e., $f(x)$ is the constant polynomial with value p. □

As a reasonable starting point for our search, we take three consecutive primes of the form

$$q,\ q + 2,\ \text{and } q + 6.$$

$q \setminus p$	3	5	7	11	13	17	19	23	29	31	37
5	2										
7	0	-									
11	2	2	6								
13	0	-	-	-							
17	2	3	3	10	11						
19	0	-	-	-	-	-					
23	2	4	-	-	-	-	-				
29	2	0	-	-	-	-	-	-			
31	0	-	-	-	-	-	-	-	-		
37	0	-	-	-	-	-	-	-	-	-	
41	2	2	5	2	6	7	8	21	11	23	22

Table 14.1: Residues of the discriminant $\Delta = 1 - 4q$ modulo an odd prime $p < q$, for the polynomial $n^2 + n + q$.
Computations of residues for $q \leq p$ are omitted. A dash indicates that it was not necessary to compute this residue, because a quadratic residue has been encountered earlier.

By Lagrange's Interpolation Theorem, there is only one polynomial f such that $f(n)$ assumes those values for $n = 0, 1, 2$:

$$f(n) = n^2 + n + q.$$

Thus we are faced with the task of finding a prime q such that $n^2 + n + q$ is prime for $0 \leq n \leq q - 2$. One thing to note is that if the number

$$n^2 + n + q$$

is composite for some $0 \leq n \leq q - 2$, then it must have a prime factor below q. Indeed, we have $n^2 + n + q < q^2$ for such n, so if $n^2 + n + q$ is composite for any $0 \leq n \leq q - 2$ then it has an odd prime factor $p < q$. In other words, we want to find an odd prime q such that

$$n^2 + n + q \not\equiv 0 \pmod{p}$$

for any odd prime $p < q$. Recall from Theorem 14.1.2 that such an incongruence holds if and only if the discriminant $\Delta = 1 - 4q$ is a quadratic nonresidue modulo p for any odd prime $p < q$. By Euler's Criterion, this observation finally narrows our search: We are looking for a prime q such that

$$(1 - 4q)^{\frac{p-1}{2}} \equiv -1 \pmod{p}$$

for any odd prime $p < q$.

We can now discover prime-generating polynomials of the form $n^2 + n + q$ by computing the integer $1 - 4q \pmod{p}$ for all odd primes $3 < p < q$. If we can find a prime p such that $1 - 4q$ is a quadratic residue mod p, we may discard that prime from the remainder of the search. This process resulted in Table 14.1, which reveals that the

following polynomials are prime-generating polynomials, in the sense that they yield prime numbers for all inputs $0 \leq n \leq q - 2$:

$$\begin{aligned} &n^2 + n + 2 && n^2 + n + 11 \\ &n^2 + n + 3 && n^2 + n + 17 \\ &n^2 + n + 5 && n^2 + n + 41. \end{aligned}$$

A deep theorem, the Baker–Heegner–Stark Theorem, shows that these are the only prime-generating polynomials of the form $n^2 + n + q$. In other words, Euler had in fact found the quadratic polynomial of this kind with the longest streak of primes! For this reason, the number 41 is sometimes called Euler's lucky prime.

Exercises

14.1 Show that a number n is square-free if and only if the only solution of the quadratic congruence $x^2 \equiv 0 \pmod{n}$ is $x = 0$.

14.2 Find the first prime which cannot be written in the form $n^2 + n + q$ for any $q \in \{2, 3, 5, 11, 17, 41\}$.

14.3 Consider an odd prime p and an integer a.

(a) Show that if $\operatorname{ord}_p(a)$ is odd, then $\left(\frac{a}{p}\right) = 1$.

(b) Show that if $\left(\frac{a}{p}\right) = -1$, then $\operatorname{ord}_p(a)$ is even.

(c) Give an example where $p \equiv 1 \pmod{4}$ and where a is an integer of even order modulo p, such that $\left(\frac{a}{p}\right) = 1$.

(d) Show that if $p \equiv 3 \pmod{4}$ and $\operatorname{ord}_p(a) = 2k$ is an even number, then we have

$$\left(\frac{a}{p}\right) = -1,$$

and $a^k \equiv -1 \pmod{p}$. Thus, in the case where $p \equiv 3 \pmod{4}$, the quadratic residues are precisely the integers of odd order, and the quadratic nonresidues are precisely the integers of even order.

(e) Show that if $p \equiv 1 \pmod{4}$ and $\operatorname{ord}_p(a)$ is even, then $\left(\frac{a}{p}\right) = 1$ if and only if $\operatorname{ord}_p(a) \mid \frac{p-1}{2}$.

(f) Combine the previous parts to show that for any odd prime p and for any integer $a \not\equiv 0 \pmod{p}$, we have the identity

$$\left(\frac{a}{p}\right) = (-1)^{\frac{p-1}{\operatorname{ord}_p(a)}}.$$

The following table lists the value $(p-1)/\operatorname{ord}_p(q)$, with p and q ranging over the first ten odd primes (note that the value 1 indicates that q is a primitive root modulo p):

$p\backslash q$	3	5	7	11	13	17	19	23	29	31
3	-	1	2	1	2	1	2	1	1	2
5	1	-	1	4	1	1	2	1	2	4
7	1	1	-	2	3	1	1	2	6	1
11	2	2	1	-	1	1	1	10	1	2
13	4	3	1	1	-	2	1	2	4	3
17	1	1	1	1	4	-	2	1	1	1
19	1	2	6	6	1	2	-	2	1	3
23	2	1	1	1	2	1	1	-	2	2
29	1	2	4	1	2	7	1	4	-	1
31	1	10	2	1	1	1	2	3	3	-

14.4 Consider an odd prime p, an integer a, and an integer m. Show that

$$a^m \equiv \left(\frac{a}{p}\right) \pmod{p} \qquad \Leftrightarrow \qquad m \equiv \frac{p-1}{2} \pmod{\operatorname{ord}_p(a)}.$$

14.5 (a) Show that $\left(\frac{-3}{p}\right) = 1$ if and only if $p \equiv 1 \pmod{3}$.

(b) For any finite set $p_1, \ldots, p_k$ of primes congruent to 1 modulo 3, use the prime divisors of

$$n = (p_1 \cdots p_k)^2 + 3$$

to show that there is a prime $q \equiv 1 \pmod{3}$ that is not already among the primes $p_1, \ldots, p_k$. Conclude that there are infinitely many primes congruent to 1 modulo 3.

14.6 (a) Show that $\left(\frac{5}{q}\right) = \left(\frac{q}{5}\right)$ holds for every prime $q \neq 5$. In other words, which two congruence classes 1, 2, 3, or 4 modulo 5 contain the primes q such that 5 to be a quadratic residue modulo q?

(b) For any finite set of primes $p_1, \ldots, p_k$ congruent to 1 modulo 5, use the prime divisors of

$$(2p_1 \cdots p_k)^4 + 5$$

to show that there is a prime $q \equiv 1 \pmod{5}$ that is not already among the primes $p_1, \ldots, p_k$. Conclude that there are infinitely many primes congruent to 1 modulo 5.

14.7 Show that the polynomial $n^2 - 79n + 1601$ returns primes for all $0 \leq n \leq 79$.

Chapter 15

Quadratic Reciprocity

15.1 The Quadratic Character of 2

In this section we consider the question for which odd primes p we have that 2 is a quadratic residue. By Exercise 14.3, this is equivalent to the question for which primes p we have

$$\frac{p-1}{\mathrm{ord}_p(2)} \equiv 0 \pmod 2 .$$

In the following table we list the quantity $\frac{p-1}{\mathrm{ord}_p(2)}$ for the odd primes below 50:

p	3	5	7	11	13	17	19	23	29	31	37	41	43	47
$\frac{p-1}{\mathrm{ord}_p(2)}$	1	1	2	1	1	2	1	2	1	6	1	2	3	2.

This table shows that 2 is a quadratic residue modulo p for the primes

$$7,\ 17,\ 23,\ 31,\ 41,\ 47,\ \ldots .$$

Unlike the case where we were seeking to characterize the primes for which -1 is a primitive root, these primes are not organized according to their residue class modulo 4. The pattern is slightly more complicated. However, we may observe that the primes for which 2 is a quadratic residue all happen to be close to a multiple of 8, while the primes for which 2 is a quadratic nonresidue are further from the multiples of 8. This turns out to be the right idea: The number 2 is a quadratic residue modulo p if and only if p is of the form $8k \pm 1$.

Note also that 2 is a primitive root if and only if $\mathrm{ord}_p(2) = p - 1$, so we see from the previous table that 2 is a primitive root modulo p for the primes

$$3,\ 5,\ 11,\ 13,\ 19,\ 29,\ 37,\ \ldots .$$

However, it is not quite true that 2 is a primitive root modulo p whenever it is a quadratic nonresidue modulo p: The first prime for which 2 is neither a quadratic residue nor a primitive root is 43.

