# Elementary Number Theory: Primes, Congruences, and Secrets

William Stein

November 16, 2011

To my wife Clarita Lefthand

# Contents

# Preface

This is a book about prime numbers, congruences, secret messages, and elliptic curves that you can read cover to cover. It grew out of undergraduate courses that the author taught at Harvard, UC San Diego, and the University of Washington.

The systematic study of number theory was initiated around 300B.C. when Euclid proved that there are infinitely many prime numbers, and also cleverly deduced the fundamental theorem of arithmetic, which asserts that every positive integer factors uniquely as a product of primes. Over a thousand years later (around 972A.D.) Arab mathematicians formulated the *congruent number problem* that asks for a way to decide whether or not a given positive integer $n$ is the area of a right triangle, all three of whose sides are rational numbers. Then another thousand years later (in 1976), Diffie and Hellman introduced the first ever public-key cryptosystem, which enabled two people to communicate secretely over a public communications channel with no predetermined secret; this invention and the ones that followed it revolutionized the world of digital communication. In the 1980s and 1990s, elliptic curves revolutionized number theory, providing striking new insights into the congruent number problem, primality testing, public-key cryptography, attacks on public-key systems, and playing a central role in Andrew Wiles' resolution of Fermat's Last Theorem.

Today, pure and applied number theory is an exciting mix of simultaneously broad and deep theory, which is constantly informed and motivated by algorithms and explicit computation. Active research is underway that promises to resolve the congruent number problem, deepen our understanding into the structure of prime numbers, and both challenge and improve

our ability to communicate securely. The goal of this book is to bring the reader closer to this world.

The reader is strongly encouraged to do every exercise in this book, checking their answers in the back (where many, but not all, solutions are given). Also, throughout the text there, are examples of calculations done using the powerful free open source mathematical software system Sage (`http://www.sagemath.org`), and the reader should try every such example and experiment with similar examples.

**Background.** The reader should know how to read and write mathematical proofs and must have know the basics of groups, rings, and fields. Thus, the prerequisites for this book are more than the prerequisites for most elementary number theory books, while still being aimed at undergraduates.

**Notation and Conventions.** We let $\mathbf{N} = \{1, 2, 3, \ldots\}$ denote the natural numbers, and use the standard notation $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ for the rings of integer, rational, real, and complex numbers, respectively. In this book, we will use the words proposition, theorem, lemma, and corollary as follows. Usually a proposition is a less important or less fundamental assertion, a theorem is a deeper culmination of ideas, a lemma is something that we will use later in this book to prove a proposition or theorem, and a corollary is an easy consequence of a proposition, theorem, or lemma. More difficult exercises are marked with a (*).

**Acknowledgements.** I would like to thank Brian Conrad, Carl Pomerance, and Ken Ribet for many clarifying comments and suggestions. Baurzhan Bektemirov, Lawrence Cabusora, and Keith Conrad read drafts of this book and made many comments, and Carl Witty commented extensively on the first two chapters. Frank Calegari used the course when teaching Math 124 at Harvard, and he and his students provided much feedback. Noam Elkies made comments and suggested Exercise 4.6. Seth Kleinerman wrote a version of Section 5.4 as a class project. Hendrik Lenstra made helpful remarks about how to present his factorization algorithm. Michael Abshoff, Sabmit Dasgupta, David Joyner, Arthur Patterson, George Stephanides, Kevin Stern, Eve Thompson, Ting-You Wang, and Heidi Williams all suggested corrections. I also benefited from conversations with Henry Cohn and David Savitt. I used Sage ([Sag08]), emacs, and LaTeX in the preparation of this book.

# 1
# Prime Numbers

Every positive integer can be written uniquely as a product of prime numbers, e.g., $100 = 2^2 \cdot 5^2$. This is surprisingly difficult to prove, as we will see below. Even more astounding is that actually *finding* a way to write certain 1,000-digit numbers as a product of primes seems out of the reach of present technology, an observation that is used by millions of people every day when they buy things online.

Since prime numbers are the building blocks of integers, it is natural to wonder how the primes are distributed among the integers.

> "There are two facts about the distribution of prime numbers. The first is that, [they are] the most arbitrary and ornery objects studied by mathematicians: they grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision."
>
> — Don Zagier [Zag75]

The Riemann Hypothesis, which is the most famous unsolved problem in number theory, postulates a very precise answer to the question of how the prime numbers are distributed.

This chapter lays the foundations for our study of the theory of numbers by weaving together the themes of prime numbers, integer factorization, and the distribution of primes. In Section 1.1, we rigorously prove that the

every positive integer is a product of primes, and give examples of specific integers for which finding such a decomposition would win one a large cash bounty. In Section 1.2, we discuss theorems about the set of prime numbers, starting with Euclid's proof that this set is infinite, and discuss the largest known prime. Finally we discuss the distribution of primes via the prime number theorem and the Riemann Hypothesis.

## 1.1 Prime Factorization

### *1.1.1 Primes*

The set of *natural numbers* is

$$\mathbf{N} = \{1, 2, 3, 4, \ldots\},$$

and the set of *integers* is

$$\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

**Definition 1.1.1** (Divides)**.** If $a, b \in \mathbf{Z}$ we say that $a$ *divides* $b$, written $a \mid b$, if $ac = b$ for some $c \in \mathbf{Z}$. In this case, we say $a$ is a *divisor* of $b$. We say that $a$ *does not divide* $b$, written $a \nmid b$, if there is no $c \in \mathbf{Z}$ such that $ac = b$.

For example, we have $2 \mid 6$ and $-3 \mid 15$. Also, all integers divide 0, and 0 divides only 0. However, 3 does not divide 7 in $\mathbf{Z}$.

*Remark* 1.1.2. The notation $b \vdots a$ for "$b$ is divisible by $a$" is common in Russian literature on number theory.

**Definition 1.1.3** (Prime and Composite)**.** An integer $n > 1$ is *prime* if the only positive divisors of $n$ are 1 and $n$. We call $n$ *composite* if $n$ is not prime.

The number 1 is neither prime nor composite. The first few primes of $\mathbf{N}$ are

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

and the first few composites are

$$4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 32, 33, 34, \ldots.$$

*Remark* 1.1.4. J. H. Conway argues in [Con97, viii] that $-1$ should be considered a prime, and in the 1914 table [Leh14], Lehmer considers 1 to be a prime. In this book, we consider neither $-1$ nor 1 to be prime.

*SAGE Example* 1.1.5. We use Sage to compute all prime numbers between $a$ and $b - 1$.

```
sage: prime_range(10,50)
[11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47]
```

We can also compute the composites in an interval.

```
sage: [n for n in range(10,30) if not is_prime(n)]
[10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28]
```

Every natural number is built, in a unique way, out of prime numbers:

**Theorem 1.1.6** (Fundamental Theorem of Arithmetic)**.** *Every natural number can be written as a product of primes uniquely up to order.*

Note that primes are the products with only one factor and 1 is the empty product.

*Remark* 1.1.7. Theorem 1.1.6, which we will prove in Section 1.1.4, is trickier to prove than you might first think. For example, unique factorization fails in the *ring*

$$\mathbf{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbf{Z}\} \subset \mathbf{C},$$

where 6 factors in two different ways:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

### *1.1.2 The Greatest Common Divisor*

We will use the notion of the greatest common divisor of two integers to prove that if $p$ is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. Proving this is the key step in our proof of Theorem 1.1.6.

**Definition 1.1.8** (Greatest Common Divisor)**.** Let

$$\gcd(a, b) = \max\{d \in \mathbf{Z} : d \mid a \text{ and } d \mid b\},$$

unless both $a$ and $b$ are 0 in which case $\gcd(0, 0) = 0$.

For example, $\gcd(1, 2) = 1$, $\gcd(6, 27) = 3$, and for any $a$, $\gcd(0, a) = \gcd(a, 0) = a$.

If $a \neq 0$, the greatest common divisor exists because if $d \mid a$ then $d \leq |a|$, and there are only $|a|$ positive integers $\leq |a|$. Similarly, the gcd exists when $b \neq 0$.

**Lemma 1.1.9.** *For any integers $a$ and $b$, we have*

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

*Proof.* We only prove that $\gcd(a, b) = \gcd(a, b - a)$, since the other cases are proved in a similar way. Suppose $d \mid a$ and $d \mid b$, so there exist integers $c_1$ and $c_2$ such that $dc_1 = a$ and $dc_2 = b$. Then $b - a = dc_2 - dc_1 = d(c_2 - c_1)$,

so $d \mid b-a$. Thus $\gcd(a,b) \leq \gcd(a,b-a)$, since the set over which we are taking the max for $\gcd(a,b)$ is a subset of the set for $\gcd(a,b-a)$. The same argument with $a$ replaced by $-a$ and $b$ replaced by $b-a$, shows that $\gcd(a,b-a) = \gcd(-a,b-a) \leq \gcd(-a,b) = \gcd(a,b)$, which proves that $\gcd(a,b) = \gcd(a,b-a)$. □

**Lemma 1.1.10.** *Suppose $a, b, n \in \mathbf{Z}$. Then $\gcd(a,b) = \gcd(a, b-an)$.*

*Proof.* By repeated application of Lemma 1.1.9, we have

$$\gcd(a,b) = \gcd(a,b-a) = \gcd(a,b-2a) = \cdots = \gcd(a,b-an).$$

□

Assume for the moment that we have already proved Theorem 1.1.6. A naive way to compute $\gcd(a,b)$ is to factor $a$ and $b$ as a product of primes using Theorem 1.1.6; then the prime factorization of $\gcd(a,b)$ can be read off from that of $a$ and $b$. For example, if $a = 2261$ and $b = 1275$, then $a = 7 \cdot \mathbf{17} \cdot 19$ and $b = 3 \cdot 5^2 \cdot \mathbf{17}$, so $\gcd(a,b) = 17$. It turns out that the greatest common divisor of two integers, even huge numbers (millions of digits), is surprisingly easy to compute using Algorithm 1.1.13 below, which computes $\gcd(a,b)$ without factoring $a$ or $b$.

To motivate Algorithm 1.1.13, we compute $\gcd(2261, 1275)$ in a different way. First, we recall a helpful fact.

**Proposition 1.1.11.** *Suppose that $a$ and $b$ are integers with $b \neq 0$. Then there exists unique integers $q$ and $r$ such that $0 \leq r < |b|$ and $a = bq + r$.*

*Proof.* For simplicity, assume that both $a$ and $b$ are positive (we leave the general case to the reader). Let $Q$ be the set of all nonnegative integers $n$ such that $a - bn$ is nonnegative. Then $Q$ is nonempty because $0 \in Q$ and $Q$ is bounded because $a - bn < 0$ for all $n > a/b$. Let $q$ be the largest element of $Q$. Then $r = a - bq < b$, otherwise $q+1$ would also be in $Q$. Thus $q$ and $r$ satisfy the existence conclusion.

To prove uniqueness, suppose that $q'$ and $r'$ also satisfy the conclusion. Then $q' \in Q$ since $r' = a - bq' \geq 0$, so $q' \leq q$, and we can write $q' = q - m$ for some $m \geq 0$. If $q' \neq q$, then $m \geq 1$ so

$$r' = a - bq' = a - b(q-m) = a - bq + bm = r + bm \geq b$$

since $r \geq 0$, a contradiction. Thus $q = q'$ and $r' = a - bq' = a - bq = r$, as claimed. □

For us, an *algorithm* is a finite sequence of instructions that can be followed to perform a specific task, such as a sequence of instructions in a computer program, which must terminate on any valid input. The word "algorithm" is sometimes used more loosely (and sometimes more precisely) than defined here, but this definition will suffice for us.

**Algorithm 1.1.12** (Division Algorithm)**.** Suppose $a$ and $b$ are integers with $b \neq 0$. This algorithm computes integers $q$ and $r$ such that $0 \leq r < |b|$ and $a = bq + r$.

We will not describe the actual steps of Algorithm 1.1.12, since it is just the familiar long division algorithm. Note that it might not be exactly the same as the standard long division algorithm you learned in school, because we make the remainder positive even when dividing a negative number by a positive number.

We use the division algorithm repeatedly to compute $\gcd(2261, 1275)$. Dividing 2261 by 1275 we find that

$$2261 = 1 \cdot 1275 + 986,$$

so $q = 1$ and $r = 986$. Notice that if a natural number $d$ divides both 2261 and 1275, then $d$ divides their difference 986 and $d$ still divides 1275. On the other hand, if $d$ divides both 1275 and 986, then it has to divide their sum 2261 as well! We have made progress:

$$\gcd(2261, 1275) = \gcd(1275, 986).$$

This equality also follows by applying Lemma 1.1.9. Repeating, we have

$$1275 = 1 \cdot 986 + 289,$$

so $\gcd(1275, 986) = \gcd(986, 289)$. Keep going:

$$\begin{aligned} 986 &= 3 \cdot 289 + 119 \\ 289 &= 2 \cdot 119 + 51 \\ 119 &= 2 \cdot 51 + 17. \end{aligned}$$

Thus $\gcd(2261, 1275) = \cdots = \gcd(51, 17)$, which is 17 because $17 \mid 51$. Thus

$$\gcd(2261, 1275) = 17.$$

Aside from some tedious arithmetic, that computation was systematic, and it was not necessary to factor any integers (which is something we do not know how to do quickly if the numbers involved have hundreds of digits).

**Algorithm 1.1.13** (Greatest Common Division)**.** Given integers $a, b$, this algorithm computes $\gcd(a, b)$.

1. [Assume $a > b > 0$] We have $\gcd(a, b) = \gcd(|a|, |b|) = \gcd(|b|, |a|)$, so we may replace $a$ and $b$ by their absolute values and hence assume $a, b \geq 0$. If $a = b$, output $a$ and terminate. Swapping if necessary, we assume $a > b$. If $b = 0$, we output $a$.
2. [Quotient and Remainder] Using Algorithm 1.1.12, write $a = bq + r$, with $0 \leq r < b$ and $q \in \mathbf{Z}$.

3. [Finished?] If $r = 0$, then $b \mid a$, so we output $b$ and terminate.
4. [Shift and Repeat] Set $a \leftarrow b$ and $b \leftarrow r$, then go to Step 2.

*Proof.* Lemmas 1.1.9–1.1.10 imply that $\gcd(a, b) = \gcd(b, r)$ so the gcd does not change in Step 4. Since the remainders form a decreasing sequence of nonnegative integers, the algorithm terminates. □

*Example* 1.1.14. Set $a = 15$ and $b = 6$.

$$\begin{aligned} 15 &= 6 \cdot 2 + 3 \qquad \gcd(15, 6) = \gcd(6, 3) \\ 6 &= 3 \cdot 2 + 0 \qquad \gcd(6, 3) = \gcd(3, 0) = 3 \end{aligned}$$

Note that we can just as easily do an example that is ten times as big, an observation that will be important in the proof of Theorem 1.1.19 below.

*Example* 1.1.15. Set $a = 150$ and $b = 60$.

$$\begin{aligned} 150 &= 60 \cdot 2 + 30 \qquad \gcd(150, 60) = \gcd(60, 30) \\ 60 &= 30 \cdot 2 + 0 \qquad \gcd(60, 30) = \gcd(30, 0) = 30 \end{aligned}$$

*SAGE Example* 1.1.16. Sage uses the `gcd` command to compute the greatest common divisor of two integers. For example,

```
sage: gcd(97,100)
1
sage: gcd(97 * 10^15, 19^20 * 97^2)
97
```

**Lemma 1.1.17.** *For any integers $a, b, n$, we have*

$$\gcd(an, bn) = \gcd(a, b) \cdot |n|.$$

*Proof.* The idea is to follow Example 1.1.15; we step through Euclid's algorithm for $\gcd(an, bn)$ and note that at every step the equation is the equation from Euclid's algorithm for $\gcd(a, b)$ but multiplied through by $n$. For simplicity, assume that both $a$ and $b$ are positive. We will prove the lemma by induction on $a + b$. The statement is true in the base case when $a + b = 2$, since then $a = b = 1$. Now assume $a, b$ are arbitrary with $a \geq b$. Let $q$ and $r$ be such that $a = bq + r$ and $0 \leq r < b$. Then by Lemmas 1.1.9–1.1.10, we have $\gcd(a, b) = \gcd(b, r)$. Multiplying $a = bq + r$ by $n$ we see that $an = bnq + rn$, so $\gcd(an, bn) = \gcd(bn, rn)$. Then

$$b + r = b + (a - bq) = a - b(q - 1) \leq a < a + b,$$

so by induction $\gcd(bn, rn) = \gcd(b, r) \cdot |n|$. Since $\gcd(a, b) = \gcd(b, r)$, this proves the lemma. □

**Lemma 1.1.18.** *Suppose $a, b, n \in \mathbf{Z}$ are such that $n \mid a$ and $n \mid b$. Then $n \mid \gcd(a, b)$.*

*Proof.* Since $n \mid a$ and $n \mid b$, there are integers $c_1$ and $c_2$, such that $a = nc_1$ and $b = nc_2$. By Lemma 1.1.17, $\gcd(a, b) = \gcd(nc_1, nc_2) = n \gcd(c_1, c_2)$, so $n$ divides $\gcd(a, b)$. □

With Algorithm 1.1.13, we can prove that if a prime divides the product of two numbers, then it has got to divide one of them. This result is the key to proving that prime factorization is unique.

**Theorem 1.1.19** (Euclid)**.** *Let $p$ be a prime and $a, b \in \mathbf{N}$. If $p \mid ab$ then $p \mid a$ or $p \mid b$.*

You might think this theorem is "intuitively obvious," but that might be because the fundamental theorem of arithmetic (Theorem 1.1.6) is deeply ingrained in your intuition. Yet Theorem 1.1.19 will be needed in our proof of the fundamental theorem of arithmetic.

*Proof of Theorem 1.1.19.* If $p \mid a$ we are done. If $p \nmid a$ then $\gcd(p, a) = 1$, since only 1 and $p$ divide $p$. By Lemma 1.1.17, $\gcd(pb, ab) = b$. Since $p \mid pb$ and, by hypothesis, $p \mid ab$, it follows (using Lemma 1.1.17) that

$$p \mid \gcd(pb, ab) = b \gcd(p, a) = b \cdot 1 = b.$$

□

### *1.1.3 Numbers Factor as Products of Primes*

In this section, we prove that every natural number factors as a product of primes. Then we discuss the difficulty of finding such a decomposition in practice. We will wait until Section 1.1.4 to prove that factorization is unique.

As a first example, let $n = 1275$. The sum of the digits of $n$ is divisible by 3, so $n$ is divisible by 3 (see Proposition 2.1.9), and we have $n = 3 \cdot 425$. The number 425 is divisible by 5, since its last digit is 5, and we have $1275 = 3 \cdot 5 \cdot 85$. Again, dividing 85 by 5, we have $1275 = 3 \cdot 5^2 \cdot 17$, which is the prime factorization of 1275. Generalizing this process proves the following proposition.

**Proposition 1.1.20.** *Every natural number is a product of primes.*

*Proof.* Let $n$ be a natural number. If $n = 1$, then $n$ is the empty product of primes. If $n$ is prime, we are done. If $n$ is composite, then $n = ab$ with $a, b < n$. By induction, $a$ and $b$ are products of primes, so $n$ is also a product of primes. □

Two questions immediately arise: (1) is this factorization unique, and (2) how quickly can we find such a factorization? Addressing (1), what if we had done something differently when breaking apart 1275 as a product of primes? Could the primes that show up be different? Let's try: we have

$1275 = 5 \cdot 255$. Now $255 = 5 \cdot 51$ and $51 = 17 \cdot 3$, and again the factorization is the same, as asserted by Theorem 1.1.6. We will prove the uniqueness of the prime factorization of any integer in Section 1.1.4.

*SAGE Example* 1.1.21. The `factor` command in Sage factors an integer as a product of primes with multiplicities. For example,

```
sage: factor(1275)
3 * 5^2 * 17
sage: factor(2007)
3^2 * 223
sage: factor(31415926535898)
2 * 3 * 53 * 73 * 2531 * 534697
```

Regarding (2), there are algorithms for integer factorization. It is a major open problem to decide how fast integer factorization algorithms can be. We say that an algorithm to factor $n$ is *polynomial time* if there is a polynomial $f(x)$ such that for any $n$ the number of steps needed by the algorithm to factor $n$ is less than $f(\log_{10}(n))$. Note that $\log_{10}(n)$ is an approximation for the number of digits of the input $n$ to the algorithm.

**Open Problem 1.1.22.** *Is there an algorithm that can factor any integer $n$ in polynomial time?*

Peter Shor [Sho97] devised a polynomial time algorithm for factoring integers on quantum computers. We will not discuss his algorithm further, except to note that in 2001 IBM researchers built a quantum computer that used Shor's algorithm to factor 15 (see [LMG+01, IBM01]). Building much larger quantum computers appears to be extremely difficult.

You can earn money by factoring certain large integers. Many cryptosystems would be easily broken if factoring certain large integers was easy. Since nobody has proven that factoring integers is difficult, one way to increase confidence that factoring is difficult is to offer cash prizes for factoring certain integers. For example, until recently there was a \$10,000 bounty on factoring the following 174-digit integer (see [RSA]):

$$\begin{aligned}&188198812920607963838697239461650439807163563379417382700763356422988859715234665485319060606504743045317388011303396716199692321205734031879550656996221305168759307650257059\end{aligned}$$

This number is known as RSA-576 since it has 576 digits when written in binary (see Section 2.3.2 for more on binary numbers). It was factored at the German Federal Agency for Information Technology Security in December 2003 (see [Wei03]):

$$\begin{aligned}&398075086424064937397125500550386491199064362342526708406385189575946388957261768583317\\ &\times\\ &472772146107435302536223071973048224632914695302097116459852171130520711256363590397527\end{aligned}$$

The previous RSA challenge was the 155-digit number

$$\begin{aligned}&10941738641570527421809707322040357612003732945449205990913842131476349984288934784717997257891267332497625752899781\\&83379707653724402714674353159335433389 7.\end{aligned}$$

It was factored on 22 August 1999 by a group of sixteen researchers in four months on a cluster of 292 computers (see [ACD+99]). They found that RSA-155 is the product of the following two 78-digit primes:

$$\begin{aligned}p &= 10263959282974110577205419657399167590071656780803806\\&\quad 68033419335217907113077 79\\q &= 10660348838016845482092722036001287867920795857598929\\&\quad 15222706082371930628086 43.\end{aligned}$$

The next RSA challenge is RSA-640:

$$\begin{aligned}&31074182404900437213507500358885679300373460228427275457201619\\&48823206440518081504556346829671723286782437916272838033415471\\&07310850191954852900733772482278352574238645401469173660247765\\&2346609,\end{aligned}$$

and its factorization was worth $20,000 until November 2005 when it was factored by F. Bahr, M. Boehm, J. Franke, and T. Kleinjun. This factorization took five months. Here is one of the prime factors (you can find the other):

$$\begin{aligned}&16347336458092538484431338838650908598417836700330923121811108\\&52389333100104508151212118167511579.\end{aligned}$$

(This team also factored a 663-bit RSA challenge integer.)

The smallest currently open challenge is RSA-704, worth $30,000:

$$\begin{aligned}&74037563479561712828046796097429573142593188889231289084936232\\&63897276503402826627689199641962511784399589433050212758537011\\&89680982867331732731089309005525051168770632990723963807867100\\&86096962537934650563796359\end{aligned}$$

*SAGE Example* 1.1.23. Using Sage, we see that the above number has 212 decimal digits and is definitely composite:

```
sage: n = 74037563479561712828046796097429573142593188889231289084936232\
...92312890849362326389727650340282662768919964196251178439958943305021275853701 1\
...84399589433050212758537011896809828673317327310893090055250511687706329907239638078671008609696253793465\
...0563796359
sage: len(n.str(2))
```

```
704
sage: len(n.str(10))
212
sage: n.is_prime()              # this is instant
False
```

These RSA numbers were factored using an algorithm called the number field sieve (see [LL93]), which is the best-known general purpose factorization algorithm. A description of how the number field sieve works is beyond the scope of this book. However, the number field sieve makes extensive use of the elliptic curve factorization method, which we will describe in Section 6.3.

### *1.1.4 The Fundamental Theorem of Arithmetic*

We are ready to prove Theorem 1.1.6 using the following idea. Suppose we have two factorizations of $n$. Using Theorem 1.1.19, we cancel common primes from each factorization, one prime at a time. At the end, we discover that the factorizations must consist of exactly the same primes. The technical details are given below.

*Proof.* If $n = 1$, then the only factorization is the empty product of primes, so suppose $n > 1$.

By Proposition 1.1.20, there exist primes $p_1, \ldots, p_d$ such that

$$n = p_1 p_2 \cdots p_d.$$

Suppose that

$$n = q_1 q_2 \cdots q_m$$

is another expression of $n$ as a product of primes. Since

$$p_1 \mid n = q_1(q_2 \cdots q_m),$$

Euclid's theorem implies that $p_1 = q_1$ or $p_1 \mid q_2 \cdots q_m$. By induction, we see that $p_1 = q_i$ for some $i$.

Now cancel $p_1$ and $q_i$, and repeat the above argument. Eventually, we find that, up to order, the two factorizations are the same. □

## 1.2 The Sequence of Prime Numbers

This section is concerned with three questions:

1. Are there infinitely many primes?
2. Given $a, b \in \mathbf{Z}$, are there infinitely many primes of the form $ax + b$?

3. How are the primes spaced along the number line?

We first show that there are infinitely many primes, then state Dirichlet's theorem that if $\gcd(a, b) = 1$, then $ax + b$ is a prime for infinitely many values of $x$. Finally, we discuss the Prime Number Theorem which asserts that there are asymptotically $x/\log(x)$ primes less than $x$, and we make a connection between this asymptotic formula and the Riemann Hypothesis.

### *1.2.1 There Are Infinitely Many Primes*

Each number on the left in the following table is prime. We will see soon that this pattern does not continue indefinitely, but something similar works.

$$\begin{aligned} 3 &= 2 + 1 \\ 7 &= 2 \cdot 3 + 1 \\ 31 &= 2 \cdot 3 \cdot 5 + 1 \\ 211 &= 2 \cdot 3 \cdot 5 \cdot 7 + 1 \\ 2311 &= 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 \end{aligned}$$

**Theorem 1.2.1** (Euclid)**.** *There are infinitely many primes.*

*Proof.* Suppose that $p_1, p_2, \ldots, p_n$ are $n$ distinct primes. We construct a prime $p_{n+1}$ not equal to any of $p_1, \ldots, p_n$, as follows. If

$$N = p_1 p_2 p_3 \cdots p_n + 1, \tag{1.2.1}$$

then by Proposition 1.1.20 there is a factorization

$$N = q_1 q_2 \cdots q_m$$

with each $q_i$ prime and $m \geq 1$. If $q_1 = p_i$ for some $i$, then $p_i \mid N$. Because of (1.2.1), we also have $p_i \mid N - 1$, so $p_i \mid 1 = N - (N - 1)$, which is a contradiction. Thus the prime $p_{n+1} = q_1$ is not in the list $p_1, \ldots, p_n$, and we have constructed our new prime. □

For example,

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509.$$

Multiplying together the first six primes and adding 1 doesn't produce a prime, but it produces an integer that is merely divisible by a new prime.

**Joke 1.2.2** (Hendrik Lenstra)**.** *There are infinitely many composite numbers. Proof.* To obtain a new composite number, multiply together the first $n$ composite numbers and don't add 1.

### *1.2.2 Enumerating Primes*

In this section we describe a sieving process that allows us to enumerate all primes up to $n$. The sieve works by first writing down all numbers up to $n$, noting that 2 is prime, and crossing off all multiples of 2. Next, note that the first number not crossed off is 3, which is prime, and cross off all multiples of 3, etc. Repeating this process, we obtain a list of the primes up to $n$. Formally, the algorithm is as follows:

**Algorithm 1.2.3** (Prime Sieve)**.** Given a positive integer $n$, this algorithm computes a list of the primes up to $n$.

1. [Initialize] Let $X = [3, 5, \ldots]$ be the list of all odd integers between 3 and $n$. Let $P = [2]$ be the list of primes found so far.
2. [Finished?] Let $p$ be the first element of $X$. If $p \geq \sqrt{n}$, append each element of $X$ to $P$ and terminate. Otherwise append $p$ to $P$.
3. [Cross Off] Set $X$ equal to the sublist of elements in $X$ that are not divisible by $p$. Go to Step 2.

For example, to list the primes $\leq 40$ using the sieve, we proceed as follows. First $P = [2]$ and

$$X = [3, 5, 7, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39].$$

We append 3 to $P$ and cross off all multiples of 3 to obtain the new list

$$X = [5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37].$$

Next we append 5 to $P$, obtaining $P = [2, 3, 5]$, and cross off the multiples of 5, to obtain $X = [7, 11, 13, 17, 19, 23, 29, 31, 37]$. Because $7^2 \geq 40$, we append $X$ to $P$ and find that the primes less than 40 are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37.$$

*Proof of Algorithm 1.2.3.* The part of the algorithm that is not clear is that when the first element $a$ of $X$ satisfies $a \geq \sqrt{n}$, then each element of $X$ is prime. To see this, suppose $m$ is in $X$, so $\sqrt{n} \leq m \leq n$ and that $m$ is divisible by no prime that is $\leq \sqrt{n}$. Write $m = \prod p_i^{e_i}$ with the $p_i$ distinct primes ordered so that $p_1 < p_2 < \ldots$. If $p_i > \sqrt{n}$ for each $i$ and there is more than one $p_i$, then $m > n$, a contradiction. Thus some $p_i$ is less than $\sqrt{n}$, which also contradicts our assumptions on $m$. □

### *1.2.3 The Largest Known Prime*

Though Theorem 1.2.1 implies that there are infinitely many primes, it still makes sense to ask the question "What is the largest *known* prime?"

A *Mersenne prime* is a prime of the form $2^q - 1$. According to [Cal] the largest known prime as of March 2007 is the 44th known Mersenne prime

$$p = 2^{32582657} - 1,$$

which has 9,808,358 decimal digits[1]. This would take over 2000 pages to print, assuming a page contains 60 lines with 80 characters per line. The Electronic Frontier Foundation has offered a $100,000 prize to the first person who finds a 10,000,000 digit prime.

Euclid's theorem implies that there definitely are infinitely many primes bigger than $p$. Deciding whether or not a number is prime is interesting, as a theoretical problem, and as a problem with applications to cryptography, as we will see in Section 2.4 and Chapter 3.

*SAGE Example* 1.2.4. We can compute the decimal expansion of $p$ in Sage, although watch out as this is a serious computation that may take around a minute on your computer. Also, do not print out $p$ or $s$ below, because both would take a very long time to scroll by.

```
sage: p = 2^32582657 - 1
sage: p.ndigits()
9808358
```

Next we convert $p$ to a decimal string and look at some of the digits.

```
sage: s = p.str(10) # this takes a long time
sage: len(s)        # s is a very long string  (long time)
9808358
sage: s[:20]        # the first 20 digits of p (long time)
'12457502601536945540'
sage: s[-20:]       # the last 20 digits       (long time)
'11752880154053967871'
```

### *1.2.4 Primes of the Form $ax + b$*

Next we turn to primes of the form $ax+b$, where $a$ and $b$ are fixed integers with $a > 1$ and $x$ varies over the natural numbers $\mathbf{N}$. We assume that $\gcd(a, b) = 1$, because otherwise there is no hope that $ax + b$ is prime infinitely often. For example, $2x + 2 = 2(x + 1)$ is only prime if $x = 0$, and is not prime for any $x \in \mathbf{N}$.

**Proposition 1.2.5.** *There are infinitely many primes of the form $4x - 1$.*

Why might this be true? We list numbers of the form $4x-1$ and underline those that are prime.

$$\underline{3}, \underline{7}, \underline{11}, 15, \underline{19}, \underline{23}, 27, \underline{31}, 35, 39, \underline{43}, \underline{47}, \ldots$$

[1]The 45th known Mersenne prime may have been found on August 23, 2008 as this book goes to press.

Not only is it plausible that underlined numbers will continue to appear indefinitely, it is something we can easily prove.

*Proof.* Suppose $p_1, p_2, \ldots, p_n$ are distinct primes of the form $4x-1$. Consider the number

$$N = 4p_1p_2\cdots p_n - 1.$$

Then $p_i \nmid N$ for any $i$. Moreover, not every prime $p \mid N$ is of the form $4x+1$; if they all were, then $N$ would be of the form $4x+1$. Since $N$ is odd, each prime divisor $p_i$ is odd so there is a $p \mid N$ that is of the form $4x-1$. Since $p \neq p_i$ for any $i$, we have found a new prime of the form $4x-1$. We can repeat this process indefinitely, so the set of primes of the form $4x-1$ cannot be finite. □

Note that this proof does not work if $4x-1$ is replaced by $4x+1$, since a product of primes of the form $4x-1$ can be of the form $4x+1$.

*Example* 1.2.6. Set $p_1 = 3$, $p_2 = 7$. Then

$$N = 4 \cdot 3 \cdot 7 - 1 = \underline{83}$$

is a prime of the form $4x-1$. Next

$$N = 4 \cdot 3 \cdot 7 \cdot 83 - 1 = \underline{6971},$$

which is again a prime of the form $4x-1$. Again,

$$N = 4 \cdot 3 \cdot 7 \cdot 83 \cdot 6971 - 1 = 48601811 = 61 \cdot \underline{796751}.$$

This time 61 is a prime, but it is of the form $4x+1 = 4 \cdot 15 + 1$. However, 796751 is prime and $796751 = 4 \cdot 199188 - 1$. We are unstoppable.

$$N = 4 \cdot 3 \cdot 7 \cdot 83 \cdot 6971 \cdot 796751 - 1 = \underline{5591} \cdot 6926049421.$$

This time the small prime, 5591, is of the form $4x-1$ and the large one is of the form $4x+1$.

**Theorem 1.2.7** (Dirichlet)**.** *Let $a$ and $b$ be integers with* $\gcd(a,b) = 1$. *Then there are infinitely many primes of the form $ax+b$.*

Proofs of this theorem typically use tools from advanced number theory, and are beyond the scope of this book (see e.g., [FT93, §VIII.4]).

### *1.2.5 How Many Primes are There?*

We saw in Section 1.2.1 that there are infinitely many primes. In order to get a sense of just how many primes there are, we consider a few warm-up questions. Then we consider some numerical evidence and state the prime number theorem, which gives an asymptotic answer to our question,

and connect this theorem with a form of the famous Riemann Hypothesis. Our discussion of counting primes in this section is very cursory; for more details, read Crandall and Pomerance's excellent book [CP01, §1.1.5].

The following vague discussion is meant to motivate a precise way to measure the number (or percentage) of primes. What percentage of natural numbers are even? Answer: Half of them. What percentage of natural numbers are of the form $4x - 1$? Answer: One fourth of them. What percentage of natural numbers are perfect squares? Answer: Zero percent of all natural numbers, in the sense that the limit of the proportion of perfect squares to all natural numbers converges to 0. More precisely,

$$\lim_{x\to\infty} \frac{\#\{n \in \mathbf{N} : n \leq x \text{ and } n \text{ is a perfect square}\}}{x} = 0,$$

since the numerator is roughly $\sqrt{x}$ and $\lim_{x\to\infty} \frac{\sqrt{x}}{x} = 0$. Likewise, it is an easy consequence of Theorem 1.2.10 that zero percent of all natural numbers are prime (see Exercise 1.4).

We are thus led to ask another question: How many positive integers $\leq x$ are perfect squares? Answer: Roughly $\sqrt{x}$. In the context of primes, we ask,

**Question 1.2.8.** How many natural numbers $\leq x$ are prime?

Let

$$\pi(x) = \#\{p \in \mathbf{N} : p \leq x \text{ is a prime}\}.$$

For example,

$$\pi(6) = \#\{2, 3, 5\} = 3.$$

Some values of $\pi(x)$ are given in Table 1.1, and Figures 1.1 and 1.2 contain graphs of $\pi(x)$. These graphs look like straight lines, which maybe bend down slightly.

*SAGE Example* 1.2.9. To compute $\pi(x)$ in Sage use the `prime_pi(x)` command:

```
sage: prime_pi(6)
3
sage: prime_pi(100)
25
sage: prime_pi(3000000)
216816
```

We can also draw a plot of $\pi(x)$ using the `plot` command:

```
sage: plot(prime_pi, 1,1000, rgbcolor=(0,0,1))
```

Gauss was an inveterate computer: he wrote in an 1849 letter that there are 216,745 primes less than 3,000,000 (this is wrong but close; the correct count is 216,816).

TABLE 1.1. Values of $\pi(x)$

| $x$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\pi(x)$ | 25 | 46 | 62 | 78 | 95 | 109 | 125 | 139 | 154 | 168 |

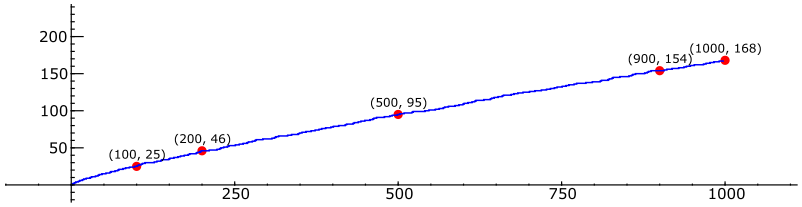



FIGURE 1.1. Graph of $\pi(x)$ for $x < 1000$

Gauss conjectured the following asymptotic formula for $\pi(x)$, which was later proved independently by Hadamard and Vallée Poussin in 1896 (but will not be proved in this book).