Theorem 15.1.1. *The Legendre symbol of* 2 *modulo* p *satisfies the identity*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

In other words, 2 *is a quadratic residue if and only if* $p \equiv \pm 1 \pmod 8$*, and it is a quadratic nonresidue if and only if* $p \equiv \pm 3 \pmod 8$*.*

Before we start the proof, recall that one of the proofs of Fermat's Little Theorem was obtained by multiplying all the numbers $1, \ldots, p-1$ with an integer a to obtain the congruence

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod p.$$

Since $(p-1)!$ isn't divisible by p, it is invertible modulo p and thus we found that $a^{p-1} \equiv 1 \pmod p$.

Proof. In this proof, we will write $P := \frac{p-1}{2}$ so that $2P = p-1$. By Euler's Criterion we have $\left(\frac{2}{p}\right) \equiv 2^P \pmod p$, which gives us the congruences

$$\left(\tfrac{2}{p}\right)P! \equiv 2^P P! \equiv 2 \cdot 4 \cdot \cdots \cdot (2P) \pmod p$$

We proceed by carefully analyzing this product of all the even positive integers below p. If we define the numbers

$$s := \left\lfloor \frac{p-1}{4} \right\rfloor \qquad \text{and} \qquad t := \left\lfloor \frac{p+1}{4} \right\rfloor,$$

so that $s + t = P$, then we can partition the set of all even positive integers below p into two subsets

$$S := \{2k \mid 1 \le k \le s\} \qquad \text{and} \qquad T := \{2k \mid s+1 \le k \le P\},$$

so that S consists precisely of all the even positive integers up to and possibly including P, and T consists of all the larger even integers up to and including $2P$. Note that the set S has s elements and the set T has exactly t elements.

The key observation is that for every integer y in T there is a unique *odd* integer $1 \le x \le P$ so that $y = p - x$, which determines a bijection. Thus, taking the product modulo p of all the even integers up to p amounts to taking the product of all the integers $1 \le x \le P$ with a sign change for every odd integer, of which there are t. In other words, we have the congruence

$$\left(\tfrac{2}{p}\right)P! \equiv 2 \cdot 4 \cdot \cdots \cdot 2P \equiv (-1)^t P! \pmod p.$$

Since $P!$ is not divisible by p we apply the cancellation law to obtain

$$\left(\frac{2}{p}\right) = (-1)^t.$$

We conclude the proof by observing that the parity of t is determined according to the following cases:

$$t \equiv \begin{cases} 0 \pmod 2 & \text{if } p \equiv \pm 1 \pmod 8 \\ 1 \pmod 2 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

□

The previous theorem is also known as the *Second Supplement* of the law of Quadratic Reciprocity. An immediate corollary of the second supplement is that

$$\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \text{ or } 3 \pmod 8 \\ -1 & \text{if } p \equiv 5 \text{ or } 7 \pmod 8 \end{cases}.$$

It is quite a useful theorem by itself: We will use it to show that there are infinitely many primes of the form $8k + r$ for any odd number $0 < r < 8$. We have already seen in Theorem 13.2.1 that there are infinitely many primes congruent to 1 modulo 2^n. In particular, there are infinitely many primes of the form $8k + 1$. It remains prove the analogous claims for the residue classes 3, 5, and 7 modulo 8. We will be following Sierpiński [Sie88].

Theorem 15.1.2. *There are infinitely many primes congruent to* 3 *modulo* 8.

Proof. If a is an odd integer, then $a^2 \equiv 1 \pmod 8$ so that

$$N_a := a^2 + 2$$

is congruent to 3 modulo 8. Not all prime divisors of N_a can be congruent to ± 1 modulo 8, since if they were, then N_a would likewise be congruent to ± 1 modulo 8. Therefore N_a has a prime divisor q so that

$$q \equiv \pm 3 \pmod 8.$$

Since $q \mid a^2 + 2$, it follows that -2 is a quadratic residue modulo q, so $q \not\equiv -3 \pmod 8$. Thus it follows that $q \equiv 3 \pmod 8$.

For a suitable choice of a, let p_n be the nth prime number, so that $p_1 = 2$, $p_2 = 3$, $p_3 = 5$ and so forth. Then we define

$$a = p_2 p_3 \cdots p_n,$$

so that any prime divisor of N_a is larger than p_n. Since N_a has a prime divisor congruent to 3 modulo 8, it follows that there are arbitrarily large such primes. □

Theorem 15.1.3. *There are infinitely many primes congruent to* 5 *modulo* 8.

Proof. If a is an odd integer, then $a^2 \equiv 1 \pmod 8$ so that

$$N_a := a^2 + 4$$

is congruent to 5 modulo 8. As before, there must be a prime divisor q of N_a so that

$$q \equiv \pm 3 \pmod 8.$$

Since $q \mid a^2 + 4$, it follows that -4 is a quadratic residue modulo q. This implies that -1 is a quadratic residue modulo q, so that $q \equiv 1 \pmod 4$. Since we assumed that $q \not\equiv 1 \pmod 8$, it follows that $q \equiv -3 \pmod 8$.

For a suitable choice of a, define

$$a = p_2 p_3 \cdots p_n$$

so that any prime divisor of N_a is larger than p_n. Since N_a has a prime divisor congruent to 5 modulo 8, it follows that there are arbitrarily large such primes. □

Theorem 15.1.4. *There are infinitely many primes congruent to* 7 *modulo* 8.

Proof. The idea of the proof is to find a natural number N of the form

$$N = 2a^2 - b^2,$$

so that $N \equiv -1 \pmod 8$ and N is relatively prime to any $m \leq n$. Assuming that $n > 1$, a suitable choice of a and b is $a := n!$ and $b := 1$. Indeed, in this case a is even, so that $2a^2 - 1 \equiv -1 \pmod 8$.

Under these assumptions, N will have a prime factor $q \not\equiv 1 \pmod 8$, and furthermore we will have

$$2a^2 \equiv b^2 \pmod q.$$

Since a^2 and b^2 are clearly quadratic residues modulo q, it follows that 2 is a quadratic residue, implying that $q \equiv \pm 1 \pmod 8$. However, since we have chosen q so that $q \not\equiv 1 \pmod 8$, it follows that $q \equiv -1 \pmod 8$. Furthermore, since q is relatively prime to any $m \leq n$, it follows that $n < q$, which shows that there are arbitrarily large primes $q \equiv -1 \pmod 8$. □

15.2 The Statement of Quadratic Reciprocity

We have already established several facts about Legendre symbols. The Legendre symbol is multiplicative, meaning that

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

The Legendre symbol $\left(\frac{-1}{p}\right)$ can be computed according to whether p is congruent to 1 or 3 modulo 4:

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4 \\ -1 & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

The Legendre symbol $\left(\frac{2}{p}\right)$ can likewise be computed according to whether p is congruent to ± 1 or ± 3 modulo 8:

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

We need one final piece to effectively compute Legendre symbols: a way to determine the Legendre symbol $\left(\frac{q}{p}\right)$ when q is an odd prime. The recipe for determining $\left(\frac{q}{p}\right)$ is given by the law of quadratic reciprocity.

To build some intuition for what the law of quadratic reciprocity should assert, we recall from Exercise 14.3 that the Legendre symbol of a modulo p can be computed by

$$\left(\frac{a}{p}\right) = (-1)^{\frac{p-1}{\operatorname{ord}_p(a)}}.$$

$p\backslash q$	3	5	7	11	13	17	19	23	29	31
3		−	+	−	+	−	+	−	−	+
5	−		−	+	−	−	+	−	+	+
7	−	−		+	−	−	−	+	+	−
11	+	+	−		−	−	−	+	−	+
13	+	−	−	−		+	−	+	+	−
17	−	−	−	−	+		+	−	−	−
19	−	+	+	+	−	+		+	−	−
23	+	−	−	−	+	−	−		+	+
29	−	+	+	−	+	−	−	+		−
31	−	+	+	−	−	−	+	−	−	

Table 15.1: The sign of the Legendre symbol $\left(\frac{q}{p}\right)$, ranging over pairs of distinct odd primes p and q.

Thus, the Legendre symbol of a modulo p depends only of the parity of the quantity

$$\frac{p-1}{\operatorname{ord}_p(a)},$$

which gives us a fairly quick way of determining the Legendre symbol $\left(\frac{q}{p}\right)$ for small odd primes p and q. We listed signs of the Legendre symbols of q modulo p in Table 15.1.