**Theorem 1.2.10** (Prime Number Theorem)**.** *The function $\pi(x)$ is asymptotic to $x/\log(x)$, in the sense that*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} = 1.$$

We do nothing more here than motivate this deep theorem with a few further observations. The theorem implies that

$$\lim_{x\to\infty} \frac{\pi(x)}{x} = \lim_{x\to\infty} \frac{1}{\log(x)} = 0,$$

so for any $a$,

$$\lim_{x\to\infty} \frac{\pi(x)}{x/(\log(x)-a)} = \lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} - \frac{a\pi(x)}{x} = 1.$$

Thus $x/(\log(x)-a)$ is also asymptotic to $\pi(x)$ for any $a$. See [CP01, §1.1.5] for a discussion of why $a = 1$ is the best choice. Table 1.2 compares $\pi(x)$ and $x/(\log(x)-1)$ for several $x < 10000$.

The record for counting primes is

$$\pi(10^{23}) = 1925320391606803968923.$$

Note that such computations are very difficult to get exactly right, so the above might be slightly wrong.

For the reader familiar with complex analysis, we mention a connection between $\pi(x)$ and the Riemann Hypothesis. The Riemann zeta function $\zeta(s)$ is a complex analytic function on $\mathbf{C} \setminus \{1\}$ that extends the function

TABLE 1.2. Comparison of $\pi(x)$ and $x/(\log(x)-1)$

| $x$ | $\pi(x)$ | $x/(\log(x)-1)$ (approx) |
|---|---|---|
| 1000 | 168 | 169.26902906044081651862562 78 |
| 2000 | 303 | 302.98887345454638780298009 94 |
| 3000 | 430 | 428.18193179752370437473857 40 |
| 4000 | 550 | 548.39220972782532641334009 85 |
| 5000 | 669 | 665.14187844865021723694558 15 |
| 6000 | 783 | 779.26988858547786268636773 74 |
| 7000 | 900 | 891.30356572233399743525677 59 |
| 8000 | 1007 | 1001.60296279477008075478428 1 |
| 9000 | 1117 | 1110.42842296318817231067501 1 |
| 10000 | 1229 | 1217.97630146155027920077570 5 |

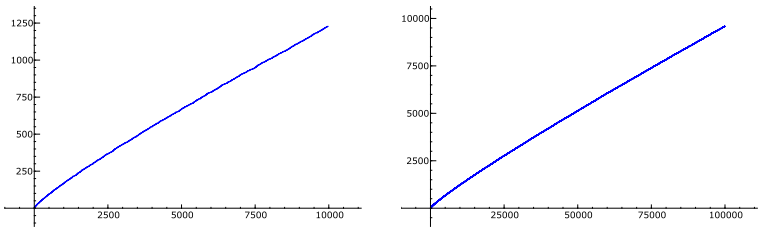



FIGURE 1.2. Graphs of $\pi(x)$ for $x < 10000$ and $x < 100000$

defined on a right half plane by $\sum_{n=1}^{\infty} n^{-s}$. The Riemann Hypothesis is the conjecture that the zeros in $\mathbf{C}$ of $\zeta(s)$ with positive real part lie on the line $\mathrm{Re}(s) = 1/2$. This conjecture is one of the Clay Math Institute million dollar millennium prize problems [Cla].

According to [CP01, §1.4.1], the Riemann Hypothesis is equivalent to the conjecture that

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\log(t)} dt$$

is a "good" approximation to $\pi(x)$, in the following precise sense.

**Conjecture 1.2.11** (Equivalent to the Riemann Hypothesis)**.** *For all* $x \geq 2.01$,

$$|\pi(x) - \mathrm{Li}(x)| \leq \sqrt{x}\log(x).$$

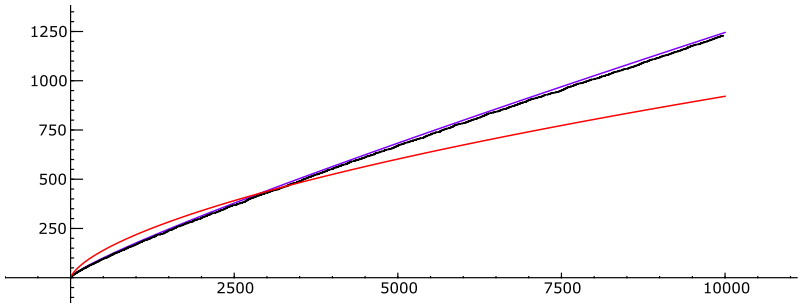
If $x = 2$, then $\pi(2) = 1$ and $\mathrm{Li}(2) = 0$, but $\sqrt{2}\log(2) = 0.9802\ldots$, so the inequality is not true for $x \geq 2$, but 2.01 is big enough. We will do nothing more to explain this conjecture, and settle for one numerical example.

*Example* 1.2.12. Let $x = 4 \cdot 10^{22}$. Then

$$\begin{aligned}
\pi(x) &= 783964159847056303858,\\
\mathrm{Li}(x) &= 783964159852157952242.7155276025801473\ldots,\\
|\pi(x) - \mathrm{Li}(x)| &= 5101648384.71552760258014\ldots,\\
\sqrt{x}\log(x) &= 10408633281397.77913344605\ldots,\\
x/(\log(x) - 1) &= 783650443647303761503.5237113087392967\ldots.
\end{aligned}$$

*SAGE Example* 1.2.13. We use Sage to graph $\pi(x)$, $\mathrm{Li}(x)$, and $\sqrt{x}\log(x)$.

```
sage: P = plot(Li, 2,10000, rgbcolor='purple')
sage: Q = plot(prime_pi, 2,10000, rgbcolor='black')
sage: R = plot(sqrt(x)*log(x),2,10000,rgbcolor='red')
sage: show(P+Q+R,xmin=0, figsize=[8,3])
```




The topmost line is $\mathrm{Li}(x)$, the next line is $\pi(x)$, and the bottom line is $\sqrt{x}\log(x)$.

For more on the prime number theorem and the Riemann hypothesis see [Zag75] and [MS08].

## 1.3 Exercises

1.1 Compute the greatest common divisor $\gcd(455, 1235)$ by hand.

1.2 Use the prime enumeration sieve to make a list of all primes up to 100.

1.3 Prove that there are infinitely many primes of the form $6x - 1$.

1.4 Use Theorem 1.2.10 to deduce that $\lim_{x\to\infty} \dfrac{\pi(x)}{x} = 0$.

1.5 Let $\psi(x)$ be the number of primes of the form $4k-1$ that are $\leq x$. Use a computer to make a conjectural guess about $\lim_{x\to\infty} \psi(x)/\pi(x)$.

1.6 So far 44 Mersenne primes $2^p - 1$ have been discovered. Give a guess, backed up by an argument, about when the next Mersenne prime might be discovered (you will have to do some online research).

1.7 (a) Let $y = 10000$. Compute $\pi(y) = \#\{\text{primes } p \leq y\}$.

(b) The prime number theorem implies $\pi(x)$ is asymptotic to $\frac{x}{\log(x)}$. How close is $\pi(y)$ to $y/\log(y)$, where $y$ is as in (a)?

1.8 Let $a, b, c, n$ be integers. Prove that

(a) if $a \mid n$ and $b \mid n$ with $\gcd(a, b) = 1$, then $ab \mid n$.

(b) if $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.

1.9 Let $a, b, c, d$, and $m$ be integers. Prove that

(a) if $a \mid b$ and $b \mid c$ then $a \mid c$.

(b) if $a \mid b$ and $c \mid d$ then $ac \mid bd$.

(c) if $m \neq 0$, then $a \mid b$ if and only if $ma \mid mb$.

(d) if $d \mid a$ and $a \neq 0$, then $|d| \leq |a|$.

1.10 In each of the following, apply the division algorithm to find $q$ and $r$ such that $a = bq + r$ and $0 \leq r < |b|$:

$$a = 300, b = 17,\ a = 729, b = 31,\ a = 300, b = -17,\ a = 389, b = 4.$$

1.11 (a) (Do this part by hand.) Compute the greatest common divisor of 323 and 437 using the algorithm described in class that involves quotients and remainders (i.e., do not just factor $a$ and $b$).

(b) Compute by any means the greatest common divisor of

$$314159265358979323846264338$$

and

$$271828182845904523536028747.$$

1.12 (a) Suppose $a$, $b$ and $n$ are positive integers. Prove that if $a^n \mid b^n$, then $a \mid b$.

(b) Suppose $p$ is a prime and $a$ and $k$ are positive integers. Prove that if $p \mid a^k$, then $p^k \mid a^k$.

1.13 (a) Prove that if a positive integer $n$ is a perfect square, then $n$ cannot be written in the form $4k + 3$ for $k$ an integer. (Hint: Compute the remainder upon division by 4 of each of $(4m)^2$, $(4m + 1)^2$, $(4m + 2)^2$, and $(4m + 3)^2$.)

(b) Prove that no integer in the sequence

$$11, 111, 1111, 11111, 111111, \ldots$$

is a perfect square. (Hint: $111 \cdots 111 = 111 \cdots 108 + 3 = 4k + 3$.)

1.14 Prove that a positive integer $n$ is prime if and only if $n$ is not divisible by any prime $p$ with $1 < p \leq \sqrt{n}$.

# 2
# The Ring of Integers Modulo $n$

A startling fact about numbers is that it takes less than a second to decide with near certainty whether or not any given 1,000 digit number $n$ is a prime, *without actually factoring* $n$. The algorithm for this involves doing some arithmetic with $n$ that works differently depending on whether $n$ is prime or composite. In particular, we do arithmetic with the set (in fact, "ring") of integers $\{0, 1, \ldots, n-1\}$ using an innovative rule for addition and multiplication, where the sum and product of two elements of that set is again in that set.

Another surprising fact is that one can almost instantly compute the last 1,000 digits of a massive multi-billion digit number like $n = 1234^{1234567890}$ without explicitly writing down all the digits of $n$. Again, this calculation involves arithmetic with the ring $\{0, 1, \ldots, n-1\}$.

This chapter is about the ring $\mathbf{Z}/n\mathbf{Z}$ of integers modulo $n$, the beautiful structure this ring has, and how to apply it to the above mentioned problems, among others. It is foundational for the rest of this book. In Section 2.1, we discuss when linear equations modulo $n$ have a solution, then introduce the Euler $\varphi$ function and prove Euler's Theorem and Wilson's theorem. In Section 2.2, we prove the Chinese Remainer Theorem, which addresses simultaneous solubility of several linear equations modulo coprime moduli. With these theoretical foundations in place, in Section 2.3, we introduce algorithms for doing powerful computations modulo $n$, including computing large powers quickly, and solving linear equations. We finish in Section 2.4 with a discussion of recognizing prime numbers using arithmetic modulo $n$.

## 2.1 Congruences Modulo $n$

**Definition 2.1.1** (Group)**.** A *group* is a set $G$ equipped with a binary operation $G \times G \to G$ (denoted by multiplication below) and an identity element $1 \in G$ such that:

1. For all $a, b, c \in G$, we have $(ab)c = a(bc)$.
2. For each $a \in G$, we have $1a = a1 = a$, and there exists $b \in G$ such that $ab = 1$.

**Definition 2.1.2** (Abelian Group)**.** An *abelian group* is a group $G$ such that $ab = ba$ for every $a, b \in G$.

**Definition 2.1.3** (Ring)**.** A *ring* $R$ is a set equipped with binary operations $+$ and $\times$ and elements $0, 1 \in R$ such that $R$ is an abelian group under $+$, and for all $a, b, c \in R$ we have

- $1a = a1 = a$
- $(ab)c = a(bc)$
- $a(b + c) = ab + ac$.

If, in addition, $ab = ba$ for all $a, b \in R$, then we call $R$ a *commutative ring*.

In this section, we define the ring $\mathbf{Z}/n\mathbf{Z}$ of integers modulo $n$, introduce the Euler $\varphi$-function, and relate it to the multiplicative order of certain elements of $\mathbf{Z}/n\mathbf{Z}$.

If $a, b \in \mathbf{Z}$ and $n \in \mathbf{N}$, we say that $a$ is *congruent to $b$ modulo $n$* if $n \mid a-b$, and write $a \equiv b \pmod{n}$. Let $n\mathbf{Z} = (n)$ be the subset of $\mathbf{Z}$ consisting of all multiples of $n$ (this is called the "ideal of $\mathbf{Z}$ generated by $n$").

**Definition 2.1.4** (Integers Modulo $n$)**.** The ring $\mathbf{Z}/n\mathbf{Z}$ of *integers modulo $n$* is the set of equivalence classes of integers modulo $n$. It is equipped with its natural ring structure:

$$(a + n\mathbf{Z}) + (b + n\mathbf{Z}) = (a + b) + n\mathbf{Z}$$

$$(a + n\mathbf{Z}) \cdot (b + n\mathbf{Z}) = (a \cdot b) + n\mathbf{Z}.$$

*Example* 2.1.5. For example,

$$\mathbf{Z}/3\mathbf{Z} = \{\{\ldots, -3, 0, 3, \ldots\}, \{\ldots, -2, 1, 4, \ldots\}, \{\ldots, -1, 2, 5, \ldots\}\}$$

*SAGE Example* 2.1.6. In Sage, we list the elements of $\mathbf{Z}/n\mathbf{Z}$ as follows:

```
sage: R = Integers(3)
sage: list(R)
[0, 1, 2]
```

We use the notation $\mathbf{Z}/n\mathbf{Z}$ because $\mathbf{Z}/n\mathbf{Z}$ is the quotient of the ring $\mathbf{Z}$ by the "ideal" $n\mathbf{Z}$ of multiples of $n$. Because $\mathbf{Z}/n\mathbf{Z}$ is the quotient of a ring by an ideal, the ring structure on $\mathbf{Z}$ induces a ring structure on $\mathbf{Z}/n\mathbf{Z}$. We often let $a$ or $a \pmod n$ denote the equivalence class $a + n\mathbf{Z}$ of $a$.

**Definition 2.1.7** (Field)**.** A *field* $K$ is a ring such that for every nonzero element $a \in K$ there is an element $b \in K$ such that $ab = 1$.

For example, if $p$ is a prime, then $\mathbf{Z}/p\mathbf{Z}$ is a field (see Exercise 2.12).

**Definition 2.1.8** (Reduction Map and Lift)**.** We call the natural reduction map $\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$, which sends $a$ to $a + n\mathbf{Z}$, *reduction modulo* $n$. We also say that $a$ is a *lift* of $a + n\mathbf{Z}$. Thus, e.g., 7 is a lift of 1 mod 3, since $7 + 3\mathbf{Z} = 1 + 3\mathbf{Z}$.

We can use that arithmetic in $\mathbf{Z}/n\mathbf{Z}$ is well defined is to derive tests for divisibility by $n$ (see Exercise 2.8).

**Proposition 2.1.9.** *A number $n \in \mathbf{Z}$ is divisible by 3 if and only if the sum of the digits of $n$ is divisible by 3.*

*Proof.* Write

$$n = a + 10b + 100c + \cdots,$$

where the digits of $n$ are $a$, $b$, $c$, etc. Since $10 \equiv 1 \pmod 3$,

$$n = a + 10b + 100c + \cdots \equiv a + b + c + \cdots \pmod 3,$$

from which the proposition follows. □

### *2.1.1 Linear Equations Modulo n*

In this section, we are concerned with how to decide whether or not a linear equation of the form $ax \equiv b \pmod n$ has a solution modulo $n$. Algorithms for *computing* solutions to $ax \equiv b \pmod n$ are the topic of Section 2.3.

First, we prove a proposition that gives a criterion under which one can cancel a quantity from both sides of a congruence.

**Proposition 2.1.10** (Cancellation)**.** *If* $\gcd(c, n) = 1$ *and*

$$ac \equiv bc \pmod n,$$

*then* $a \equiv b \pmod n$.

*Proof.* By definition

$$n \mid ac - bc = (a - b)c.$$

Since $\gcd(n, c) = 1$, it follows from Theorem 1.1.6 that $n \mid a - b$, so

$$a \equiv b \pmod n,$$

as claimed. □

When $a$ has a multiplicative inverse $a'$ in $\mathbf{Z}/n\mathbf{Z}$ (i.e., $aa' \equiv 1 \pmod{n}$) then the equation $ax \equiv b \pmod{n}$ has a unique solution $x \equiv a'b \pmod{n}$. Thus, it is of interest to determine the units in $\mathbf{Z}/n\mathbf{Z}$, i.e., the elements which have a multiplicative inverse.

We will use complete sets of residues to prove that the units in $\mathbf{Z}/n\mathbf{Z}$ are exactly the $a \in \mathbf{Z}/n\mathbf{Z}$ such that $\gcd(\tilde{a}, n) = 1$ for any lift $\tilde{a}$ of $a$ to $\mathbf{Z}$ (it doesn't matter which lift).

**Definition 2.1.11** (Complete Set of Residues)**.** We call a subset $R \subset \mathbf{Z}$ of size $n$ whose reductions modulo $n$ are pairwise distinct a *complete set of residues* modulo $n$. In other words, a complete set of residues is a choice of representative for each equivalence class in $\mathbf{Z}/n\mathbf{Z}$.

For example,

$$R = \{0, 1, 2, \ldots, n-1\}$$

is a complete set of residues modulo $n$. When $n = 5$, $R = \{0, 1, -1, 2, -2\}$ is a complete set of residues.

**Lemma 2.1.12.** *If $R$ is a complete set of residues modulo $n$ and $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$, then $aR = \{ax : x \in R\}$ is also a complete set of residues modulo $n$.*

*Proof.* If $ax \equiv ax' \pmod{n}$ with $x, x' \in R$, then Proposition 2.1.10 implies that $x \equiv x' \pmod{n}$. Because $R$ is a complete set of residues, this implies that $x = x'$. Thus the elements of $aR$ have distinct reductions modulo $n$. It follows, since $\#aR = n$, that $aR$ is a complete set of residues modulo $n$. □

**Proposition 2.1.13** (Units)**.** *If $\gcd(a, n) = 1$, then the equation $ax \equiv b \pmod{n}$ has a solution, and that solution is unique modulo $n$.*

*Proof.* Let $R$ be a complete set of residues modulo $n$, so there is a unique element of $R$ that is congruent to $b$ modulo $n$. By Lemma 2.1.12, $aR$ is also a complete set of residues modulo $n$, so there is a unique element $ax \in aR$ that is congruent to $b$ modulo $n$, and we have $ax \equiv b \pmod{n}$. □

Algebraically, this proposition asserts that if $\gcd(a, n) = 1$, then the map $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$ given by left multiplication by $a$ is a bijection.

*Example* 2.1.14. Consider the equation $2x \equiv 3 \pmod{7}$, and the complete set $R = \{0, 1, 2, 3, 4, 5, 6\}$ of coset representatives. We have

$$2R = \{0, 2, 4, 6, 8 \equiv 1, 10 \equiv 3, 12 \equiv 5\},$$

so $2 \cdot 5 \equiv 3 \pmod{7}$.

When $\gcd(a, n) \neq 1$, then the equation $ax \equiv b \pmod{n}$ may or may not have a solution. For example, $2x \equiv 1 \pmod{4}$ has no solution, but $2x \equiv 2 \pmod{4}$ does, and in fact it has more than one mod 4 ($x = 1$ and $x = 3$). Generalizing Proposition 2.1.13, we obtain the following more general criterion for solvability.

**Proposition 2.1.15** (Solvability)**.** *The equation* $ax \equiv b \pmod{n}$ *has a solution if and only if* $\gcd(a, n)$ *divides* $b$*.*

*Proof.* Let $g = \gcd(a, n)$. If there is a solution $x$ to the equation $ax \equiv b \pmod{n}$, then $n \mid (ax - b)$. Since $g \mid n$ and $g \mid a$, it follows that $g \mid b$.

Conversely, suppose that $g \mid b$. Then $n \mid (ax - b)$ if and only if

$$\frac{n}{g} \mid \left(\frac{a}{g}x - \frac{b}{g}\right).$$

Thus $ax \equiv b \pmod{n}$ has a solution if and only if $\frac{a}{g}x \equiv \frac{b}{g} \pmod{\frac{n}{g}}$ has a solution. Since $\gcd(a/g, n/g) = 1$, Proposition 2.1.13 implies this latter equation does have a solution. □

In Chapter 4, we will study quadratic reciprocity, which gives a nice criterion for whether or not a quadratic equation modulo $n$ has a solution.

### *2.1.2 Euler's Theorem*

Let $(\mathbf{Z}/n\mathbf{Z})^*$ denote the set of elements $[x] \in \mathbf{Z}/n\mathbf{Z}$ such that $\gcd(x, n) = 1$.

The set $(\mathbf{Z}/n\mathbf{Z})^*$ is a group, called the *group of units of the ring* $\mathbf{Z}/n\mathbf{Z}$; it will be of great interest to us. Each element of this group has an order, and Lagrange's theorem from group theory implies that each element of $(\mathbf{Z}/n\mathbf{Z})^*$ has an order that divides the order of $(\mathbf{Z}/n\mathbf{Z})^*$. In elementary number theory, this fact goes by the monicker "Fermat's Little Theorem" when $n$ is prime and "Euler's Theorem" in general, and we reprove it from basic principles in this section.

**Definition 2.1.16** (Order of an Element)**.** Let $n \in \mathbf{N}$ and $x \in \mathbf{Z}$ and suppose that $\gcd(x, n) = 1$. The *order* of $x$ modulo $n$ is the smallest $m \in \mathbf{N}$ such that

$$x^m \equiv 1 \pmod{n}.$$

To show that the definition makes sense, we verify that such an $m$ exists. Consider $x, x^2, x^3, \ldots$ modulo $n$. There are only finitely many residue classes modulo $n$, so we must eventually find two integers $i, j$ with $i < j$ such that

$$x^j \equiv x^i \pmod{n}.$$

Since $\gcd(x, n) = 1$, Proposition 2.1.10 implies that we can cancel $x$'s and conclude that

$$x^{j-i} \equiv 1 \pmod{n}.$$

*SAGE Example* 2.1.17. Use `x.multiplicative_order()` to compute the order of an element of $\mathbf{Z}/n\mathbf{Z}$ in Sage.

```
sage: R = Integers(10)
sage: a = R(3)                  # create an element of Z/10Z
sage: a.multiplicative_order()
4
```

Notice that the powers of $a$ are periodic with period 4, i.e., there are four powers and they repeat:

```
sage: [a^i for i in range(15)]
[1, 3, 9, 7, 1, 3, 9, 7, 1, 3, 9, 7, 1, 3, 9]
```

The command `range(n)` we use above returns the list of integers between 0 and $n-1$, inclusive.

**Definition 2.1.18** (Euler's $\varphi$-function)**.** For $n \in \mathbf{N}$, let

$$\varphi(n) = \#\{a \in \mathbf{N} : a \leq n \text{ and } \gcd(a, n) = 1\}.$$

For example,

$$\begin{aligned}\varphi(1) &= \#\{1\} = 1,\\ \varphi(2) &= \#\{1\} = 1,\\ \varphi(5) &= \#\{1, 2, 3, 4\} = 4,\\ \varphi(12) &= \#\{1, 5, 7, 11\} = 4.\end{aligned}$$

Also, if $p$ is any prime number then

$$\varphi(p) = \#\{1, 2, \ldots, p-1\} = p - 1.$$

In Section 2.2.1, we prove that if $\gcd(m, r) = 1$, then $\varphi(mr) = \varphi(m)\varphi(r)$. This will yield an easy way to compute $\varphi(n)$ in terms of the prime factorization of $n$.

*SAGE Example* 2.1.19. Use the `euler_phi(n)` command to compute $\varphi(n)$ in Sage:

```
sage: euler_phi(2007)
1332
```

**Theorem 2.1.20** (Euler's Theorem)**.** *If* $\gcd(x, n) = 1$*, then*

$$x^{\varphi(n)} \equiv 1 \pmod{n}.$$

*Proof.* As mentioned above, Euler's Theorem has the following group-theoretic interpretation. The set of units in $\mathbf{Z}/n\mathbf{Z}$ is a group

$$(\mathbf{Z}/n\mathbf{Z})^* = \{a \in \mathbf{Z}/n\mathbf{Z} : \gcd(a, n) = 1\}$$

that has order $\varphi(n)$. The theorem then asserts that the order of an element of $(\mathbf{Z}/n\mathbf{Z})^*$ divides the order $\varphi(n)$ of $(\mathbf{Z}/n\mathbf{Z})^*$. This is a special case of

the more general fact (Lagrange's Theorem) that if $G$ is a finite group and $g \in G$, then the order of $g$ divides the cardinality of $G$.

We now give an elementary proof of the theorem. Let

$$P = \{a : 1 \leq a \leq n \text{ and } \gcd(a, n) = 1\}.$$

In the same way that we proved Lemma 2.1.12, we see that the reductions modulo $n$ of the elements of $xP$ are the same as the reductions of the elements of $P$. Thus

$$\prod_{a \in P} (xa) \equiv \prod_{a \in P} a \pmod{n},$$

since the products are over the same numbers modulo $n$. Now cancel the $a$'s on both sides to get

$$x^{\#P} \equiv 1 \pmod{n},$$

as claimed. □

*SAGE Example* 2.1.21. We illustrate Euler's Theorem using Sage. The `Mod(x,n)` command returns the equivalence class of $x$ in $\mathbf{Z}/n\mathbf{Z}$.

```
sage: n = 20
sage: k = euler_phi(n); k
8
sage: [Mod(x,n)^k for x in range(n) if gcd(x,n) == 1]
[1, 1, 1, 1, 1, 1, 1, 1]
```

### *2.1.3 Wilson's Theorem*

The following characterization of prime numbers, from the 1770s, is called "Wilson's Theorem," though it was first proved by Lagrange.

**Proposition 2.1.22** (Wilson's Theorem)**.** *An integer $p > 1$ is prime if and only if $(p-1)! \equiv -1 \pmod{p}$.*

For example, if $p = 3$, then $(p-1)! = 2 \equiv -1 \pmod{3}$. If $p = 17$, then

$$(p-1)! = 20922789888000 \equiv -1 \pmod{17}.$$

But if $p = 15$, then

$$(p-1)! = 87178291200 \equiv 0 \pmod{15},$$

so 15 is composite. Thus Wilson's theorem could be viewed as a primality test, though, from a computational point of view, it is probably one of the world's *least efficient* primality tests since computing $(n-1)!$ takes so many steps.

*Proof.* The statement is clear when $p = 2$, so henceforth we assume that $p > 2$. We first assume that $p$ is prime and prove that $(p-1)! \equiv -1 \pmod{p}$. If $a \in \{1, 2, \ldots, p-1\}$, then the equation

$$ax \equiv 1 \pmod{p}$$

has a unique solution $a' \in \{1, 2, \ldots, p-1\}$. If $a = a'$, then $a^2 \equiv 1 \pmod{p}$, so $p \mid a^2 - 1 = (a-1)(a+1)$, so $p \mid (a-1)$ or $p \mid (a+1)$, so $a \in \{1, p-1\}$. We can thus pair off the elements of $\{2, 3, \ldots, p-2\}$, each with their inverse. Thus

$$2 \cdot 3 \cdot \cdots \cdot (p-2) \equiv 1 \pmod{p}.$$

Multiplying both sides by $p-1$ proves that $(p-1)! \equiv -1 \pmod{p}$.

Next, we assume that $(p-1)! \equiv -1 \pmod{p}$ and prove that $p$ must be prime. Suppose not, so that $p \geq 4$ is a composite number. Let $\ell$ be a prime divisor of $p$. Then $\ell < p$, so $\ell \mid (p-1)!$. Also, by assumption,

$$\ell \mid p \mid ((p-1)! + 1).$$

This is a contradiction, because a prime can not divide a number $a$ and also divide $a + 1$, since it would then have to divide $(a+1) - a = 1$. □

*Example* 2.1.23. We illustrate the key step in the above proof in the case $p = 17$. We have

$$2{\cdot}3 \cdots 15 = (2{\cdot}9){\cdot}(3{\cdot}6){\cdot}(4{\cdot}13){\cdot}(5{\cdot}7){\cdot}(8{\cdot}15){\cdot}(10{\cdot}12){\cdot}(14{\cdot}11) \equiv 1 \pmod{17},$$

where we have paired up the numbers $a, b$ for which $ab \equiv 1 \pmod{17}$.

*SAGE Example* 2.1.24. We use Sage to create a table of triples; the first column contains $n$, the second column contains $(n-1)!$ modulo $n$, and the third contains $-1$ modulo $n$. Notice that the first columns contains a prime precisely when the second and third columns are equal. (The ... notation indicates a multi-line command in Sage; you should not type the dots in explicitly.)

```
sage: for n in range(1,10):
...    print n, factorial(n-1) % n, -1 % n
1 0 0
2 1 1
3 2 2
4 2 3
5 4 4
6 0 5
7 6 6
8 0 7
9 0 8
```

## 2.2 The Chinese Remainder Theorem

In this section, we prove the Chinese Remainder Theorem, which gives conditions under which a system of linear equations is guaranteed to have a solution. In the 4th century a Chinese mathematician asked the following:

**Question 2.2.1.** There is a quantity whose number is unknown. Repeatedly divided by 3, the remainder is 2; by 5 the remainder is 3; and by 7 the remainder is 2. What is the quantity?

In modern notation, Question 2.2.1 asks us to find a positive integer solution to the following system of three equations:

$$\begin{aligned} x &\equiv 2 \pmod 3 \\ x &\equiv 3 \pmod 5 \\ x &\equiv 2 \pmod 7 \end{aligned}$$

The Chinese Remainder Theorem asserts that a solution exists, and the proof gives a method to find one. (See Section 2.3 for the necessary algorithms.)

**Theorem 2.2.2** (Chinese Remainder Theorem)**.** *Let $a, b \in \mathbf{Z}$ and $n, m \in \mathbf{N}$ such that $\gcd(n, m) = 1$. Then there exists $x \in \mathbf{Z}$ such that*

$$\begin{aligned} x &\equiv a \pmod m, \\ x &\equiv b \pmod n. \end{aligned}$$

*Moreover $x$ is unique modulo $mn$.*

*Proof.* If we can solve for $t$ in the equation

$$a + tm \equiv b \pmod n,$$

then $x = a + tm$ will satisfy both congruences. To see that we can solve, subtract $a$ from both sides and use Proposition 2.1.13 together with our assumption that $\gcd(n, m) = 1$ to see that there is a solution.

For uniqueness, suppose that $x$ and $y$ solve both congruences. Then $z = x - y$ satisfies $z \equiv 0 \pmod m$ and $z \equiv 0 \pmod n$, so $m \mid z$ and $n \mid z$. Since $\gcd(n, m) = 1$, it follows that $nm \mid z$, so $x \equiv y \pmod{nm}$. □

**Algorithm 2.2.3** (Chinese Remainder Theorem)**.** Given coprime integers $m$ and $n$ and integers $a$ and $b$, this algorithm find an integer $x$ such that $x \equiv a \pmod m$ and $x \equiv b \pmod n$.

1. [Extended GCD] Use Algorithm 2.3.7 below to find integers $c, d$ such that $cm + dn = 1$.
2. [Answer] Output $x = a + (b - a)cm$ and terminate.

*Proof.* Since $c \in \mathbf{Z}$, we have $x \equiv a \pmod{m}$, and using that $cm + dn = 1$, we have $a + (b-a)cm \equiv a + (b-a) \equiv b \pmod{n}$. □

Now we can answer Question 2.2.1. First, we use Theorem 2.2.2 to find a solution to the pair of equations

$$\begin{aligned} x &\equiv 2 \pmod{3}, \\ x &\equiv 3 \pmod{5}. \end{aligned}$$

Set $a = 2$, $b = 3$, $m = 3$, $n = 5$. Step 1 is to find a solution to $t \cdot 3 \equiv 3 - 2 \pmod{5}$. A solution is $t = 2$. Then $x = a + tm = 2 + 2 \cdot 3 = 8$. Since any $x'$ with $x' \equiv x \pmod{15}$ is also a solution to those two equations, we can solve all three equations by finding a solution to the pair of equations

$$\begin{aligned} x &\equiv 8 \pmod{15} \\ x &\equiv 2 \pmod{7}. \end{aligned}$$

Again, we find a solution to $t \cdot 15 \equiv 2 - 8 \pmod{7}$. A solution is $t = 1$, so

$$x = a + tm = 8 + 15 = 23.$$

Note that there are other solutions. Any $x' \equiv x \pmod{3 \cdot 5 \cdot 7}$ is also a solution; e.g., $23 + 3 \cdot 5 \cdot 7 = 128$.

*SAGE Example* 2.2.4. The `CRT(a,b,m,n)` command in Sage computes an integer $x$ such that $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$. For example,

```
sage: CRT(2,3, 3, 5)
-7
```

The `CRT_list` command computes a number that reduces to several numbers modulo coprime moduli. We use it to answer Question 2.2.1:

```
sage: CRT_list([2,3,2], [3,5,7])
23
```

### *2.2.1 Multiplicative Functions*

Recall from Definition 2.1.18 that the *Euler $\varphi$-function* is

$$\varphi(n) = \#\{a : 1 \le a \le n \text{ and } \gcd(a, n) = 1\}.$$

**Lemma 2.2.5.** *Suppose that $m, n \in \mathbf{N}$ and $\gcd(m, n) = 1$. Then the map*

$$\psi : (\mathbf{Z}/mn\mathbf{Z})^* \to (\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*. \tag{2.2.1}$$

*defined by*

$$\psi(c) = (c \bmod m,\ c \bmod n)$$

*is a bijection.*

*Proof.* We first show that $\psi$ is injective. If $\psi(c) = \psi(c')$, then $m \mid c - c'$ and $n \mid c - c'$, so $nm \mid c - c'$ because $\gcd(n, m) = 1$. Thus $c = c'$ as elements of $(\mathbf{Z}/mn\mathbf{Z})^*$.

Next we show that $\psi$ is surjective, i.e., that every element of $(\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*$ is of the form $\psi(c)$ for some $c$. Given $a$ and $b$ with $\gcd(a, m) = 1$ and $\gcd(b, n) = 1$, Theorem 2.2.2 implies that there exists $c$ with $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$. We may assume that $1 \leq c \leq nm$, and since $\gcd(a, m) = 1$ and $\gcd(b, n) = 1$, we must have $\gcd(c, nm) = 1$. Thus $\psi(c) = (a, b)$. □

**Definition 2.2.6** (Multiplicative Function)**.** A function $f : \mathbf{N} \to \mathbf{C}$ is *multiplicative* if, whenever $m, n \in \mathbf{N}$ and $\gcd(m, n) = 1$, we have

$$f(mn) = f(m) \cdot f(n).$$

**Proposition 2.2.7** (Multiplicativity of $\varphi$)**.** *The function $\varphi$ is multiplicative.*

*Proof.* The map $\psi$ of Lemma 2.2.5 is a bijection, so the set on the left in (2.2.1) has the same size as the product set on the right in (2.2.1). Thus

$$\varphi(mn) = \varphi(m) \cdot \varphi(n).$$

□

The proposition is helpful in computing $\varphi(n)$, at least if we assume we can compute the factorization of $n$ (see Section 3.4.1 for a connection between factoring $n$ and computing $\varphi(n)$). For example,

$$\varphi(12) = \varphi(2^2) \cdot \varphi(3) = 2 \cdot 2 = 4.$$

Also, for $n \geq 1$, we have

$$\varphi(p^n) = p^n - \frac{p^n}{p} = p^n - p^{n-1} = p^{n-1}(p - 1), \tag{2.2.2}$$

since $\varphi(p^n)$ is the number of numbers less than $p^n$ minus the number of those that are divisible by $p$. Thus, e.g.,

$$\varphi(389 \cdot 11^2) = 388 \cdot (11^2 - 11) = 388 \cdot 110 = 42680.$$

## 2.3 Quickly Computing Inverses and Huge Powers

This section is about how to solve the equation $ax \equiv 1 \pmod{n}$ when we know it has a solution, and how to efficiently compute $a^m \pmod{n}$. We also discuss a simple probabilistic primality test that relies on our ability to compute $a^m \pmod{n}$ quickly. All three of these algorithms are of fundamental importance to the cryptography algorithms of Chapter 3.

### *2.3.1 How to Solve $ax \equiv 1 \pmod{n}$*

Suppose $a, n \in \mathbf{N}$ with $\gcd(a, n) = 1$. Then by Proposition 2.1.13 the equation $ax \equiv 1 \pmod{n}$ has a unique solution. How can we find it?

**Proposition 2.3.1** (Extended Euclidean Representation)**.** *Suppose $a, b \in \mathbf{Z}$ and let $g = \gcd(a, b)$. Then there exists $x, y \in \mathbf{Z}$ such that*

$$ax + by = g.$$

*Remark* 2.3.2. If $e = cg$ is a multiple of $g$, then $cax + cby = cg = e$, so $e = (cx)a + (cy)b$ can also be written in terms of $a$ and $b$.

*Proof of Proposition 2.3.1.* Let $g = \gcd(a, b)$. Then $\gcd(a/g, b/g) = 1$, so by Proposition 2.1.15, the equation

$$\frac{a}{g} \cdot x \equiv 1 \left( \mod \frac{b}{g} \right) \tag{2.3.1}$$

has a solution $x \in \mathbf{Z}$. Multiplying (2.3.1) through by $g$ yields $ax \equiv g \pmod{b}$, so there exists $y$ such that $b \cdot (-y) = ax - g$. Then $ax + by = g$, as required. □

Given $a, b$ and $g = \gcd(a, b)$, our proof of Proposition 2.3.1 gives a way to explicitly find $x, y$ such that $ax + by = g$, assuming one knows an algorithm to solve linear equations modulo $n$. Since we do not know such an algorithm, we now discuss a way to explicitly find $x$ and $y$. This algorithm will in fact enable us to solve linear equations modulo $n$. To solve $ax \equiv 1 \pmod{n}$ when $\gcd(a, n) = 1$, use the Algorithm 2.3.7 to find $x$ and $y$ such that $ax + ny = 1$. Then $ax \equiv 1 \pmod{n}$.

*Example* 2.3.3. Suppose $a = 5$ and $b = 7$. The steps of Algorithm 1.1.13 to compute $\gcd(5, 7)$ are as follows. Here we underline certain numbers, because it clarifies the subsequent back substitution we will use to find $x$ and $y$.

$$\begin{aligned} \underline{7} &= 1 \cdot \underline{5} + \underline{2} && \text{so } \underline{2} = \underline{7} - \underline{5} \\ \underline{5} &= 2 \cdot \underline{2} + \underline{1} && \text{so } \underline{1} = \underline{5} - 2 \cdot \underline{2} = \underline{5} - 2(\underline{7} - \underline{5}) = 3 \cdot \underline{5} - 2 \cdot \underline{7} \end{aligned}$$

On the right, we have back-substituted in order to write each partial remainder as a linear combination of $a$ and $b$. In the last step, we obtain $\gcd(a, b)$ as a linear combination of $a$ and $b$, as desired.

*Example* 2.3.4. That example was not too complicated, so we try another one. Let $a = 130$ and $b = 61$. We have

$$\begin{aligned} \underline{130} &= 2 \cdot \underline{61} + \underline{8} & \underline{8} &= \underline{130} - 2 \cdot \underline{61} \\ \underline{61} &= 7 \cdot \underline{8} + \underline{5} & \underline{5} &= -7 \cdot \underline{130} + 15 \cdot \underline{61} \\ \underline{8} &= 1 \cdot \underline{5} + \underline{3} & \underline{3} &= 8 \cdot \underline{130} - 17 \cdot \underline{61} \\ \underline{5} &= 1 \cdot \underline{3} + \underline{2} & \underline{2} &= -15 \cdot \underline{130} + 32 \cdot \underline{61} \\ \underline{3} &= 1 \cdot \underline{2} + \underline{1} & \underline{1} &= 23 \cdot \underline{130} - 49 \cdot \underline{61} \end{aligned}$$

Thus $x = 23$ and $y = -49$ is a solution to $130x + 61y = 1$.

*Example* 2.3.5. This example is just like Example 2.3.4 above, except we make the notation on the right more compact.

$$\begin{aligned} \underline{130} &= 2 \cdot \underline{61} + \underline{8} & \underline{8} &= (1, -2) \\ \underline{61} &= 7 \cdot \underline{8} + \underline{5} & \underline{5} &= (-7, 15) = (0, 1) - 7(1, -2) \\ \underline{8} &= 1 \cdot \underline{5} + \underline{3} & \underline{3} &= (8, -17) = (1, -2) - (-7, 15) \\ \underline{5} &= 1 \cdot \underline{3} + \underline{2} & \underline{2} &= (-15, 32) = (-7, 15) - (8, -17) \\ \underline{3} &= 1 \cdot \underline{2} + \underline{1} & \underline{1} &= (23, -49) = (8, -17) - (-15, 32) \end{aligned}$$

Notice at each step that the vector on the right is just the vector from two steps ago minus a multiple of the vector from one step ago, where the multiple is the cofficient of what we divide by.

*SAGE Example* 2.3.6. The `xgcd(a,b)` command computes the greatest common divisor $g$ of $a$ and $b$ along with $x, y$ such that $ax + by = g$.

```
sage: xgcd(5,7)
(1, -4, 3)
sage: xgcd(130,61)
(1, 23, -49)
```

**Algorithm 2.3.7** (Extended Euclidean Algorithm)**.** Suppose $a$ and $b$ are integers and let $g = \gcd(a, b)$. This algorithm finds $g$, $x$ and $y$ such that $ax + by = g$. We describe only the steps when $a > b \geq 0$, since one can easily reduce to this case.

1. [Initialize] Set $x = 1$, $y = 0$, $r = 0$, $s = 1$.
2. [Finished?] If $b = 0$, set $g = a$ and terminate.
3. [Quotient and Remainder] Use Algorithm 1.1.12 to write $a = qb + c$ with $0 \leq c < b$.
4. [Shift] Set $(a, b, r, s, x, y) = (b, c, x - qr, y - qs, r, s)$ and go to Step 2. (This shift step is nicely illustrated in Example 2.3.5.)

*Proof.* This algorithm is the same as Algorithm 1.1.13, except that we keep track of extra variables $x, y, r, s$, so it terminates and when it terminates $d = \gcd(a, b)$. We omit the rest of the inductive proof that the algorithm is correct, and instead refer the reader to [Knu97, §1.2.1]. □

**Algorithm 2.3.8** (Inverse Modulo $n$)**.** Suppose $a$ and $n$ are integers and $\gcd(a, n) = 1$. This algorithm finds an $x$ such that $ax \equiv 1 \pmod{n}$.

1. [Compute Extended GCD] Use Algorithm 2.3.7 to compute integers $x, y$ such that $ax + ny = \gcd(a, n) = 1$.
2. [Finished] Output $x$.

*Proof.* Reduce $ax+ny = 1$ modulo $n$ to see that $x$ satisfies $ax \equiv 1 \pmod{n}$. □

*Example* 2.3.9. Solve $17x \equiv 1 \pmod{61}$. First, we use Algorithm 2.3.7 to find $x, y$ such that $17x + 61y = 1$:

$$\begin{aligned} \underline{61} &= 3 \cdot \underline{17} + \underline{10} & \underline{10} &= \underline{61} - 3 \cdot \underline{17} \\ \underline{17} &= 1 \cdot \underline{10} + \underline{7} & \underline{7} &= -\underline{61} + 4 \cdot \underline{17} \\ \underline{10} &= 1 \cdot \underline{7} + \underline{3} & \underline{3} &= 2 \cdot \underline{61} - 7 \cdot \underline{17} \\ \underline{3} &= 2 \cdot \underline{3} + \underline{1} & \underline{1} &= -5 \cdot \underline{61} + 18 \cdot \underline{17} \end{aligned}$$

Thus $17 \cdot 18 + 61 \cdot (-5) = 1$ so $x = 18$ is a solution to $17x \equiv 1 \pmod{61}$.

*SAGE Example* 2.3.10. Sage implements the above algorithm for quickly computing inverses modulo $n$. For example,

```
sage: a = Mod(17, 61)
sage: a^(-1)
18
```

### *2.3.2 How to Compute $a^m \pmod{n}$*

Let $a$ and $n$ be integers, and $m$ a nonnegative integer. In this section, we describe an efficient algorithm to compute $a^m \pmod{n}$. For the cryptography applications in Chapter 3, $m$ will have hundreds of digits.

The naive approach to computing $a^m \pmod{n}$ is to simply compute $a^m = a \cdot a \cdots a \pmod{n}$ by repeatedly multiplying by $a$ and reducing modulo $m$. Note that after each arithmetic operation is completed, we reduce the result modulo $n$ so that the sizes of the numbers involved do not get too large. Nonetheless, this algorithm is horribly inefficient because it takes $m - 1$ multiplications, which is huge if $m$ has hundreds of digits.

A much more efficient algorithm for computing $a^m \pmod{n}$ involves writing $m$ in binary, then expressing $a^m$ as a product of expressions $a^{2^i}$, for various $i$. These latter expressions can be computed by repeatedly squaring $a^{2^i}$. This more clever algorithm is not "simpler," but it is vastly more efficient since the number of operations needed grows with the number of binary digits of $m$, whereas with the naive algorithm in the previous paragraph, the number of operations is $m - 1$.

**Algorithm 2.3.11** (Write a number in binary)**.** Let $m$ be a nonnegative integer. This algorithm writes $m$ in binary, so it finds $\varepsilon_i \in \{0, 1\}$ such that $m = \sum_{i=0}^{r} \varepsilon_i 2^i$ with each $\varepsilon_i \in \{0, 1\}$.

1. [Initialize] Set $i = 0$.
2. [Finished?] If $m = 0$, terminate.
3. [Digit] If $m$ is odd, set $\varepsilon_i = 1$, otherwise $\varepsilon_i = 0$. Increment $i$.

4. [Divide by 2] Set $m = \left\lfloor \frac{m}{2} \right\rfloor$, the greatest integer $\leq m/2$. Goto Step 2.

*SAGE Example* 2.3.12. To write a number in binary using Sage, use the `str` command:

```
sage: 100.str(2)
'1100100'
```

Notice the above is the correct binary expansion:

```
sage: 0*2^0 + 0*2^1 + 1*2^2 + 0*2^3 + 0*2^4 + 1*2^5 + 1*2^6
100
```

**Algorithm 2.3.13** (Compute Power)**.** Let $a$ and $n$ be integers and $m$ a nonnegative integer. This algorithm computes $a^m$ modulo $n$.

1. [Write in Binary] Write $m$ in binary using Algorithm 2.3.11, so $a^m = \prod_{\varepsilon_i=1} a^{2^i} \pmod{n}$.
2. [Compute Powers] Compute $a$, $a^2$, $a^{2^2} = (a^2)^2$, $a^{2^3} = (a^{2^2})^2$, etc., up to $a^{2^r}$, where $r+1$ is the number of binary digits of $m$.
3. [Multiply Powers] Multiply together the $a^{2^i}$ such that $\varepsilon_i = 1$, always working modulo $n$.

*Example* 2.3.14. We can compute the last 2 digits of $7^{91}$, by finding $7^{91} \pmod{100}$. First, because $\gcd(7, 100) = 1$, we have by Theorem 2.1.20 that $7^{\varphi(100)} \equiv 1 \pmod{100}$. Because $\varphi$ is multiplicative,

$$\varphi(100) = \varphi(2^2 \cdot 5^2) = (2^2 - 2) \cdot (5^2 - 5) = 40.$$

Thus $7^{40} \equiv 1 \pmod{100}$, hence

$$7^{91} \equiv 7^{40+40+11} \equiv 7^{11} \pmod{100}.$$

We now compute $7^{11} \pmod{100}$ using the above algorithm. First, write 11 in binary by repeatedly dividing by 2.

$$\begin{aligned} 11 &= 5 \cdot 2 + 1 \\ 5 &= 2 \cdot 2 + 1 \\ 2 &= 1 \cdot 2 + 0 \\ 1 &= 0 \cdot 2 + 1 \end{aligned}$$

So in binary, $(11)_2 = 1011$, which we check:

$$11 = 1 \cdot 8 + 1 \cdot 2 + 1.$$

Next, compute $a, a^2, a^4, a^8$ and output $a^8 \cdot a^2 \cdot a$. We have

$$\begin{aligned} a &= 7 \\ a^2 &\equiv 49 \\ a^4 &\equiv 49^2 \equiv 1 \\ a^8 &\equiv 1^2 \equiv 1 \end{aligned}$$

Note: it is easiest to square 49 by working modulo 4 and 25 and using the Chinese Remainder Theorem. Finally,

$$7^{91} \equiv 7^{11} \equiv a^8 \cdot a^2 \cdot a \equiv 1 \cdot 49 \cdot 7 \equiv 43 \pmod{100}.$$

*SAGE Example* 2.3.15. Sage implements the above algorithm for computing powers efficiently. For example,

```
sage: Mod(7,100)^91
43
```

We can also, of course, directly compute $7^{91}$ in Sage, though we would not want to do this by hand:

```
sage: 7^91
80153343160247310515380886994816022539378033762994852
00750196460484168019074 3
```

## 2.4 Primality Testing

**Theorem 2.4.1** (Pseudoprimality)**.** *An integer $p > 1$ is prime if and only if for* every $a \not\equiv 0 \pmod{p}$,

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

*Proof.* If $p$ is prime, then the statement follows from Proposition 2.1.22. If $p$ is composite, then there is a divisor $a$ of $p$ with $2 \leq a < p$. If $a^{p-1} \equiv 1 \pmod{p}$, then $p \mid a^{p-1} - 1$. Since $a \mid p$, we have $a \mid a^{p-1} - 1$, hence there exists an integer $k$ such that $ak = a^{p-1} - 1$. Subtracting, we see that $a^{p-1} - ak = 1$, so $a(a^{p-2} - k) = 1$. This implies that $a \mid 1$, which is a contradiction since $a \geq 2$. □

Suppose $n \in \mathbf{N}$. Using Theorem 2.4.1 and Algorithm 2.3.13, we can either quickly prove that $n$ is not prime, or convince ourselves that $n$ is likely prime (but not quickly prove that $n$ is prime). For example, if $2^{n-1} \not\equiv 1 \pmod{n}$, then we have proved that $n$ is not prime. On the other hand, if $a^{n-1} \equiv 1 \pmod{n}$ for a few $a$, it "seems likely" that $n$ is prime, and we loosely refer to such a number that seems prime for several bases as a *pseudoprime*.

There are composite numbers $n$ (called *Carmichael numbers*) with the amazing property that $a^{n-1} \equiv 1 \pmod{n}$ for *all* $a$ with $\gcd(a, n) = 1$. The first Carmichael number is 561, and it is a theorem that there are infinitely many such numbers ([AGP94]).

*Example* 2.4.2. Is $p = 323$ prime? We compute $2^{322} \pmod{323}$. Making a table as above, we have

| $i$ | $m$ | $\varepsilon_i$ | $2^{2^i} \bmod 323$ |
|---|---|---|---|
| 0 | 322 | 0 | 2 |
| 1 | 161 | 1 | 4 |
| 2 | 80 | 0 | 16 |
| 3 | 40 | 0 | 256 |
| 4 | 20 | 0 | 290 |
| 5 | 10 | 0 | 120 |
| 6 | 5 | 1 | 188 |
| 7 | 2 | 0 | 137 |
| 8 | 1 | 1 | 35 |

Thus

$$2^{322} \equiv 4 \cdot 188 \cdot 35 \equiv 157 \pmod{323},$$

so 323 is not prime, though this computation gives no information about how 323 factors as a product of primes. In fact, one finds that $323 = 17 \cdot 19$.

*SAGE Example* 2.4.3. It's possible to easily prove that a large number is composite, but the proof does not easily yield a factorization. For example if

$$n = 954680934860934509834095834098509348509384590 83,$$

then $2^{n-1} \not\equiv 1 \pmod{n}$, so $n$ is composite.

```
sage: n = 954680934860934509834095834098509348509384590 83
sage: Mod(2,n)^(n-1)
341734441392655538708302663785984070692486872 41
```

Note that factoring $n$ actually takes much longer than the above computation (which was essentially instant).

```
sage: factor(n)                 # takes up to a few seconds.
1610302526747 * 592858123864154884463971917910238 89
```

Another practical primality test is the Miller-Rabin test, which has the property that each time it is run on a number $n$ it either correctly asserts that the number is definitely not prime, or that it is probably prime, and the probability of correctness goes up with each successive call. If Miller-Rabin is called $m$ times on $n$ and in each case claims that $n$ is probably prime, then one can in a precise sense bound the probability that $n$ is composite in terms of $m$.

We state the Miller-Rabin algorithm precisely, but do not prove anything about the probability that it will succeed.

**Algorithm 2.4.4** (Miller-Rabin Primality Test)**.** Given an integer $n \geq 5$ this algorithm outputs either true or false. If it outputs true, then $n$ is "probably prime," and if it outputs false, then $n$ is definitely composite.

1. [Split Off Power of 2] Compute the unique integers $m$ and $k$ such that $m$ is odd and $n - 1 = 2^k \cdot m$.
2. [Random Base] Choose a random integer $a$ with $1 < a < n$.
3. [Odd Power] Set $b = a^m \pmod{n}$. If $b \equiv \pm 1 \pmod{n}$ output true and terminate.
4. [Even Powers] If $b^{2^r} \equiv -1 \pmod{n}$ for any $r$ with $1 \leq r \leq k-1$, output true and terminate. Otherwise output false.

If Miller-Rabin outputs true for $n$, we can call it again with $n$ and if it again outputs true then the probability that we have incorrectly determined that $n$ is prime (when $n$ is actually composite) decreases.

*Proof.* We will prove that the algorithm is correct, but will prove nothing about how likely the algorithm is to assert that a composite is prime. We must prove that if the algorithm pronounces an integer $n$ composite, then $n$ really is composite. Thus suppose $n$ is prime, yet the algorithm pronounces $n$ composite. Then $a^m \not\equiv \pm 1 \pmod{n}$, and for all $r$ with $1 \leq r \leq k-1$ we have $a^{2^r m} \not\equiv -1 \pmod{n}$. Since $n$ is prime and $2^{k-1}m = (n-1)/2$, Proposition 4.2.1 implies that $a^{2^{k-1}m} \equiv \pm 1 \pmod{n}$, so by our hypothesis $a^{2^{k-1}m} \equiv 1 \pmod{n}$. But then $(a^{2^{k-2}m})^2 \equiv 1 \pmod{n}$, so by Proposition 2.5.3 (which is proved right after it is stated, and whose proof does not depend on this argument), we have $a^{2^{k-2}m} \equiv \pm 1 \pmod{n}$. Again, by our hypothesis, this implies $a^{2^{k-2}m} \equiv 1 \pmod{n}$. Repeating this argument inductively, we see that $a^m \equiv \pm 1 \pmod{n}$, which contradicts our hypothesis on $a$. □

Until recently it was an open problem to give an algorithm (with proof) that decides whether or not any integer is prime in time bounded by a polynomial in the number of digits of the integer. Agrawal, Kayal, and Saxena recently found the first polynomial-time primality test (see [AKS02]). We will not discuss their algorithm further, because for our applications to cryptography Miller-Rabin or pseudoprimality tests will be sufficient. See [Sho05, Ch. 21] for a book that gives a detailed exposition of this algorithm.

*SAGE Example* 2.4.5. The `is_prime` command uses a combination of techniques to determine (provably correctly!) whether or not an integer is prime.

```
sage: n = 954680934860934509834095834098509348509384590 83
sage: is_prime(n)
False
```

We use the `is_prime` function to make a table of the first few Mersenne primes (see Section 1.2.3).

```
sage: for p in primes(100):
...   if is_prime(2^p - 1):
...       print p, 2^p - 1
2 3
3 7
5 31
7 127
13 8191
17 131071
19 524287
31 2147483647
61 2305843009213693951
89 618970019642690137449562111
```

There is a specialized test for primality of Mersenne numbers called the Lucas-Lehmer test. This remarkably simple algorithm determines provably correctly whether or not a number $2^p - 1$ is prime. We implement it in a few lines of code and use the Lucas-Lehmer test to check for primality of two Mersenne numbers:

```
sage: def is_prime_lucas_lehmer(p):
...    s = Mod(4, 2^p - 1)
...    for i in range(3, p+1):
...        s = s^2 - 2
...    return s == 0
sage: # Check primality of 2^9941 - 1
sage: is_prime_lucas_lehmer(9941)
True
sage: # Check primality of 2^next_prime(1000)-1
sage: is_prime_lucas_lehmer(next_prime(1000))
False
```

For more on Mersenne primes, see the Great Internet Mersenne Prime Search (GIMPS) project at `http://www.mersenne.org/`.

## 2.5 The Structure of $(\mathbf{Z}/p\mathbf{Z})^*$

This section is about the structure of the group $(\mathbf{Z}/p\mathbf{Z})^*$ of units modulo a prime number $p$. The main result is that this group is always cyclic. We will use this result later in Chapter 4 in our proof of quadratic reciprocity.

**Definition 2.5.1** (Primitive root)**.** A *primitive root* modulo an integer $n$ is an element of $(\mathbf{Z}/n\mathbf{Z})^*$ of order $\varphi(n)$.

We will prove that there is a primitive root modulo every prime $p$. Since the unit group $(\mathbf{Z}/p\mathbf{Z})^*$ has order $p-1$, this implies that $(\mathbf{Z}/p\mathbf{Z})^*$ is a cyclic group, a fact that will be extremely useful, since it completely determines the structure of $(\mathbf{Z}/p\mathbf{Z})^*$ as a group.

If $n$ is an odd prime power, then there is a primitive root modulo $n$ (see Exercise 2.28), but there is no primitive root modulo the prime power $2^3$, and hence none mod $2^n$ for $n \geq 3$ (see Exercise 2.27).

Section 2.5.1 is the key input to our proof that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic; here we show that for every divisor $d$ of $p-1$ there are exactly $d$ elements of $(\mathbf{Z}/p\mathbf{Z})^*$ whose order divides $d$. We then use this result in Section 2.5.2 to produce an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order $q^r$ when $q^r$ is a prime power that exactly divides $p-1$ (i.e., $q^r$ divides $p-1$, but $q^{r+1}$ does not divide $p-1$), and multiply together these elements to obtain an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order $p-1$.

*SAGE Example* 2.5.2. Use the `primitive_root` command to compute the smallest positive integer that is a primitive root modulo $n$. For example, below we compute primitive roots modulo $p$ for each prime $p < 20$.

```
sage: for p in primes(20):
...    print p, primitive_root(p)
2 1
3 2
5 2
7 3
11 2
13 2
17 3
19 2
```

### *2.5.1 Polynomials over* $\mathbf{Z}/p\mathbf{Z}$

The polynomials $x^2 - 1$ has four roots in $\mathbf{Z}/8\mathbf{Z}$, namely 1, 3, 5, and 7. In contrast, the following proposition shows that a polynomial of degree $d$ over a field, such as $\mathbf{Z}/p\mathbf{Z}$, can have at most $d$ roots.

**Proposition 2.5.3** (Root Bound)**.** *Let $f \in k[x]$ be a nonzero polynomial over a field $k$. Then there are at most* $\deg(f)$ *elements $\alpha \in k$ such that $f(\alpha) = 0$.*

*Proof.* We prove the proposition by induction on $\deg(f)$. The cases in which $\deg(f) \leq 1$ are clear. Write $f = a_n x^n + \cdots a_1 x + a_0$. If $f(\alpha) = 0$, then

$$\begin{aligned} f(x) &= f(x) - f(\alpha) \\ &= a_n(x^n - \alpha^n) + \cdots + a_1(x - \alpha) + a_0(1 - 1) \\ &= (x - \alpha)(a_n(x^{n-1} + \cdots + \alpha^{n-1}) + \cdots + a_2(x + \alpha) + a_1) \\ &= (x - \alpha)g(x), \end{aligned}$$

for some polynomial $g(x) \in k[x]$. Next, suppose that $f(\beta) = 0$ with $\beta \neq \alpha$. Then $(\beta - \alpha)g(\beta) = 0$, so, since $\beta - \alpha \neq 0$ and $k$ is a field, we have $g(\beta) = 0$. By our inductive hypothesis, $g$ has at most $n-1$ roots, so there are at most $n - 1$ possibilities for $\beta$. It follows that $f$ has at most $n$ roots. □

*SAGE Example* 2.5.4. We use Sage to find the roots of a polynomials over $\mathbf{Z}/13\mathbf{Z}$.

```
sage: R.<x> = PolynomialRing(Integers(13))
sage: f = x^15 + 1
sage: f.roots()
[(12, 1), (10, 1), (4, 1)]
sage: f(12)
0
```

The output of the roots command above lists each root along with its multiplicity (which is 1 in each case above).

**Proposition 2.5.5.** *Let $p$ be a prime number and let $d$ be a divisor of $p - 1$. Then $f = x^d - 1 \in (\mathbf{Z}/p\mathbf{Z})[x]$ has exactly $d$ roots in $\mathbf{Z}/p\mathbf{Z}$.*

*Proof.* Let $e = (p - 1)/d$. We have

$$\begin{aligned} x^{p-1} - 1 &= (x^d)^e - 1 \\ &= (x^d - 1)((x^d)^{e-1} + (x^d)^{e-2} + \cdots + 1) \\ &= (x^d - 1)g(x), \end{aligned}$$

where $g \in (\mathbf{Z}/p\mathbf{Z})[x]$ and $\deg(g) = de - d = p - 1 - d$. Theorem 2.1.20 implies that $x^{p-1} - 1$ has exactly $p - 1$ roots in $\mathbf{Z}/p\mathbf{Z}$, since every nonzero element of $\mathbf{Z}/p\mathbf{Z}$ is a root! By Proposition 2.5.3, $g$ has *at most* $p - 1 - d$ roots and $x^d - 1$ has at most $d$ roots. Since a root of $(x^d - 1)g(x)$ is a root of either $x^d - 1$ or $g(x)$ and $x^{p-1} - 1$ has $p - 1$ roots, $g$ must have exactly $p - 1 - d$ roots and $x^d - 1$ must have exactly $d$ roots, as claimed. □

*SAGE Example* 2.5.6. We use Sage to illustrate the proposition.

```
sage: R.<x> = PolynomialRing(Integers(13))
sage: f = x^6 + 1
sage: f.roots()
[(11, 1), (8, 1), (7, 1), (6, 1), (5, 1), (2, 1)]
```

We pause to reemphasize that the analog of Proposition 2.5.5 is false when $p$ is replaced by a composite integer $n$, since a root mod $n$ of a product of two polynomials need not be a root of either factor. For example, $f = x^2 - 1 = (x-1)(x+1) \in \mathbf{Z}/15\mathbf{Z}[x]$ has the four roots 1, 4, 11, and 14.

### *2.5.2 Existence of Primitive Roots*

Recall from Section 2.1.2 that the *order* of an element $x$ in a finite group is the smallest $m \geq 1$ such that $x^m = 1$. In this section, we prove that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic by using the results of Section 2.5.1 to produce an element of $(\mathbf{Z}/p\mathbf{Z})^*$ of order $d$ for each prime power divisor $d$ of $p-1$, and then we multiply these together to obtain an element of order $p-1$.

We will use the following lemma to assemble elements of each order dividing $p-1$ to produce an element of order $p-1$.

**Lemma 2.5.7.** *Suppose $a, b \in (\mathbf{Z}/n\mathbf{Z})^*$ have orders $r$ and $s$, respectively, and that $\gcd(r,s) = 1$. Then $ab$ has order $rs$.*

*Proof.* This is a general fact about commuting elements of any group; our proof only uses that $ab = ba$ and nothing special about $(\mathbf{Z}/n\mathbf{Z})^*$. Since

$$(ab)^{rs} = a^{rs}b^{rs} = 1,$$

the order of $ab$ is a divisor of $rs$. Write this divisor as $r_1 s_1$ where $r_1 \mid r$ and $s_1 \mid s$. Raise both sides of the equation

$$a^{r_1 s_1} b^{r_1 s_1} = (ab)^{r_1 s_1} = 1$$

to the power $r_2 = r/r_1$ to obtain

$$a^{r_1 r_2 s_1} b^{r_1 r_2 s_1} = 1.$$

Since $a^{r_1 r_2 s_1} = (a^{r_1 r_2})^{s_1} = 1$, we have

$$b^{r_1 r_2 s_1} = 1,$$

so $s \mid r_1 r_2 s_1$. Since $\gcd(s, r_1 r_2) = \gcd(s, r) = 1$, it follows that $s = s_1$. Similarly $r = r_1$, so the order of $ab$ is $rs$. □

**Theorem 2.5.8** (Primitive Roots)**.** *There is a primitive root modulo any prime $p$. In particular, the group $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic.*

*Proof.* The theorem is true if $p = 2$, since 1 is a primitive root, so we may assume $p > 2$. Write $p-1$ as a product of distinct prime powers $q_i^{n_i}$:

$$p - 1 = q_1^{n_1} q_2^{n_2} \cdots q_r^{n_r}.$$

By Proposition 2.5.5, the polynomial $x^{q_i^{n_i}} - 1$ has exactly $q_i^{n_i}$ roots, and the polynomial $x^{q_i^{n_i - 1}} - 1$ has exactly $q_i^{n_i - 1}$ roots. There are $q_i^{n_i} - q_i^{n_i - 1} =$

$q_i^{n_i-1}(q_i-1)$ elements $a \in \mathbf{Z}/p\mathbf{Z}$ such that $a^{q_i^{n_i}} = 1$ but $a^{q_i^{n_i-1}} \neq 1$; each of these elements has order $q_i^{n_i}$. Thus for each $i = 1, \ldots, r$, we can choose an $a_i$ of order $q_i^{n_i}$. Then, using Lemma 2.5.7 repeatedly, we see that

$$a = a_1 a_2 \cdots a_r$$

has order $q_1^{n_1} \cdots q_r^{n_r} = p-1$, so $a$ is a primitive root modulo $p$. □

*Example* 2.5.9. We illustrate the proof of Theorem 2.5.8 when $p = 13$. We have

$$p - 1 = 12 = 2^2 \cdot 3.$$

The polynomial $x^4 - 1$ has roots $\{1, 5, 8, 12\}$ and $x^2 - 1$ has roots $\{1, 12\}$, so we may take $a_1 = 5$. The polynomial $x^3 - 1$ has roots $\{1, 3, 9\}$, and we set $a_2 = 3$. Then $a = 5 \cdot 3 = 15 \equiv 2$ is a primitive root. To verify this, note that the successive powers of 2 (mod 13) are

$$2,\ 4,\ 8,\ 3,\ 6,\ 12,\ 11,\ 9,\ 5,\ 10,\ 7,\ 1.$$

*Example* 2.5.10. Theorem 2.5.8 is false if, for example, $p$ is replaced by a power of 2 bigger than 4. For example, the four elements of $(\mathbf{Z}/8\mathbf{Z})^*$ each have order dividing 2, but $\varphi(8) = 4$.

**Theorem 2.5.11** (Primitive Roots mod $p^n$)**.** *Let $p^n$ be a power of an odd prime. Then there is a primitive root modulo $p^n$.*

The proof is left as Exercise 2.28.

**Proposition 2.5.12** (Number of primitive roots)**.** *If there is a primitive root modulo $n$, then there are exactly $\varphi(\varphi(n))$ primitive roots modulo $n$.*

*Proof.* The primitive roots modulo $n$ are the generators of $(\mathbf{Z}/n\mathbf{Z})^*$, which by assumption is cyclic of order $\varphi(n)$. Thus they are in bijection with the generators of any cyclic group of order $\varphi(n)$. In particular, the number of primitive roots modulo $n$ is the same as the number of elements of $\mathbf{Z}/\varphi(n)\mathbf{Z}$ with additive order $\varphi(n)$. An element of $\mathbf{Z}/\varphi(n)\mathbf{Z}$ has additive order $\varphi(n)$ if and only if it is coprime to $\varphi(n)$. There are $\varphi(\varphi(n))$ such elements, as claimed. □

*Example* 2.5.13. For example, there are $\varphi(\varphi(17)) = \varphi(16) = 2^4 - 2^3 = 8$ primitive roots mod 17, namely $3, 5, 6, 7, 10, 11, 12, 14$. The $\varphi(\varphi(9)) = \varphi(6) = 2$ primitive roots modulo 9 are 2 and 5. There are no primitive roots modulo 8, even though $\varphi(\varphi(8)) = \varphi(4) = 2 > 0$.

### *2.5.3 Artin's Conjecture*

**Conjecture 2.5.14** (Emil Artin)**.** *Suppose $a \in \mathbf{Z}$ is not $-1$ or a perfect square. Then there are infinitely many primes $p$ such that $a$ is a primitive root modulo $p$.*

There is no single integer $a$ such that Artin's conjecture is known to be true. For any given $a$, Pieter [Mor93] proved that there are infinitely many $p$ such that the order of $a$ is divisible by the largest prime factor of $p-1$. Hooley [Hoo67] proved that something called the Generalized Riemann Hypothesis implies Conjecture 2.5.14.

*Remark* 2.5.15. Artin conjectured more precisely that if $N(x,a)$ is the number of primes $p \leq x$ such that $a$ is a primitive root modulo $p$, then $N(x,a)$ is asymptotic to $C(a)\pi(x)$, where $C(a)$ is a positive constant that depends only on $a$ and $\pi(x)$ is the number of primes up to $x$.

### *2.5.4 Computing Primitive Roots*

Theorem 2.5.8 does not suggest an efficient algorithm for finding primitive roots. To actually find a primitive root mod $p$ in practice, we try $a = 2$, then $a = 3$, etc., until we find an $a$ that has order $p-1$. Computing the order of an element of $(\mathbf{Z}/p\mathbf{Z})^*$ requires factoring $p-1$, which we do not know how to do quickly in general, so finding a primitive root modulo $p$ for large $p$ seems to be a difficult problem.

**Algorithm 2.5.16** (Primitive Root)**.** Given a prime $p$, this algorithm computes the smallest positive integer $a$ that generates $(\mathbf{Z}/p\mathbf{Z})^*$.

1. [$p = 2$?] If $p = 2$ output 1 and terminate. Otherwise set $a = 2$.
2. [Prime Divisors] Compute the prime divisors $p_1, \ldots, p_r$ of $p-1$.
3. [Generator?] If for every $p_i$, we have $a^{(p-1)/p_i} \not\equiv 1 \pmod{p}$, then $a$ is a generator of $(\mathbf{Z}/p\mathbf{Z})^*$, so output $a$ and terminate.
4. [Try next] Set $a = a + 1$ and go to Step 3.

*Proof.* Let $a \in (\mathbf{Z}/p\mathbf{Z})^*$. The order of $a$ is a divisor $d$ of the order $p-1$ of the group $(\mathbf{Z}/p\mathbf{Z})^*$. Write $d = (p-1)/n$, for some divisor $n$ of $p-1$. If $a$ is not a generator of $(\mathbf{Z}/p\mathbf{Z})^*$, then since $n \mid (p-1)$, there is a prime divisor $p_i$ of $p-1$ such that $p_i \mid n$. Then

$$a^{(p-1)/p_i} = (a^{(p-1)/n})^{n/p_i} \equiv 1 \pmod{p}.$$

Conversely, if $a$ is a generator, then $a^{(p-1)/p_i} \not\equiv 1 \pmod{p}$ for any $p_i$. Thus the algorithm terminates with Step 3 if and only if the $a$ under consideration is a primitive root. By Theorem 2.5.8, there is at least one primitive root, so the algorithm terminates. □

## 2.6 Exercises

2.1 Prove that for any positive integer $n$, the set $(\mathbf{Z}/n\mathbf{Z})^*$ under multiplication modulo $n$ is a group.

2.2 Compute the following gcd's using Algorithm 1.1.13:

$$\gcd(15, 35) \quad \gcd(247, 299) \quad \gcd(51, 897) \quad \gcd(136, 304)$$

2.3 Use Algorithm 2.3.7 to find $x$, $y \in \mathbf{Z}$ such that $2261x + 1275y = 17$.

2.4 Prove that if $a$ and $b$ are integers and $p$ is a prime, then $(a+b)^p \equiv a^p + b^p \pmod{p}$. You may assume that the binomial coefficient

$$\frac{p!}{r!(p-r)!}$$

is an integer.

2.5 (a) Prove that if $x, y$ is a solution to $ax+by = d$, with $d = \gcd(a, b)$, then for all $c \in \mathbf{Z}$,

$$x' = x + c \cdot \frac{b}{d}, \qquad y' = y - c \cdot \frac{a}{d} \tag{2.6.1}$$

is also a solution to $ax + by = d$.

(b) Find two distinct solutions to $2261x + 1275y = 17$.

(c) Prove that all solutions are of the form (2.6.1) for some $c$.

2.6 Let $f(x) = x^2 + ax + b \in \mathbf{Z}[x]$ be a quadratic polynomial with integer coefficients, for example, $f(x) = x^2 + x + 6$. Formulate a conjecture about when the set

$$\{f(n) : n \in \mathbf{Z} \text{ and } f(n) \text{ is prime}\}$$

is infinite. Give numerical evidence that supports your conjecture.

2.7 Find four complete sets of residues modulo 7, where the $i$th set satisfies the $i$th condition: (1) nonnegative, (2) odd, (3) even, (4) prime.

2.8 Find rules in the spirit of Proposition 2.1.9 for divisibility of an integer by 5, 9, and 11, and prove each of these rules using arithmetic modulo a suitable $n$.

2.9 (*) *(The following problem is from the 1998 Putnam Competition.)* Define a sequence of decimal integers $a_n$ as follows: $a_1 = 0$, $a_2 = 1$, and $a_{n+2}$ is obtained by writing the digits of $a_{n+1}$ immediately followed by those of $a_n$. For example, $a_3 = 10$, $a_4 = 101$, and $a_5 = 10110$. Determine the $n$ such that $a_n$ is a multiple of 11, as follows:

(a) Find the smallest integer $n > 1$ such that $a_n$ is divisible by 11.

(b) Prove that $a_n$ is divisible by 11 if and only if $n \equiv 1 \pmod{6}$.

2.10 Find an integer $x$ such that $37x \equiv 1 \pmod{101}$.

2.11 What is the order of 2 modulo 17?

2.12 Let $p$ be a prime. Prove that $\mathbf{Z}/p\mathbf{Z}$ is a field.

2.13 Find an $x \in \mathbf{Z}$ such that $x \equiv -4 \pmod{17}$ and $x \equiv 3 \pmod{23}$.

2.14 Prove that if $n > 4$ is composite then

$$(n-1)! \equiv 0 \pmod{n}.$$

2.15 For what values of $n$ is $\varphi(n)$ odd?

2.16 (a) Prove that $\varphi$ is multiplicative as follows. Suppose $m, n$ are positive integers and $\gcd(m,n) = 1$. Show that the natural map $\psi : \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is an injective homomorphism of rings, hence bijective by counting, then look at unit groups.

(b) Prove conversely that if $\gcd(m,n) > 1$, then the natural map $\psi : \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is not an isomorphism.

2.17 Seven competitive math students try to share a huge hoard of stolen math books equally between themselves. Unfortunately, six books are left over, and in the fight over them, one math student is expelled. The remaining six math students, still unable to share the math books equally since two are left over, again fight, and another is expelled. When the remaining five share the books, one book is left over, and it is only after yet another math student is expelled that an equal sharing is possible. What is the minimum number of books that allows this to happen?

2.18 Show that if $p$ is a positive integer such that both $p$ and $p^2 + 2$ are prime, then $p = 3$.

2.19 Let $\varphi : \mathbf{N} \to \mathbf{N}$ be the Euler $\varphi$ function.

(a) Find all natural numbers $n$ such that $\varphi(n) = 1$.

(b) Do there exist natural numbers $m$ and $n$ such that $\varphi(mn) \neq \varphi(m) \cdot \varphi(n)$?

2.20 Find a formula for $\varphi(n)$ directly in terms of the prime factorization of $n$.

2.21 (a) Prove that if $\varphi : G \to H$ is a group homomorphism, then $\ker(\varphi)$ is a subgroup of $G$.