The patterns in this table might not be immediately apparent. However, those who did Exercise 14.6 might spot that column 5 is identical to row 5. Looking for more identical columns and rows, we create a new table by comparing the entry in row i and column j in Table 15.1 with the entry in row j and column i. When these two entries are equal, we mark it with a solid black circle (●), and if they are different we mark it with an open circle (○). This results in the following table:

$p\backslash q$	3	5	7	11	13	17	19	23	29	31
3		●	○	○	●	●	○	○	●	○
5	●		●	●	●	●	●	●	●	●
7	○	●		○	●	●	○	○	●	○
11	○	●	○		●	●	○	○	●	○
13	●	●	●	●		●	●	●	●	●
17	●	●	●	●	●		●	●	●	●
19	○	●	○	○	●	●		○	●	○
23	○	●	○	○	●	●	○		●	○
29	●	●	●	●	●	●	●	●		●
31	○	●	○	○	●	●	○	○	●	

Using this table, we quickly spot that columns 5, 13, 17, and 29 are identical to rows 5, 13, 17, and 29, respectively, which means that these columns were also identical to

their respective rows in Table 15.1. Something special is going on with the primes

$$5,\ 13,\ 17,\ 29,$$

and something seems to be off about the primes

$$3,\ 7,\ 11,\ 19,\ 23,\ 31.$$

We are very familiar with this grouping of the primes by now: The first set of primes are all congruent to 1 modulo 4, while the second set of primes are all congruent to 3 modulo 4. These observations suggest the following:

1. If at least one of p and q is congruent to 1 modulo 4, then we have
$$\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right).$$
2. If both p and q are congruent to 3 modulo 4, then we have
$$\left(\frac{q}{p}\right) = -\left(\frac{p}{q}\right).$$

Legendre and Gauss spotted these patterns towards the end of the 18th century. They formulated it succinctly as follows:

Theorem 15.2.1 (The Law of Quadratic Reciprocity)**.** *For any two distinct odd primes p and q we have*
$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Example 15.2.2. The law of quadratic reciprocity is incredibly effective in computations of Legendre symbols. For example, since 17 and 101 are congruent to 1 modulo 4, we have
$$\left(\frac{17}{101}\right) = \left(\frac{101}{17}\right) = \left(\frac{16}{17}\right) = \left(\frac{4}{17}\right)\left(\frac{4}{17}\right) = 1.$$
Indeed, it turns out that $44^2 \equiv 17$ and $57^2 \equiv 17$ modulo 101.

On the other hand, both 23 and 43 are congruent to 3 modulo 4, so the law of quadratic reciprocity tells us that

$$\left(\frac{23}{43}\right) = -\left(\frac{43}{23}\right) = -\left(\frac{20}{23}\right) = -\left(\frac{4}{23}\right)\left(\frac{5}{23}\right) = -\left(\frac{5}{23}\right) = -\left(\frac{23}{5}\right) = -\left(\frac{3}{5}\right) = 1.$$

In this case, it turns out that $18^2 \equiv 23$ and $25^2 \equiv 23$ modulo 43.

As a final example, for good measure, note that 31 and 83 are both congruent to 3 modulo 4. Using the law of quadratic reciprocity we obtain that

$$\left(\frac{31}{83}\right) = -\left(\frac{83}{31}\right) = -\left(\frac{21}{31}\right) = -\left(\frac{3}{31}\right)\left(\frac{7}{31}\right) = \left(\frac{31}{3}\right)\left(\frac{31}{7}\right) = \left(\frac{1}{3}\right)\left(\frac{3}{7}\right) = -\left(\frac{7}{3}\right) = -1.$$

15.3 Gauss's Lemma

While it is customary to consider the standard residue classes modulo an odd prime p to be the integers $0 \leq r < p$, we have now the occasion to consider the complete residue system of integers between $-\frac{p-1}{2}$ and $\frac{p-1}{2}$.

Definition 15.3.1. The *least absolute residue* of an integer a modulo a natural number n is the unique integer

$$-\frac{n}{2} < x \leq \frac{n}{2}$$

such that $x \equiv a \pmod{n}$. We will write $\lfloor a \rceil_n$ for the least absolute residue of a modulo n, and we will write $\|\lfloor a \rceil\|_n$ for its absolute value.

Thus, the least absolute residue of an integer a modulo an odd prime p, which is what we will be using this concept for, is the unique integer

$$-\frac{p-1}{2} \leq x \leq \frac{p-1}{2}$$

such that $x \equiv a \pmod{p}$. For example, the least absolute residue of 4 modulo 7 is -3, the least absolute residue of 6 modulo 13 is 6, and the least absolute residue of 12 modulo 13 is -1. In the following table, we list the least absolute residues modulo 11:

a	0	1	2	3	4	5	6	7	8	9	10
$\lfloor a \rceil_{11}$	0	1	2	3	4	5	-5	-4	-3	-2	-1.

If we consider an integer a relatively prime to an odd prime p, then multiplication by a defines a bijection on $\mathbb{Z}/p\mathbb{Z}$. It follows that the function

$$x \mapsto \lfloor ax \rceil_p : \{0, \ldots, p-1\} \to \{-\tfrac{p-1}{2}, \ldots, \tfrac{p-1}{2}\}$$

is also a bijection. Restricting this bijection to the set $\{1, \ldots, \frac{p-1}{2}\}$, we obtain an injective function

$$x \mapsto \lfloor ax \rceil_p : \{1, \ldots, \tfrac{p-1}{2}\} \to \{-\tfrac{p-1}{2}, \ldots, \tfrac{p-1}{2}\}.$$

The following lemma shows that every $1 \leq x \leq \frac{p-1}{2}$ occurs exactly once as the absolute value of any number of the form $\lfloor ax \rceil_p$.

Lemma 15.3.2. *The function*

$$r_a : \{1, \ldots, \tfrac{p-1}{2}\} \to \{1, \ldots, \tfrac{p-1}{2}\}$$

given by $r_a(x) := \|\lfloor ax \rceil\|_p$ is a bijection.

Proof. Since r_a is a function between two sets of finite size $\frac{p-1}{2}$, it suffices to show that r_a is injective. To this end, consider x and y such that $r_a(x) = r_a(y)$. Since the function $x \mapsto \lfloor ax \rceil_p$ is injective, this implies that

$$\lfloor ax \rceil_p = -\lfloor ay \rceil_p$$

or, equivalently, that $ax - ay$ is divisible by p. Since a is relatively prime to p, this implies that $x - y$ is divisible by p, i.e., that $x = y$. □

Lemma 15.3.3 (Gauss's Lemma)**.** *Consider an odd prime p and an integer such that $p \nmid a$. Define μ to be the number of elements x among the residue classes*

$$a,\ 2a,\ 3a,\ \ldots,\ \tfrac{p-1}{2}a \pmod{p},$$

such that $\lfloor ax \rceil_p$ is negative. Then we have

$$\left(\frac{a}{p}\right) = (-1)^{\mu}.$$

Proof. Write $P := (p-1)/2$. Then we have that $p \nmid P!$, so suffices to show that

$$\left(\frac{a}{p}\right) P! \equiv (-1)^{\mu} P! \pmod{p}.$$

By Euler's Criterion we have $\left(\frac{a}{p}\right) \equiv a^P \pmod{p}$, so we obtain that

$$\left(\frac{a}{p}\right) P! \equiv a^P P! \pmod{p}.$$

Redistributing the P factors of a of the exponent a^P over $P!$, and using Lemma 15.3.3, we obtain:

$$a^P P! = (1a)(2a)\cdots(Pa) \equiv (-1)^{\mu} P! \pmod{p}. \qquad \square$$

15.4 Eisenstein's Proof of The Law of Quadratic Reciprocity

Gauss was the first to prove the law of quadratic reciprocity in 1796, when he was only 19 years old. This proof was included in his seminal work, *Disquisitiones Arithmeticae*, published in 1801. Gauss was very fond of this result, and throughout his life-time he would publish seven different proofs.

The proof we will present here is due to Gotthold Eisenstein, which he published in 1844. It is celebrated for its simplicity and for the insightful perspective it provides on geometric aspects of the quadratic reciprocity law.

In Eisenstein's proof of the law of quadratic reciprocity, we will make use of the *floor function* $x \mapsto \lfloor x \rfloor$, which is a function from the rational numbers or the real numbers, to the integers. The integer

$$\lfloor x \rfloor$$

is the largest integer below x. In other words, its defining property is that the floor function satisfies the logical equivalence

$$k \leq \lfloor x \rfloor \qquad \Leftrightarrow \qquad k \leq x$$

for every integer k and every rational or real number x. Concretely, we have the following examples:

$$\left\lfloor \frac{3}{2} \right\rfloor = 1, \qquad \left\lfloor \frac{29}{31} \right\rfloor = 0, \qquad \text{and} \qquad \left\lfloor -\frac{1}{4} \right\rfloor = -1.$$

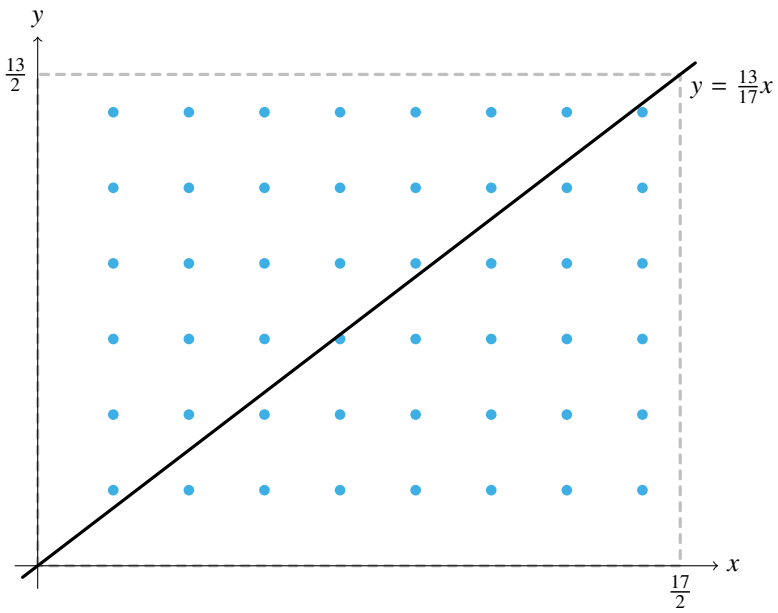


Figure 15.1: In Eisenstein's proof of quadratic reciprocity, we count the $\frac{p-1}{2}$ by $\frac{q-1}{2}$ lattice points in the marked rectangle as the sum of the number lattice points below the diagonal and the number of lattice points above the diagonal.