(b) Prove that $\ker(\varphi)$ is *normal*, i.e., if $a \in G$ and $b \in \ker(\varphi)$, then $a^{-1}ba \in \ker(\varphi)$.

2.22 Is the set $\mathbf{Z}/5\mathbf{Z} = \{0, 1, 2, 3, 4\}$ with binary operation multiplication modulo 5 a group?

2.23 Find all *four* solutions to the equation

$$x^2 - 1 \equiv 0 \pmod{35}.$$

2.24 Prove that for any positive integer $n$ the fraction $(12n+1)/(30n+2)$ is in reduced form.

2.25 Suppose $a$ and $b$ are positive integers.

(a) Prove that $\gcd(2^a - 1,\ 2^b - 1) = 2^{\gcd(a,b)} - 1$.

(b) Does it matter if 2 is replaced by an arbitrary prime $p$?

(c) What if 2 is replaced by an arbitrary positive integer $n$?

2.26 For every positive integer $b$, show that there exists a positive integer $n$ such that the polynomial $x^2 - 1 \in (\mathbf{Z}/n\mathbf{Z})[x]$ has at least $b$ roots.

2.27 (a) Prove that there is no primitive root modulo $2^n$ for any $n \geq 3$.

(b) (*) Prove that $(\mathbf{Z}/2^n\mathbf{Z})^*$ is generated by $-1$ and 5.

2.28 Let $p$ be an odd prime.

(a) (*) Prove that there is a primitive root modulo $p^2$. (Hint: Use that if $a, b$ have orders $n, m$, with $\gcd(n, m) = 1$, then $ab$ has order $nm$.)

(b) Prove that for any $n$, there is a primitive root modulo $p^n$.

(c) Explicitly find a primitive root modulo 125.

2.29 (*) In terms of the prime factorization of $n$, characterize the integers $n$ such that there is a primitive root modulo $n$.

2.30 Compute the last two digits of $3^{45}$.

2.31 Find the integer $a$ such that $0 \leq a < 113$ and

$$102^{70} + 1 \equiv a^{37} \pmod{113}.$$

2.32 Find the proportion of primes $p < 1000$ such that 2 is a primitive root modulo $p$.

2.33 Find a prime $p$ such that the smallest primitive root modulo $p$ is 37.

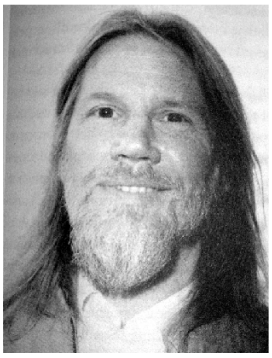
# 3
# Public-key Cryptography

In the 1970s, techniques from number theory changed the world forever by providing, for the first time ever, a way for two people to communicate secret messages under the assumption that *all* of their communication is intercepted and read by an adversary. This idea has stood the test of time. In fact, whenever you buy something online, you use such a system, which typically involves working in the ring of integers modulo $n$. This chapter tells the story of several such systems.

## 3.1 Playing with Fire

I recently watched a TV show called La Femme Nikita about a woman named Nikita who is forced to be an agent for a shady anti-terrorist organization called Section One. Nikita has strong feelings for fellow agent Michael, and she most trusts Walter, Section One's ex-biker gadgets and explosives expert. Often Nikita's worst enemies are her superiors and coworkers at Section One. A synopsis for a Season Three episode is as follows:

> PLAYING WITH FIRE
>
> On a mission to secure detonation chips from a terrorist organization's heavily armed base camp, Nikita is captured as a hostage by the enemy. Or so it is made to look. Michael and Nikita have actually created the scenario in order to secretly rendezvous with each other. The ruse works, but when Birkoff



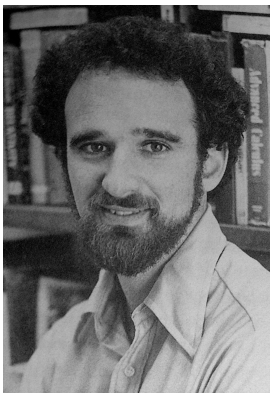


FIGURE 3.1. Diffie and Hellman (photos from [Sin99])

> [Section One's master hacker] accidentally discovers encrypted messages between Michael and Nikita sent with Walter's help, Birkoff is forced to tell Madeline. Suspecting that Michael and Nikita may be planning a coup d'état, Operations and Madeline use a second team of operatives to track Michael and Nikita's next secret rendezvous... killing them if necessary.

What sort of encryption might Walter have helped them to use? I let my imagination run free, and this is what I came up with. After being captured at the base camp, Nikita is given a phone by her captors in hopes that she'll use it and they'll be able to figure out what she is really up to. Everyone is eagerly listening in on her calls.

*Remark* 3.1.1. In this book, we will assume a method is available for producing random integers. Methods for generating random integers are involved and interesting, but we will not discuss them in this book. For an in-depth treatment of random numbers, see [Knu98, Ch. 3].

Nikita remembers a conversation with Walter about a public-key cryptosystem called the "Diffie-Hellman key exchange." She remembers that it allows two people to agree on a secret key in the presence of eavesdroppers. Moreover, Walter mentioned that though Diffie-Hellman was the first ever public-key exchange system, it is still in common use today (for example, in OpenSSH protocol version 2, see `http://www.openssh.com/`).

Nikita pulls out her handheld computer and phone, calls up Michael, and they do the following, which is *wrong* (try to figure out what is wrong as you read it).

1. Together they choose a big prime number $p$ and a number $g$ with $1 < g < p$.

2. Nikita *secretly* chooses an integer $n$.

3. Michael *secretly* chooses an integer $m$.
4. Nikita tells Michael $ng \pmod p$.
5. Michael tells $mg \pmod p$ to Nikita.
6. The "secret key" is $s = nmg \pmod p$, which both Nikita and Michael can easily compute.

Here's a very simple example with small numbers that illustrates what Michael and Nikita do. (They really used much larger numbers.)

1. $p = 97$, $g = 5$
2. $n = 31$
3. $m = 95$
4. $ng \equiv 58 \pmod{97}$
5. $mg \equiv 87 \pmod{97}$
6. $s = nmg = 78 \pmod{97}$

Nikita and Michael are foiled because everyone easily figures out $s$:

1. Everyone knows $p$, $g$, $ng \pmod p$, and $mg \pmod p$.
2. Using Algorithm 2.3.7, anyone can easily find $a, b \in \mathbf{Z}$ such that $ag + bp = 1$, which exists because $\gcd(g, p) = 1$.
3. Then, $ang \equiv n \pmod p$, so everyone knows Nikita's secret key $n$, and hence can easily compute the shared secret $s$.

To taunt her, Nikita's captors give her a paragraph from a review of Diffie and Hellman's 1976 paper "New Directions in Cryptography" [DH76]:

> "The authors discuss some recent results in communications theory [...] The first [method] has the feature that an unauthorized 'eavesdropper' will find it computationally infeasible to decipher the message [...] They propose a couple of techniques for implementing the system, but the reviewer was unconvinced."

## 3.2 The Diffie-Hellman Key Exchange

As night darkens Nikita's cell, she reflects on what has happened. Upon realizing that she mis-remembered how the system works, she phones Michael and they do the following:

1. Together Michael and Nikita choose a 200-digit integer $p$ that is likely to be prime (see Section 2.4), and choose a number $g$ with $1 < g < p$.
2. Nikita *secretly* chooses an integer $n$.
3. Michael *secretly* chooses an integer $m$.
4. Nikita computes $g^n \pmod p$ on her handheld computer and tells Michael the resulting number over the phone.
5. Michael tells Nikita $g^m \pmod p$.
6. The shared secret key is then
$$s \equiv (g^n)^m \equiv (g^m)^n \equiv g^{nm} \pmod p,$$
which both Nikita and Michael can compute.

Here is a simplified example that illustrates what they did, that involves only relatively simple arithmetic.

1. $p = 97$, $g = 5$
2. $n = 31$
3. $m = 95$
4. $g^n \equiv 7 \pmod p$
5. $g^m \equiv 39 \pmod p$
6. $s \equiv (g^n)^m \equiv 14 \pmod p$

### *3.2.1 The Discrete Log Problem*

Nikita communicates with Michael by encrypting everything using their agreed upon secret key (for example, using a standard symmetric cipher such as AES, Arcfour, Cast128, 3DES, or Blowfish). In order to understand the conversation, the eavesdropper needs $s$, but it takes a long time to compute $s$ given only $p$, $g$, $g^n$, and $g^m$. One way would be to compute $n$ from knowledge of $g$ and $g^n$; this is possible, but appears to be "computationally infeasible," in the sense that it would take too long to be practical.

Let $a$, $b$, and $n$ be real numbers with $a, b > 0$ and $n \geq 0$. Recall that the "log to the base $b$" function is characterized by

$$\log_b(a) = n \text{ if and only if } a = b^n.$$

We use the $\log_b$ function in algebra to solve the following problem: Given a base $b$ and a power $a$ of $b$, find an exponent $n$ such that

$$a = b^n.$$

That is, given $a = b^n$ and $b$, find $n$.

*SAGE Example* 3.2.1. The number $a = 19683$ is the $n$th power of $b = 3$ for some $n$. We quickly find that

$$n = \log_3(19683) = \log(19683)/\log(3) = 9.$$

```
sage: log(19683.0)
9.88751059801299
sage: log(3.0)
1.09861228866811
sage: log(19683.0) / log(3.0)
9.00000000000000
```

Sage can quickly compute a numerical approximation for $\log(x)$, for any $x$, by computing a partial sum of an appropriate rapidly-converging infinite series (at least for $x$ in a certain range).

The discrete log problem is the analog of computing $\log_b(a)$ but where both $b$ and $a$ are elements of a finite group.

**Problem 3.2.2** (Discrete Log Problem)**.** Let $G$ be a finite group, for example, $G = (\mathbf{Z}/p\mathbf{Z})^*$. Given $b \in G$ and a power $a$ of $b$, find a positive integer $n$ such that $b^n = a$.

As far as we know, finding discrete logarithms in $(\mathbf{Z}/p\mathbf{Z})^*$ when $p$ is large is "very difficult" in practice. Over the years, many people have been very motivated to try. For example, if Nikita's captors could efficiently solve Problem 3.2.2, then they could read the messages she exchanges with Michael. Unfortunately, we have no formal proof that computing discrete logarithms on a classical computer is difficult. Also, Peter Shor [Sho97] showed that if one could build a sufficiently complicated quantum computer, it could solve the discrete logarithm problem in time bounded by a polynomial function of the number of digits of $\#G$.

It is easy to give an inefficient algorithm that solves the discrete log problem. Simply try $b^1$, $b^2$, $b^3$, etc., until we find an exponent $n$ such that $b^n = a$. For example, suppose $a = 18$, $b = 5$, and $p = 23$. Working modulo 23, we have

$$b^1 = 5,\ b^2 = 2,\ b^3 = 10,\ \ldots,\ b^{12} = 18,$$

so $n = 12$. When $p$ is large, computing the discrete log this way soon becomes impractical, because increasing the number of digits of the modulus makes the computation take vastly longer.

*SAGE Example* 3.2.3. Perhaps part of the reason that computing discrete logarithms is difficult, is that the logarithm in the real numbers is continuous, but the (minimum) logarithm of a number mod $n$ bounces around at random. We illustrate this exotic behavior in Figure 3.2.

This draws the continuous plot.

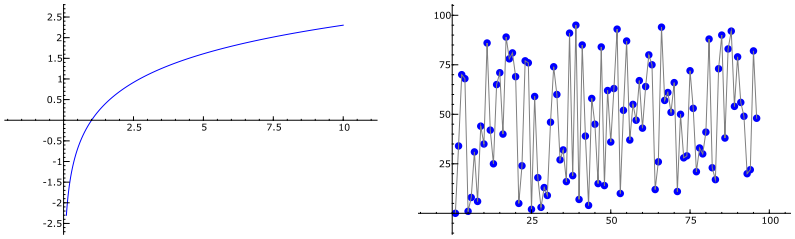



FIGURE 3.2. Graphs of the continuous log and of the discrete log modulo 53. Which picture looks easier to predict?

```
sage: plot(log, 0.1,10, rgbcolor=(0,0,1))
```

This draws the discrete plot.

```
sage: p = 53
sage: R = Integers(p)
sage: a = R.multiplicative_generator()
sage: v = sorted([(a^n, n) for n in range(p-1)])
sage: G = plot(point(v,pointsize=50,rgbcolor=(0,0,1)))
sage: H = plot(line(v,rgbcolor=(0.5,0.5,0.5)))
sage: G + H
```

### *3.2.2 Realistic Diffie-Hellman Example*

In this section, we present an example that uses bigger numbers. First, we prove a proposition that we can use to choose a prime $p$ in such a way that it is easy to find a $g \in (\mathbf{Z}/p\mathbf{Z})^*$ with order $p-1$. We have already seen in Section 2.5 that for every prime $p$ there exists an element $g$ of order $p-1$, and we gave Algorithm 2.5.16 for finding a primitive root for any prime. The significance of Proposition 3.2.4 below is that it suggests an algorithm for finding a primitive root that is easier to use in practice when $p$ is large, because it does not require factoring $p-1$. Of course, one could also just use a random $g$ for Diffie-Hellman; it is not essential that $g$ generates $(\mathbf{Z}/p\mathbf{Z})^*$.

**Proposition 3.2.4.** *Suppose $p$ is a prime such that $(p-1)/2$ is also prime. Then each element of $(\mathbf{Z}/p\mathbf{Z})^*$ has order one of $1$, $2$, $(p-1)/2$, or $p-1$.*

*Proof.* Since $p$ is prime, the group $(\mathbf{Z}/p\mathbf{Z})^*$ is of order $p-1$. By assumption, the prime factorization of $p-1$ is $2 \cdot ((p-1)/2)$. Let $a \in (\mathbf{Z}/p\mathbf{Z})^*$. Then by Theorem 2.1.20, $a^{p-1} = 1$, so the order of $a$ is a divisor of $p-1$, which proves the proposition. □

Given a prime $p$ with $(p-1)/2$ prime, find an element of order $p-1$ as follows. If 2 has order $p-1$, we are done. If not, 2 has order $(p-1)/2$ since 2 does not have order either 1 or 2. Then $-2$ has order $p-1$.

Let $p = 934509830948509384509834096 11$. Then $p$ is prime, but $(p-1)/2$ is not. So we keep adding 2 to $p$ and testing pseudoprimality using algorithms from Section 2.4 until we find that the next pseudoprime after $p$ is

$$q = 93450983094850938450983409623.$$

It turns out that $q$ pseudoprime and $(q-1)/2$ is also pseudoprime. We find that 2 has order $(q-1)/2$, so $g = -2$ has order $q-1$ modulo $q$, and is hence a generator of $(\mathbf{Z}/q\mathbf{Z})^*$, at least assuming that $q$ is really prime.

The secret random numbers generated by Nikita and Michael are

$$n = 183199223755318591716133791 81$$

and

$$m = 82335836243866695680141440300.$$

Nikita sends

$$g^n = 45416776270485369791375944998 \in (\mathbf{Z}/p\mathbf{Z})^*$$

to Michael, and Michael sends

$$g^m = 15048074151770884271824225393 \in (\mathbf{Z}/p\mathbf{Z})^*$$

to Nikita. They agree on the secret key

$$g^{nm} = 85771409470770521212346739540 \in (\mathbf{Z}/p\mathbf{Z})^*.$$

*SAGE Example* 3.2.5. We illustrate the above computations using Sage.

```
sage: q = 93450983094850938450983409623
sage: q.is_prime()
True
sage: is_prime((q-1)//2)
True
sage: g = Mod(-2, q)
sage: g.multiplicative_order()
93450983094850938450983409622
sage: n = 183199223755318591716133791 81
sage: m = 82335836243866695680141440300
sage: g^n
45416776270485369791375944998
sage: g^m
15048074151770884271824225393
sage: (g^n)^m
85771409470770521212346739540
sage: (g^m)^n
85771409470770521212346739540
```

### *3.2.3 The Man in the Middle Attack*

Since their first system was broken, instead of talking on the phone, Michael and Nikita can now only communicate via text messages. One of her captors, The Man, is watching each of the transmissions; moreover, he can intercept messages and send false messages. When Nikita sends a message to Michael announcing $g^n \pmod p$, The Man intercepts this message, and sends his own number $g^t \pmod p$ to Michael. Eventually, Michael and The Man agree on the secret key $g^{tm} \pmod p$, and Nikita and The Man agree on the key $g^{tn} \pmod p$. When Nikita sends a message to Michael she unwittingly uses the secret key $g^{tn} \pmod p$; The Man then intercepts it, decrypts it, changes it, and re-encrypts it using the key $g^{tm} \pmod p$, and sends it on to Michael. This is bad because now The Man can read every message sent between Michael and Nikita, and moreover, he can change them in transmission in subtle ways.

One way to get around this attack is to use a digital signature scheme based on the RSA cryptosystem. We will not discuss digital signatures further in this book, but will discuss RSA in the next section.

## 3.3 The RSA Cryptosystem

The Diffie-Hellman key exchange has drawbacks. As discussed in Section 3.2.3, it is susceptible to the man in the middle attack. This section is about the RSA public-key cryptosystem of Rivest, Shamir, and Adleman [RSA78], which is an alternative to Diffie-Hellman that is more flexible in some ways.

We first describe the RSA cryptosystem, then discuss several ways to attack it. It is important to be aware of such weaknesses, in order to avoid foolish mistakes when implementing RSA. We barely scratched the surface here of the many possible attacks on specific implementations of RSA or other cryptosystems.

### *3.3.1 How RSA works*

The fundamental idea behind RSA is to try to construct a trap-door or one-way function on a set $X$. This is an invertible function

$$E : X \to X$$

such that it is easy for Nikita to compute $E^{-1}$, but extremely difficult for anybody else to do so.

Here is how Nikita makes a one-way function $E$ on the set of integers modulo $n$.

1. Using a method hinted at in Section 2.4, Nikita picks two large primes $p$ and $q$, and lets $n = pq$.

2. It is then easy for Nikita to compute

$$\varphi(n) = \varphi(p) \cdot \varphi(q) = (p-1) \cdot (q-1).$$

3. Nikita next chooses a random integer $e$ with

$$1 < e < \varphi(n) \text{ and } \gcd(e, \varphi(n)) = 1.$$

4. Nikita uses the algorithm from Section 2.3.2 to find a solution $x = d$ to the equation

$$ex \equiv 1 \pmod{\varphi(n)}.$$

5. Finally, Nikita defines a function $E : \mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$ by

$$E(x) = x^e \in \mathbf{Z}/n\mathbf{Z}.$$

Note that anybody can compute $E$ fairly quickly using the repeated-squaring algorithm from Section 2.3.2. Nikita's *public key* is the pair of integers $(n, e)$, which is just enough information for people to easily compute $E$. Nikita knows a number $d$ such that $ed \equiv 1 \pmod{\varphi(n)}$, so, as we will see, she can quickly compute $E^{-1}$.

To send Nikita a message, proceed as follows. Encode your message, in some way, as a sequence of numbers modulo $n$ (see Section 3.3.2)

$$m_1, \ldots, m_r \in \mathbf{Z}/n\mathbf{Z},$$

then send

$$E(m_1), \ldots, E(m_r)$$

to Nikita. (Recall that $E(m) = m^e$ for $m \in \mathbf{Z}/n\mathbf{Z}$.)

When Nikita receives $E(m_i)$, she finds each $m_i$ by using that $E^{-1}(m) = m^d$, a fact that follows from Proposition 3.3.1

**Proposition 3.3.1** (Decryption Key)**.** *Let $n$ be an integer that is a product of distinct primes and let $d, e \in \mathbf{N}$ be such that $p-1 \mid de-1$ for each prime $p \mid n$. Then $a^{de} \equiv a \pmod{n}$ for all $a \in \mathbf{Z}$.*

*Proof.* Since $n \mid a^{de} - a$, if and only if $p \mid a^{de} - a$ for each prime divisor $p$ of $n$, it suffices to prove that $a^{de} \equiv a \pmod{p}$ for each prime divisor $p$ of $n$. If $\gcd(a, p) \neq 1$, then $a \equiv 0 \pmod{p}$, so $a^{de} \equiv a \pmod{p}$. If $\gcd(a, p) = 1$, then Theorem 2.1.20 asserts that $a^{p-1} \equiv 1 \pmod{p}$. Since $p - 1 \mid de - 1$, we have $a^{de-1} \equiv 1 \pmod{p}$ as well. Multiplying both sides by $a$ shows that $a^{de} \equiv a \pmod{p}$. □

Thus to decrypt $E(m_i)$ Nikita computes

$$E(m_i)^d = (m_i^e)^d = m_i.$$

*SAGE Example* 3.3.2. We implement the RSA cryptosystem using Sage. The `rsa` function creates a key with (at most) the given number of bits, i.e., if `bits` equals 20, it creates a key $n = pq$ such that $n$ is approximately $2^{20}$. Typical real-life cryptosystems would choose keys that are 512, 1024, or 2048 bits long. Try generating large keys yourself using Sage; how long does it take?

```
sage: def rsa(bits):
...    # only prove correctness up to 1024 bits
...    proof = (bits <= 1024)
...    p = next_prime(ZZ.random_element(2**(bits//2 +1)),
...                proof=proof)
...    q = next_prime(ZZ.random_element(2**(bits//2 +1)),
...                proof=proof)
...    n = p * q
...    phi_n = (p-1) * (q-1)
...    while True:
...        e = ZZ.random_element(1,phi_n)
...        if gcd(e,phi_n) == 1: break
...    d = lift(Mod(e,phi_n)^(-1))
...    return e, d, n
...
sage: def encrypt(m,e,n):
...    return lift(Mod(m,n)^e)
...
sage: def decrypt(c,d,n):
...    return lift(Mod(c,n)^d)
...
sage: e,d,n = rsa(20)
sage: c = encrypt(123, e, n)
sage: decrypt(c, d, n)
123
```

### *3.3.2 Encoding a Phrase in a Number*

In order to use the RSA cryptosystem to encrypt messages, it is necessary to encode them as a sequence of numbers of size less than $n = pq$. We now describe a simple way to do this. Note that in any actual deployed implementation, it is crucial that you add extra random characters ("salt") at the beginning of each block of the message, so that the same plain text encodes differently each time. This helps thwart chosen plain text attacks.

Suppose $s$ is a sequence of capital letters and spaces, and that $s$ does not begin with a space. We encode $s$ as a number in base 27 as follows: a single space corresponds to 0, the letter $A$ to 1, $B$ to 2, ..., $Z$ to 26. Thus "RUN

NIKITA" is a number written in base 27.

$$\begin{aligned}\text{RUN NIKITA} \quad \leftrightarrow \quad & 27^9 \cdot 18 + 27^8 \cdot 21 + 27^7 \cdot 14 + 27^6 \cdot 0 + 27^5 \cdot 14 \\ & + 27^4 \cdot 9 + 27^3 \cdot 11 + 27^2 \cdot 9 + 27 \cdot 20 + 1 \\ & = 143338425831991 \text{ (in decimal).}\end{aligned}$$

To recover the letters from the decimal number, repeatedly divide by 27 and read off the letter corresponding to each remainder.

$$\begin{array}{rcrcrl} 143338425831991 & = & 5308830586370 \cdot 27 & + & 1 & \text{“A”} \\ 5308830586370 & = & 196623355050 \cdot 27 & + & 20 & \text{“T”} \\ 196623355050 & = & 7282346483 \cdot 27 & + & 9 & \text{“I”} \\ 7282346483 & = & 269716536 \cdot 27 & + & 11 & \text{“K”} \\ 269716536 & = & 9989501 \cdot 27 & + & 9 & \text{“I”} \\ 9989501 & = & 369981 \cdot 27 & + & 14 & \text{“N”} \\ 369981 & = & 13703 \cdot 27 & + & 0 & \text{“ ”} \\ 13703 & = & 507 \cdot 27 & + & 14 & \text{“N”} \\ 507 & = & 18 \cdot 27 & + & 21 & \text{“U”} \\ 18 & = & 0 \cdot 27 & + & 18 & \text{“R”} \end{array}$$

If $27^k \leq n$, then any sequence of $k$ letters can be encoded as above using a positive integer $\leq n$. Thus if we can encrypt integers of size at most $n$, then we must break our message up into blocks of size at most $\log_{27}(n)$.

*SAGE Example* 3.3.3. We use Sage to implement conversion between a string and a number, though in a bit more generally than in the toy illustration above (which used only base 27). The input string s on a computer is stored in a format called ASCII, so each "letter" corresponds to an integer between 0 and 255, inclusive. This number is obtained from the letter using the `ord` command.

```
sage: def encode(s):
...     s = str(s)      # make input a string
...     return sum(ord(s[i])*256^i for i in range(len(s)))
sage: def decode(n):
...     n = Integer(n)  # make input an integer
...     v = []
...     while n != 0:
...         v.append(chr(n % 256))
...         n //= 256    # this replaces n by floor(n/256).
...     return ''.join(v)
sage: m = encode('Run Nikita!'); m
403547690147146494219687 22
sage: decode(m)
'Run Nikita!'
```

### *3.3.3 Some Complete Examples*

To make the arithmetic easier to follow, we use small prime numbers $p$ and $q$ and encrypt the single letter "X" using the RSA cryptosystem. First, we compute the parameters of an RSA cryptosystem.

1. Choose $p$ and $q$: Let $p = 17$, $q = 19$, so $n = pq = 323$.
2. Compute $\varphi(n)$:
$$\begin{aligned}\varphi(n) &= \varphi(p \cdot q) = \varphi(p) \cdot \varphi(q) = (p-1)(q-1) \\ &= pq - p - q + 1 = 323 - 17 - 19 + 1 = 288.\end{aligned}$$
3. Randomly choose an $e < 288$: We choose $e = 95$.
4. Solve
$$95x \equiv 1 \pmod{288}.$$
Using the GCD algorithm, we find that $d = 191$ solves the equation.

We have thus computed the parameters of an RSA public key cryptosystem. The public key is $(323, 95)$, so the encryption function is

$$E(x) = x^{95},$$

and the decryption function is $D(x) = x^{191}$.

Next, we encrypt the letter "X". It is encoded as the number 24, since X is the 24th letter of the alphabet. We have

$$E(24) = 24^{95} = 294 \in \mathbf{Z}/323\mathbf{Z}.$$

To decrypt, we compute $E^{-1}$:

$$E^{-1}(294) = 294^{191} = 24 \in \mathbf{Z}/323\mathbf{Z}.$$

This next example illustrates RSA but with bigger numbers. Let

$$p = 738873402423833494183027176953,\ q = 37877768068656628823782 73.$$

Then,

$$n = p \cdot q = 279868753691091597012726360634791146094855419785354216 9$$

and

$$\begin{aligned}\varphi(n) &= (p-1)(q-1) \\ &= 279868753691091597012726286747072126030819435194398694 4.\end{aligned}$$

Using a pseudo-random number generator on a computer, the author randomly chose the integer

$$e = 14839591948662041793485360102847166554421390249157206 99.$$

Then,

$$d = 21133679284963054695413483870886329734578023587816108 03$$

Since $\log_{27}(n) \approx 38.04$, we can encode then encrypt single blocks of up to 38 letters. Let's encrypt the string `RUN NIKITA`, which encodes as $m = 143338425831991$. We have

$$\begin{aligned} E(m) &= m^e \\ &= 150455443299656813339308887860094810177372680087887399 0. \end{aligned}$$

*Remark* 3.3.4. In practice, one usually choses $e$ to be small, since that does not seem to reduce the security of RSA, and makes the key size smaller. For example, in the OpenSSL documentation (see `http://www.openssl.org/`) about their implementation of RSA, it states that "The exponent is an odd number, typically 3, 17 or 65537."

## 3.4 Attacking RSA

Suppose Nikita's public key is $(n, e)$ and her decryption key is $d$, so $ed \equiv 1 \pmod{\varphi(n)}$. If somehow we compute the factorization $n = pq$, then we can compute $\varphi(n) = (p-1)(q-1)$ and hence compute $d$. Thus, if we can factor $n$ then we can break the corresponding RSA public-key cryptosystem.

### *3.4.1 Factoring $n$ Given $\varphi(n)$*

Suppose $n = pq$. Given $\varphi(n)$, it is very easy to compute $p$ and $q$. We have

$$\varphi(n) = (p-1)(q-1) = pq - (p+q) + 1,$$

so we know both $pq = n$ and $p + q = n + 1 - \varphi(n)$. Thus, we know the polynomial

$$x^2 - (p+q)x + pq = (x-p)(x-q)$$

whose roots are $p$ and $q$. These roots can be found using the quadratic formula.

*Example* 3.4.1. The number $n = pq = 31615577110997599711$ is a product of two primes, and $\varphi(n) = 31615577098574867424$. We have

$$\begin{aligned} f &= x^2 - (n + 1 - \varphi(n))x + n \\ &= x^2 - 12422732288x + 31615577110997599711 \\ &= (x - 3572144239)(x - 8850588049), \end{aligned}$$

where the factorization step is easily accomplished using the quadratic formula:

$$\begin{aligned}&\frac{-b+\sqrt{b^2-4ac}}{2a}\\ &\quad=\frac{12422732288+\sqrt{12422732288^2-4\cdot 31615577110997599711}}{2}\\ &\quad=8850588049.\end{aligned}$$

We conclude that $n = 3572144239 \cdot 8850588049$.

*SAGE Example* 3.4.2. The following Sage function factors $n = pq$ given $n$ and $\varphi(n)$.

```
sage: def crack_rsa(n, phi_n):
...     R.<x> = PolynomialRing(QQ)
...     f = x^2 - (n+1 -phi_n)*x + n
...     return [b for b, _ in f.roots()]
sage: crack_rsa(31615577110997599711, 31615577098574867424)
[8850588049, 3572144239]
```

### *3.4.2 When p and q are Close*

Suppose that $p$ and $q$ are "close" to each other. Then it is easy to factor $n$ using a factorization method of Fermat called the *Fermat Factorization Method*.

Suppose $n = pq$ with $p > q$. Then,

$$n = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2.$$

Since $p$ and $q$ are "close,"

$$s = \frac{p-q}{2}$$

is small,

$$t = \frac{p+q}{2}$$

is only slightly larger than $\sqrt{n}$, and $t^2 - n = s^2$ is a perfect square. So, we just try

$$t = \lceil\sqrt{n}\rceil, \quad t = \lceil\sqrt{n}\rceil + 1, \quad t = \lceil\sqrt{n}\rceil + 2, \ldots$$

until $t^2 - n$ is a perfect square $s^2$. (Here $\lceil x\rceil$ denotes the least integer $n \geq x$.) Then

$$p = t + s, \qquad q = t - s.$$

*Example* 3.4.3. Suppose $n = 23360947609$. Then

$$\sqrt{n} = 152842.88\ldots.$$

If $t = 152843$, then $\sqrt{t^2 - n} = 187.18\ldots$.
If $t = 152844$, then $\sqrt{t^2 - n} = 583.71\ldots$.
If $t = 152845$, then $\sqrt{t^2 - n} = 804 \in \mathbf{Z}$.
Thus $s = 804$. We find that $p = t + s = 153649$ and $q = t - s = 152041$.

*SAGE Example* 3.4.4. We implement the above algorithm for factoring an RSA modulus $n = pq$, when one of $p$ and $q$ is close to $\sqrt{n}$.

```
sage: def crack_when_pq_close(n):
...    t = Integer(ceil(sqrt(n)))
...    while True:
...       k = t^2 - n
...       if k > 0:
...          s = Integer(int(round(sqrt(t^2 - n))))
...          if s^2 + n == t^2:
...             return t+s, t-s
...
...       t += 1
...
sage: crack_when_pq_close(23360947609)
(153649, 152041)
```

For example, you might think that choosing a random prime, and the next prime after would be a good idea, but instead it creates an easy-to-crack cryptosystem.

```
sage: p = next_prime(2^128); p
340282366920938463463374607431768211507
sage: q = next_prime(p)
sage: crack_when_pq_close(p*q)
(340282366920938463463374607431768211537,
      340282366920938463463374607431768211507)
```

### *3.4.3 Factoring n Given d*

In this section, we show that finding the decryption key $d$ for an RSA cryptosystem is, in practice, at least as difficult as factoring $n$. We give a probabilistic algorithm that given a decryption key determines the factorization of $n$.

Consider an RSA cryptosystem with modulus $n$ and encryption key $e$. Suppose we somehow finding an integer $d$ such that

$$a^{ed} \equiv a \pmod{n}$$

for all $a$. Then $m = ed - 1$ satisfies $a^m \equiv 1 \pmod{n}$ for all $a$ that are coprime to $n$. As we saw in Section 3.4.1, knowing $\varphi(n)$ leads directly to a factorization of $n$. Unfortunately, knowing $d$ does not seem to lead easily to

a factorization of $n$. However, there is a probabilistic procedure that, given an $m$ such that $a^m \equiv 1 \pmod{n}$, will find a factorization of $n$ with "high probability" (we will not analyze the probability here).

**Algorithm 3.4.5** (Probabilistic Algorithm to Factor $n$)**.** Let $n = pq$ be the product of two distinct odd primes, and suppose $m$ is an integer such that $a^m \equiv 1 \pmod{n}$ for all $a$ coprime to $n$. This probabilistic algorithm factors $n$ with "high probability." In the steps below, $a$ always denotes an integer coprime to $n = pq$.

1. [Divide out powers of 2] If $m$ is even and $a^{m/2} \equiv 1 \pmod{n}$ for several randomly chosen $a$, set $m = m/2$, and go to Step 1, otherwise let $a$ be such that $a^{m/2} \not\equiv 1 \pmod{n}$.
2. [Compute GCD] Choose a random $a$ and compute $g = \gcd(a^{m/2} - 1, n)$.
3. [Terminate?] If $g$ is a proper divisor of $n$, output $g$ and terminate. Otherwise go to Step 2.

Before giving the proof, we introduce some more terminology from algebra.

**Definition 3.4.6** (Group Homomorphism)**.** Let $G$ and $H$ be groups. A map $\varphi : G \to H$ is a *group homomorphism* if for all $a, b \in G$ we have $\varphi(ab) = \varphi(a)\varphi(b)$. A group homomorphism is called *surjective* if for every $c \in H$ there is $a \in G$ such that $\varphi(a) = c$. The *kernel* of a group homomorphism $\varphi : G \to H$ is the set $\ker(\varphi)$ of elements $a \in G$ such that $\varphi(a) = 1$. A group homomorphism is *injective* if $\ker(\varphi) = \{1\}$.

**Definition 3.4.7** (Subgroup)**.** If $G$ is a group and $H$ is a subset of $G$, then $H$ is a *subgroup* if $H$ is a group under the group operation on $G$.

For example, if $\varphi : G \to H$ is a group homomorphism, then $\ker(\varphi)$ is a subgroup of $G$ (see Exercise 2.21).

We now return to discussing Algorithm 3.4.5. In Step 1, note that $m$ is even since $(-1)^m \equiv 1 \pmod{n}$, so it makes sense to consider $m/2$. It is not practical to determine whether or not $a^{m/2} \equiv 1 \pmod{n}$ for all $a$, because it would require doing a computation for too many $a$. Instead, we try a few random $a$; if $a^{m/2} \equiv 1 \pmod{n}$ for the $a$ we check, we divide $m$ by 2. Also note that if there exists even a single $a$ such that $a^{m/2} \not\equiv 1 \pmod{n}$, then half the $a$ have this property, since then $a \mapsto a^{m/2}$ is a surjective homomorphism $(\mathbf{Z}/n\mathbf{Z})^* \to \{\pm 1\}$ and the kernel has index 2.

Proposition 2.5.3 implies that if $x^2 \equiv 1 \pmod{p}$ then $x = \pm 1 \pmod{p}$. In Step 2, since $(a^{m/2})^2 \equiv 1 \pmod{n}$, we also have $(a^{m/2})^2 \equiv 1 \pmod{p}$ and $(a^{m/2})^2 \equiv 1 \pmod{q}$, so $a^{m/2} \equiv \pm 1 \pmod{p}$ and $a^{m/2} \equiv \pm 1 \pmod{q}$. Since $a^{m/2} \not\equiv 1 \pmod{n}$, there are three possibilities for these signs, so with positive probability one of the following two possibilities occurs:

1. $a^{m/2} \equiv +1 \pmod{p}$ and $a^{m/2} \equiv -1 \pmod{q}$

2. $a^{m/2} \equiv -1 \pmod{p}$ and $a^{m/2} \equiv +1 \pmod{q}$.

The only other possibility is that both signs are $-1$. In the first case,

$$p \mid a^{m/2} - 1 \qquad \text{but} \qquad q \nmid a^{m/2} - 1,$$

so $\gcd(a^{m/2} - 1, pq) = p$, and we have factored $n$. Similarly, in the second case, $\gcd(a^{m/2} - 1, pq) = q$, and we again factor $n$.