Eisenstein's Proof of The Law of Quadratic Reciprocity. By Gauss's Lemma it suffices to show that

$$(-1)^{\mu(q,p)}(-1)^{\mu(p,q)} = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

In other words, we have to show that

$$\mu(q,p) + \mu(p,q) \equiv \frac{p-1}{2}\frac{q-1}{2} \pmod 2.$$

This suggests that there should be a way of splitting a set of $\frac{p-1}{2}\frac{q-1}{2}$ points into two subsets S and T, so that their respective number of elements have the same parities as the quantities $\mu(q,p)$ and $\mu(p,q)$. There is indeed such a way. Figure 15.1 we shows Eisenstein's setup in case of the primes $p = 17$ and $q = 13$.

Consider the set of $\frac{p-1}{2}$ by $\frac{q-1}{2}$ lattice points in the positive quadrant of the plane; that is, the points with positive integer coordinates within the rectangle spanned by the four points

$$(0,0),\ (\tfrac{p}{2},0),\ (\tfrac{p}{2},\tfrac{q}{2}),\ \text{and}\ (0,\tfrac{q}{2}).$$

We partition these lattice points by the diagonal of the rectangle: The set of lattice points below the diagonal is S, and the set of lattice points above the diagonal is T. Since the diagonal of this rectangle is given by the line $y = \frac{q}{p}x$, there is no point with positive integer coordinates on the diagonal, so there is no ambiguity as to which part a lattice point belongs to.

Now observe that the number $|S|$ of lattice points in the set S is given by

$$|S| = \sum_{k=1}^{\frac{p-1}{2}} \left\lfloor \frac{kq}{p} \right\rfloor.$$

This formula is obtained by first recognizing that for each $1 \leq k \leq \frac{p-1}{2}$, a lattice point of the form (k, i) is in the set S if and only if

$$1 \leq i \leq \frac{q}{p}k.$$

The number of such lattice points is $\lfloor \frac{kq}{p} \rfloor$, so the total number of elements in S is obtained by summing up all of these totals. Similarly, the number $|T|$ of lattice points in the set T is given by

$$|T| = \sum_{k=1}^{\frac{q-1}{2}} \left\lfloor \frac{kp}{q} \right\rfloor.$$

Thus, it remains to prove that

$$\sum_{k=1}^{\frac{p-1}{2}} \left\lfloor \frac{kq}{p} \right\rfloor \equiv \mu(q,p) \pmod 2 \quad \text{and} \quad \sum_{k=1}^{\frac{q-1}{2}} \left\lfloor \frac{kp}{q} \right\rfloor \equiv \mu(p,q) \pmod 2.$$

We will verify these congruences separately in the following lemma. □

Lemma 15.4.1. *Consider an odd prime p and an odd integer a not divisible by p, and let $\mu(a,p)$ be the quantity defined in Gauss's Lemma. Then*

$$\sum_{k=1}^{\frac{p-1}{2}} \left\lfloor \frac{ka}{p} \right\rfloor \equiv \mu(a,p) \pmod 2.$$

Proof. For each k there are unique integers q_k and r_k such that $ka = q_k p + r_k$, where

$$-\frac{p-1}{2} \leq r_k \leq \frac{p-1}{2}.$$

It follows that

$$\left\lfloor \frac{ka}{p} \right\rfloor = \begin{cases} q_k & \text{if } r_k > 0 \\ q_k - 1 & \text{if } r_k < 0. \end{cases}$$

Adding up the integers $\lfloor \frac{ka}{p} \rfloor$ therefore amounts to adding up the integers q_k and subtracting the number of integers k such that r_k is negative:

$$\sum_{k=1}^{\frac{p-1}{2}} \left\lfloor \frac{ka}{p} \right\rfloor = \sum_{k=1}^{\frac{p-1}{2}} q_k - \mu(a,p).$$

To complete the proof, it therefore suffices to show that

$$\sum_{k=1}^{\frac{p-1}{2}} q_k \equiv 0 \pmod 2 .$$

To prove this, recall from Lemma 15.3.2 that the map $k \mapsto |r_k|$ is a permutation of the set $\{1, \ldots, \frac{p-1}{2}\}$. In other words, every integer from 1 to $\frac{p-1}{2}$ occurs exactly once as an integer of the form $|r_k|$. Since the parity of an integer is unchanged by taking its absolute value, we have

$$\sum_{k=1}^{\frac{p-1}{2}} r_k \equiv \sum_{k=1}^{\frac{p-1}{2}} |r_k| \equiv \sum_{k=1}^{\frac{p-1}{2}} k \pmod 2 .$$

On the other hand, observe that since a and p are assumed to be odd, we have the congruences

$$k \equiv ka = q_k p + r_k \equiv q_k + r_k \pmod 2 .$$

Thus we obtain the congruences

$$\sum_{k=1}^{\frac{p-1}{2}} k \equiv \sum_{k=1}^{\frac{p-1}{2}} q_k + \sum_{k=1}^{\frac{p-1}{2}} r_k \equiv \sum_{k=1}^{\frac{p-1}{2}} q_k + \sum_{k=1}^{\frac{p-1}{2}} k \pmod 2 .$$

Subtracting $\sum_k k$ from both sides, we find that $\sum_k q_k \equiv 0 \pmod 2$. □

Exercises

15.1 Compute the values of the following Legendre symbols:

$$\left(\frac{101}{163}\right), \quad \left(\frac{137}{359}\right), \quad \left(\frac{113}{479}\right), \quad \text{and} \quad \left(\frac{139}{953}\right).$$

15.2 Determine whether the following quadratic congruences have solutions:

(a) $x^2 + x + 1 \equiv 0 \pmod{13}$.
(b) $x^2 + x + 3 \equiv 0 \pmod 7$.
(c) $2x^2 + 3x + 5 \equiv 0 \pmod{17}$.
(d) $x^2 + 4x + 8 \equiv 0 \pmod{23}$.
(e) $2x^2 + 3x + 4 \equiv 0 \pmod{11}$.
(f) $4x^2 + 7x + 14 \equiv 0 \pmod{31}$.

15.3 Show that for any prime $p > 3$, the Legendre symbol of 3 is determined according to the following cases:

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod{12} \\ -1 & \text{if } p \equiv \pm 5 \pmod{12} . \end{cases}$$

15.4 (a) For any integer $a > 2$, let p be a prime factor of $a^4 - a^2 + 1$. Show that

$$\operatorname{ord}_p(a) = 12.$$

(b) Use a suitable choice of a in

$$N_a := a^6 + 1$$

to show that there are infinitely many primes congruent to 1 modulo 12.

(c) For each r in the set $\{5, 7, 11\}$, show that there are infinitely many primes congruent to r modulo 12.

Chapter 16

Arithmetic Functions

16.1 Multiplicative Functions

A function f defined on the positive integers with output in any number system, such as the integers, real numbers, or complex numbers, is called an *arithmetic function*. For example, Euler's totient function ϕ, which returns for each n the number of elements $0 < k < n$ relatively prime to n, is an arithmetic function. Some other important arithmetic functions include the *number of divisors function* τ and the *sum of divisors function* σ:

$$\tau(n) := \sum_{d|n} 1 \qquad \text{and} \qquad \sigma(n) := \sum_{d|n} d.$$

Note that the number of divisors and the sum of divisors are not defined for $n = 0$. Arithmetic functions are typically only defined on the positive integers $\{1, 2, 3, \ldots\}$. Even for functions which are defined on 0, such as Euler's totient function, when we consider them as arithmetic functions we will consider their domain of definition to be the positive integers.

Other functions, which are worth naming are the *constant function* $n \mapsto 1$, which we will simply denote by 1, and the *identity function* $n \mapsto n$, which we will denote by id. In the pages that follow, we shall encounter two more arithmetic functions: the *unit function* ε and the *Möbius function* μ.

Definition 16.1.1. An arithmetic function f is said to be *multiplicative* if for any two relatively prime natural numbers m and n we have

$$f(mn) = f(m)f(n).$$

Proposition 16.1.2. *If f is a nonzero multiplicative function, then $f(1) = 1$.*

Proof. Since f is assumed to be nonzero, there is a natural number n such that $f(n) \neq 0$. Thus,

$$f(n) = f(1n) = f(1)f(n)$$

by the assumption that f is multiplicative. Since multiplication by a nonzero number is injective, it follows that $f(1) = 1$. □

It follows that the constant function 1 is the only constant multiplicative function. By the following proposition, we see that any multiplicative function is determined by its values on prime powers.

Proposition 16.1.3. *If f is a multiplicative function and $n = p_1^{m_1} \cdots p_k^{m_k}$, where the primes $p_1, \ldots, p_k$ are distinct, then*

$$f(n) = f(p_1^{m_1}) \cdots f(p_k^{m_k}).$$

Proof. Since the primes $p_1, \ldots, p_k$ are assumed to be distinct, it follows that the prime powers $p_1^{m_1}, \ldots, p_k^{m_k}$ are pairwise relatively prime, hence the claim follows by induction on k. □

The following theorem provides a useful way of proving that a function is multiplicative.

Theorem 16.1.4. *Suppose f is a multiplicative function, and define the function F by*

$$F(n) := \sum_{d|n} f(d).$$

Proof. Consider two relatively prime natural numbers m and n. Then there is a bijection $(c, d) \mapsto cd$ from the set

$$\{c \mid c \text{ divides } m\} \times \{d \mid d \text{ divides n}\}$$

to the set of divisors of mn. For the inverse function, define the function

$$e \mapsto (\gcd(e, m), \gcd(e, n)),$$

which sends a divisor e of mn to a pair of divisors of m and n. To see that this is indeed an inverse, note that if $c \mid m$ and $d \mid n$, then $\gcd(d, m) = 1$ so that

$$c = \gcd(cd, m).$$

Similarly, we have that $\gcd(c, n) = 1$ so that $d = \gcd(cd, n)$. This shows that

$$(c, d) = (\gcd(cd, m), \gcd(cd, n)).$$

We also have to show that $\gcd(e, m)\gcd(e, n) = e$ for any divisor e of mn. Write $c = \gcd(e, m)$ and $d = \gcd(e, n)$. Since c and d are relatively prime and since both are divisors of e, it follows that $cd \mid e$. Now, suppose that $kcd = e$. Since $\gcd(kcd, m) = c$ it follows that $\gcd(k, m) = 1$. Similarly, $\gcd(k, n) = 1$. However, if p is any prime divisor of k, which divides mn, it follows that $p \mid m$ or $p \mid n$. Since this is impossible, it follows that $k = 1$. This completes the proof that the map $(c, d) \mapsto cd$ is a bijection.

Using this bijection, we find that

$$\begin{aligned} F(mn) = \sum_{e|mn} f(e) &= \sum_{c|m} \sum_{d|n} f(cd) \\ &= \sum_{c|m} \sum_{d|n} f(c)f(d) = \sum_{c|m} f(c) \sum_{d|n} f(d) = F(m)F(n). \end{aligned}$$ □

Theorem 16.1.5. *The functions τ and σ are multiplicative.*

Proof. The functions τ and σ are defined by

$$\tau(n) := \sum_{d|n} 1, \qquad \text{and} \qquad \sigma(n) := \sum_{d|n} d.$$

Clearly the constant function $n \mapsto 1$ and the identity function $n \mapsto n$ are multiplicative, so it follows from Theorem 16.1.4 that τ and σ are multiplicative. □

The previous theorem implies that if $n = p_1^{m_1} \cdots p_k^{m_k}$ is a product of distinct primes, then

$$\tau(n) = \prod_{i=1}^{k} (m_i + 1), \qquad \text{and} \qquad \sigma(n) = \prod_{i=1}^{k} \frac{p_i^{m_1+1} - 1}{p_i - 1}.$$

Indeed, for any prime power p^m, the divisors of p^m are the prime powers p^i where $0 \le i \le m$. There are $m + 1$ powers of p in this range, and their sum is computed by the formula for the geometric series

$$1 + p + \cdots + p^m = \frac{p^{m+1} - 1}{p - 1}.$$

16.2 Dirichlet Convolution

The set of arithmetic functions with values taken in a fixed, chosen number system such as the integers, the real numbers, or the complex numbers, possesses an interesting algebraic structure that can be used effectively to obtain many algebraic results and identities about numbers. The most important operation on arithmetic functions is *Dirichlet convolution*, a way of combining two arithmetic functions f and g into a new arithmetic function $f * g$.

Definition 16.2.1. Consider two arithmetic functions f and g. We define their *Dirichlet convolution* $f * g$ by

$$(f * g)(n) := \sum_{d|n} f(d) g\left(\frac{n}{d}\right).$$

Theorem 16.2.2. *Dirichlet convolution is associative and commutative, meaning that*

$$(f * g) * h = f * (g * h), \qquad \textit{and} \qquad f * g = g * f,$$

respectively.

Proof. Another way of writing the Dirichlet convolution of f and g is

$$(f * g)(n) = \sum_{\substack{c,d \\ cd=n}} f(c) g(d).$$

Since this expression is symmetric in f and g, we see immediately that the Dirichlet convolution is commutative. Furthermore, to see that the Dirichlet convolution is associative we make the following calculation:

$$\begin{aligned}
((f*g)*h)(n) &= \sum_{\substack{c,d\\cd=n}} (f*g)(c)h(d) \\
&= \sum_{\substack{c,d\\cd=n}} \sum_{\substack{a,b\\ab=c}} f(a)g(b)h(d) \\
&= \sum_{\substack{a,b,d\\abd=n}} f(a)g(b)h(d) \\
&= \sum_{\substack{a,e\\ae=n}} \sum_{\substack{b,d\\bd=e}} f(a)g(b)h(d) \\
&= \sum_{\substack{a,e\\ae=n}} f(a)(g*h)(e) \\
&= (f*(g*h))(n). \qquad \square
\end{aligned}$$

Theorem 16.2.3. *Let ε be the function given by*

$$\varepsilon(n) := \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $\varepsilon * f = f$ and $f * \varepsilon = f$ for any arithmetic function f.

Proof. Consider an arithmetic function f. Then

$$(\varepsilon * f)(n) = \sum_{d|n} \varepsilon(d) f\left(\frac{n}{d}\right) = f(n),$$

since the only term contributing to the sum is $d = 1$. The fact that $f * \varepsilon = f$ follows by commutativity of the Dirichlet convolution. $\square$

16.3 The Möbius Inversion Formula

Definition 16.3.1. The *Möbius function* μ is defined by

$$\mu(n) := \begin{cases} (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes,} \\ 0 & \text{if } p^2 \mid n \text{ for some prime } p. \end{cases}$$

For example, $\mu(3 \cdot 7 \cdot 17^3) = 0$ because the input is divisible by 17^2. On the other hand, $\mu(3 \cdot 7 \cdot 17) = (-1)^3 = -1$ because the input is the product of three distinct primes. The following table lists the values of $\mu(n)$ for the first 15 natural numbers

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\mu(n)$	0	1	−1	−1	0	−1	1	−1	0	0	1	−1	0	−1	1

Theorem 16.3.2. *The Möbius function is multiplicative, and it satisfies*

$$\sum_{d|n} \mu(d) = \varepsilon(n).$$

In other words, $\mu * 1 = \varepsilon$.

Proof. It is immediate from the definition that the Möbius function is multiplicative. Likewise, the arithmetic function

$$F(n) := \sum_{d|n} \mu(d)$$

is multiplicative by Theorem 16.1.4. Since multiplicative functions are completely determined by their values on prime powers, and ε is the multiplicative function given by

$$\varepsilon(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise,} \end{cases}$$

it suffices to show that $F(p^k) = 0$ for any $k \geq 1$. To see this, we simply evaluate

$$\sum_{d|p^k} \mu(p^k) = \mu(p^0) + \mu(p^1) + 0 = 1 - 1 = 0. \qquad \square$$

Remark 16.3.3. The previous theorem could be viewed as a recurrence relation that is satisfied by the Möbius function:

$$\mu(n) = -\sum_{\substack{d|n \\ d \neq n}} \mu(d)$$

Gian-Carlo Rota used this perspective to make the theory of arithmetic functions and the Möbius function applicable to a wide class of partially ordered sets [Rot64].

Theorem 16.3.4 (The Möbius Inversion Formula)**.** *Consider two arithmetic functions* f *and* g*. Then we have*

$$g(n) = \sum_{d|n} f(d) \qquad \Leftrightarrow \qquad f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right).$$

In other words, if $g = 1 * f$ *then* $f = \mu * g$.

Proof. If $g = 1 * f$, then

$$\begin{aligned} \mu * g &= \mu * (1 * f) \\ &= (\mu * 1) * f \\ &= \varepsilon * f \\ &= f. \end{aligned}$$

Conversely, if $f = \mu * g$, then

$$\begin{aligned} 1 * f &= 1 * (\mu * g) \\ &= (1 * \mu) * g \\ &= \varepsilon * g \\ &= g. \end{aligned}$$

□

Corollary 16.3.5. *Euler's totient function satisfies the identity*

$$\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}$$

In other words, $\phi = \mu * \mathrm{id}$.