*Example* 3.4.8. Somehow we discover that the RSA cryptosystem with

$$n = 32295194023343 \qquad \text{and} \qquad e = 29468811804857$$

has decryption key $d = 11127763319273$. We use this information and Algorithm 3.4.5 to factor $n$. If

$$m = ed - 1 = 327921963064646896263108960,$$

then $\varphi(pq) \mid m$, so $a^m \equiv 1 \pmod{n}$ for all $a$ coprime to $n$. For each $a \leq 20$ we find that $a^{m/2} \equiv 1 \pmod{n}$, so we replace $m$ with

$$\frac{m}{2} = 163960981532323448131554480.$$

Again, we find with this new $m$ that for each $a \leq 20$, $a^{m/2} \equiv 1 \pmod{n}$, so we replace $m$ by 819804907661617240657772 40. Yet again, for each $a \leq 20$, $a^{m/2} \equiv 1 \pmod{n}$, so we replace $m$ by 409902453830808620328886 20. This is enough, since $2^{m/2} \equiv 4015382800099 \pmod{n}$. Then,

$$\gcd(2^{m/2} - 1, n) = \gcd(4015382800098, 32295194023343) = 737531,$$

and we have found a factor of $n$. Dividing, we find that

$$n = 737531 \cdot 43788253.$$

*SAGE Example* 3.4.9. We implement Algorithm 3.4.5 in Sage.

```
sage: def crack_given_decrypt(n, m):
...     n = Integer(n); m = Integer(m);  # some type checking
...     # Step 1: divide out powers of 2
...     while True:
...         if is_odd(m): break
...         divide_out = True
...         for i in range(5):
...            a = randrange(1,n)
...            if gcd(a,n) == 1:
...               if Mod(a,n)^(m//2) != 1:
...                   divide_out = False
...                   break
```

```
...         if divide_out:
...             m = m//2
...         else:
...             break
...     # Step 2: Compute GCD
...     while True:
...         a = randrange(1,n)
...         g = gcd(lift(Mod(a, n)^(m//2)) - 1, n)
...         if g != 1 and g != n:
...             return g
...
```

We show how to verify Example 3.4.8 using Sage.

```
sage: n=32295194023343; e=29468811804857; d=11127763319273
sage: crack_given_decrypt(n, e*d - 1)
737531
sage: factor(n)
737531 * 43788253
```

We try a much larger example.

```
sage: e = 22601762315966221465875845336488389513
sage: d = 31940292321834506197902778067109010093
sage: n = 26849492403959099246944675130990465673
sage: p = crack_given_decrypt(n, e*d - 1)
sage: p   # random output (could be other prime divisor)
13432418150982799907
sage: n % p
0
```

### *3.4.4 Further Remarks*

If one were to implement an actual RSA cryptosystem, there are many additional tricks and ideas to keep in mind. For example, one can add some extra random letters to each block of text, so that a given string will encrypt differently each time it is encrypted. This makes it more difficult for an attacker who knows the encrypted and plaintext versions of one message to gain information about subsequent encrypted messages. In any particular implementation, there might be attacks that would be devastating in practice, but which would not require factorization of the RSA modulus.

RSA is in common use, for example, it is used in OpenSSH protocol version 1 (see `http://www.openssh.com/`).

We will consider the ElGamal cryptosystem in Sections 6.4.2. It has a similar flavor to RSA, but is more flexible in some ways.

Probably the best general purpose attack on RSA is the number field sieve, which is a general algorithm for factoring integers of the form $pq$. A description of the sieve is beyond the scope of this book. The elliptic curve method is another related general algorithm that we will discuss in detail in Section 6.3.

*SAGE Example* 3.4.10. Here is a simple example of using a variant of the number field sieve (called the quadratic sieve) in Sage to factor an RSA key with about 192 bits:

```
sage: set_random_seed(0)
sage: p = next_prime(randrange(2^96))
sage: q = next_prime(randrange(2^97))
sage: n = p * q
sage: qsieve(n)
([6340271405786663791648052309,
  461023131085921802863987571 59], '')
```

## 3.5 Exercises

3.1 This problem concerns encoding phrases using numbers using the encoding of Section 3.3.2. What is the longest that an arbitrary sequence of letters (no spaces) can be if it must fit in a number that is less than $10^{20}$?

3.2 Suppose Michael creates an RSA cryptosystem with a very large modulus $n$ for which the factorization of $n$ cannot be found in a reasonable amount of time. Suppose that Nikita sends messages to Michael by representing each alphabetic character as an integer between 0 and 26 (`A` corresponds to 1, `B` to 2, etc., and a space ␣ to 0), then encrypts each number *separately* using Michael's RSA cryptosystem. Is this method secure? Explain your answer.

3.3 For any $n \in \mathbf{N}$, let $\sigma(n)$ be the sum of the divisors of $n$; for example, $\sigma(6) = 1 + 2 + 3 + 6 = 12$ and $\sigma(10) = 1 + 2 + 5 + 10 = 18$. Suppose that $n = pqr$ with $p$, $q$, and $r$ distinct primes. Devise an "efficient" algorithm that given $n$, $\varphi(n)$ and $\sigma(n)$, computes the factorization of $n$. For example, if $n = 105$, then $p = 3$, $q = 5$, and $r = 7$, so the input to the algorithm would be

$$n = 105, \qquad \varphi(n) = 48, \qquad \text{and} \quad \sigma(n) = 192,$$

and the output would be 3, 5, and 7.

3.4 You and Nikita wish to agree on a secret key using the Diffie-Hellman key exchange. Nikita announces that $p = 3793$ and $g = 7$. Nikita

secretly chooses a number $n < p$ and tells you that $g^n \equiv 454 \pmod{p}$. You choose the random number $m = 1208$. What is the secret key?

3.5 You see Michael and Nikita agree on a secret key using the Diffie-Hellman key exchange. Michael and Nikita choose $p = 97$ and $g = 5$. Nikita chooses a random number $n$ and tells Michael that $g^n \equiv 3 \pmod{97}$, and Michael chooses a random number $m$ and tells Nikita that $g^m \equiv 7 \pmod{97}$. Brute force crack their code: What is the secret key that Nikita and Michael agree upon? What is $n$? What is $m$?

3.6 In this problem, you will "crack" an RSA cryptosystem. What is the secret decoding number $d$ for the RSA cryptosystem with public key $(n, e) = (5352381469067, 4240501142039)$?

3.7 Nikita creates an RSA cryptosystem with public key

$$(n, e) = (1433811615146881, 329222149569169).$$

In the following two problems, show the steps you take to factor $n$. (Don't simply factor $n$ directly using a computer.)

(a) Somehow you discover that $d = 116439879930113$. Show how to use the probabilistic algorithm of Section 3.4.3 to factor $n$.

(b) In part (a) you found that the factors $p$ and $q$ of $n$ are very close. Show how to use the Fermat Factorization Method of Section 3.4.2 to factor $n$.

# 4
# Quadratic Reciprocity

A linear equation

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

has a solution if and only if $\gcd(a, n)$ divides $b$ (see Proposition 2.1.15). This chapter is about some amazing mathematics motivated by the search for a criterion for whether or not a given quadratic equation

$$ax^2 + bx + c \equiv 0 \pmod{n}$$

has a solution. In many cases, the Chinese Remainder Theorem and the quadratic formula reduce this to the key question of whether a given integer $a$ is a perfect square modulo a prime $p$.

The Quadratic Reciprocity Law of Gauss provides a precise answer to the following question: For which primes $p$ is the image of $a$ in $(\mathbf{Z}/p\mathbf{Z})^*$ a perfect square? A deep fact, which we will completely prove in this chapter, is that the answer depends only on the reduction of $p$ modulo $4a$. Thus *to decide if $a$ is a square modulo $p$, one only needs to consider the residue of $p$ modulo $4a$*, which is extremely surprising. It turns out that this "reciprocity law" goes to the heart of modern number theory and touches on advanced topics such as class field theory and the Langlands program.

There are over a hundred proofs of the Quadratic Reciprocity Law (see [Lem] for a long list). In this chapter, we give two proofs. The first, which we give in Section 4.3, is completely elementary and involves keeping track of integer points in intervals. It is satisfying because one can understand every detail without much abstraction, but it might be unsatisfying if you find it difficult to conceptualize what is going on. In contrast, our second

proof, which we give in Section 4.4, is more abstract and uses a conceptual development of properties of Gauss sums. You should read Sections 4.1 and 4.2, then at least one of Section 4.3 or Section 4.4, depending on your taste and how much abstract algebra you know.

In Section 4.5, we return to the computational question of actually finding square roots and solving quadratic equations in practice.

## 4.1 Statement of the Quadratic Reciprocity Law

In this section, we state the Quadratic Reciprocity Law.

**Definition 4.1.1** (Quadratic Residue)**.** Fix a prime $p$. An integer $a$ not divisible by $p$ is a *quadratic residue* modulo $p$ if $a$ is a square modulo $p$; otherwise, $a$ is a *quadratic nonresidue*.

For example, the squares modulo 5 are

$$1^2 = 1, \quad 2^2 = 4, \quad 3^2 = 4, \quad 4^2 = 1, \qquad \pmod 5$$

so 1 and 4 are both quadratic residues and 2 and 3 are quadratic nonresidues.

The quadratic reciprocity theorem is the deepest theorem that we will prove in this book. It connects the question of whether or not $a$ is a quadratic residue modulo $p$ to the question of whether $p$ is a quadratic residue modulo each of the prime divisors of $a$. To express it precisely, we introduce some new notation.

**Definition 4.1.2** (Legendre Symbol)**.** Let $p$ be an odd prime and let $a$ be an integer. Set

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } \gcd(a,p) \neq 1, \\ +1 & \text{if } a \text{ is a quadratic residue, and} \\ -1 & \text{if } a \text{ is a quadratic nonresidue.} \end{cases}$$

We call this symbol the *Legendre Symbol*.

For example, we have

$$\left(\frac{1}{5}\right) = 1, \quad \left(\frac{2}{5}\right) = -1, \quad \left(\frac{3}{5}\right) = -1, \quad \left(\frac{4}{5}\right) = 1, \quad \left(\frac{5}{5}\right) = 1.$$

This notation is well entrenched in the literature even though it is also the notation for "$a$ divided by $p$;" be careful not to confuse the two.

*SAGE Example* 4.1.3. Use the `legendre_symbol` command to compute the Legendre symbol in Sage.

```
sage: legendre_symbol(2,3)
-1
sage: legendre_symbol(1,3)
1
sage: legendre_symbol(3,5)
-1
sage: legendre_symbol(Mod(3,5), 5)
-1
```

Since $\left(\frac{a}{p}\right)$ only depends on $a \pmod p$, it makes sense to define $\left(\frac{a}{p}\right)$ for $a \in \mathbf{Z}/p\mathbf{Z}$ to be $\left(\frac{\tilde{a}}{p}\right)$ for any lift $\tilde{a}$ of $a$ to $\mathbf{Z}$.

Recall (see Definition 3.4.6) that a group homomorphism $\varphi : G \to H$ is a map such that for every $a, b \in G$ we have $\varphi(ab) = \varphi(a)\varphi(b)$. Moreover, we say that $\varphi$ is surjective if for every $c \in H$ there is an $a \in G$ with $\varphi(a) = c$. The next lemma explains how the quadratic residue symbol defines a surjective group homomorphism.

**Lemma 4.1.4.** *The map $\psi : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$ given by $\psi(a) = \left(\frac{a}{p}\right)$ is a surjective group homomorphism.*

*Proof.* By Theorem 2.5.8, primitive roots exist, so there is $g \in (\mathbf{Z}/p\mathbf{Z})^*$ such that the elements of $(\mathbf{Z}/p\mathbf{Z})^*$ are

$$g, g^2, \ldots, g^{(p-1)/2}, g^{(p+1)/2}, \ldots, g^{p-1} = 1.$$

Since $p-1$ is even, the squares of elements of $(\mathbf{Z}/p\mathbf{Z})^*$ are

$$g^2, g^4, \ldots, g^{(p-1)/2 \cdot 2} = 1, g^{p+1} = g^2, \ldots, g^{2(p-1)}.$$

Note that the powers of $g$ starting with $g^{p+1} = g^2$ all appeared earlier on the list. Thus, the perfect squares in $(\mathbf{Z}/p\mathbf{Z})^*$ are exactly the powers $g^n$ with $n = 2, 4, \ldots, p-1$, even, and the nonsquares the powers $g^n$ with $n = 1, 3, \ldots, p-2$, odd. It follows that $\psi$ is a homomorphism since an odd plus an odd is even, the sum of two evens is even, and odd plus an even is odd. Moreover, since $g$ is not a square, $\psi(g) = -1$, so $\psi$ is surjective. □

*Remark* 4.1.5. We rephrase the above proof in the language of group theory. The group $G = (\mathbf{Z}/p\mathbf{Z})^*$ of order $p-1$ is a cyclic group. Since $p$ is odd, $p-1$ is even, so the subgroup $H$ of squares of elements of $G$ has index 2 in $G$. (See Exercise 4.2 for why $H$ is a subgroup.) Since $\left(\frac{a}{p}\right) = 1$ if and only if $a \in H$, we see that $\psi$ is the composition $G \to G/H \cong \{\pm 1\}$, where we identify the nontrivial element of $G/H$ with $-1$.

*Remark* 4.1.6. We can alternatively prove that $\psi$ is surjective without using that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic, as follows. If $a \in (\mathbf{Z}/p\mathbf{Z})^*$ is a square, say $a \equiv b^2$

TABLE 4.1. When is 5 a square modulo $p$?

| $p$ | $\left(\frac{5}{p}\right)$ | $p \bmod 5$ | $p$ | $\left(\frac{5}{p}\right)$ | $p \bmod 5$ |
|---|---|---|---|---|---|
| 7 | $-1$ | 2 | 29 | 1 | 4 |
| 11 | 1 | 1 | 31 | 1 | 1 |
| 13 | $-1$ | 3 | 37 | $-1$ | 2 |
| 17 | $-1$ | 2 | 41 | 1 | 1 |
| 19 | 1 | 4 | 43 | $-1$ | 3 |
| 23 | $-1$ | 3 | 47 | $-1$ | 2 |

$\pmod p$, then $a^{(p-1)/2} = b^{p-1} \equiv 1 \pmod p$, so $a$ is a root of $f = x^{(p-1)/2} - 1$. By Proposition 2.5.3, the polynomial $f$ has at most $(p-1)/2$ roots. Thus, there must be an $a \in (\mathbf{Z}/p\mathbf{Z})^*$ that is not a root of $f$, and for that $a$, we have $\psi(a) = \left(\frac{a}{p}\right) = -1$, and trivially $\psi(1) = 1$, so the map $\psi$ is surjective. Note that this argument does not prove that $\psi$ is a homomorphism.

The symbol $\left(\frac{a}{p}\right)$ only depends on the residue class of $a$ modulo $p$, so making a table of values $\left(\frac{a}{5}\right)$ for many values of $a$ would be easy. Would it be easy to make a table of $\left(\frac{5}{p}\right)$ for many $p$? Perhaps, since there *appears* to be a simple pattern in Table 4.1. It seems that $\left(\frac{5}{p}\right)$ depends only on the congruence class of $p$ modulo 5. More precisely, $\left(\frac{5}{p}\right) = 1$ if and only if $p \equiv 1, 4 \pmod 5$, i.e., $\left(\frac{5}{p}\right) = 1$ if and only if $p$ is a square modulo 5.

Based on similar observations, in the 18th century various mathematicians found a conjectural explanation for the mystery suggested by Table 4.1. Finally, on April 8, 1796, at the age of 19, Gauss proved the following theorem.

**Theorem 4.1.7** (Gauss's Quadratic Reciprocity Law)**.** *Suppose $p$ and $q$ are distinct odd primes. Then*

$$\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}} \left(\frac{q}{p}\right).$$

*Also*

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} \qquad \text{and} \qquad \left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

We will give two proofs of Gauss's formula relating $\left(\frac{p}{q}\right)$ to $\left(\frac{q}{p}\right)$. The first elementary proof is in Section 4.3, and the second more algebraic proof is in Section 4.4.

In our example, Gauss's theorem implies that

$$\left(\frac{5}{p}\right) = (-1)^{2\cdot\frac{p-1}{2}}\left(\frac{p}{5}\right) = \left(\frac{p}{5}\right) = \begin{cases} +1 & \text{if } p \equiv 1,4 \pmod 5 \\ -1 & \text{if } p \equiv 2,3 \pmod 5. \end{cases}$$

As an application, the following example illustrates how to answer questions like "is $a$ a square modulo $b$" using Theorem 4.1.7.

*Example* 4.1.8. Is 69 a square modulo the prime 389? We have

$$\left(\frac{69}{389}\right) = \left(\frac{3\cdot 23}{389}\right) = \left(\frac{3}{389}\right)\cdot\left(\frac{23}{389}\right) = (-1)\cdot(-1) = 1.$$

Here

$$\left(\frac{3}{389}\right) = \left(\frac{389}{3}\right) = \left(\frac{2}{3}\right) = -1,$$

and

$$\begin{aligned}\left(\frac{23}{389}\right) &= \left(\frac{389}{23}\right) = \left(\frac{21}{23}\right) = \left(\frac{-2}{23}\right) \\ &= \left(\frac{-1}{23}\right)\left(\frac{2}{23}\right) = (-1)^{\frac{23-1}{2}}\cdot 1 = -1.\end{aligned}$$

Thus 69 is a square modulo 389.

*SAGE Example* 4.1.9. We could also do this computation in Sage as follows:

```
sage: legendre_symbol(69,389)
1
```

Though we know that 69 is a square modulo 389, we don't know an explicit $x$ such that $x^2 \equiv 69 \pmod{389}$! This is reminiscent of how we proved using Theorem 2.1.20 that certain numbers are composite without knowing a factorization.

*Remark* 4.1.10. The Jacobi symbol is an extension of the Legendre symbol to composite moduli. For more details, see Exercise 4.9.

## 4.2 Euler's Criterion

Let $p$ be an odd prime and $a$ an integer not divisible by $p$. Euler used the existence of primitive roots to show that $\left(\frac{a}{p}\right)$ is congruent to $a^{(p-1)/2}$ modulo $p$. We will use this fact repeatedly below in both proofs of Theorem 4.1.7.

**Proposition 4.2.1** (Euler's Criterion)**.** *We have* $\left(\frac{a}{p}\right) = 1$ *if and only if*

$$a^{(p-1)/2} \equiv 1 \pmod p.$$

*Proof.* The map $\varphi : (\mathbf{Z}/p\mathbf{Z})^* \to (\mathbf{Z}/p\mathbf{Z})^*$ given by $\varphi(a) = a^{(p-1)/2}$ is a group homomorphism, since powering is a group homomorphism of any abelian group (see Exercise 4.2). Let $\psi : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$ be the homomorphism $\psi(a) = \left(\frac{a}{p}\right)$ of Lemma 4.1.4. If $a \in \ker(\psi)$, then $a = b^2$ for some $b \in (\mathbf{Z}/p\mathbf{Z})^*$, so

$$\varphi(a) = a^{(p-1)/2} = (b^2)^{(p-1)/2} = b^{p-1} = 1.$$

Thus $\ker(\psi) \subset \ker(\varphi)$. By Lemma 4.1.4, $\ker(\psi)$ has index 2 in $(\mathbf{Z}/p\mathbf{Z})^*$, i.e., $\#(\mathbf{Z}/p\mathbf{Z})^* = 2 \cdot \#\ker(\psi)$. Since the kernel of a homomorphism is a group, and the order of a subgroup divides the order of the group, we have either $\ker(\varphi) = \ker(\psi)$ or $\varphi = 1$. If $\varphi = 1$, the polynomial $x^{(p-1)/2} - 1$ has $p - 1$ roots in the field $\mathbf{Z}/p\mathbf{Z}$, which contradicts Proposition 2.5.3. Thus $\ker(\varphi) = \ker(\psi)$, which proves the proposition. □

*SAGE Example* 4.2.2. From a computational point of view, Corollary 4.2.3 provides a convenient way to compute $\left(\frac{a}{p}\right)$, which we illustrate in Sage:

```
sage: def kr(a, p):
...    if Mod(a,p)^((p-1)//2) == 1:
...       return 1
...    else:
...       return -1
sage: for a in range(1,5):
...    print a, kr(a,5)
1 1
2 -1
3 -1
4 1
```

**Corollary 4.2.3.** *The equation* $x^2 \equiv a \pmod{p}$ *has no solution if and only if* $a^{(p-1)/2} \equiv -1 \pmod{p}$. *Thus* $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$.

*Proof.* This follows from Proposition 4.2.1 and the fact that the polynomial $x^2 - 1$ has no roots besides $+1$ and $-1$ (which follows from Proposition 2.5.5). □

As additional computational motivation for the value of Corollary 4.2.3, note that to evaluate $\left(\frac{a}{p}\right)$ using Theorem 4.1.7 would not be practical if $a$ and $p$ are both very large, because it would require factoring $a$. However, Corollary 4.2.3 provides a method for evaluating $\left(\frac{a}{p}\right)$ without factoring $a$.

*Example* 4.2.4. Suppose $p = 11$. By squaring each element of $(\mathbf{Z}/11\mathbf{Z})^*$, we see that the squares modulo 11 are $\{1, 3, 4, 5, 9\}$. We compute $a^{(p-1)/2} = a^5$

for each $a \in (\mathbf{Z}/11\mathbf{Z})^*$ and get

$$1^5 = 1,\ 2^5 = -1,\ 3^5 = 1,\ 4^5 = 1,\ 5^5 = 1,$$
$$6^5 = -1,\ 7^5 = -1,\ 8^5 = -1,\ 9^5 = 1,\ 10^5 = -1.$$

Thus the $a$ with $a^5 = 1$ are $\{1, 3, 4, 5, 9\}$, just as Proposition 4.2.1 predicts.

*Example* 4.2.5. We determine whether or not 3 is a square modulo the prime $p = 726377359$.

```
sage: p = 726377359
sage: Mod(3, p)^((p-1)//2)
726377358
```

so

$$3^{(p-1)/2} \equiv -1 \pmod{726377359}.$$

Thus 3 is not a square modulo $p$. This computation wasn't difficult, but it would have been tedious by hand. Since 3 is small, the Quadratic Reciprocity Law provides a way to answer this question, which could easily be carried out by hand:

$$\left(\frac{3}{726377359}\right) = (-1)^{(3-1)/2\cdot(726377359-1)/2}\left(\frac{726377359}{3}\right)$$
$$= (-1)\cdot\left(\frac{1}{3}\right) = -1.$$

## 4.3 First Proof of Quadratic Reciprocity

Our first proof of quadratic reciprocity is elementary. The proof involves keeping track of integer points in intervals. Proving Gauss's lemma is the first step; this lemma computes $\left(\frac{a}{p}\right)$ in terms of the number of integers of a certain type that lie in a certain interval. We next prove Lemma 4.3.3, which controls how the parity of the number of integer points in an interval changes when an endpoint of the interval is changed. We then prove that $\left(\frac{a}{p}\right)$ depends only on $p$ modulo $4a$ by applying Gauss's Lemma and keeping careful track of intervals as they are rescaled and their endpoints are changed. Finally, in Section 4.3.2, we use some basic algebra to deduce the Quadratic Reciprocity Law using the tools we've just developed. Our proof follows the one given in [Dav99] closely.

**Lemma 4.3.1** (Gauss's Lemma)**.** *Let $p$ be an odd prime and let $a$ be an integer $\not\equiv 0 \pmod{p}$. Form the numbers*

$$a,\ 2a,\ 3a,\ \ldots,\ \frac{p-1}{2}a$$

*and reduce them modulo $p$ to lie in the interval $(-\frac{p}{2},\ \frac{p}{2})$, i.e., for each of the above products $k \cdot a$ find a number in the interval $(-\frac{p}{2},\ \frac{p}{2})$ that is congruent to $k \cdot a$ modulo $p$. Let $\nu$ be the number of negative numbers in the resulting set. Then*

$$\left(\frac{a}{p}\right) = (-1)^{\nu}.$$

*Proof.* In defining $\nu$, we expressed each number in

$$S = \left\{a, 2a, \ldots, \frac{p-1}{2}a\right\}$$

as congruent to a number in the set

$$\left\{1, -1, 2, -2, \ldots, \frac{p-1}{2}, -\frac{p-1}{2}\right\}.$$

No number $1, 2, \ldots, \frac{p-1}{2}$ appears more than once, with either choice of sign, because if it did then either two elements of $S$ are congruent modulo $p$ or 0 is the sum of two elements of $S$, and both events are impossible (the former case cannot occur because of cancellation modulo $p$, and in the latter case we would have $ka + ja \equiv 0 \pmod{p}$ for $1 \leq k, j \leq (p-1)/2$, so $k + j \equiv 0 \pmod{p}$, a contradiction). The resulting set must be of the form

$$T = \left\{\varepsilon_1 \cdot 1, \varepsilon_2 \cdot 2, \ldots, \varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right\},$$

where each $\varepsilon_i$ is either $+1$ or $-1$. Multiplying together the elements of $S$ and of $T$, we see that

$$(1a) \cdot (2a) \cdot (3a) \cdots \left(\frac{p-1}{2}a\right) \equiv (\varepsilon_1 \cdot 1) \cdot (\varepsilon_2 \cdot 2) \cdots \left(\varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right) \pmod{p},$$

so

$$a^{(p-1)/2} \equiv \varepsilon_1 \cdot \varepsilon_2 \cdots \varepsilon_{(p-1)/2} \pmod{p}.$$

The lemma then follows from Proposition 4.2.1, since $\left(\frac{a}{p}\right) = a^{(p-1)/2}$. □

*SAGE Example* 4.3.2. We illustrate Gauss's Lemma using Sage. The `gauss` function below prints out a list of the normalized numbers appearing in the statement of Gauss's Lemma, and returns $(-1)^{\nu}$. In each case below, $(-1)^{\nu} = \left(\frac{a}{p}\right)$.

```
sage: def gauss(a, p):
...    # make the list of numbers reduced modulo p
```

```
...     v = [(n*a)%p for n in range(1, (p-1)//2 + 1)]
...     # normalize them to be in the range -p/2 to p/2
...     v = [(x if (x < p/2) else x - p) for x in v]
...     # sort and print the resulting numbers
...     v.sort()
...     print v
...     # count the number that are negative
...     num_neg = len([x for x in v if x < 0])
...     return (-1)^num_neg
sage: gauss(2, 13)
[-5, -3, -1, 2, 4, 6]
-1
sage: legendre_symbol(2,13)
-1
sage: gauss(4, 13)
[-6, -5, -2, -1, 3, 4]
1
sage: legendre_symbol(4,13)
1
sage: gauss(2,31)
[-15, -13, -11, -9, -7, -5, -3, -1, 2, 4, 6, 8, 10, 12, 14]
1
sage: legendre_symbol(2,31)
1
```

### *4.3.1 Euler's Proposition*

For rational numbers $a, b \in \mathbf{Q}$, let

$$(a, b) \cap \mathbf{Z} = \{x \in \mathbf{Z} : a \leq x \leq b\}$$

be the set of integers between $a$ and $b$. The following lemma will help us to keep track of how many integers lie in certain intervals.

**Lemma 4.3.3.** *Let $a, b \in \mathbf{Q}$. Then for any integer $n$,*

$$\#\left((a, b) \cap \mathbf{Z}\right) \equiv \#\left((a, b + 2n) \cap \mathbf{Z}\right) \pmod 2$$

*and*

$$\#\left((a, b) \cap \mathbf{Z}\right) \equiv \#\left((a - 2n, b) \cap \mathbf{Z}\right) \pmod 2,$$

*provided that each interval involved in the congruence is nonempty.*

Note that if one of the intervals is empty, then the statement may be false; for example, if $(a, b) = (-1/2, 1/2)$ and $n = -1$, then $\#((a, b) \cap \mathbf{Z}) = 1$ but $\#(a, b - 2) \cap \mathbf{Z} = 0$.

*Proof.* Let $\lceil x \rceil$ denotes the least integer $\geq x$. Since $n > 0$,

$$(a, b + 2n) = (a, b) \cup [b, b + 2n),$$

where the union is disjoint. There are $2n$ integers

$$\lceil b \rceil, \lceil b \rceil + 1, \ldots, \lceil b \rceil + 2n - 1$$

in the interval $[b, b + 2n)$, so the first congruence of the lemma is true in this case. We also have

$$(a, b - 2n) = (a, b) \text{ minus } [b - 2n, b)$$

and $[b-2n, b)$ contains exactly $2n$ integers, so the lemma is also true when $n$ is negative. The statement about $\#\left((a - 2n, b) \cap \mathbf{Z}\right)$ is proved in a similar manner. □

Once we have proved the following proposition, it will be easy to deduce the Quadratic Reciprocity Law.

**Proposition 4.3.4** (Euler)**.** *Let $p$ be an odd prime and let $a$ be a positive integer with $p \nmid a$. If $q$ is a prime with $q \equiv \pm p \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.*

*Proof.* We will apply Lemma 4.3.1 to compute $\left(\frac{a}{p}\right)$. Let

$$S = \left\{a, 2a, 3a, \ldots, \frac{p-1}{2}a\right\}$$

and

$$I = \left(\frac{1}{2}p, p\right) \cup \left(\frac{3}{2}p, 2p\right) \cup \cdots \cup \left(\left(b - \frac{1}{2}\right)p, bp\right),$$

where $b = \frac{1}{2}a$ or $\frac{1}{2}(a-1)$, whichever is an integer.

We check that every element of $S$ that is equivalent modulo $p$ to something in the interval $(-\frac{p}{2}, 0)$ lies in $I$. First suppose that $b = \frac{1}{2}a$. Then

$$bp = \frac{1}{2}ap = \frac{p}{2}a > \frac{p-1}{2}a,$$

so each element of $S$ that is equivalent modulo $p$ to an element of $(-\frac{p}{2}, 0)$ lies in $I$. Next suppose that $b = \frac{1}{2}(a-1)$. Then

$$bp + \frac{p}{2} = \frac{a-1}{2}p + \frac{p}{2} = \frac{p-1+a}{2} > \frac{p-1}{2}a,$$

so $((b - \frac{1}{2})p, bp)$ is the last interval that could contain an element of $S$ that reduces to $(-\frac{p}{2}, 0)$. Note that the integer endpoints of $I$ are not in $S$, since

those endpoints are divisible by $p$, but no element of $S$ is divisible by $p$. Thus, by Lemma 4.3.1,

$$\left(\frac{a}{p}\right) = (-1)^{\#(S\cap I)}.$$

To compute $\#(S \cap I)$, first rescale by $a$ to see that

$$\#(S\cap I) = \#\left(\frac{1}{a}S \cap \frac{1}{a}I\right) = \#\left(\mathbf{Z}\cap \frac{1}{a}I\right),$$

where

$$\frac{1}{a}I = \left(\left(\frac{p}{2a}, \frac{p}{a}\right) \cup \left(\frac{3p}{2a}, \frac{2p}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)p}{2a}, \frac{bp}{a}\right)\right),$$

$\frac{1}{a}S = \{1, 2, 3, 4, \ldots, (p-1)/2\}$, and the second equality is because $\frac{1}{a}I \subset (0, (p-1)/2 + 1/2]$, since

$$\frac{pb}{a} \le \frac{p\frac{a}{2}}{a} = \frac{p}{2} = \frac{p-1}{2} + \frac{1}{2}.$$

Write $p = 4ac + r$, and let

$$J = \left(\left(\frac{r}{2a}, \frac{r}{a}\right) \cup \left(\frac{3r}{2a}, \frac{2r}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)r}{2a}, \frac{br}{a}\right)\right).$$

The only difference between $\frac{1}{a}I$ and $J$ is that the endpoints of intervals are changed by addition of an even integer, since

$$\frac{r}{2a} - \frac{p}{2a} = \frac{p}{2a} - 2c - \frac{p}{2a} = -2c.$$

By Lemma 4.3.3,

$$\nu = \#\left(\mathbf{Z}\cap\frac{1}{a}I\right) \equiv \#(\mathbf{Z}\cap J) \pmod 2.$$

Thus $\left(\frac{a}{p}\right) = (-1)^\nu$ depends only on $r$ and $a$, i.e., only on $p$ modulo $4a$. Thus if $q \equiv p \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.

If $q \equiv -p \pmod{4a}$, then the only change in the above computation is that $r$ is replaced by $4a - r$. This changes $J$ into

$$K = \left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right) \cup \left(6 - \frac{3r}{2a}, 8 - \frac{2r}{a}\right) \cup \cdots \cup \left(4b - 2 - \frac{(2b-1)r}{2a}, 4b - \frac{br}{a}\right).$$

Thus $K$ is the same as $-J$, except even integers have been added to the endpoints. By Lemma 4.3.3,

$$\#(K \cap \mathbf{Z}) \equiv \# \left( \frac{1}{a} I \cap \mathbf{Z} \right) \pmod 2,$$

so $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$ again, which completes the proof. □

The following more careful analysis in the special case when $a = 2$ helps illustrate the proof of the above lemma, and the result is frequently useful in computations. For an alternative proof of the proposition, see Exercise 4.6.

**Proposition 4.3.5** (Legendre Lymbol of 2)**.** *Let $p$ be an odd prime. Then*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

*Proof.* When $a = 2$, the set $S = \{a, 2a, \ldots, 2 \cdot \frac{p-1}{2}\}$ is

$$\{2, 4, 6, \ldots, p-1\}.$$

We must count the parity of the number of elements of $S$ that lie in the interval $I = (\frac{p}{2}, p)$. Writing $p = 8c + r$, we have

$$\begin{aligned} \#\left(I \cap S\right) &= \# \left( \frac{1}{2} I \cap \mathbf{Z} \right) = \# \left( \left( \frac{p}{4}, \frac{p}{2} \right) \cap \mathbf{Z} \right) \\ &= \# \left( \left( 2c + \frac{r}{4}, 4c + \frac{r}{2} \right) \cap \mathbf{Z} \right) \equiv \# \left( \left( \frac{r}{4}, \frac{r}{2} \right) \cap \mathbf{Z} \right) \pmod 2, \end{aligned}$$

where the last equality comes from Lemma 4.3.3. The possibilities for $r$ are $1, 3, 5, 7$. When $r = 1$, the cardinality is 0; when $r = 3, 5$ it is 1; and when $r = 7$ it is 2. □

### *4.3.2 Proof of Quadratic Reciprocity*

It is now straightforward to deduce the Quadratic Reciprocity Law.

*First Proof of Theorem 4.1.7.* First suppose that $p \equiv q \pmod 4$. By swapping $p$ and $q$ if necessary, we may assume that $p > q$, and write $p - q = 4a$. Since $p = 4a + q$,

$$\left(\frac{p}{q}\right) = \left(\frac{4a+q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{4}{q}\right)\left(\frac{a}{q}\right) = \left(\frac{a}{q}\right),$$

and

$$\left(\frac{q}{p}\right) = \left(\frac{p-4a}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right) \cdot \left(\frac{a}{p}\right).$$

Proposition 4.3.4 implies that $\left(\frac{a}{q}\right) = \left(\frac{a}{p}\right)$, since $p \equiv q \pmod{4a}$. Thus

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}},$$

where the last equality is because $\frac{p-1}{2}$ is even if and only if $\frac{q-1}{2}$ is even.

Next suppose that $p \not\equiv q \pmod{4}$, so $p \equiv -q \pmod{4}$. Write $p+q = 4a$. We have

$$\left(\frac{p}{q}\right) = \left(\frac{4a-q}{q}\right) = \left(\frac{a}{q}\right), \quad \text{and} \quad \left(\frac{q}{p}\right) = \left(\frac{4a-p}{p}\right) = \left(\frac{a}{p}\right).$$

Since $p \equiv -q \pmod{4a}$, Proposition 4.3.4 implies that $\left(\frac{a}{q}\right) = \left(\frac{a}{p}\right)$. Since $(-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}} = 1$, the proof is complete. □

## 4.4 A Proof of Quadratic Reciprocity Using Gauss Sums

In this section, we present a beautiful proof of Theorem 4.1.7 using algebraic identities satisfied by sums of "roots of unity." The objects we introduce in the proof are of independent interest, and provide a powerful tool to prove higher-degree analogs of quadratic reciprocity. (For more on higher reciprocity, see [IR90]. See also Section 6 of [IR90], on which the proof below is modeled.)

**Definition 4.4.1** (Root of Unity)**.** An $n$th *root of unity* is a complex number $\zeta$ such that $\zeta^n = 1$. A root of unity $\zeta$ is a *primitive* $n$th root of unity if $n$ is the smallest positive integer such that $\zeta^n = 1$.

For example, $-1$ is a primitive second root of unity, and $\zeta = \frac{\sqrt{-3}-1}{2}$ is a primitive cube root of unity. More generally, for any $n \in \mathbf{N}$ the complex number

$$\zeta_n = \cos(2\pi/n) + i\sin(2\pi/n)$$

is a primitive $n$th root of unity (this follows from the identity $e^{i\theta} = \cos(\theta) + i\sin(\theta)$). For the rest of this section, we fix an odd prime $p$ and the primitive $p$th root $\zeta = \zeta_p$ of unity.

*SAGE Example* 4.4.2. In Sage, use the `CyclotomicField` command to create an exact $p$th root of $\zeta$ unity. Expressions in $\zeta$ are always re-expressed as polynomials in $\zeta$ of degree at most $p-1$.

```
sage: K.<zeta> = CyclotomicField(5)
sage: zeta^5
1
```

```
sage: 1/zeta
-zeta^3 - zeta^2 - zeta - 1
```

**Definition 4.4.3** (Gauss Sum)**.** Fix an odd prime $p$. The *Gauss sum* associated to an integer $a$ is

$$g_a = \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) \zeta^{an},$$

where $\zeta = \zeta_p = \cos(2\pi/p) + i\sin(2\pi/p) = e^{2\pi i/p}$.

Note that $p$ is implicit in the definition of $g_a$. If we were to change $p$, then the Gauss sum $g_a$ associated to $a$ would be different. The definition of $g_a$ also depends on our choice of $\zeta$; we've chosen $\zeta = \zeta_p$, but could have chosen a different $\zeta$ and then $g_a$ could be different.

*SAGE Example* 4.4.4. We define a `gauss_sum` function and compute the Gauss sum $g_2$ for $p = 5$:

```
sage: def gauss_sum(a,p):
...   K.<zeta> = CyclotomicField(p)
...   return sum(legendre_symbol(n,p) * zeta^(a*n)
...              for n in range(1,p))
sage: g2 = gauss_sum(2,5); g2
2*zeta^3 + 2*zeta^2 + 1
sage: g2.complex_embedding()
-2.2360679775 + 3.33066907388e-16*I
sage: g2^2
5
```

Here, $g_2$ is initially output as a polynomial in $\zeta_5$, so there is no loss of precision. The `complex_embedding` command shows some embedding of $g_2$ into the complex numbers, which is only correct to about the first 15 digits. Note that $g_2^2 = 5$, so $g_2 = -\sqrt{5}$.

We compute a graphical representation of the Gauss sum $g_2$ as follows (see Figure 4.1):

```
zeta = CDF(exp(2*pi*I/5))
v = [legendre_symbol(n,5) * zeta^(2*n) for n in range(1,5)]
S = sum([point(tuple(z), pointsize=100) for z in v])
show(S + point(tuple(sum(v)), pointsize=100, rgbcolor='red'))
```

Figure 4.1 illustrates the Gauss sum $g_2$ for $p = 5$. The Gauss sum is obtained by adding the points on the unit circle, with signs as indicated, to obtain the real number $-\sqrt{5}$. This suggests the following proposition, whose proof will require some work.

**Proposition 4.4.5** (Gauss Sum)**.** *For any $a$ not divisible by $p$,*

$$g_a^2 = (-1)^{(p-1)/2} p.$$

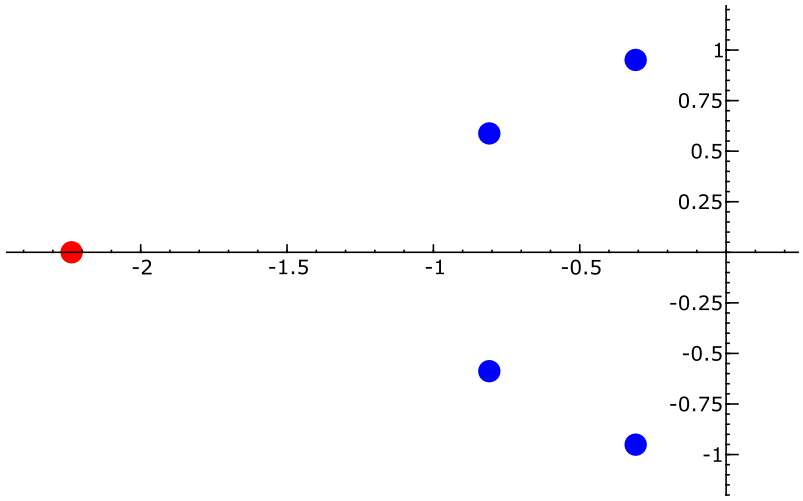



FIGURE 4.1. The red dot is the Gauss sum $g_2$ for $p = 5$

*SAGE Example* 4.4.6. We illustrate using Sage that the proposition is correct for $p = 7$ and $p = 13$:

```
sage: [gauss_sum(a, 7)^2 for a in range(1,7)]
[-7, -7, -7, -7, -7, -7]
sage: [gauss_sum(a, 13)^2 for a in range(1,13)]
[13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13]
```

In order to prove the proposition, we introduce a few lemmas.

**Lemma 4.4.7.** *For any integer $a$,*

$$\sum_{n=0}^{p-1} \zeta^{an} = \begin{cases} p & \text{if } a \equiv 0 \pmod{p}, \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* If $a \equiv 0 \pmod{p}$, then $\zeta^a = 1$, so the sum equals the number of summands, which is $p$. If $a \not\equiv 0 \pmod{p}$, then we use the identity

$$x^p - 1 = (x-1)(x^{p-1} + \cdots + x + 1)$$

with $x = \zeta^a$. We have $\zeta^a \neq 1$, so $\zeta^a - 1 \neq 0$ and

$$\sum_{n=0}^{p-1} \zeta^{an} = \frac{\zeta^{ap} - 1}{\zeta^a - 1} = \frac{1-1}{\zeta^a - 1} = 0.$$

□

**Lemma 4.4.8.** *If $x$ and $y$ are arbitrary integers, then*

$$\sum_{n=0}^{p-1} \zeta^{(x-y)n} = \begin{cases} p & \text{if } x \equiv y \pmod{p}, \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* This follows from Lemma 4.4.7 by setting $a = x - y$. □

**Lemma 4.4.9.** *We have $g_0 = 0$.*

*Proof.* By definition

$$g_0 = \sum_{n=0}^{p-1} \left(\frac{n}{p}\right). \tag{4.4.1}$$

By Lemma 4.1.4, the map

$$\left(\frac{\cdot}{p}\right) : (\mathbf{Z}/p\mathbf{Z})^* \to \{\pm 1\}$$

is a surjective homomorphism of groups. Thus, half the elements of $(\mathbf{Z}/p\mathbf{Z})^*$ map to $+1$ and half map to $-1$ (the subgroup that maps to $+1$ has index 2). Since $\left(\frac{0}{p}\right) = 0$, the sum (4.4.1) is 0. □

**Lemma 4.4.10.** *For any integer $a$,*

$$g_a = \left(\frac{a}{p}\right) g_1.$$

*Proof.* When $a \equiv 0 \pmod{p}$, the lemma follows from Lemma 4.4.9, so suppose that $a \not\equiv 0 \pmod{p}$. Then,

$$\left(\frac{a}{p}\right) g_a = \left(\frac{a}{p}\right) \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^{an} = \sum_{n=0}^{p-1} \left(\frac{an}{p}\right) \zeta^{an} = \sum_{m=0}^{p-1} \left(\frac{m}{p}\right) \zeta^{m} = g_1.$$

Here, we use that multiplication by $a$ is an automorphism of $\mathbf{Z}/p\mathbf{Z}$. Finally, multiply both sides by $\left(\frac{a}{p}\right)$ and use that $\left(\frac{a}{p}\right)^2 = 1$. □

We have enough lemmas to prove Proposition 4.4.5.

*Proof of Proposition 4.4.5.* We evaluate the sum $\sum_{a=0}^{p-1} g_a g_{-a}$ in two different ways. By Lemma 4.4.10, since $a \not\equiv 0 \pmod{p}$ we have

$$g_a g_{-a} = \left(\frac{a}{p}\right) g_1 \left(\frac{-a}{p}\right) g_1 = \left(\frac{-1}{p}\right) \left(\frac{a}{p}\right)^2 g_1^2 = (-1)^{(p-1)/2} g_1^2,$$

where the last step follows from Proposition 4.2.1 and that $\left(\frac{a}{p}\right) \in \{\pm 1\}$. Thus

$$\sum_{a=0}^{p-1} g_a g_{-a} = (p-1)(-1)^{(p-1)/2} g_1^2. \tag{4.4.2}$$

On the other hand, by definition

$$\begin{aligned} g_a g_{-a} &= \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^{an} \cdot \sum_{m=0}^{p-1} \left(\frac{m}{p}\right) \zeta^{-am} \\ &= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an} \zeta^{-am} \\ &= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an-am}. \end{aligned}$$

Let $\delta(n,m) = 1$ if $n \equiv m \pmod{p}$ and 0 otherwise. By Lemma 4.4.8,

$$\begin{aligned} \sum_{a=0}^{p-1} g_a g_{-a} &= \sum_{a=0}^{p-1} \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \zeta^{an-am} \\ &= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) \sum_{a=0}^{p-1} \zeta^{an-am} \\ &= \sum_{n=0}^{p-1} \sum_{m=0}^{p-1} \left(\frac{n}{p}\right) \left(\frac{m}{p}\right) p\delta(n,m) \\ &= \sum_{n=0}^{p-1} \left(\frac{n}{p}\right)^2 p \\ &= p(p-1). \end{aligned}$$

Equate (4.4.2) and the above equality, then cancel $(p-1)$ to see that

$$g_1^2 = (-1)^{(p-1)/2} p.$$

Since $a \not\equiv 0 \pmod{p}$, we have $\left(\frac{a}{p}\right)^2 = 1$, so by Lemma 4.4.10,

$$g_a^2 = \left(\frac{a}{p}\right)^2 g_1^2 = g_1^2,$$

and the proposition is proved. □

### *4.4.1 Proof of Quadratic Reciprocity*

We are now ready to prove Theorem 4.1.7 using Gauss sums.

*Proof.* Let $q$ be an odd prime with $q \neq p$. Set $p^* = (-1)^{(p-1)/2} p$ and recall that Proposition 4.4.5 asserts that $p^* = g^2$, where $g = g_1 = \sum_{n=0}^{p-1} \left(\frac{n}{p}\right) \zeta^n$.

Proposition 4.2.1 implies that

$$(p^*)^{(q-1)/2} \equiv \left(\frac{p^*}{q}\right) \pmod{q}.$$

We have $g^{q-1} = (g^2)^{(q-1)/2} = (p^*)^{(q-1)/2}$, so multiplying both sides of the displayed equation by $g$ yields a congruence

$$g^q \equiv g\left(\frac{p^*}{q}\right) \pmod{q}. \tag{4.4.3}$$

But wait, what does this congruence mean, given that $g^q$ is not an integer? It means that the difference $g^q - g\left(\frac{p^*}{q}\right)$ is a multiple of $q$ in the ring $\mathbf{Z}[\zeta]$ of all polynomials in $\zeta$ with coefficients in $\mathbf{Z}$.

The ring $\mathbf{Z}[\zeta]/(q)$ has characteristic $q$, so if $x, y \in \mathbf{Z}[\zeta]$, then $(x+y)^q \equiv x^q + y^q \pmod{q}$. Applying this to (4.4.3), we see that

$$g^q = \left(\sum_{n=0}^{p-1}\left(\frac{n}{p}\right)\zeta^n\right)^q \equiv \sum_{n=0}^{p-1}\left(\frac{n}{p}\right)^q \zeta^{nq} \equiv \sum_{n=0}^{p-1}\left(\frac{n}{p}\right)\zeta^{nq} \equiv g_q \pmod{q}.$$

By Lemma 4.4.10,

$$g^q \equiv g_q \equiv \left(\frac{q}{p}\right) g \pmod{q}.$$

Combining this with (4.4.3) yields

$$\left(\frac{q}{p}\right) g \equiv \left(\frac{p^*}{q}\right) g \pmod{q}.$$

Since $g^2 = p^*$ and $p \neq q$, we can cancel $g$ from both sides to find that $\left(\frac{q}{p}\right) \equiv \left(\frac{p^*}{q}\right) \pmod{q}$. Since both residue symbols are $\pm 1$ and $q$ is odd, it follows that $\left(\frac{q}{p}\right) = \left(\frac{p^*}{q}\right)$. Finally, we note using Corollary 4.2.3 that

$$\left(\frac{p^*}{q}\right) = \left(\frac{(-1)^{(p-1)/2}p}{q}\right) = \left(\frac{-1}{q}\right)^{(p-1)/2}\left(\frac{p}{q}\right) = (-1)^{\frac{q-1}{2}\cdot\frac{p-1}{2}} \cdot \left(\frac{p}{q}\right).$$

□

## 4.5 Finding Square Roots

We return in this section to the question of computing square roots. If $K$ is a field in which $2 \neq 0$, and $a, b, c \in K$, with $a \neq 0$, then the two solutions to the quadratic equation $ax^2 + bx + c = 0$ are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Now assume $K = \mathbf{Z}/p\mathbf{Z}$, with $p$ an odd prime. Using Theorem 4.1.7, we can decide whether or not $b^2 - 4ac$ is a perfect square in $\mathbf{Z}/p\mathbf{Z}$, and hence whether or not $ax^2 + bx + c = 0$ has a solution in $\mathbf{Z}/p\mathbf{Z}$. However, Theorem 4.1.7 says nothing about how to actually find a solution when there is one. Also note that for this problem we do *not* need the full Quadratic Reciprocity Law; in practice, deciding whether an element of $\mathbf{Z}/p\mathbf{Z}$ is a perfect square with Proposition 4.2.1 is quite fast, in view of Section 2.3.

Suppose $a \in \mathbf{Z}/p\mathbf{Z}$ is a nonzero quadratic residue. If $p \equiv 3 \pmod 4$, then $b = a^{\frac{p+1}{4}}$ is a square root of $a$ because

$$b^2 = a^{\frac{p+1}{2}} = a^{\frac{p-1}{2}+1} = a^{\frac{p-1}{2}} \cdot a = \left(\frac{a}{p}\right) \cdot a = a.$$

We can compute $b$ in time polynomial in the number of digits of $p$ using the powering algorithm of Section 2.3.

Suppose next that $p \equiv 1 \pmod 4$. Unfortunately, we do not know a deterministic algorithm that takes $a$ and $p$ as input, outputs a square root of $a$ modulo $p$ when one exists, and is polynomial-time in $\log(p)$.

*Remark* 4.5.1. There is an algorithm due to Schoof [Sch85] that computes the square root of $a$ in time $O((\sqrt{(|a|)}^{1/2+\varepsilon} \cdot \log(p))^9)$. This beautiful algorithm (which makes use of elliptic curves) is not polynomial time in the sense described above, since for large $a$ it takes exponentially longer than for small $a$.

We next describe a probabilistic algorithm to compute a square root of $a$ modulo $p$, which is very quick in practice. Recall the notion of ring from Definition 2.1.3. We will also need the notion of ring homomorphism and isomorphism.

**Definition 4.5.2** (Homomorphism of Rings)**.** Let $R$ and $S$ be rings. A *homomorphism of rings* $\varphi : R \to S$ is a map such that for all $a, b \in R$, we have

- $\varphi(ab) = \varphi(a)\varphi(b)$,
- $\varphi(a + b) = \varphi(a) + \varphi(b)$, and
- $\varphi(1) = 1$.

An *isomorphism* $\varphi : R \to S$ of rings is a ring homomorphism that is bijective.

Consider the ring

$$R = (\mathbf{Z}/p\mathbf{Z})[x]/(x^2 - a)$$

defined as follows. We have

$$R = \{u + v\alpha : u, v \in \mathbf{Z}/p\mathbf{Z}\}$$

with multiplication defined by

$$(u + v\alpha)(z + w\alpha) = (uz + awv) + (uw + vz)\alpha.$$

Here $\alpha$ corresponds to the class of $x$ in $R$.

*SAGE Example* 4.5.3. We define and work with the ring $R$ above in Sage as follows (for $p = 13$):

```
sage: S.<x> = PolynomialRing(GF(13))
sage: R.<alpha> = S.quotient(x^2 - 3)
sage: (2+3*alpha)*(1+2*alpha)
7*alpha + 7
```

Let $b$ and $c$ be the square roots of $a$ in $\mathbf{Z}/p\mathbf{Z}$ (though we cannot easily compute $b$ and $c$ yet, we can consider them in order to deduce an algorithm to find them). We have ring homomorphisms $f : R \to \mathbf{Z}/p\mathbf{Z}$ and $g : R \to \mathbf{Z}/p\mathbf{Z}$ given by $f(u + v\alpha) = u + vb$ and $g(u + v\alpha) = u + vc$. Together, these define a ring isomorphism

$$\varphi : R \longrightarrow \mathbf{Z}/p\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}$$

given by $\varphi(u + v\alpha) = (u + vb, u + vc)$. Choose in some way a random element $z$ of $(\mathbf{Z}/p\mathbf{Z})^*$, and define $u, v \in \mathbf{Z}/p\mathbf{Z}$ by

$$u + v\alpha = (1 + z\alpha)^{\frac{p-1}{2}},$$

where we compute $(1+z\alpha)^{\frac{p-1}{2}}$ quickly using an analog of the binary powering algorithm of Section 2.3.2. If $v = 0$, we try again with another random $z$. If $v \neq 0$, we can quickly find the desired square roots $b$ and $c$ as follows. The quantity $u+vb$ is a $(p-1)/2$ power in $\mathbf{Z}/p\mathbf{Z}$, so it equals either 0, 1, or $-1$, so $b = -u/v$, $(1 - u)/v$, or $(-1 - u)/v$, respectively. Since we know $u$ and $v$, we can try each of $-u/v$, $(1 - u)/v$, and $(-1 - u)/v$ and see which is a square root of $a$.

*Example* 4.5.4. Continuing Example 4.1.8, we find a square root of 69 modulo 389. We apply the algorithm described above in the case $p \equiv 1 \pmod 4$. We first choose the random $z = 24$ and find that $(1 + 24\alpha)^{194} = -1$. The coefficient of $\alpha$ in the power is 0, and we try again with $z = 51$. This time, we have $(1 + 51\alpha)^{194} = 239\alpha = u + v\alpha$. The inverse of 239 in $\mathbf{Z}/389\mathbf{Z}$ is 153, so we consider the following three possibilities for a square root of 69:

$$-\frac{u}{v} = 0 \qquad \frac{1-u}{v} = 153 \qquad -\frac{1-u}{v} = -153.$$

Thus, 153 and $-153$ are the square roots of 69 in $\mathbf{Z}/389\mathbf{Z}$.

*SAGE Example* 4.5.5. We implement the above algorithm in Sage and illustrate it with some examples.

```
sage: def find_sqrt(a, p):
...    assert (p-1)%4 == 0
...    assert legendre_symbol(a,p) == 1
...    S.<x> = PolynomialRing(GF(p))
...    R.<alpha> = S.quotient(x^2 - a)
...    while True:
...        z = GF(p).random_element()
...        w = (1 + z*alpha)^((p-1)//2)
...        (u, v) = (w[0], w[1])
...        if v != 0: break
...    if (-u/v)^2 == a: return -u/v
...    if ((1-u)/v)^2 == a: return (1-u)/v
...    if ((-1-u)/v)^2 == a: return (-1-u)/v
...
sage: b = find_sqrt(3,13)
sage: b                         # random: either 9 or 3
9
sage: b^2
3
sage: b = find_sqrt(3,13)
sage: b                         # see, it's random
4
sage: find_sqrt(5,389)          # random: either 303 or 86
303
sage: find_sqrt(5,389)          # see, it's random
86
```

## 4.6 Exercises

4.1 Calculate the following by hand: $\left(\frac{3}{97}\right)$, $\left(\frac{3}{389}\right)$, $\left(\frac{22}{11}\right)$, and $\left(\frac{5!}{7}\right)$.

4.2 Let $G$ be an abelian group, and let $n$ be a positive integer.

(a) Prove that the map $\varphi : G \to G$ given by $\varphi(x) = x^n$ is a group homomorphism.

(b) Prove that the subset $H$ of $G$ of squares of elements of $G$ is a subgroup.

4.3 Use Theorem 4.1.7 to show that for $p \geq 5$ prime,

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 11 \pmod{12}, \\ -1 & \text{if } p \equiv 5, 7 \pmod{12}. \end{cases}$$

4.4 (*) Use that $(\mathbf{Z}/p\mathbf{Z})^*$ is cyclic to give a direct proof that $\left(\frac{-3}{p}\right) = 1$ when $p \equiv 1 \pmod 3$. (Hint: There is an element $c \in (\mathbf{Z}/p\mathbf{Z})^*$ of order 3. Show that $(2c+1)^2 = -3$.)

4.5 (*) If $p \equiv 1 \pmod 5$, show directly that $\left(\frac{5}{p}\right) = 1$ by the method of Exercise 4.4. (Hint: Let $c \in (\mathbf{Z}/p\mathbf{Z})^*$ be an element of order 5. Show that $(c+c^4)^2 + (c+c^4) - 1 = 0$, etc.)

4.6 (*) Let $p$ be an odd prime. In this exercise, you will prove that $\left(\frac{2}{p}\right) = 1$ if and only if $p \equiv \pm 1 \pmod 8$.

(a) Prove that
$$x = \frac{1-t^2}{1+t^2}, \qquad y = \frac{2t}{1+t^2}$$
is a parameterization of the set of solutions to $x^2 + y^2 \equiv 1 \pmod p$, in the sense that the solutions $(x, y) \in \mathbf{Z}/p\mathbf{Z}$ are in bijection with the $t \in \mathbf{Z}/p\mathbf{Z} \cup \{\infty\}$ such that $1+t^2 \not\equiv 0 \pmod p$. Here, $t = \infty$ corresponds to the point $(-1, 0)$. (Hint: if $(x_1, y_1)$ is a solution, consider the line $y = t(x+1)$ through $(x_1, y_1)$ and $(-1, 0)$, and solve for $x_1, y_1$ in terms of $t$.)

(b) Prove that the number of solutions to $x^2 + y^2 \equiv 1 \pmod p$ is $p+1$ if $p \equiv 3 \pmod 4$ and $p-1$ if $p \equiv 1 \pmod 4$.

(c) Consider the set $S$ of pairs $(a, b) \in (\mathbf{Z}/p\mathbf{Z})^* \times (\mathbf{Z}/p\mathbf{Z})^*$ such that $a + b = 1$ and $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = 1$. Prove that $\#S = (p+1-4)/4$ if $p \equiv 3 \pmod 4$ and $\#S = (p-1-4)/4$ if $p \equiv 1 \pmod 4$. Conclude that $\#S$ is odd if and only if $p \equiv \pm 1 \pmod 8$.

(d) The map $\sigma(a, b) = (b, a)$ that swaps coordinates is a bijection of the set $S$. It has exactly one fixed point if and only if there is an $a \in \mathbf{Z}/p\mathbf{Z}$ such that $2a = 1$ and $\left(\frac{a}{p}\right) = 1$. Also, prove that $2a = 1$ has a solution $a \in \mathbf{Z}/p\mathbf{Z}$ with $\left(\frac{a}{p}\right) = 1$ if and only if $\left(\frac{2}{p}\right) = 1$.

(e) Finish by showing that $\sigma$ has exactly one fixed point if and only if $\#S$ is odd, i.e., if and only if $p \equiv \pm 1 \pmod 8$.

Remark: The method of proof of this exercise can be generalized to give a proof of the full Quadratic Reciprocity Law.

4.7 How many natural numbers $x < 2^{13}$ satisfy the equation
$$x^2 \equiv 5 \pmod{2^{13} - 1}?$$
You may assume that $2^{13} - 1$ is prime.

4.8 Find the natural number $x < 97$ such that $x \equiv 4^{48} \pmod{97}$. Note that 97 is prime.

4.9 In this problem, we will formulate an analog of quadratic reciprocity for a symbol like $\left(\frac{a}{q}\right)$, but without the restriction that $q$ be a prime. Suppose $n$ is an odd positive integer, which we factor as $\prod_{i=1}^{k} p_i^{e_i}$. We define the Jacobi symbol $\left(\frac{a}{n}\right)$ as follows:

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{e_i}.$$

(a) Give an example to show that $\left(\frac{a}{n}\right) = 1$ need not imply that $a$ is a perfect square modulo $n$.

(b) (*) Let $n$ be odd and $a$ and $b$ be integers. Prove that the following holds:

i. $\left(\frac{a}{n}\right)\left(\frac{b}{n}\right) = \left(\frac{ab}{n}\right)$. (Thus $a \mapsto \left(\frac{a}{n}\right)$ induces a homomorphism from $(\mathbf{Z}/n\mathbf{Z})^*$ to $\{\pm 1\}$.)

ii. $\left(\frac{-1}{n}\right) \equiv n \pmod{4}$.

iii. $\left(\frac{2}{n}\right) = 1$ if $n \equiv \pm 1 \pmod{8}$ and $-1$ otherwise.

iv. Assume $a$ is positive and odd. Then $\left(\frac{a}{n}\right) = (-1)^{\frac{a-1}{2}\cdot\frac{n-1}{2}}\left(\frac{n}{a}\right)$

4.10 (*) Prove that for any $n \in \mathbf{Z}$, the integer $n^2 + n + 1$ does not have any divisors of the form $6k - 1$.

# 5
# Continued Fractions

The golden ratio $\frac{1+\sqrt{5}}{2}$ is equal to the infinite fraction

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}},$$

and the fraction

$$\frac{103993}{33102} = 3.14159265301190260407\ldots$$

is an excellent approximation to $\pi$. Both of these observations are explained by continued fractions.

Continued fractions are theoretically beautiful and provide tools that yield powerful algorithms for solving problems in number theory. For example, continued fractions provide a fast way to write a prime—even a hundred digit prime—as a sum of two squares, when possible.

Continued fractions are thus a beautiful algorithmic and conceptual tool in number theory that has many applications. For example, they provide a surprisingly efficient way to recognize a rational number given just the first few digits of its decimal expansion, and they give a sense in which $e$ is "less complicated" than $\pi$ (see Example 5.3.4 and Section 5.4).

In Section 5.2, we study continued fractions of finite length and lay the foundations for our later investigations. In Section 5.3, we give the continued fraction procedure, which associates to a real number $x$ a continued fraction that converges to $x$. In Section 5.5, we characterize (eventually)

periodic continued fractions as the continued fractions of nonrational roots of quadratic polynomials, then discuss an unsolved mystery concerning continued fractions of roots of irreducible polynomials of degree greater than 2. We conclude the chapter with applications of continued fractions to recognizing approximations to rational numbers (Section 5.6) and writing integers as sums of two squares (Section 5.7).

The reader is encouraged to read more about continued fractions in [HW79, Ch. X], [Khi63], [Bur89, §13.3], and [NZM91, Ch. 7].

## 5.1 The Definition

A *continued fraction* is an expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}.$$

In this book, we will assume that the $a_i$ are real numbers and $a_i > 0$ for $i \geq 1$, and the expression may or may not go on indefinitely. More general notions of continued fractions have been extensively studied, but they are beyond the scope of this book. We will be most interested in the case when the $a_i$ are all integers.

We denote the continued fraction displayed above by

$$[a_0, a_1, a_2, \ldots].$$

For example,

$$[1, 2] = 1 + \frac{1}{2} = \frac{3}{2},$$

$$\begin{aligned}[3, 7, 15, 1, 292] &= 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292}}}} \\ &= \frac{103993}{33102} = 3.14159265301190260407\ldots,\end{aligned}$$

and

$$[2,1,2,1,1,4,1,1,6] = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \frac{1}{6}}}}}}}}$$
$$= \frac{1264}{465}$$
$$= 2.71827956989247311827956 98\ldots$$

The second two examples were chosen to foreshadow that continued fractions can be used to obtain good rational approximations to irrational numbers. Note that the first approximates $\pi$, and the second $e$.

## 5.2 Finite Continued Fractions

This section is about continued fractions of the form $[a_0, a_1, \ldots, a_m]$ for some $m \geq 0$. We give an inductive definition of numbers $p_n$ and $q_n$ such that for all $n \leq m$

$$[a_0, a_1, \ldots, a_n] = \frac{p_n}{q_n}. \tag{5.2.1}$$

We then give related formulas for the determinants of the $2 \times 2$ matrices $\left(\begin{smallmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} p_n & p_{n-2} \\ q_n & q_{n-2} \end{smallmatrix}\right)$, which we will repeatedly use to deduce properties of the sequence of partial convergents $[a_0, \ldots, a_k]$. We will use Algorithm 1.1.13 to prove that every rational number is represented by a continued fraction, as in (5.2.1).

**Definition 5.2.1** (Finite Continued Fraction)**.** A *finite continued fraction* is an expression

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \frac{1}{a_n}}}}$$

where each $a_m$ is a real number and $a_m > 0$ for all $m \geq 1$.

**Definition 5.2.2** (Simple Continued Fraction)**.** A *simple continued fraction* is a finite or infinite continued fraction in which the $a_i$ are all integers.

To get a feeling for continued fractions, observe that

$$\begin{aligned}
[a_0] &= a_0, \\
[a_0, a_1] &= a_0 + \frac{1}{a_1} = \frac{a_0 a_1 + 1}{a_1}, \\
[a_0, a_1, a_2] &= a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_0 a_1 a_2 + a_0 + a_2}{a_1 a_2 + 1}.
\end{aligned}$$

Also,

$$\begin{aligned}
[a_0, a_1, \ldots, a_{n-1}, a_n] &= \left[a_0, a_1, \ldots, a_{n-2}, a_{n-1} + \frac{1}{a_n}\right] \\
&= a_0 + \frac{1}{[a_1, \ldots, a_n]} \\
&= [a_0, [a_1, \ldots, a_n]].
\end{aligned}$$

*SAGE Example* 5.2.3. The `continued_fraction` command computes continued fractions:

```
sage: continued_fraction(17/23)
[0, 1, 2, 1, 5]
sage: continued_fraction(e)
[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1,
 12, 1, 1, 11]
```

Use the optional second argument `bits = n` to determine the precision (in bits) of the input number that is used to compute the continued fraction.

```
sage: continued_fraction(e, bits=20)
[2, 1, 2, 1, 1, 4, 1, 1, 6]
sage: continued_fraction(e, bits=30)
[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1]
```

You can obtain the value of a continued fraction and even do arithmetic with continued fractions:

```
sage: a = continued_fraction(17/23); a
[0, 1, 2, 1, 5]
sage: a.value()
17/23
sage: b = continued_fraction(6/23); b
[0, 3, 1, 5]
sage: a + b
[1]
```

### 5.2.1 Partial Convergents

Fix a finite continued fraction $[a_0, \ldots, a_m]$. We do not assume at this point that the $a_i$ are integers.

**Definition 5.2.4** (Partial convergents)**.** For $0 \leq n \leq m$, the $n$th *convergent* of the continued fraction $[a_0, \ldots, a_m]$ is $[a_0, \ldots, a_n]$. These convergents for $n < m$ are also called *partial convergents*.

For each $n$ with $-2 \leq n \leq m$, define real numbers $p_n$ and $q_n$ as follows:

$$\begin{array}{llll} p_{-2} = 0, & p_{-1} = 1, & p_0 = a_0, & \cdots \quad p_n = a_n p_{n-1} + p_{n-2} \quad \cdots, \\ q_{-2} = 1, & q_{-1} = 0, & q_0 = 1, & \cdots \quad q_n = a_n q_{n-1} + q_{n-2} \quad \cdots. \end{array}$$

**Proposition 5.2.5** (Partial Convergents)**.** *For $n \geq 0$ with $n \leq m$ we have*

$$[a_0, \ldots, a_n] = \frac{p_n}{q_n}.$$

*Proof.* We use induction. The assertion is obvious when $n = 0, 1$. Suppose the proposition is true for all continued fractions of length $n - 1$. Then

$$\begin{aligned} [a_0, \ldots, a_n] &= [a_0, \ldots, a_{n-2}, a_{n-1} + \frac{1}{a_n}] \\ &= \frac{\left(a_{n-1} + \frac{1}{a_n}\right) p_{n-2} + p_{n-3}}{\left(a_{n-1} + \frac{1}{a_n}\right) q_{n-2} + q_{n-3}} \\ &= \frac{(a_{n-1}a_n + 1)p_{n-2} + a_n p_{n-3}}{(a_{n-1}a_n + 1)q_{n-2} + a_n q_{n-3}} \\ &= \frac{a_n(a_{n-1}p_{n-2} + p_{n-3}) + p_{n-2}}{a_n(a_{n-1}q_{n-2} + q_{n-3}) + q_{n-2}} \\ &= \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}} \\ &= \frac{p_n}{q_n}. \end{aligned}$$

□

*SAGE Example* 5.2.6. If `c` is a continued fraction, use `c.convergents()` to compute a list of the partial convergents of `c`.

```
sage: c = continued_fraction(pi,bits=33); c
[3, 7, 15, 1, 292, 2]
sage: c.convergents()
[3, 22/7, 333/106, 355/113, 103993/33102, 208341/66317]
```

As we will see, the convergents of a continued fraction are the best rational approximations to the value of the continued fraction. In the example above, the listed convergents are the best rational approximations of $\pi$ with given denominator size.

**Proposition 5.2.7.** *For $n \geq 0$ with $n \leq m$ we have*

$$p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1} \tag{5.2.2}$$

*and*

$$p_n q_{n-2} - q_n p_{n-2} = (-1)^n a_n. \tag{5.2.3}$$

*Equivalently,*

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}}$$

*and*

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}}.$$

*Proof.* The case for $n = 0$ is obvious from the definitions. Now suppose $n > 0$ and the statement is true for $n - 1$. Then

$$\begin{aligned} p_n q_{n-1} - q_n p_{n-1} &= (a_n p_{n-1} + p_{n-2}) q_{n-1} - (a_n q_{n-1} + q_{n-2}) p_{n-1} \\ &= p_{n-2} q_{n-1} - q_{n-2} p_{n-1} \\ &= -(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) \\ &= -(-1)^{n-2} = (-1)^{n-1}. \end{aligned}$$

This completes the proof of (5.2.2). For (5.2.3), we have

$$\begin{aligned} p_n q_{n-2} - p_{n-2} q_n &= (a_n p_{n-1} + p_{n-2}) q_{n-2} - p_{n-2}(a_n q_{n-1} + q_{n-2}) \\ &= a_n (p_{n-1} q_{n-2} - p_{n-2} q_{n-1}) \\ &= (-1)^n a_n. \end{aligned}$$

□

*Remark* 5.2.8. Expressed in terms of matrices, the proposition asserts that the determinant of $\left(\begin{smallmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{smallmatrix}\right)$ is $(-1)^{n-1}$, and of $\left(\begin{smallmatrix} p_n & p_{n-2} \\ q_n & q_{n-2} \end{smallmatrix}\right)$ is $(-1)^n a_n$.

*SAGE Example* 5.2.9. We use Sage to verify Proposition 5.2.7 for the first few terms of the continued fraction of $\pi$.

```
sage: c = continued_fraction(pi); c
[3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 3]
sage: for n in range(-1, len(c)):
...    print c.pn(n)*c.qn(n-1) - c.qn(n)*c.pn(n-1),
1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1
sage: for n in range(len(c)):
...    print c.pn(n)*c.qn(n-2) - c.qn(n)*c.pn(n-2),
3 -7 15 -1 292 -1 1 -1 2 -1 3 -1 14 -3
```

**Corollary 5.2.10** (Convergents in lowest terms)**.** *If $[a_0, a_1, \ldots, a_m]$ is a simple continued fraction, so each $a_i$ is an integer, then the $p_n$ and $q_n$ are integers and the fraction $p_n/q_n$ is in lowest terms.*

*Proof.* It is clear that the $p_n$ and $q_n$ are integers, from the formula that defines them. If $d$ is a positive divisor of both $p_n$ and $q_n$, then $d \mid (-1)^{n-1}$, so $d = 1$. □

*SAGE Example* 5.2.11. We illustrate Corollary 5.2.10 using Sage.

```
sage: c = continued_fraction([1,2,3,4,5])
sage: c.convergents()
[1, 3/2, 10/7, 43/30, 225/157]
sage: [c.pn(n) for n in range(len(c))]
[1, 3, 10, 43, 225]
sage: [c.qn(n) for n in range(len(c))]
[1, 2, 7, 30, 157]
```

### *5.2.2 The Sequence of Partial Convergents*

Let $[a_0, \ldots, a_m]$ be a continued fraction and for $n \leq m$ let

$$c_n = [a_0, \ldots, a_n] = \frac{p_n}{q_n}$$

denote the $n$th convergent. Recall that by definition of continued fraction, $a_n > 0$ for $n > 0$, which gives the partial convergents of a continued fraction additional structure. For example, the partial convergents of $[2, 1, 2, 1, 1, 4, 1, 1, 6]$ are

$$2, 3, 8/3, 11/4, 19/7, 87/32, 106/39, 193/71, 1264/465.$$

To make the size of these numbers clearer, we approximate them using decimals. We also underline every other number, to illustrate some extra structure.

$$\underline{2}, 3, \underline{2.66667}, 2.75000, \underline{2.71429}, 2.71875, \underline{2.71795}, 2.71831, \underline{2.71828}$$

The underlined numbers are smaller than all of the nonunderlined numbers, and the sequence of underlined numbers is strictly increasing, whereas the nonunderlined numbers strictly decrease.

*SAGE Example* 5.2.12. Figure 5.1 illustrates the above pattern on another continued fraction using Sage.

```
sage: c = continued_fraction([1,1,1,1,1,1,1,1])
sage: v = [(i, c.pn(i)/c.qn(i)) for i in range(len(c))]
sage: P = point(v, rgbcolor=(0,0,1), pointsize=40)
sage: L = line(v, rgbcolor=(0.5,0.5,0.5))
sage: L2 = line([(0,c.value()),(len(c)-1,c.value())], \
...       thickness=0.5, rgbcolor=(0.7,0,0))
sage: (L+L2+P).show(xmin=0,ymin=1)
```

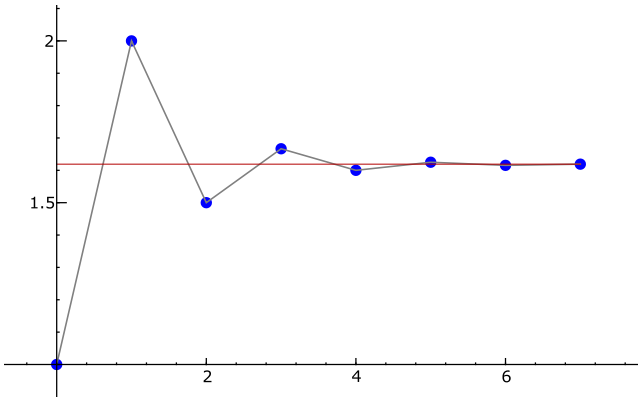



FIGURE 5.1. Graph of a Continued Fraction

We next prove that this extra structure is a general phenomenon.

**Proposition 5.2.13** (How Convergents Converge)**.** *The even indexed convergents $c_{2n}$ increase strictly with $n$, and the odd indexed convergents $c_{2n+1}$ decrease strictly with $n$. Also, the odd indexed convergents $c_{2n+1}$ are greater than all of the even indexed convergents $c_{2m}$.*

*Proof.* The $a_n$ are positive for $n \geq 1$, so the $q_n$ are positive. By Proposition 5.2.7, for $n \geq 2$,

$$c_n - c_{n-2} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}},$$

which proves the first claim.