Proof. By Theorem 13.3.1 we have $\mathrm{id} = 1 * \phi$, so we have $\phi = \mu * \mathrm{id}$ by the Möbius inversion formula. □

We also give a direct proof of Corollary 16.3.5, offering a second perspective on the totient function.

Direct proof of Corollary 16.3.5. Note that Euler's totient function ϕ can be written in the form

$$\phi(n) = \sum_{k=1}^{n} \left\lfloor \frac{1}{\gcd(k,n)} \right\rfloor.$$

Then it follows from Theorem 16.3.2 that

$$\phi(n) = \sum_{k=1}^{n} \sum_{d|\gcd(k,n)} \mu(d) = \sum_{k=1}^{n} \sum_{\substack{d|n \\ d|k}} \mu(d) = \sum_{d|n} \sum_{\substack{k=1 \\ d|k}}^{n} \mu(d) = \sum_{d|n} \mu(d) \frac{n}{d}.$$

The last step follows, since there are exactly $\frac{n}{d}$ multiples of d among the numbers from 1 to n. □

16.4 Dirichlet Inverses

Definition 16.4.1. Consider an arithmetic function f such that $f(1)$ is nonzero. Then we define the *Dirichlet inverse* f^{-1} of f by

$$f^{-1}(1) := \frac{1}{f(1)} \qquad \text{and} \qquad f^{-1}(n) = \frac{-1}{f(1)} \sum_{\substack{d|n \\ d \neq n}} f^{-1}(d) f\left(\frac{n}{d}\right) \qquad (\text{for } n > 1).$$

Theorem 16.4.2. *For every arithmetic function f such that $f(1)$ is nonzero, the arithmetic function f^{-1} is the inverse of f with respect to Dirichlet involution, in the sense that*

$$f^{-1} * f = \varepsilon \qquad \textit{and} \qquad f * f^{-1} = \varepsilon.$$

Furthermore, f^{-1} is the unique such arithmetic function.

Proof. To show that $(f * f^{-1})(n) = \varepsilon(n)$ for every $n \geq 1$, there are two cases to consider. In the case where $n = 1$, we have

$$(f * f^{-1})(1) = f(1)f^{-1}(1) = f(1)\frac{1}{f(1)} = 1 = \varepsilon(1).$$

In the case where $n > 1$, we have

$$(f * f^{-1})(n) = \sum_{d|n} f(d)f^{-1}\left(\frac{n}{d}\right) = f(1)f^{-1}(n) + \sum_{\substack{d|n \\ d\neq n}} f(d)f^{-1}\left(\frac{n}{d}\right),$$

which is 0 by the definition of f^{-1}. This completes the proof that $f * f^{-1} = \varepsilon$. The equality $f^{-1} * f = \varepsilon$ follows by commutativity of Dirichlet convolution.

Uniqueness of the inverse follows from a purely algebraic argument. Suppose that g is another arithmetic function such that $f * g = \varepsilon$. Then we obtain

$$f^{-1} = f^{-1} * \varepsilon = f^{-1} * (f * g) = (f^{-1} * f) * g = \varepsilon * g = g. \qquad \square$$

Example 16.4.3. The Dirichlet inverse of the constant function 1 is the Möbius function μ. Indeed, we saw in Theorem 16.3.2 that $\mu * 1 = \varepsilon$.

Remark 16.4.4. If $f^{-1} = g$ for some arithmetic functions f and g, then it follows that $g^{-1} = f$. Indeed, the equations

$$f * g = \varepsilon \qquad \text{and} \qquad g * f = \varepsilon$$

simultaneously imply that g is the inverse of f and f is the inverse of g. Thus we have that $1^{-1} = \mu$ and $\mu^{-1} = 1$.

Exercises

16.1 Show that the number $\tau(n)$ of divisors of n is odd if and only if n is a square number.

16.2 Compute $\mu * \mu$.

16.3 Prove that

$$\sum_{k=1}^{n} \gcd(k, n)\mu(\gcd(k, n)) = \mu(n).$$

16.4 (a) Show that if two out of three arithmetic functions f, g, and $f * g$ are multiplicative, then so is the third.

(b) Show that if f is an arithmetic function such that $f(1)$ is nonzero, then f is multiplicative if and only if f^{-1} is multiplicative.

16.5 Let f be a multiplicative function.

(a) Show that

$$f^{-1}(n) = \mu(n)f(n)$$

for every square-free natural number n.

(b) Show that

$$f^{-1}(p) = -f(p) \qquad \text{and} \qquad f^{-1}(p^2) = f(p^2) - f(p)^2$$

for every prime number p.

(c) Show that

$$f^{-1}(p^m) = \sum_{k=1}^{m} (-1)^k \left(\sum_{\substack{i_1+\cdots+i_k=m \\ i_1,\ldots,i_k \geq 1}} f(p^{i_1}) \cdots f(p^{i_k}) \right).$$

16.6 Define *Jordan's totient function* $J_k(n)$ by

$$J_k(n) = \#\{(x_1, \ldots, x_k) \mid 1 \leq x_1, \ldots, x_k \leq n \text{ and } \gcd(x_1, \ldots, x_k, n) = 1\}.$$

In other words, Jordan's totient function counts the number of k-tuples $(x_1, \ldots, x_k)$ of positive integers not exceeding n such that if d simultaneously divides all the integers $x_1, \ldots, x_k$ and n, then $d = 1$. Note that $J_1(n)$ is just Euler's totient function ϕ.

(a) Find $J_2(n)$ for $1 \leq n \leq 12$.

(b) Show that for any divisor d of n, the number of elements in the set

$$\left\{ (x_1, \ldots, x_k) \,\middle|\, 1 \leq x_1, \ldots, x_k \leq n \text{ and } \gcd(x_1, \ldots, x_k, n) = \frac{n}{d} \right\}$$

is exactly $J_k(d)$.

(c) Show that

$$\sum_{d|n} J_k(d) = n^k \qquad \text{and} \qquad \sum_{d|n} \mu(d) \left(\frac{n}{d}\right)^k = J_k(n).$$

(d) Show that J_k is a multiplicative function.

(e) Show that

$$J_k(p^m) = p^{km} - p^{k(m-1)}$$

for any prime p and any exponent $m \geq 1$.

(f) Show that

$$J_k(n) = n^k \prod_{p|n} \left(1 - \frac{1}{p^k}\right).$$

Chapter 17

The Distribution of the Prime Numbers

In this chapter we will study the function $\pi(x)$, which is given by

$$\pi(x) := |\{p \leq x \mid p \text{ prime}\}|$$

17.1 The Bachmann–Landau Notation for Asymptotic Growth

In order to analyze the distribution of primes, we will make use of some handy notation for the asymptotic growth of function, which is nowadays commonly called the *Big-O notation*. The idea is due to Paul Bachmann, who used it in 1894 in his book *Analytische Zahlentheorie*, and was later popularized by Edmund Landau, who also started using the *Little-O notation*.

Definition 17.1.1. We say that f is *of the order at most* g if there is a number M such that

$$\frac{|f(x)|}{g(x)} < M$$

for all sufficiently large x. If f is of the order at most g, we write $f = O(g)$ or $f(x) = O(g(x))$.

Definition 17.1.2. We say that f is *of strictly smaller order than* g if the condition

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$$

holds. If f is of strictly smaller order than g, we write $f = o(g)$ or $f(x) = o(g(x))$.

Definition 17.1.3. We say that f is *asymptotically equal to* g if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

If f is asymptotically equal to g, we write $f \sim g$ or $f(x) \sim g(x)$.

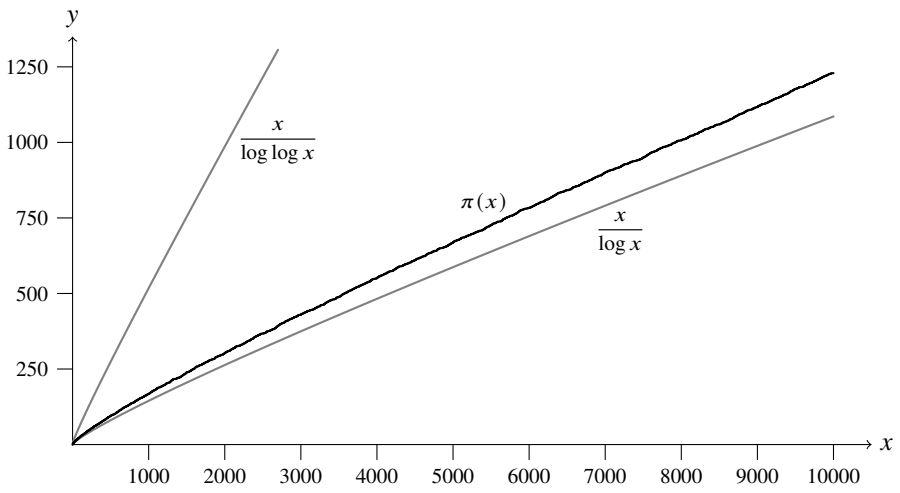


Figure 17.1: The prime counting function $\pi(x)$ compared to the functions $x/\log x$ and $x/\log\log x$, where x ranges from 1 to 10,000. There are 1,229 primes below 10,000.

17.2 An Elementary Estimate of the Prime Counting Function

In this section we will prove an elementary estimate for the growth of the prime counting function:

$$\pi(x) = O\left(\frac{x}{\log\log x}\right)$$

While this estimate is not the strongest known estimate for the prime counting function, which is given by the prime number theorem as the asymptotic relation

$$\pi(x) \sim \frac{x}{\log x},$$

it can be proven by fairly elementary means and it will be quite useful in our future investigations in the distribution of the primes. In Figure 17.1 we have plotted the values of $\pi(x)$, $x/\log x$ and $x/\log\log x$ for comparison.

First, we will use the sieve of Eratosthenes and the inclusion-exclusion principle to obtain an exact expression for $\pi(x)$, which will serve as the starting point of our estimates of $\pi(x)$. Recall that if $n \le x$ is not divisible by any prime less than $\sqrt{x}$, then n is prime. Thus, if we eliminate all the multiples of the primes less than $\sqrt{x}$ from the set of integers $\{2, \ldots, \lfloor x \rfloor\}$, then the set of primes $\le x$ remains. The inclusion-exclusion principle allows us to compute the number of integers that are a multiple of a prime below $\sqrt{x}$ by the formula

$$\sum_{k=1}^{\lfloor\sqrt{x}\rfloor} (-1)^{k+1} \left(\sum_{2 \le p_1 < \cdots < p_k \le \sqrt{x}} \left\lfloor \frac{x}{p_1 \cdots p_k} \right\rfloor \right)$$

However, this count also includes all the primes below $\sqrt{x}$, of which there are $\pi(\sqrt{x})$. Thus, the exact number of primes below x is given by the formula:

$$\pi(x) = \pi\big(\sqrt{x}\big) + \lfloor x \rfloor - 1 + \sum_{k=1}^{\lfloor \sqrt{x} \rfloor} (-1)^k \left(\sum_{2 \leq p_1 < \cdots < p_k \leq \sqrt{x}} \left\lfloor \frac{x}{p_1 \cdots p_k} \right\rfloor \right).$$

In order to facilitate estimation of the quantity $\pi(x)$, we will count the number of integers less than x not divisible by any of the primes $p_1, \ldots, p_r$, postponing the choice of r. Let $A(x, r)$ be the number of integers in the set $\{2, \ldots, \lfloor x \rfloor\}$ that are not divisible by any of the primes $p_1, \ldots, p_r$. Then, as before, the number $A(x, r)$ can be computed exactly by the formula

$$A(x, r) = r + \lfloor x \rfloor - 1 + \sum_{k=1}^{r} (-1)^k \left(\sum_{1 \leq i_1 < \cdots i_k \leq r} \left\lfloor \frac{x}{p_{i_1} \cdots p_{i_k}} \right\rfloor \right).$$

We also introduce the number $B(x, r)$, which is expressed similarly except that we omit all uses of the floor function:

$$B(x, r) = r + x - 1 + \sum_{k=1}^{r} (-1)^k \left(\sum_{1 \leq i_1 < \cdots i_k \leq r} \frac{x}{p_{i_1} \cdots p_{i_k}} \right).$$

Since we have the inequalities $0 \leq x - \lfloor x \rfloor < 1$ by the definition of the floor function, we can estimate the difference $|B(x, r) - A(x, r)|$ by

$$|B(x, r) - A(x, r)| < 1 + \sum_{k=1}^{r} \left(\sum_{1 \leq i_1 < \cdots i_k \leq r} 1 \right) = 2^r.$$

In the last equation, we used the fact that the sum of all the binomial coefficients under r is 2^r, and that the number of strictly increasing sequences $1 \leq i_1 < \cdots < i_k \leq r$ is exactly the binomial coefficient $\binom{r}{k}$, since any subset of $\{1, \ldots, r\}$ uniquely determines such a sequence. It follows that $A(x, r) < B(x, r) + 2^r$. Furthermore, we have $\pi(x) \leq r + A(x, r)$, which shows that

$$\pi(x) < 2^{r+1} + \sum_{k=1}^{r} (-1)^k \left(\sum_{1 \leq i_1 < \cdots i_k \leq r} \frac{x}{p_{i_1} \cdots p_{i_k}} \right),$$

where the term 2^{r+1} comes from the term $r + 2^r$, which is clearly strictly smaller.

On the other hand, expanding the product gives the identity

$$\prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) = \sum_{k=0}^{r} (-1)^k \left(\sum_{1 \leq i_1 < \cdots i_k \leq r} \frac{1}{p_{i_1} \cdots p_{i_k}} \right).$$

Thus, we find that

$$\pi(x) \leq 2^{r+1} + x \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right)$$

In order to make further progress in estimating the prime counting function, we need a bound on the Euler product $\prod_{i=1}^{r} \big(1 - \frac{1}{p_i}\big)$, which we will establish in the following theorem.

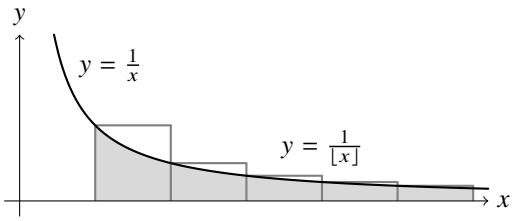


Figure 17.2: The sum of unit fractions can be estimated from below by computing the area under the graph $x \mapsto \frac{1}{x}$.

Theorem 17.2.1. *For any natural number $n \geq 2$, we have*

$$\prod_{p \leq n} \left(1 - \frac{1}{p}\right) < \frac{1}{\log n}.$$

Proof. The statement of the theorem is equivalent to:

$$\log n < \prod_{p \leq n} \frac{1}{1 - p^{-1}}.$$

Note that the factors on the right-hand side are the sums of the geometric series $1 + p^{-1} + p^{-2} + \cdots$. Thus, we can evaluate this product as

$$\prod_{p \leq n} \frac{1}{1 - p^{-1}} = \sum_{k=0}^{\infty} \left(\sum_{2 \leq p_1 \leq \cdots \leq p_k \leq n} \frac{1}{p_1 \cdots p_k} \right).$$

In the sum on the right-hand side, we see exactly one occurrence of every unit fraction $\frac{1}{m}$ where all the prime factors of m are at most n. This sum is certainly strictly larger than the sum $1 + \cdots + \frac{1}{n}$, so that

$$\sum_{m=1}^{n} \frac{1}{m} < \prod_{p \leq n} \frac{1}{1 - p^{-1}}.$$

In order to establish a lower bound for the sum of the unit fractions $\frac{1}{m}$ over $1 \leq m \leq n$, we will use that the inequality $\frac{1}{x} \leq \frac{1}{\lfloor x \rfloor}$ holds for all $x \geq 1$. Thus we see that we can estimate the sum of the unit fractions $\frac{1}{n}$ from below by computing the area under the graph of the function $x \mapsto \frac{1}{x}$, as indicated in Figure 17.2 by integrating the function $x \mapsto \frac{1}{x}$. Doing so, we obtain:

$$\log n < \int_1^{n+1} \frac{1}{x} \, \mathrm{d}x < \sum_{m=1}^{n} \frac{1}{m}.$$

Combining these two inequalities, we obtain the desired strict inequality

$$\log n < \prod_{p \leq n} \frac{1}{1 - p^{-1}}.$$ □

We are now ready to prove the estimate for the prime counting function stated at the beginning of this section.