Suppose for the sake of contradiction that there exist integers $r$ and $m$ such that $c_{2m+1} < c_{2r}$. Proposition 5.2.7 implies that for $n \geq 1$,

$$c_n - c_{n-1} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}}$$

has sign $(-1)^{n-1}$, so for all $s \geq 0$ we have $c_{2s+1} > c_{2s}$. Thus it is impossible that $r = m$. If $r < m$, then by what we proved in the first paragraph, $c_{2m+1} < c_{2r} < c_{2m}$, a contradiction (with $s = m$). If $r > m$, then $c_{2r+1} < c_{2m+1} < c_{2r}$, which is also a contradiction (with $s = r$). □

### *5.2.3 Every Rational Number is Represented*

**Proposition 5.2.14** (Rational Continued Fractions)**.** *Every nonzero rational number can be represented by a simple continued fraction.*

*Proof.* Without loss of generality, we may assume that the rational number is $a/b$, with $b \geq 1$ and $\gcd(a, b) = 1$. Algorithm 1.1.13 gives:

$$\begin{aligned} a &= b \cdot a_0 + r_1, & 0 < r_1 < b \\ b &= r_1 \cdot a_1 + r_2, & 0 < r_2 < r_1 \\ &\cdots \\ r_{n-2} &= r_{n-1} \cdot a_{n-1} + r_n, & 0 < r_n < r_{n-1} \\ r_{n-1} &= r_n \cdot a_n + 0. \end{aligned}$$

Note that $a_i > 0$ for $i > 0$ (also $r_n = 1$, since $\gcd(a, b) = 1$). Rewrite the equations as follows:

$$\begin{aligned} a/b &= a_0 + r_1/b = a_0 + 1/(b/r_1), \\ b/r_1 &= a_1 + r_2/r_1 = a_1 + 1/(r_1/r_2), \\ r_1/r_2 &= a_2 + r_3/r_2 = a_2 + 1/(r_2/r_3), \\ &\cdots \\ r_{n-1}/r_n &= a_n. \end{aligned}$$

It follows that

$$\frac{a}{b} = [a_0, a_1, \ldots, a_n].$$

□

The proof of Proposition 5.2.14 leads to an algorithm for computing the continued fraction of a rational number.

A nonzero rational number can be represented in exactly two ways; for example, $2 = [1, 1] = [2]$ (see Exercise 5.2).

## 5.3 Infinite Continued Fractions

This section begins with the continued fraction procedure, which associates a sequence $a_0, a_1, \ldots$ of integers to a real number $x$. After giving several examples, we prove that $x = \lim_{n\to\infty}[a_0, a_1, \ldots, a_n]$ by proving that the odd and even partial convergents become arbitrarily close to each other. We also show that if $a_0, a_1, \ldots$ is any infinite sequence of positive integers, then the sequence of $c_n = [a_0, a_1, \ldots, a_n]$ converges. More generally, if $a_n$ is an arbitrary sequence of positive reals such that $\sum_{n=0}^{\infty} a_n$ diverges then $(c_n)$ converges.

### *5.3.1 The Continued Fraction Procedure*

Let $x \in \mathbf{R}$ and write

$$x = a_0 + t_0$$

with $a_0 \in \mathbf{Z}$ and $0 \leq t_0 < 1$. We call the number $a_0$ the *floor* of $x$, and we also sometimes write $a_0 = \lfloor x \rfloor$. If $t_0 \neq 0$, write

$$\frac{1}{t_0} = a_1 + t_1$$

with $a_1 \in \mathbf{N}$ and $0 \leq t_1 < 1$. Thus $t_0 = \frac{1}{a_1+t_1} = [0, a_1 + t_1]$, which is a continued fraction expansion of $t_0$, which need not be simple. Continue in this manner so long as $t_n \neq 0$ writing

$$\frac{1}{t_n} = a_{n+1} + t_{n+1}$$

with $a_{n+1} \in \mathbf{N}$ and $0 \leq t_{n+1} < 1$. We call this procedure, which associates to a real number $x$ the sequence of integers $a_0, a_1, a_2, \ldots$, the *continued fraction process*.

*Example* 5.3.1. Let $x = \frac{8}{3}$. Then $x = 2 + \frac{2}{3}$, so $a_0 = 2$ and $t_0 = \frac{2}{3}$. Then $\frac{1}{t_0} = \frac{3}{2} = 1 + \frac{1}{2}$, so $a_1 = 1$ and $t_1 = \frac{1}{2}$. Then $\frac{1}{t_1} = 2$, so $a_2 = 2$, $t_2 = 0$, and the sequence terminates. Notice that

$$\frac{8}{3} = [2, 1, 2],$$

so the continued fraction procedure produces the continued fraction of $\frac{8}{3}$.

*Example* 5.3.2. Let $x = \frac{1+\sqrt{5}}{2}$. Then

$$x = 1 + \frac{-1+\sqrt{5}}{2},$$

so $a_0 = 1$ and $t_0 = \frac{-1+\sqrt{5}}{2}$. We have

$$\frac{1}{t_0} = \frac{2}{-1+\sqrt{5}} = \frac{-2-2\sqrt{5}}{-4} = \frac{1+\sqrt{5}}{2},$$

so $a_1 = 1$ and $t_1 = \frac{-1+\sqrt{5}}{2}$. Likewise, $a_n = 1$ for all $n$. As we will see below, the following exciting equality makes sense.

$$\frac{1+\sqrt{5}}{2} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1+\cdots}}}}}$$

*SAGE Example* 5.3.3. The equality of Example 5.3.2 is consistent with the following Sage calculation:

```
sage: def cf(bits):
...   x = (1 + sqrt(RealField(bits)(5))) / 2
...   return continued_fraction(x)
sage: cf(10)
[1, 1, 1, 1, 1, 1, 1, 3]
sage: cf(30)
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 2]
sage: cf(50)
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

*Example* 5.3.4. Suppose $x = e = 2.71828182\ldots$. Using the continued fraction procedure, we find that

$$a_0, a_1, a_2, \ldots = 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots$$

For example, $a_0 = 2$ is the floor of 2. Subtracting 2 and inverting, we obtain $1/0.718\ldots = 1.3922\ldots$, so $a_1 = 1$. Subtracting 1 and inverting yields $1/0.3922\ldots = 2.5496\ldots$, so $a_2 = 2$. We will prove in Section 5.4 that the continued fraction of $e$ obeys a simple pattern.

The 5th partial convergent of the continued fraction of $e$ is

$$[a_0, a_1, a_2, a_3, a_4, a_5] = \frac{87}{32} = 2.71875,$$

which is a good rational approximation to $e$, in the sense that

$$\left|\frac{87}{32} - e\right| = 0.000468\ldots.$$

Note that $0.000468\ldots < 1/32^2 = 0.000976\ldots$, which illustrates the bound in Corollary 5.3.11.

Let's do the same thing with $\pi = 3.14159265358979\ldots$. Applying the continued fraction procedure, we find that the continued fraction of $\pi$ is

$$a_0, a_1, a_2, \ldots = 3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, \ldots$$

The first few partial convergents are

$$3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \ldots$$

These are good rational approximations to $\pi$; for example,

$$\frac{103993}{33102} = 3.14159265301\ldots.$$

Notice that the continued fraction of $e$ exhibits a nice pattern (see Section 5.4 for a proof), whereas the continued fraction of $\pi$ exhibits no pattern

that is obvious to the author. The continued fraction of $\pi$ has been extensively studied, and over 20 million terms have been computed. The data suggests that every integer appears infinitely often as a partial convergent. For much more about the continued fraction of $\pi$, or of any other sequence in this book, type the first few terms of the sequence into [Slo].

### *5.3.2 Convergence of Infinite Continued Fractions*

**Lemma 5.3.5.** *For every $n$ such that $a_n$ is defined, we have*

$$x = [a_0, a_1, \ldots, a_n + t_n],$$

*and if $t_n \neq 0$, then $x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}]$.*

*Proof.* We use induction. The statements are both true when $n = 0$. If the second statement is true for $n-1$, then

$$\begin{aligned} x &= \left[a_0, a_1, \ldots, a_{n-1}, \frac{1}{t_{n-1}}\right] \\ &= [a_0, a_1, \ldots, a_{n-1}, a_n + t_n] \\ &= \left[a_0, a_1, \ldots, a_{n-1}, a_n, \frac{1}{t_n}\right]. \end{aligned}$$

Similarly, the first statement is true for $n$ if it is true for $n-1$. □

**Theorem 5.3.6** (Continued Fraction Limit)**.** *Let $a_0, a_1, \ldots$ be a sequence of integers such that $a_n > 0$ for all $n \geq 1$, and for each $n \geq 0$, set $c_n = [a_0, a_1, \ldots a_n]$. Then $\lim_{n\to\infty} c_n$ exists.*

*Proof.* For any $m \geq n$, the number $c_n$ is a partial convergent of $[a_0, \ldots, a_m]$. By Proposition 5.2.13, the even convergents $c_{2n}$ form a strictly *increasing* sequence and the odd convergents $c_{2n+1}$ form a strictly *decreasing* sequence. Moreover, the even convergents are all $\leq c_1$ and the odd convergents are all $\geq c_0$. Hence $\alpha_0 = \lim_{n\to\infty} c_{2n}$ and $\alpha_1 = \lim_{n\to\infty} c_{2n+1}$ both exist, and $\alpha_0 \leq \alpha_1$. Finally, by Proposition 5.2.7

$$|c_{2n} - c_{2n-1}| = \frac{1}{q_{2n} \cdot q_{2n-1}} \leq \frac{1}{2n(2n-1)} \to 0,$$

so $\alpha_0 = \alpha_1$. □

We define

$$[a_0, a_1, \ldots] = \lim_{n\to\infty} c_n.$$

*Example* 5.3.7. We illustrate the theorem with $x = \pi$. As in the proof of Theorem 5.3.6, let $c_n$ be the $n$th partial convergent to $\pi$. The $c_n$ with $n$ odd converge down to $\pi$

$$c_1 = 3.1428571\ldots,\ c_3 = 3.1415929\ldots,\ c_5 = 3.1415926\ldots$$

whereas the $c_n$ with $n$ even converge up to $\pi$

$$c_2 = 3.1415094\ldots,\ c_4 = 3.1415926\ldots,\ c_6 = 3.1415926\ldots.$$

**Theorem 5.3.8.** *Let $a_0, a_1, a_2, \ldots$ be a sequence of real numbers such that $a_n > 0$ for all $n \geq 1$, and for each $n \geq 0$, set $c_n = [a_0, a_1, \ldots a_n]$. Then $\lim_{n\to\infty} c_n$ exists if and only if the sum $\sum_{n=0}^{\infty} a_n$ diverges.*

*Proof.* We only prove that if $\sum a_n$ diverges, then $\lim_{n\to\infty} c_n$ exists. A proof of the converse can be found in [Wal48, Ch. 2, Thm. 6.1].

Let $q_n$ be the sequence of "denominators" of the partial convergents, as defined in Section 5.2.1, so $q_{-2} = 1$, $q_{-1} = 0$, and for $n \geq 0$, we have

$$q_n = a_n q_{n-1} + q_{n-2}.$$

As we saw in the proof of Theorem 5.3.6, the limit $\lim_{n\to\infty} c_n$ exists provided that the sequence $\{q_n q_{n-1}\}$ diverges to positive infinity.

For $n$ even,

$$\begin{aligned} q_n &= a_n q_{n-1} + q_{n-2} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + q_{n-4} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + a_{n-4} q_{n-5} + q_{n-6} \\ &= a_n q_{n-1} + a_{n-2} q_{n-3} + \cdots + a_2 q_1 + q_0 \end{aligned}$$

and for $n$ odd,

$$q_n = a_n q_{n-1} + a_{n-2} q_{n-3} + \cdots + a_1 q_0 + q_{-1}.$$

Since $a_n > 0$ for $n > 0$, the sequence $\{q_n\}$ is increasing, so $q_i \geq 1$ for all $i \geq 0$. Applying this fact to the above expressions for $q_n$, we see that for $n$ even

$$q_n \geq a_n + a_{n-2} + \cdots + a_2,$$

and for $n$ odd

$$q_n \geq a_n + a_{n-2} + \cdots + a_1.$$

If $\sum a_n$ diverges, then at least one of $\sum a_{2n}$ or $\sum a_{2n+1}$ must diverge. The above inequalities then imply that at least one of the sequences $\{q_{2n}\}$ or $\{q_{2n+1}\}$ diverge to infinity. Since $\{q_n\}$ is an increasing sequence, it follows that $\{q_n q_{n-1}\}$ diverges to infinity. □

*Example* 5.3.9. Let $a_n = \frac{1}{n\log(n)}$ for $n \geq 2$ and $a_0 = a_1 = 0$. By the integral test, $\sum a_n$ diverges, so by Theorem 5.3.8, the continued fraction $[a_0, a_1, a_2, \ldots]$ converges. This convergence is very slow, since, e.g.

$$[a_0, a_1, \ldots, a_{9999}] = 0.5750039671012225425930\ldots$$

yet

$$[a_0, a_1, \ldots, a_{10000}] = 0.7169153932917378550424\ldots.$$

**Theorem 5.3.10.** *Let $x \in \mathbf{R}$ be a real number. Then $x$ is the value of the (possibly infinite) simple continued fraction $[a_0, a_1, a_2, \ldots]$ produced by the continued fraction procedure.*

*Proof.* If the sequence is finite, then some $t_n = 0$ and the result follows by Lemma 5.3.5. Suppose the sequence is infinite. By Lemma 5.3.5,

$$x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}].$$

By Proposition 5.2.5 (which we apply in a case when the partial quotients of the continued fraction are not integers), we have

$$x = \frac{\frac{1}{t_n} \cdot p_n + p_{n-1}}{\frac{1}{t_n} \cdot q_n + q_{n-1}}.$$

Thus, if $c_n = [a_0, a_1, \ldots, a_n]$, then

$$\begin{aligned} x - c_n &= x - \frac{p_n}{q_n} \\ &= \frac{\frac{1}{t_n}p_nq_n + p_{n-1}q_n - \frac{1}{t_n}p_nq_n - p_nq_{n-1}}{q_n\left(\frac{1}{t_n}q_n + q_{n-1}\right)}. \\ &= \frac{p_{n-1}q_n - p_nq_{n-1}}{q_n\left(\frac{1}{t_n}q_n + q_{n-1}\right)} \\ &= \frac{(-1)^n}{q_n\left(\frac{1}{t_n}q_n + q_{n-1}\right)}. \end{aligned}$$

Thus

$$\begin{aligned} |x - c_n| &= \frac{1}{q_n\left(\frac{1}{t_n}q_n + q_{n-1}\right)} \\ &< \frac{1}{q_n(a_{n+1}q_n + q_{n-1})} \\ &= \frac{1}{q_n \cdot q_{n+1}} \leq \frac{1}{n(n+1)} \to 0. \end{aligned}$$

In the inequality, we use that $a_{n+1}$ is the integer part of $\frac{1}{t_n}$, and is hence $\leq \frac{1}{t_n} < 1$, since $t_n < 1$. □

This corollary follows from the proof of Theorem 5.3.10.

**Corollary 5.3.11** (Convergence of continued fraction)**.** *Let $a_0, a_1, \ldots$ define a simple continued fraction, and let $x = [a_0, a_1, \ldots] \in \mathbf{R}$ be its value. Then for all $m$,*

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}}.$$

**Proposition 5.3.12.** *If $x$ is a rational number, then the sequence $a_0, a_1, \ldots$ produced by the continued fraction procedure terminates.*

*Proof.* Let $[b_0, b_1, \ldots, b_m]$ be the continued fraction representation of $x$ that we obtain using Algorithm 1.1.13, so the $b_i$ are the partial quotients at each step. If $m = 0$, then $x$ is an integer, so we may assume $m > 0$. Then

$$x = b_0 + 1/[b_1, \ldots, b_m].$$

If $[b_1, \ldots, b_m] = 1$, then $m = 1$ and $b_1 = 1$, which will not happen using Algorithm 1.1.13, since it would give $[b_0{+}1]$ for the continued fraction of the integer $b_0 + 1$. Thus $[b_1, \ldots, b_m] > 1$, so in the continued fraction algorithm we choose $a_0 = b_0$ and $t_0 = 1/[b_1, \ldots, b_m]$. Repeating this argument enough times proves the claim. □

## 5.4 The Continued Fraction of $e$

The continued fraction expansion of $e$ begins $[2, 1, 2, 1, 1, 4, 1, 1, 6, \ldots]$. The obvious pattern in fact does continue, as Euler proved in 1737 (see [Eul85]), and we will prove in this section. As an application, Euler gave a proof that $e$ is irrational by noting that its continued fraction is infinite.

The proof we give below draws heavily on the proof in [Coh], which describes a slight variant of a proof of Hermite (see [Old70]). The continued fraction representation of $e$ is also treated in the German book [Per57], but the proof requires substantial background from elsewhere in that text.

### *5.4.1 Preliminaries*

First, we write the continued fraction of $e$ in a slightly different form. Instead of $[2, 1, 2, 1, 1, 4, \ldots]$, we can start the sequence of coefficients

$$[1, 0, 1, 1, 2, 1, 1, 4, \ldots]$$

to make the pattern the same throughout. (Everywhere else in this chapter we assume that the partial quotients $a_n$ for $n \geq 1$ are positive, but

temporarily relax that condition here and allow $a_1 = 0$.) The numerators and denominators of the convergents given by this new sequence satisfy a simple recurrence. Using $r_i$ as a stand-in for $p_i$ or $q_i$, we have

$$\begin{aligned} r_{3n} &= r_{3n-1} + r_{3n-2} \\ r_{3n-1} &= r_{3n-2} + r_{3n-3} \\ r_{3n-2} &= 2(n-1)r_{3n-3} + r_{3n-4}. \end{aligned}$$

Our first goal is to collapse these three recurrences into one recurrence that only makes mention of $r_{3n}$, $r_{3n-3}$, and $r_{3n-6}$. We have

$$\begin{aligned} r_{3n} &= r_{3n-1} + r_{3n-2} \\ &= (r_{3n-2} + r_{3n-3}) + (2(n-1)r_{3n-3} + r_{3n-4}) \\ &= (4n-3)r_{3n-3} + 2r_{3n-4}. \end{aligned}$$

This same method of simplification also shows us that

$$r_{3n-3} = 2r_{3n-7} + (4n-7)r_{3n-6}.$$

To get rid of $2r_{3n-4}$ in the first equation, we make the substitutions

$$\begin{aligned} 2r_{3n-4} &= 2(r_{3n-5} + r_{3n-6}) \\ &= 2((2(n-2)r_{3n-6} + r_{3n-7}) + r_{3n-6}) \\ &= (4n-6)r_{3n-6} + 2r_{3n-7}. \end{aligned}$$

Substituting for $2r_{3n-4}$ and then $2r_{3n-7}$, we finally have the needed collapsed recurrence,

$$r_{3n} = 2(2n-1)r_{3n-3} + r_{3n-6}.$$

### *5.4.2 Two Integral Sequences*

We define the sequences $x_n = p_{3n}$, $y_n = q_{3n}$. Since the $3n$-convergents will converge to the same real number that the $n$ convergents do, $x_n/y_n$ also converges to the limit of the continued fraction. Each sequence $\{x_n\}$, $\{y_n\}$ will obey the recurrence relation derived in the previous section (where $z_n$ is a stand-in for $x_n$ or $y_n$):

$$z_n = 2(2n-1)z_{n-1} + z_{n-2}, \text{ for all } n \geq 2. \tag{5.4.1}$$

The two sequences can be found in Table 5.1. (The initial conditions $x_0 = 1$, $x_1 = 3$, $y_0 = y_1 = 1$ are taken straight from the first few convergents of the original continued fraction.) Notice that since we are skipping several convergents at each step, the ratio $x_n/y_n$ converges to $e$ very quickly.

TABLE 5.1. Convergents

| $n$ | 0 | 1 | 2 | 3 | 4 | $\cdots$ |
|---|---|---|---|---|---|---|
| $x_n$ | 1 | 3 | 19 | 193 | 2721 | $\cdots$ |
| $y_n$ | 1 | 1 | 7 | 71 | 1001 | $\cdots$ |
| $x_n/y_n$ | 1 | 3 | $2.714\ldots$ | $2.71830\ldots$ | $2.7182817\ldots$ | $\cdots$ |

### *5.4.3 A Related Sequence of Integrals*

Now, we define a sequence of real numbers $T_0, T_1, T_2, \ldots$ by the following integrals:

$$T_n = \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt.$$

Below, we compute the first two terms of this sequence explicitly. (When we compute $T_1$, we are doing the integration by parts $u = t(t-1)$, $dv = e^t dt$. Since the integral runs from 0 to 1, the boundary condition is 0 when evaluated at each of the endpoints. This vanishing will be helpful when we do the integral in the general case.)

$$\begin{aligned}
T_0 &= \int_0^1 e^t dt = e - 1,\\
T_1 &= \int_0^1 t(t-1)e^t dt\\
&= -\int_0^1 ((t-1)+t)e^t dt\\
&= -(t-1)e^t\Big|_0^1 - te^t\Big|_0^1 + 2\int_0^1 e^t dt\\
&= -1 - e + 2(e-1) = e - 3.
\end{aligned}$$

The reason that we defined this series now becomes apparent: $T_0 = y_0 e - x_0$ and $T_1 = y_1 e - x_1$. In general, it will be true that $T_n = y_n e - x_n$. We will now prove this fact.

It is clear that if $T_n$ were to satisfy the same recurrence that the $x_i$ and $y_i$ do in (5.4.1), then the above statement holds by induction. (The initial conditions are correct, as needed.) So, we simplify $T_n$ by integrating by

parts twice in succession:

$$\begin{aligned}
T_n &= \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt \\
&= -\int_0^1 \frac{t^{n-1}(t-1)^n + t^n(t-1)^{n-1}}{(n-1)!}\ e^t dt \\
&= \int_0^1 \Big( \frac{t^{n-2}(t-1)^n}{(n-2)!} + n\frac{t^{n-1}(t-1)^{n-1}}{(n-1)!} \\
&\qquad\qquad + n\frac{t^{n-1}(t-1)^{n-1}}{(n-1)!} + \frac{t^n(t-1)^{n-2}}{(n-2)!} \Big) e^t dt \\
&= 2nT_{n-1} + \int_0^1 \frac{t^{n-2}(t-1)^{n-2}}{(n-2)!}(2t^2-2t+1)\ e^t dt \\
&= 2nT_{n-1} + 2\int_0^1 \frac{t^{n-1}(t-1)^{n-1}}{(n-2)!}\ e^t dt + \int_0^1 \frac{t^{n-2}(t-1)^{n-2}}{(n-2)!}\ e^t dt \\
&= 2nT_{n-1} + 2(n-1)T_{n-1} + T_{n-2} \\
&= 2(2n-1)T_{n-1} + T_{n-2},
\end{aligned}$$

which is the desired recurrence.

Therefore, $T_n = y_n e - x_n$. To conclude the proof, we consider the limit as $n$ approaches infinity:

$$\lim_{n\to\infty} \int_0^1 \frac{t^n(t-1)^n}{n!}\ e^t dt = 0,$$

by inspection, and therefore

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \lim_{n\to\infty} (e - \frac{T_n}{y_n}) = e.$$

Therefore, the ratio $x_n/y_n$ approaches $e$, and the continued fraction expansion $[2, 1, 2, 1, 1, 4, 1, 1, \ldots]$ does in fact converge to $e$.

### *5.4.4 Extensions of the Argument*

The method of proof of this section generalizes to show that the continued fraction expansion of $e^{1/n}$ is

$$[1,\ (n-1),\ 1,\ 1,\ (3n-1),\ 1,\ 1,\ (5n-1),\ 1,\ 1,\ (7n-1), \ldots]$$

for all $n \in \mathbf{N}$ (see Exercise 5.6).

## 5.5 Quadratic Irrationals

The main result of this section is that the continued fraction expansion of a number is eventually repeating if and only if the number is a quadratic

irrational. This can be viewed as an analog for continued fractions of the familiar fact that the decimal expansion of $x$ is eventually repeating if and only if $x$ is rational. The proof that continued fractions of quadratic irrationals eventually repeats is surprisingly difficult and involves an interesting finiteness argument. Section 5.5.2 emphasizes our striking ignorance about continued fractions of real roots of irreducible polynomials over $\mathbf{Q}$ of degree bigger than 2.

**Definition 5.5.1** (Quadratic Irrational)**.** A *quadratic irrational* is a real number $\alpha \in \mathbf{R}$ that is irrational and satisfies a quadratic polynomial with coefficients in $\mathbf{Q}$.

Thus, for example, $(1+\sqrt{5})/2$ is a quadratic irrational. Recall that

$$\frac{1+\sqrt{5}}{2} = [1, 1, 1, \ldots].$$

The continued fraction of $\sqrt{2}$ is $[1, 2, 2, 2, 2, 2, \ldots]$, and the continued fraction of $\sqrt{389}$ is

$$[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38, \ldots].$$

Does the $[1, 2, 1, 1, 1, 1, 2, 1, 38]$ pattern repeat over and over again?

*SAGE Example* 5.5.2. We compute more terms of the continued fraction expansion of $\sqrt{389}$ using Sage:

```
sage: def cf_sqrt_d(d, bits):
...   x = sqrt(RealField(bits)(d))
...   return continued_fraction(x)
sage: cf_sqrt_d(389,50)
[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38]
sage: cf_sqrt_d(389,100)
[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38,
 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1,
 2, 1, 1]
```

### *5.5.1 Periodic Continued Fractions*

**Definition 5.5.3** (Periodic Continued Fraction)**.** A *periodic continued fraction* is a continued fraction $[a_0, a_1, \ldots, a_n, \ldots]$ such that

$$a_n = a_{n+h}$$

for some fixed positive integer $h$ and all sufficiently large $n$. We call the minimal such $h$ the *period of the continued fraction.*

*Example* 5.5.4. Consider the periodic continued fraction $[1,2,1,2,\ldots] = [\overline{1,2}]$. What does it converge to? We have

$$[\overline{1,2}] = 1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1+\cdots}}}},$$

so if $\alpha = [\overline{1,2}]$ then

$$\alpha = 1 + \cfrac{1}{2 + \cfrac{1}{\alpha}} = 1 + \cfrac{1}{\cfrac{2\alpha+1}{\alpha}} = 1 + \frac{\alpha}{2\alpha+1} = \frac{3\alpha+1}{2\alpha+1}$$

Thus $2\alpha^2 - 2\alpha - 1 = 0$, so

$$\alpha = \frac{1+\sqrt{3}}{2}.$$

**Theorem 5.5.5** (Periodic Characterization)**.** *An infinite simple continued fraction is periodic if and only if it represents a quadratic irrational.*

*Proof.* ($\Longrightarrow$) First suppose that

$$[a_0, a_1, \ldots, a_n, \overline{a_{n+1}, \ldots, a_{n+h}}]$$

is a periodic continued fraction. Set $\alpha = [a_{n+1}, a_{n+2}, \ldots]$. Then

$$\alpha = [a_{n+1}, \ldots, a_{n+h}, \alpha],$$

so by Proposition 5.2.5

$$\alpha = \frac{\alpha p_{n+h} + p_{n+h-1}}{\alpha q_{n+h} + q_{n+h-1}}.$$

Here we use that $\alpha$ is the last partial quotient. Thus, $\alpha$ satisfies a quadratic equation with coefficients in $\mathbf{Q}$. Computing as in Example 5.5.4 and rationalizing the denominators, and using that the $a_i$ are all integers, shows that

$$\begin{aligned}[a_0, a_1, \ldots] &= [a_0, a_1, \ldots, a_n, \alpha] \\ &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{\alpha}}}\end{aligned}$$

is of the form $c + d\alpha$, with $c, d \in \mathbf{Q}$, so $[a_0, a_1, \ldots]$ also satisfies a quadratic polynomial over $\mathbf{Q}$.

The continued fraction procedure applied to the value of an infinite simple continued fraction yields that continued fraction back, so by Proposition 5.3.12, $\alpha \notin \mathbf{Q}$ because it is the value of an infinite continued fraction.

($\Longleftarrow$) Suppose $\alpha \in \mathbf{R}$ is an irrational number that satisfies a quadratic equation

$$a\alpha^2 + b\alpha + c = 0 \tag{5.5.1}$$

with $a, b, c \in \mathbf{Z}$ and $a \neq 0$. Let $[a_0, a_1, \ldots]$ be the continued fraction expansion of $\alpha$. For each $n$, let

$$r_n = [a_n, a_{n+1}, \ldots],$$

so

$$\alpha = [a_0, a_1, \ldots, a_{n-1}, r_n].$$

We will prove periodicity by showing that the set of $r_n$'s is finite. If we have shown finiteness, then there exists $n, h > 0$ such that $r_n = r_{n+h}$, so

$$\begin{aligned}
[a_0, \ldots, a_{n-1}, r_n] &= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, r_{n+h}] \\
&= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, r_n] \\
&= [a_0, \ldots, a_{n-1}, a_n, \ldots, a_{n+h-1}, a_n, \ldots, a_{n+h-1}, r_{n+h}] \\
&= [a_0, \ldots, a_{n-1}, \overline{a_n, \ldots, a_{n+h-1}}].
\end{aligned}$$

It remains to show there are only finitely many distinct $r_n$. We have

$$\alpha = \frac{p_n}{q_n} = \frac{r_n p_{n-1} + p_{n-2}}{r_n q_{n-1} + q_{n-2}}.$$

Substituting this expression for $\alpha$ into the quadratic equation (5.5.1), we see that

$$A_n r_n^2 + B_n r_n + C_n = 0,$$

where

$$\begin{aligned}
A_n &= a p_{n-1}^2 + b p_{n-1} q_{n-1} + c q_{n-1}^2, \\
B_n &= 2a p_{n-1} p_{n-2} + b(p_{n-1} q_{n-2} + p_{n-2} q_{n-1}) + 2c q_{n-1} q_{n-2}, \text{ and} \\
C_n &= a p_{n-2}^2 + b p_{n-2} q_{n-2} + c q_{n-2}^2.
\end{aligned}$$

Note that $A_n, B_n, C_n \in \mathbf{Z}$, that $C_n = A_{n-1}$, and that

$$B_n^2 - 4A_n C_n = (b^2 - 4ac)(p_{n-1} q_{n-2} - q_{n-1} p_{n-2})^2 = b^2 - 4ac.$$

Recall from the proof of Theorem 5.3.10 that

$$\left| \alpha - \frac{p_{n-1}}{q_{n-1}} \right| < \frac{1}{q_n q_{n-1}}.$$

Thus,

$$|\alpha q_{n-1} - p_{n-1}| < \frac{1}{q_n} < \frac{1}{q_{n-1}},$$

so

$$p_{n-1} = \alpha q_{n-1} + \frac{\delta}{q_{n-1}} \qquad \text{with } |\delta| < 1.$$

Hence,

$$\begin{aligned} A_n &= a\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right)^2 + b\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right) q_{n-1} + c q_{n-1}^2 \\ &= (a\alpha^2 + b\alpha + c) q_{n-1}^2 + 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta \\ &= 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta. \end{aligned}$$

Thus,

$$|A_n| = \left|2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta\right| < 2|a\alpha| + |a| + |b|.$$

We conclude that there are only finitely many possibilities for the integer $A_n$. Also,

$$|C_n| = |A_{n-1}| \quad \text{and} \quad |B_n| = \sqrt{b^2 - 4(ac - A_n C_n)},$$

so there are only finitely many triples $(A_n, B_n, C_n)$, and hence only finitely many possibilities for $r_n$ as $n$ varies, which completes the proof. (The proof above closely follows [HW79, Thm. 177, pg.144–145].) □

### *5.5.2 Continued Fractions of Algebraic Numbers of Higher Degree*

**Definition 5.5.6** (Algebraic Number)**.** An *algebraic number* is a root of a polynomial $f \in \mathbf{Q}[x]$.

**Open Problem 5.5.7.** *Give a simple description of the complete continued fractions expansion of the algebraic number $\sqrt[3]{2}$. It begins*

$$[1, 3, 1, 5, 1, 1, 4, 1, 1, 8, 1, 14, 1, 10, 2, 1, 4, 12, 2, 3, 2, 1, 3, 4, 1, 1, 2, 14,$$
$$3, 12, 1, 15, 3, 1, 4, 534, 1, 1, 5, 1, 1, \ldots]$$

The author does not see a pattern, and the 534 reduces his confidence that he will. Lang and Trotter (see [LT72]) analyzed many terms of the continued fraction of $\sqrt[3]{2}$ statistically, and their work suggests that $\sqrt[3]{2}$ has an "unusual" continued fraction; later work in [LT74] suggests that maybe it does not.

**Khintchine (see [Khi63, pg. 59])**

> No properties of the representing continued fractions, analogous to those which have just been proved, are known for algebraic numbers of higher degree [as of 1963]. [...] It is of interest to point out that up till the present time *no continued fraction development of an algebraic number of higher degree than the second is known* [emphasis added]. It is not even known if such a development has bounded elements. Generally speaking the problems associated with the continued fraction expansion of algebraic numbers of degree higher than the second are extremely difficult and virtually unstudied.

**Richard Guy (see [Guy94, pg. 260])**

> Is there an algebraic number of degree greater than two whose simple continued fraction has unbounded partial quotients? Does every such number have unbounded partial quotients?

Baum and Sweet [BS76] answered the analog of Richard Guy's question, but with algebraic numbers replaced by elements of a field $K$ other than $\mathbf{Q}$. (The field $K$ is $\mathbf{F}_2((1/x))$, the field of Laurent series in the variable $1/x$ over the finite field with two elements. An element of $K$ is a polynomial in $x$ plus a formal power series in $1/x$.) They found an $\alpha$ of degree 3 over $K$ whose continued fraction has all terms of bounded degree, and other elements of various degrees greater than 2 over $K$ whose continued fractions have terms of unbounded degree.

## 5.6 Recognizing Rational Numbers

Suppose that somehow you can compute approximations to some rational number, and want to figure what the rational number probably is. Computing the approximation to high enough precision to find a period in the decimal expansion is not a good approach, because the period can be huge (see below). A much better approach is to compute the simple continued fraction of the approximation, and truncate it before a large partial quotient $a_n$, then compute the value of the truncated continued fraction. This results in a rational number that has a relatively small numerator and denominator, and is close to the approximation of the rational number, since the tail end of the continued fraction is at most $1/a_n$.

We begin with a contrived example, which illustrates how to recognize a rational number. Let

$$x = 9495/3847 = 2.468157005458799064205874707564335846113 85\ldots.$$

The continued fraction of the truncation 2.468157005458799064 is

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 328210621945, 2, 1, 1, 1, \ldots]$$

We have

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1] = \frac{9495}{3847}.$$

Notice that no repetition is evident in the digits of $x$ given above, though we know that the decimal expansion of $x$ must be eventually periodic, since all decimal expansions of rational numbers are eventually periodic. In fact, the length of the period of the decimal expansion of $1/3847$ is 3846, which is the order of 10 modulo 3847 (see Exercise 5.7).

For a slightly less contrived application of this idea, suppose $f(x) \in \mathbf{Z}[x]$ is a polynomial with integer coefficients, and we know for some reason that one root of $f$ is a rational number. We can find that rational number, by using Newton's method to approximate each root, and continued fractions to decide whether each root is a rational number (we can substitute the value of the continued fraction approximation into $f$ to see if it is actually a root). One could also use the well-known Rational Root Theorem, which asserts that any rational root $n/d$ of $f$, with $n, d \in \mathbf{Z}$ coprime, has the property that $n$ divides the constant term of $f$ and $d$ the leading coefficient of $f$. However, using that theorem to find $n/d$ would require factoring the constant and leading terms of $f$, which could be completely impractical if they have a few hundred digits (see Section 1.1.3). In contrast, Newton's method and continued fractions should quickly find $n/d$, assuming the degree of $f$ isn't too large.

For example, suppose $f = 3847x^2 - 14808904x + 36527265$. To apply Newton's method, let $x_0$ be a guess for a root of $f$. Iterate using the recurrence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Choosing $x_0 = 0$, approximations of the first two iterates are

$$x_1 = 2.466574501394566404103909378,$$

and

$$x_2 = 2.468157004807401923043166846.$$

The continued fraction of the approximations $x_1$ and $x_2$ are

$$[2, 2, 6, 1, 47, 2, 1, 4, 3, 1, 5, 8, 2, 3]$$

and

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 103, 8, 1, 2, 3, \ldots].$$

Truncating the continued fraction of $x_2$ before 103 gives

$$[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1],$$

which evaluates to 9495/3847, which is a rational root of $f$.

*SAGE Example* 5.6.1. We do the above calculation using SAGE. First we implement the Newton iteration:

```
sage: def newton_root(f, iterates=2, x0=0, prec=53):
...     x = RealField(prec)(x0)
...     R = PolynomialRing(ZZ,'x')
...     f = R(f)
...     g = f.derivative()
...     for i in range(iterates):
...         x = x - f(x)/g(x)
...     return x
```

Next we run the Newton iteration, and compute the continued fraction of the result:

```
sage: a = newton_root(3847*x^2 - 14808904*x + 36527265); a
2.46815700480740
sage: cf = continued_fraction(a); cf
[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 103, 8, 1, 2, 3, 1, 1]
```

We truncate the continued fraction and compute its value.

```
sage: c = cf[:12]; c
[2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1]
sage: c.value()
9495/3847
```

Another computational application of continued fractions, which we can only hint at, is that there are functions in certain parts of advanced number theory (that are beyond the scope of this book) that take rational values at certain points, and which can only be computed efficiently via approximations; using continued fractions as illustrated above to evaluate such functions is crucial.

## 5.7 Sums of Two Squares

In this section, we apply continued fractions to prove the following theorem.

**Theorem 5.7.1.** *A positive integer $n$ is a sum of two squares if and only if all prime factors of $p \mid n$ such that $p \equiv 3 \pmod 4$ have even exponent in the prime factorization of $n$.*

We first consider some examples. Notice that $5 = 1^2 + 2^2$ is a sum of two squares, but 7 is not a sum of two squares. Since 2001 is divisible by 3 (because $2 + 1$ is divisible by 3), but not by 9 (since $2 + 1$ is not), Theorem 5.7.1 implies that 2001 is not a sum of two squares. The theorem also implies that $2 \cdot 3^4 \cdot 5 \cdot 7^2 \cdot 13$ is a sum of two squares.

*SAGE Example* 5.7.2. We use Sage to write a short program that *naively* determines whether or not an integer $n$ is a sum of two squares, and if so returns $a, b$ such that $a^2 + b^2 = n$.