Theorem 17.2.2.

$$\pi(x) = O\left(\frac{x}{\log\log x}\right).$$

Proof. By the previous theorem and the remarks preceding it, we have the inequality

$$\pi(x) \leq 2^{r+1} + \frac{x}{\log p_r}.$$

Since $r < p_r$, we obtain that

$$\pi(x) < 2^{r+1} + \frac{x}{\log r}.$$

A suitable value for r is now $\lfloor \log x \rfloor$, for which we obtain

$$\pi(x) \leq \frac{x}{\log\log x} + 2 \cdot 2^{\log x} = \frac{x}{\log\log x} + 2 \cdot x^{\log 2}.$$

The function $x^{\log 2}$ is of strictly smaller order than $\frac{x}{\log\log x}$. Therefore we have

$$\pi(x) = O\left(\frac{x}{\log\log x}\right). \qquad \square$$

17.3 Chebyshev's Theorem

Theorem 17.3.1. *There are positive constants c_1 and c_2 such that for $x \geq 2$, we have*

$$c_1 \frac{x}{\log x} < \pi(x) < c_2 \frac{x}{\log x}.$$

Proof. From Corollary 5.4.6 it follows that the central binomial coefficient $\binom{2n}{n}$ divides the least common multiple of the digits $1, \ldots, 2n$, and therefore we have

$$\binom{2n}{n} \leq \operatorname{lcm}(1, \ldots, 2n).$$

We therefore obtain the inequalities

$$n^{\pi(2n)-\pi(n)} \leq \prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq \operatorname{lcm}(1, \ldots, 2n) \leq (2n)^{\pi(2n)}.$$

It follows that

$$(\pi(2n) - \pi(n)) \log n \leq \log \binom{2n}{n} \leq \pi(2n) \log(2n).$$

Since $2^n \leq \binom{2n}{n} < 4^n$, we conclude from the above inequalities that

$$\pi(2n) - \pi(n) \leq 2 \log 2 \frac{n}{\log n},$$

and

$$\pi(2n) \geq \frac{n \log 2}{\log(2n)}.$$

This last inequality gives us for $x \geq 4$ that

$$\pi(x) \geq \pi\left(2\left\lfloor \frac{x}{2} \right\rfloor\right) \geq \frac{\lfloor x/2 \rfloor \log 2}{\log(2\lfloor x/2 \rfloor)} \geq c_1 \frac{x}{\log x}.$$

Thus, it remains to show that there is a positive constant c_2 such that

$$\pi(x) \leq c_2 \frac{x}{\log x}.$$

We will use the inequality $\pi(2n) - \pi(n) \leq c_2' \frac{n}{\log n}$ to set up a "telescoping argument". Note that for any positive y we have

$$\pi(y) - \pi(y/2) = \pi(y) - \pi(\lfloor y/2 \rfloor) \leq 1 + \pi(2\lfloor y/2 \rfloor) - \pi(\lfloor y/2 \rfloor).$$

Thus, there is a positive constant c_2' such that

$$\pi(y) - \pi(y/2) < c_2'' \frac{y}{\log y}$$

for any $y \geq 2$.

Using the fact that $\pi(z) \leq z$, we get that

$$\begin{aligned} \pi(y) \log y - \pi(y/2) \log \frac{y}{2} &= (\pi(y) - \pi(y/2)) \log y + \pi(y/2) \log 2 \\ &< \log y \cdot c_2'' \frac{y}{\log y} + y/2 \\ &< c_2''' y \end{aligned}$$

Now we choose $y = x/2^m$ with $1 \leq 2^m \leq x/2$. Then this becomes

$$\pi\left(\frac{x}{2^m}\right) \log \frac{x}{2^m} - \pi\left(\frac{x}{2^{m+1}}\right) \log \frac{x}{2^{m+1}} < c_2''' \frac{x}{2^m}$$

Summing over all such ms, we find that

$$\pi(x) \log x - \pi\left(\frac{x}{2^{\mu+1}}\right) \log \frac{x}{2^{\mu+1}} < c_2 x,$$

where $2^\mu \leq x/2 < 2^{\mu+1}$. This means, however, that $x/2^{\mu+1} < 2$, implying that $\pi(x/2^{\mu+1}) = 0$. Thus we conclude that

$$\pi(x) \log x < c_2 x. \qquad \square$$

Bibliography

[And94] George E. Andrews. *Number Theory*. Dover Books on Mathematics. Dover Publications, 1994. ISBN: 9780486682525.

[Ber34] B. Berggren. "Pytagoreiska triangular". Swedish. In: *Tidskrift för elementär matematik, fysik och kemi* 17 (1934), pp. 129–139.

[Ber45] Joseph Bertrand. "Mémoire sur le nombre de valeurs que peut prendre une fonction quand on y permute les lettres qu'elle renferme". French. In: *Journal de l'École Royale Polytechnique* 18.30 (1845). Available at https://books.google.com/books?id=WTa-qRIWckoC&pg=PA123, pp. 123–140.

[CW03] Daniel Cass and Gerald Wildenberg. "Math Bite: A Novel Proof of the Infinitude of Primes, Revisited". In: *Mathematics Magazine* 76.3 (2003), p. 203. DOI: 10.1080/0025570X.2003.11953179. URL: https://www.tandfonline.com/doi/abs/10.1080/0025570X.2003.11953179.

[Che50] Pafnuty Chebyshev. "Mémoire sur les nombres premiers". French. In: *Journal de Mathématiques Pures et Appliquées* 17 (1850), pp. 366–390.

[Dio10] Diophantus. *Diophantus of Alexandria: A Study in the History of Greek Algebra*. Ed. by Thomas Little Heath. Includes an English translation of *Arithmetica*. Cambridge: Cambridge University Press, 1910.

[Erd32] Paul Erdős. "Beweis eines Satzes von Tschebyschef". In: *Acta Scientiarum Mathematicarum (Szeged)* 5 (1932), pp. 194–198.

[Euc08] Euclid. *The Elements*. Trans. by Thomas Heath. Reprint of the 2nd edition. New York: Dover Publications, 1908. ISBN: 9780486612887.

[Eul88] Leonhard Euler. *Introductio in Analysin Infinitorum*. Trans. by John D. Blanton. Springer-Verlag, 1988.

[Far09] Bakir Farhi. "An Identity Involving the Least Common Multiple of Binomial Coefficients and Its Application". In: *American Mathematical Monthly* 116.9 (2009), pp. 836–839.

[Fib25] Leonardo Fibonacci. *Liber Quadratorum*. Latin. English translation available at https://api.pageplace.de/preview/DT0400.9780080886503_A23591681/preview-9780080886503_A23591681.pdf. 1225.

[Fou25] OEIS Foundation. *A006992: Bertrand primes*. https://oeis.org/A006992. The On-Line Encyclopedia of Integer Sequences. 2025.

[Fur55] Hillel Furstenberg. "On the Infinitude of Primes". In: *American Mathematical Monthly* 62.5 (1955), p. 353. DOI: 10.2307/2307043. URL: https://doi.org/10.2307/2307043.

[Gau86] J. Carl Friedrich Gauss. *Disquisitiones Arithmeticae*. English. Trans. by Jeffrey R. O. Wright. New Haven: Yale University Press, 1986.

[Gua23] The Guardian. "US teens say they have new proof for 2,000-year-old mathematical theorem". In: (Mar. 2023). URL: https://www.theguardian.com/us-news/2023/mar/24/new-orleans-pythagoras-theorem-trigonometry-prove.

[Har+08] Godfrey H. Hardy et al. *An Introduction to the Theory of Numbers*. 6th. With a foreword by Andrew Wiles. Oxford: Oxford University Press, 2008. ISBN: 9780199219865.

[Hen96] Kurt Hensel. "Ueber den grössten gemeinsamen Theiler aller Zahlen, welche durch eine ganze Function von n Veränderlichen darstellbar sind". In: *Journal für die reine und angewandte Mathematik* 116 (1896), pp. 350–356. ISSN: 0075-4102; 1435-5345/e.

[JJ24] Ne'Kiya Jackson and Calcea Johnson. "Five or Ten New Proofs of the Pythagorean Theorem". In: *The American Mathematical Monthly* 131.9 (2024), pp. 739–752. DOI: 10.1080/00029890.2024.2370240.

[Kum52] Ernst Eduard Kummer. "Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen". In: *Journal für die reine und angewandte Mathematik* 44 (1852), pp. 93–146. DOI: 10.1515/crll.1852.44.93.

[Lag70] Joseph-Louis Lagrange. "Réflexions sur la résolution algébrique des équations". In: *Histoire de l'Académie Royale des Sciences et Belles-Lettres de Berlin* (1770), pp. 123–192.

[Leg98] Adrien-Marie Legendre. *Essai sur la théorie des nombres*. Paris: Duprat, 1798. URL: https://openlibrary.org/books/OL24146260M/Essai_sur_la_th%C3%A9orie_des_nombres.

[Leh33] Derrick H. Lehmer. "On Euler's totient function". In: *Bulletin of the American Mathematical Society* 38 (1933), pp. 745–751. DOI: 10.1090/s0002-9904-1932-05521-5.

[LeV56a] William J. LeVeque. *Topics in Number Theory, Volume 1*. Reading, MA: Addison-Wesley, 1956. ISBN: 9780201042252.

[LeV56b] William J. LeVeque. *Topics in Number Theory, Volume 2*. Reading, MA: Addison-Wesley, 1956. ISBN: 9780201042269.

[Rot64] Gian-Carlo Rota. "On the Foundations of Combinatorial Theory I. Theory of Möbius Functions". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2.4 (1964), pp. 340–368. DOI: 10.1007/BF00531932.

[Sai06] Filip Saidak. "A New Proof of Euclid's Theorem". In: *The American Mathematical Monthly* 113.10 (2006), pp. 937–938. DOI: 10.2307/27642094.

[Sie88] Wacław Sierpiński. *Elementary Theory of Numbers*. Ed. by A. Schinzel. Second English Edition. Amsterdam: North-Holland, 1988. ISBN: 978-0-444-86662-2.

[Sil12] Joseph H. Silverman. *A Friendly Introduction to Number Theory*. 4th. Pearson, 2012. ISBN: 978-0321816191.