```
sage: def sum_of_two_squares_naive(n):
...    for i in range(int(sqrt(n))):
...        if is_square(n - i^2):
...            return i, (Integer(n-i^2)).sqrt()
...    return "%s is not a sum of two squares"%n
```

We next use our function in a couple of cases.

```
sage: sum_of_two_squares_naive(23)
'23 is not a sum of two squares'
sage: sum_of_two_squares_naive(389)
(10, 17)
sage: sum_of_two_squares_naive(2007)
'2007 is not a sum of two squares'
sage: sum_of_two_squares_naive(2008)
'2008 is not a sum of two squares'
sage: sum_of_two_squares_naive(2009)
(28, 35)
sage: 28^2 + 35^2
2009
sage: sum_of_two_squares_naive(2*3^4*5*7^2*13)
(189, 693)
```

**Definition 5.7.3** (Primitive)**.** A representation $n = x^2 + y^2$ is *primitive* if $x$ and $y$ are coprime.

**Lemma 5.7.4.** *If $n$ is divisible by a prime $p \equiv 3 \pmod 4$, then $n$ has no primitive representations.*

*Proof.* Suppose $n$ has a primitive representation, $n = x^2 + y^2$, and let $p$ be any prime factor of $n$. Then

$$p \mid x^2 + y^2 \quad \text{and} \quad \gcd(x, y) = 1,$$

so $p \nmid x$ and $p \nmid y$. Since $\mathbf{Z}/p\mathbf{Z}$ is a field, we may divide by $y^2$ in the equation $x^2 + y^2 \equiv 0 \pmod p$ to see that $(x/y)^2 \equiv -1 \pmod p$. Thus the Legendre symbol $\left(\frac{-1}{p}\right)$ equals $+1$. However, by Proposition 4.2.1,

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$$

so $\left(\frac{-1}{p}\right) = 1$ if and only if $(p-1)/2$ is even, which is to say $p \equiv 1 \pmod 4$. □

*Proof of Theorem 5.7.1* ($\Longrightarrow$). Suppose that $p \equiv 3 \pmod 4$ is a prime, that $p^r \mid n$ but $p^{r+1} \nmid n$ with $r$ odd, and that $n = x^2 + y^2$. Letting $d = \gcd(x, y)$, we have

$$x = dx', \quad y = dy', \quad \text{and} \quad n = d^2 n'$$

with $\gcd(x', y') = 1$ and

$$(x')^2 + (y')^2 = n'.$$

Because $r$ is odd, $p \mid n'$, so Lemma 5.7.4 implies that $\gcd(x', y') > 1$, which is a contradiction. □

To prepare for our proof of the implication ($\Longleftarrow$) of Theorem 5.7.1, we reduce the problem to the case when $n$ is prime. Write $n = n_1^2 n_2$, where $n_2$ has no prime factors $p \equiv 3 \pmod 4$. It suffices to show that $n_2$ is a sum of two squares, since

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1 x_2 - y_1 y_2)^2 + (x_1 y_2 + x_2 y_1)^2, \tag{5.7.1}$$

so a product of two numbers that are sums of two squares is also a sum of two squares. Since $2 = 1^2 + 1^2$ is a sum of two squares, it suffices to show that any prime $p \equiv 1 \pmod 4$ is a sum of two squares.

**Lemma 5.7.5.** *If $x \in \mathbf{R}$ and $n \in \mathbf{N}$, then there is a fraction $\dfrac{a}{b}$ in lowest terms such that $0 < b \leq n$ and*

$$\left| x - \frac{a}{b} \right| \leq \frac{1}{b(n+1)}.$$

*Proof.* Consider the continued fraction $[a_0, a_1, \ldots]$ of $x$. By Corollary 5.3.11, for each $m$

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}}.$$

Since $q_{m+1} \geq q_m + 1$ and $q_0 = 1$, either there exists an $m$ such that $q_m \leq n < q_{m+1}$, or the continued fraction expansion of $x$ is finite and $n$ is larger than the denominator of the rational number $x$, in which case we take $\frac{a}{b} = x$ and are done. In the first case,

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}} \leq \frac{1}{q_m \cdot (n+1)},$$

so $\dfrac{a}{b} = \dfrac{p_m}{q_m}$ satisfies the conclusion of the lemma. □

*Proof of Theorem 5.7.1* ($\Longleftarrow$). As discussed above, it suffices to prove that any prime $p \equiv 1 \pmod 4$ is a sum of two squares. Since $p \equiv 1 \pmod 4$,

$$(-1)^{(p-1)/2} = 1,$$

Proposition 4.2.1 implies that $-1$ is a square modulo $p$; i.e., there exists $r \in \mathbf{Z}$ such that $r^2 \equiv -1 \pmod{p}$. Lemma 5.7.5, with $n = \lfloor \sqrt{p} \rfloor$ and $x = -\frac{r}{p}$, implies that there are integers $a, b$ such that $0 < b < \sqrt{p}$ and

$$\left| -\frac{r}{p} - \frac{a}{b} \right| \leq \frac{1}{b(n+1)} < \frac{1}{b\sqrt{p}}.$$

Letting $c = rb + pa$, we have that

$$|c| < \frac{pb}{b\sqrt{p}} = \frac{p}{\sqrt{p}} = \sqrt{p}$$

so

$$0 < b^2 + c^2 < 2p.$$

But $c \equiv rb \pmod{p}$, so

$$b^2 + c^2 \equiv b^2 + r^2 b^2 \equiv b^2(1 + r^2) \equiv 0 \pmod{p}.$$

Thus $b^2 + c^2 = p$. □

*Remark* 5.7.6. Our proof of Theorem 5.7.1 leads to an efficient algorithm to compute a representation of any $p \equiv 1 \pmod{4}$ as a sum of two squares.

*SAGE Example* 5.7.7. We next use Sage and Theorem 5.7.1 to give an efficient algorithm for writing a prime $p \equiv 1 \pmod{4}$ as a sum of two squares. First we implement the algorithm that comes out of the proof of the theorem.

```
sage: def sum_of_two_squares(p):
...     p = Integer(p)
...     assert p%4 == 1, "p must be 1 modulo 4"
...     r = Mod(-1,p).sqrt().lift()
...     v = continued_fraction(-r/p)
...     n = floor(sqrt(p))
...     for x in v.convergents():
...         c = r*x.denominator() + p*x.numerator()
...         if -n <= c and c <= n:
...             return (abs(x.denominator()),abs(c))
```

Next we use the algorithm to write the first 10-digit prime $\equiv 1 \pmod{4}$ as a sum of two squares:

```
sage: p = next_prime(next_prime(10^10))
sage: sum_of_two_squares(p)
(55913, 82908)
```

The above calculation was essentially instantanoues. If instead we use the naive algorithm from before, it takes several seconds to write $p$ as a sum of two squares.

```
sage: sum_of_two_squares_naive(p)
(55913, 82908)
```

## 5.8 Exercises

5.1 If $c_n = p_n/q_n$ is the $n$th convergent of $[a_0, a_1, \ldots, a_n]$ and $a_0 > 0$, show that

$$[a_n, a_{n-1}, \ldots, a_1, a_0] = \frac{p_n}{p_{n-1}}$$

and

$$[a_n, a_{n-1}, \ldots, a_2, a_1] = \frac{q_n}{q_{n-1}}.$$

(Hint: In the first case, notice that $\dfrac{p_n}{p_{n-1}} = a_n + \dfrac{p_{n-2}}{p_{n-1}} = a_n + \dfrac{1}{\frac{p_{n-1}}{p_{n-2}}}$.)

5.2 Show that every nonzero rational number can be represented in exactly two ways by a finite simple continued fraction. (For example, 2 can be represented by $[1, 1]$ and $[2]$, and $1/3$ by $[0, 3]$ and $[0, 2, 1]$.)

5.3 Evaluate the infinite continued fraction $[2, \overline{1, 2, 1}]$.

5.4 Determine the infinite continued fraction of $\frac{1+\sqrt{13}}{2}$.

5.5 Let $a_0 \in \mathbf{R}$ and $a_1, \ldots, a_n$ and $b$ be positive real numbers. Prove that

$$[a_0, a_1, \ldots, a_n + b] < [a_0, a_1, \ldots, a_n]$$

if and only if $n$ is odd.

5.6 (*) Extend the method presented in the text to show that the continued fraction expansion of $e^{1/k}$ is

$$[1,\ (k-1),\ 1,\ 1,\ (3k-1),\ 1,\ 1,\ (5k-1),\ 1,\ 1,\ (7k-1), \ldots]$$

for all $k \in \mathbf{N}$.

(a) Compute $p_0$, $p_3$, $q_0$, and $q_3$ for the above continued fraction. Your answers should be in terms of $k$.

(b) Condense three steps of the recurrence for the numerators and denominators of the above continued fraction. That is, produce a simple recurrence for $r_{3n}$ in terms of $r_{3n-3}$ and $r_{3n-6}$ whose coefficients are polynomials in $n$ and $k$.

(c) Define a sequence of real numbers by

$$T_n(k) = \frac{1}{k^n} \int_0^{1/k} \frac{(kt)^n (kt-1)^n}{n!}\ e^t dt.$$

i. Compute $T_0(k)$, and verify that it equals $q_0 e^{1/k} - p_0$.

ii. Compute $T_1(k)$, and verify that it equals $q_3 e^{1/k} - p_3$.

iii. Integrate $T_n(k)$ by parts twice in succession, as in Section 5.4, and verify that $T_n(k)$, $T_{n-1}(k)$, and $T_{n-2}(k)$ satisfy the recurrence produced in part 6b, for $n \geq 2$.

(d) Conclude that the continued fraction

$$[1, (k-1), 1, 1, (3k-1), 1, 1, (5k-1), 1, 1, (7k-1), \ldots]$$

represents $e^{1/k}$.

5.7 Let $d$ be an integer that is coprime to 10. Prove that the decimal expansion of $\frac{1}{d}$ has a period equal to the order of 10 modulo $d$. (Hint: For every positive integer $r$, we have $\frac{1}{1-10^r} = \sum_{n\geq 1} 10^{-rn}$.)

5.8 Find a positive integer that has at least three different representations as the sum of two squares, disregarding signs and the order of the summands.

5.9 Show that if a natural number $n$ is the sum of two rational squares it is also the sum of two integer squares.

5.10 (*) Let $p$ be an odd prime. Show that $p \equiv 1, 3 \pmod 8$ if and only if $p$ can be written as $p = x^2 + 2y^2$ for some choice of integers $x$ and $y$.

5.11 Prove that of any four consecutive integers, at least one is not representable as a sum of two squares.

# 6
# Elliptic Curves

Elliptic curves are number theoretic objects that are central to both pure and applied number theory. Deep problems in number theory such as the congruent number problem—which integers are the area of a right triangle with rational side lengths?—translate naturally into questions about elliptic curves. Other questions, such as the famous Birch and Swinnerton-Dyer conjecture, describe mysterious structure that mathematicians expect elliptic curves to have. One can also associate finite abelian groups to elliptic curves, and in many cases these groups are well suited to the construction of cryptosystems. In particular, elliptic curves are widely believed to provide good security with smaller key sizes, something that is useful in many applications, for example, if we are going to print an encryption key on a postage stamp, it is helpful if the key is short! Morover, there is a way to use elliptic curves to factor integers, which plays a crucial role in sophisticated attacks on the RSA public-key cryptosystem of Section 3.3.

This chapter is a brief introduction to elliptic curves that builds on the ideas of Chapters 1–3 and introduces several deep theorems and ideas that we will not prove. In Section 6.1, we define elliptic curves and draw some pictures of them, and then in Section 6.2 we describe how to put a group structure on the set of points on an elliptic curve. Sections 6.3 and 6.4 are about how to apply elliptic curves to two cryptographic problems—constructing public-key cryptosystems and factoring integers. Finally, in Section 6.5, we consider elliptic curves over the rational numbers, and explain a deep connection between elliptic curves and a 1,000-year old unsolved problem.

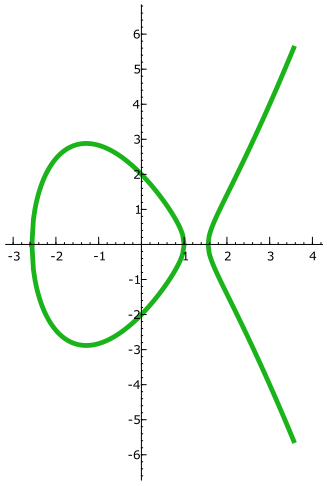



FIGURE 6.1. The elliptic curve $y^2 = x^3 - 5x + 4$ over $\mathbf{R}$

## 6.1 The Definition

**Definition 6.1.1** (Elliptic Curve)**.** An *elliptic curve* over a field $K$ is a curve defined by an equation of the form

$$y^2 = x^3 + ax + b,$$

where $a, b \in K$ and $-16(4a^3 + 27b^2) \neq 0$.

The condition that $-16(4a^3 + 27b^2) \neq 0$ implies that the curve has no "singular points," which will be essential for the applications we have in mind (see Exercise 6.1).

*SAGE Example* 6.1.2. We use the `EllipticCurve` command to create an elliptic curve over the rational field $\mathbf{Q}$ and draw the plot in Figure 6.1.

```
sage: E = EllipticCurve([-5, 4])
sage: E
Elliptic Curve defined by y^2  = x^3 - 5*x + 4
over Rational Field
sage: P = E.plot(thickness=4,rgbcolor=(0.1,0.7,0.1))
sage: P.show(figsize=[4,6])
```

We will use elliptic curves over finite fields to factor integers in Section 6.3 and to construct cryptosystems in Section 6.4. The following Sage code creates an elliptic curve over the finite field of order 37 and plots it, as illustrated in Figure 6.2.

```
sage: E = EllipticCurve(GF(37), [1,0])
sage: E
Elliptic Curve defined by y^2  = x^3 + x over
Finite Field of size 37
sage: E.plot(pointsize=45)
```

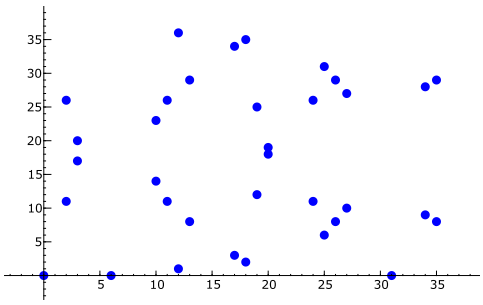



FIGURE 6.2. The elliptic curve $y^2 = x^3 + x$ over $\mathbf{Z}/37\mathbf{Z}$

In Section 6.2, we will put a natural abelian group structure on the set

$$E(K) = \{(x, y) \in K \times K : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\}$$

of $K$-rational points on an elliptic curve $E$ over $K$. Here, $\mathcal{O}$ may be thought of as a point on $E$ "at infinity." Figure 6.2 contains a plot of the points of $y^2 = x^3 + x$ over the finite field $\mathbf{Z}/37\mathbf{Z}$, though note that we do not explicitly draw the point at $\mathcal{O}$ at infinity.

*Remark* 6.1.3. If $K$ has characteristic 2 (i.e., we have $1+1 = 0$ in $K$), then for any choice of $a, b$, the quantity $-16(4a^3 + 27b^2) \in K$ is 0, so according to Definition 6.1.1 there are no elliptic curves over $K$. There is a similar problem in characteristic 3. If we instead consider equations of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

we obtain a more general definition of elliptic curves, which correctly allows for elliptic curves in characteristics 2 and 3; these elliptic curves are popular in cryptography because arithmetic on them is often easier to efficiently implement on a computer.

## 6.2 The Group Structure on an Elliptic Curve

Let $E$ be an elliptic curve over a field $K$, given by an equation $y^2 = x^3 + ax + b$. We begin by defining a binary operation $+$ on $E(K)$.

**Algorithm 6.2.1** (Elliptic Curve Group Law)**.** Given $P_1, P_2 \in E(K)$, this algorithm computes a third point $R = P_1 + P_2 \in E(K)$.

1. [Is $P_i = \mathcal{O}$?] If $P_1 = \mathcal{O}$ set $R = P_2$ or if $P_2 = \mathcal{O}$ set $R = P_1$ and terminate. Otherwise write $(x_i, y_i) = P_i$.
2. [Negatives] If $x_1 = x_2$ and $y_1 = -y_2$, set $R = \mathcal{O}$ and terminate.
3. [Compute $\lambda$] Set $\lambda = \begin{cases} (3x_1^2 + a)/(2y_1) & \text{if } P_1 = P_2, \\ (y_1 - y_2)/(x_1 - x_2) & \text{otherwise.} \end{cases}$
4. [Compute Sum] Then $R = \left(\lambda^2 - x_1 - x_2, -\lambda x_3 - \nu\right)$, where $\nu = y_1 - \lambda x_1$ and $x_3 = \lambda^2 - x_1 - x_2$ is the $x$-coordinate of $R$.

Note that in Step 3, if $P_1 = P_2$, then $y_1 \neq 0$; otherwise, we would have terminated in the previous step.

**Theorem 6.2.2.** *The binary operation + defined in Algorithm 6.2.1 endows the set $E(K)$ with an abelian group structure, with identity $\mathcal{O}$.*

Before discussing why the theorem is true, we reinterpret + geometrically, so that it will be easier for us to visualize. We obtain the sum $P_1 + P_2$ by finding the third point $P_3$ of intersection between $E$ and the line $L$ determined by $P_1$ and $P_2$, then reflecting $P_3$ about the $x$-axis. (This description requires suitable interpretation in cases 1 and 2, and when $P_1 = P_2$.) This is illustrated in Figure 6.3, in which $(0, 2) + (1, 0) = (3, 4)$ on $y^2 = x^3 - 5x + 4$.

*SAGE Example* 6.2.3. We create the elliptic curve $y^2 = x^3 - 5x + 4$ in Sage, then add together $P = (1, 0)$ and $Q = (0, 2)$. We also compute $P + P$, which is the point $\mathcal{O}$ at infinity, which is represented in Sage by $(0 : 1 : 0)$, and compute the sum $P + Q + Q + Q + Q$, which is surprisingly large.

```
sage: E = EllipticCurve([-5,4])
sage: P = E([1,0]); Q = E([0,2])
sage: P + Q
(3 : 4 : 1)
sage: P + P
(0 : 1 : 0)
sage: P + Q + Q + Q + Q
(350497/351649 : 16920528/208527857 : 1)
```

To further clarify the above geometric interpretation of the group law, we prove the following proposition.

**Proposition 6.2.4** (Geometric Group Law)**.** *Suppose $P_i = (x_i, y_i)$, $i = 1, 2$ are distinct points on an elliptic curve $y^2 = x^3 + ax + b$, and that $x_1 \neq x_2$. Let $L$ be the unique line through $P_1$ and $P_2$. Then $L$ intersects the graph of $E$ at exactly one other point*

$$Q = \left(\lambda^2 - x_1 - x_2, \quad \lambda x_3 + \nu\right),$$

*where $\lambda = (y_1 - y_2)/(x_1 - x_2)$ and $\nu = y_1 - \lambda x_1$.*

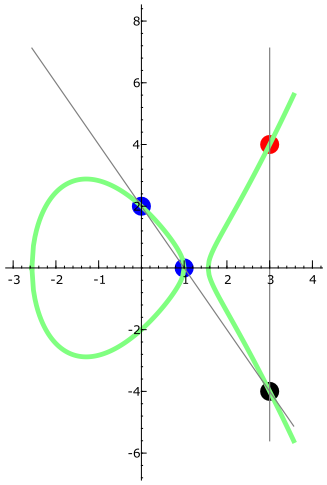



FIGURE 6.3. The Group Law: $(1,0)+(0,2)=(3,4)$ on $y^2=x^3-5x+4$

*Proof.* The line $L$ through $P_1$, $P_2$ is $y=y_1+(x-x_1)\lambda$. Substituting this into $y^2=x^3+ax+b$, we get

$$(y_1+(x-x_1)\lambda)^2=x^3+ax+b.$$

Simplifying, we get $f(x)=x^3-\lambda^2x^2+\cdots=0$, where we omit the coefficients of $x$ and the constant term since they will not be needed. Since $P_1$ and $P_2$ are in $L\cap E$, the polynomial $f$ has $x_1$ and $x_2$ as roots. By Proposition 2.5.3, the polynomial $f$ can have at most three roots. Writing $f=\prod(x-x_i)$ and equating terms, we see that $x_1+x_2+x_3=\lambda^2$. Thus, $x_3=\lambda^2-x_1-x_2$, as claimed. Also, from the equation for $L$ we see that $y_3=y_1+(x_3-x_1)\lambda=\lambda x_3+\nu$, which completes the proof. □

To prove Theorem 6.2.2 means to show that $+$ satisfies the three axioms of an abelian group with $\mathcal{O}$ as identity element: existence of inverses, commutativity, and associativity. The existence of inverses follows immediately from the definition, since $(x,y)+(x,-y)=\mathcal{O}$. Commutativity is also clear from the definition of group law, since in Parts 1–3, the recipe is unchanged if we swap $P_1$ and $P_2$; in Part 4 swapping $P_1$ and $P_2$ does not change the line determined by $P_1$ and $P_2$, so by Proposition 6.2.4 it does not change the sum $P_1+P_2$.

It is more difficult to prove that $+$ satisfies the associative axiom, i.e., that $(P_1+P_2)+P_3=P_1+(P_2+P_3)$. This fact can be understood from at least three points of view. One is to reinterpret the group law geometrically (extending Proposition 6.2.4 to all cases), and thus transfer the problem to a question in plane geometry. This approach is beautifully explained

with exactly the right level of detail in [ST92, §I.2]. Another approach is to use the formulas that define + to reduce associativity to checking specific algebraic identities; this is something that would be extremely tedious to do by hand, but can be done using a computer (also tedious). A third approach (see [Sil86] or [Har77]) is to develop a general theory of "divisors on algebraic curves," from which associativity of the group law falls out as a natural corollary. The third approach is the best, because it opens up many new vistas; however, we will not pursue it further because it is beyond the scope of this book.

*SAGE Example* 6.2.5. In the following Sage session, we use the formula from Algorithm 6.2.1 to verify that the group law holds for any choice of points $P_1, P_2, P_3$ on any elliptic curve over $\mathbf{Q}$ such that the points $P_1, P_2, P_3, P_1 + P_2, P_2 + P_3$ are all distinct and nonzero. We define a polynomial ring $R$ in 8 variables.

```
sage: R.<x1,y1,x2,y2,x3,y3,a,b> = QQ[]
```

We define the relations the $x_i$ will satisfy, and a quotient ring $Q$ in which those relations are satisfied. (Quotients of polynomial rings are a generalization of the construction $\mathbf{Z}/n\mathbf{Z}$ that may be viewed as the quotient of the ring $\mathbf{Z}$ of integers by the relation that sets $n$ to equal 0.)

```
sage: rels = [y1^2 - (x1^3 + a*x1 + b),
...           y2^2 - (x2^3 + a*x2 + b),
...           y3^2 - (x3^3 + a*x3 + b)]
...
sage: Q = R.quotient(rels)
```

We define the group operation, which assumes the points are distinct.

```
sage: def op(P1,P2):
...       x1,y1 = P1;  x2,y2 = P2
...       lam = (y1 - y2)/(x1 - x2); nu  = y1 - lam*x1
...       x3 = lam^2 - x1 - x2; y3 = -lam*x3 - nu
...       return (x3, y3)
```

We define three points, add them together via $P_1 + (P_2 + P_3)$ and $(P_1 + (P_2 + P_3))$, and observe that the results are the same modulo the relations.

```
sage: P1 = (x1,y1); P2 = (x2,y2); P3 = (x3,y3)
sage: Z = op(P1, op(P2,P3)); W = op(op(P1,P2),P3)
sage: (Q(Z[0].numerator()*W[0].denominator() -
...          Z[0].denominator()*W[0].numerator())) == 0
True
sage: (Q(Z[1].numerator()*W[1].denominator() -
...          Z[1].denominator()*W[1].numerator())) == 0
True
```

## 6.3 Integer Factorization Using Elliptic Curves

In 1987, Hendrik Lenstra published the landmark paper [Len87] that introduces and analyzes the Elliptic Curve Method (ECM), which is a powerful algorithm for factoring integers using elliptic curves. Lenstra's method is also described in [ST92, §IV.4], [Dav99, §VIII.5], and [Coh93, §10.3].

Lenstra's algorithm is well suited for finding "medium-sized" factors of an integer $N$, which today means between 10 to 40 decimal digits. The ECM method is not *directly* used for factoring RSA challenge numbers (see Section 1.1.3), but it is used on auxiliary numbers as a crucial step in the "number field sieve," which is the best known algorithm for hunting for such factorizations. Also, implementation of ECM typically requires little memory.

H. Lenstra

### *6.3.1 Pollard's $(p-1)$-Method*

Lenstra's discovery of ECM was inspired by Pollard's $(p-1)$-method, which we describe in this section.

**Definition 6.3.1** (Power Smooth)**.** Let $B$ be a positive integer. If $n$ is a positive integer with prime factorization $n = \prod p_i^{e_i}$, then $n$ is *$B$-power smooth* if $p_i^{e_i} \leq B$ for all $i$.

For example, $30 = 2{\cdot}3{\cdot}5$ is $B$ power smooth for $B = 5, 7$, but $150 = 2{\cdot}3{\cdot}5^2$ is not 5-power smooth (it is $B = 25$-power smooth).

We will use the following algorithm in both the Pollard $p-1$ and elliptic curve factorization methods.

**Algorithm 6.3.2** (Least Common Multiple of First $B$ Integers)**.** Given a positive integer $B$, this algorithm computes the least common multiple of the positive integers up to $B$.

1. [Sieve] Using, for example, the prime sieve (Algorithm 1.2.3), compute a list $P$ of all primes $p \leq B$.
2. [Multiply] Compute and output the product $\prod_{p \in P} p^{\lfloor \log_p(B) \rfloor}$.

*Proof.* Set $m = \operatorname{lcm}(1, 2, \ldots, B)$. Then,

$$\operatorname{ord}_p(m) = \max(\{\operatorname{ord}_p(n) : 1 \leq n \leq B\}) = \operatorname{ord}_p(p^r),$$

where $p^r$ is the largest power of $p$ that satisfies $p^r \leq B$. Since $p^r \leq B < p^{r+1}$, we have $r = \lfloor \log_p(B) \rfloor$. □

*SAGE Example* 6.3.3. We implement Algorithm 6.3.2 in Sage and compute the least common multiple for $B = 100$ using both the above algorithm and a naive algorithm. We use `math.log` below so that $\log_p(B)$ is computed quickly using double precision numbers.

```
sage: def lcm_upto(B):
...       return prod([p^int(math.log(B)/math.log(p))
...                    for p in prime_range(B+1)])
sage: lcm_upto(10^2)
69720375229712477164533808935312303556800
sage: LCM([1..10^2])
69720375229712477164533808935312303556800
```

Algorithm 6.3.2 as implemented above in Sage takes about a second for $B = 10^6$.

Let $N$ be a positive integer that we wish to factor. We use the Pollard $(p-1)$-method to look for a nontrivial factor of $N$ as follows. First, we choose a positive integer $B$, usually with at most six digits. Suppose that there is a prime divisor $p$ of $N$ such that $p-1$ is $B$-power smooth. We try to find $p$ using the following strategy. If $a > 1$ is an integer not divisible by $p$, then by Theorem 2.1.20,

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

Let $m = \operatorname{lcm}(1, 2, 3, \ldots, B)$, and observe that our assumption that $p-1$ is $B$-power smooth implies that $p-1 \mid m$, so

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

Thus

$$p \mid \gcd(a^m - 1, N) > 1.$$

If $\gcd(a^m-1, N) < N$ also then $\gcd(a^m-1, N)$ is a nontrivial factor of $N$. If $\gcd(a^m - 1, N) = N$, then $a^m \equiv 1 \pmod{q^r}$ for every prime power divisor $q^r$ of $N$. In this case, repeat the above steps but with a smaller choice of $B$ or possibly a different choice of $a$. Also, it is a good idea to check from the start whether or not $N$ is not a perfect power $M^r$ and, if so, replace $N$ by $M$. We formalize the algorithm as follows:

**Algorithm 6.3.4** (Pollard $p-1$ Method)**.** Given a positive integer $N$ and a bound $B$, this algorithm attempts to find a nontrivial factor $g$ of $N$. (Each prime $p \mid g$ is likely to have the property that $p-1$ is $B$-power smooth.)

1. [Compute lcm] Use Algorithm 6.3.2 to compute $m = \operatorname{lcm}(1, 2, \ldots, B)$.
2. [Initialize] Set $a = 2$.
3. [Power and gcd] Compute $x = a^m - 1 \pmod{N}$ and $g = \gcd(x, N)$.
4. [Finished?] If $g \neq 1$ or $N$, output $g$ and terminate.

5. [Try Again?] If $a < 10$ (say), replace $a$ by $a + 1$ and go to step 3. Otherwise, terminate.

For fixed $B$, Algorithm 6.3.4 often splits $N$ when $N$ is divisible by a prime $p$ such that $p - 1$ is $B$-power smooth. Approximately 15 percent of primes $p$ in the interval from $10^{15}$ and $10^{15} + 10000$ are such that $p - 1$ is $10^6$ power smooth, so the Pollard method with $B = 10^6$ already fails nearly 85 percent of the time at finding 15-digit primes in this range (see also Exercise 6.10). We will not analyze Pollard's method further, since it was mentioned here only to set the stage for the elliptic curve factorization method.

The following examples illustrate the Pollard $(p - 1)$-method.

*Example* 6.3.5. In this example, Pollard works perfectly. Let $N = 5917$. We try to use the Pollard $p - 1$ method with $B = 5$ to split $N$. We have $m = \text{lcm}(1, 2, 3, 4, 5) = 60$; taking $a = 2$, we have

$$2^{60} - 1 \equiv 3416 \pmod{5917}$$

and

$$\gcd(2^{60} - 1, 5917) = \gcd(3416, 5917) = 61,$$

so 61 is a factor of 5917.

*Example* 6.3.6. In this example, we replace $B$ with a larger integer. Let $N = 779167$. With $B = 5$ and $a = 2$, we have

$$2^{60} - 1 \equiv 710980 \pmod{779167},$$

and $\gcd(2^{60} - 1, 779167) = 1$. With $B = 15$, we have

$$m = \text{lcm}(1, 2, \ldots, 15) = 360360,$$

$$2^{360360} - 1 \equiv 584876 \pmod{779167},$$

and

$$\gcd(2^{360360} - 1, N) = 2003,$$

so 2003 is a nontrivial factor of 779167.

*Example* 6.3.7. In this example, we replace $B$ by a smaller integer. Let $N = 4331$. Suppose $B = 7$, so $m = \text{lcm}(1, 2, \ldots, 7) = 420$,

$$2^{420} - 1 \equiv 0 \pmod{4331},$$

and $\gcd(2^{420} - 1, 4331) = 4331$, so we do not obtain a factor of 4331. If we replace $B$ by 5, Pollard's method works:

$$2^{60} - 1 \equiv 1464 \pmod{4331},$$

and $\gcd(2^{60} - 1, 4331) = 61$, so we split 4331.

*Example* 6.3.8. In this example, $a = 2$ does not work, but $a = 3$ does. Let $N = 187$. Suppose $B = 15$, so $m = \text{lcm}(1, 2, \ldots, 15) = 360360$,

$$2^{360360} - 1 \equiv 0 \pmod{187},$$

and $\gcd(2^{360360} - 1, 187) = 187$, so we do not obtain a factor of 187. If we replace $a = 2$ by $a = 3$, then Pollard's method works:

$$3^{360360} - 1 \equiv 66 \pmod{187},$$

and $\gcd(3^{360360} - 1, 187) = 11$. Thus $187 = 11 \cdot 17$.

*SAGE Example* 6.3.9. We implement the Pollard $(p-1)$-method in Sage and use our implementation to do all of the above examples.

```
sage: def pollard(N, B=10^5, stop=10):
...       m = prod([p^int(math.log(B)/math.log(p))
...                 for p in prime_range(B+1)])
...       for a in [2..stop]:
...           x = (Mod(a,N)^m - 1).lift()
...           if x == 0: continue
...           g = gcd(x, N)
...           if g != 1 or g != N: return g
...       return 1
sage: pollard(5917,5)
61
sage: pollard(779167,5)
1
sage: pollard(779167,15)
2003
sage: pollard(4331,7)
1
sage: pollard(4331,5)
61
sage: pollard(187, 15, 2)
1
sage: pollard(187, 15)
11
```

### *6.3.2 Motivation for the Elliptic Curve Method*

Fix a positive integer $B$. If $N = pq$ with $p$ and $q$ prime, and we assume that $p - 1$ and $q - 1$ are not $B$-power smooth, then the Pollard $(p-1)$-method is unlikely to work. For example, let $B = 20$ and suppose that $N = 59 \cdot 101 = 5959$. Note that neither $59 - 1 = 2 \cdot 29$ nor $101 - 1 = 4 \cdot 25$ is $B$-power smooth. With $m = \text{lcm}(1, 2, 3, \ldots, 20) = 232792560$, we have

$$2^m - 1 \equiv 5944 \pmod{N},$$

and $\gcd(2^m - 1, N) = 1$, so we do not find a factor of $N$.

As remarked above, the problem is that $p-1$ is not 20-power smooth for either $p = 59$ or $p = 101$. However, notice that $p - 2 = 3 \cdot 19$ is 20-power smooth. Lenstra's ECM replaces $(\mathbf{Z}/p\mathbf{Z})^*$, which has order $p - 1$, by the group of points on an elliptic curve $E$ over $\mathbf{Z}/p\mathbf{Z}$. It is a theorem that

$$\#E(\mathbf{Z}/p\mathbf{Z}) = p + 1 \pm s$$

for some nonnegative integer $s < 2\sqrt{p}$ (see [Sil86, §V.1] for a proof). Also, every value of $s$ subject to this bound occurs, as one can see using "complex multiplication theory." For example, if $E$ is the elliptic curve

$$y^2 = x^3 + x + 54$$

over $\mathbf{Z}/59\mathbf{Z}$, then by enumerating points one sees that $E(\mathbf{Z}/59\mathbf{Z})$ is cyclic of order 57. The set of numbers $59 + 1 \pm s$ for $s \leq 15$ contains 14 numbers that are $B$-power smooth for $B = 20$, which illustrates that working with an elliptic curve gives us more flexibility. For example, $60 = 59 + 1 + 0$ is 5-power smooth and $70 = 59 + 1 + 10$ is 7-power smooth.

### *6.3.3 Lenstra's Elliptic Curve Factorization Method*

**Algorithm 6.3.10** (Elliptic Curve Factorization Method)**.** Given a positive integer $N$ and a bound $B$, this algorithm attempts to find a nontrivial factor $g$ of $N$ or outputs "Fail."

1. [Compute lcm] Use Algorithm 6.3.2 to compute $m = \operatorname{lcm}(1, 2, \ldots, B)$.
2. [Choose Random Elliptic Curve] Choose a random $a \in \mathbf{Z}/N\mathbf{Z}$ such that $4a^3 + 27 \in (\mathbf{Z}/N\mathbf{Z})^*$. Then $P = (0, 1)$ is a point on the elliptic curve $y^2 = x^3 + ax + 1$ over $\mathbf{Z}/N\mathbf{Z}$.
3. [Compute Multiple] Attempt to compute $mP$ using an elliptic curve analog of Algorithm 2.3.13. If at some point we cannot compute a sum of points because some denominator in Step 3 of Algorithm 6.2.1 is not coprime to $N$, we compute the greatest common divisor $g$ of this denominator with $N$. If $g$ is a nontrivial divisor, output it. If every denominator is coprime to $N$, output "Fail."

If Algorithm 6.3.10 fails for one random elliptic curve, there is an option that is unavailable with Pollard's $(p-1)$-method—we may repeat the above algorithm with a different elliptic curve. With Pollard's method we always work with the group $(\mathbf{Z}/N\mathbf{Z})^*$, but here we can try many groups $E(\mathbf{Z}/N\mathbf{Z})$ for many curves $E$. As mentioned above, the number of points on $E$ over $\mathbf{Z}/p\mathbf{Z}$ is of the form $p + 1 - t$ for some $t$ with $|t| < 2\sqrt{p}$; Algorithm 6.3.10 thus has a chance if $p + 1 - t$ is $B$-power smooth for some $t$ with $|t| < 2\sqrt{p}$.

### *6.3.4 Examples*

For simplicity, we use an elliptic curve of the form

$$y^2 = x^3 + ax + 1,$$

which has the point $P = (0, 1)$ already on it.

We factor $N = 5959$ using the elliptic curve method. Let

$$m = \text{lcm}(1, 2, \ldots, 20) = 232792560 = 11011110000000100001111110000_2,$$

where $x_2$ means $x$ is written in binary. First, we choose $a = 1201$ at random and consider $y^2 = x^3 + 1201x + 1$ over $\mathbf{Z}/5959\mathbf{Z}$. Using the formula for $P + P$ from Algorithm 6.2.1 we compute $2^i \cdot P = 2^i \cdot (0, 1)$ for $i \in B = \{4, 5, 6, 7, 8, 13, 21, 22, 23, 24, 26, 27\}$. Then $\sum_{i \in B} 2^i P = mP$. It turns out that during no step of this computation does a number not coprime to 5959 appear in any denominator, so we do not split $N$ using $a = 1201$. Next, we try $a = 389$ and at some stage in the computation we add $P = (2051, 5273)$ and $Q = (637, 1292)$. When computing the group law explicitly, we try to compute $\lambda = (y_1 - y_2)/(x_1 - x_2)$ in $(\mathbf{Z}/5959\mathbf{Z})^*$, but we fail since $x_1 - x_2 = 1414$ and $\gcd(1414, 5959) = 101$. We thus find a nontrivial factor 101 of 5959.

*SAGE Example* 6.3.11. We implement elliptic curve factorization in Sage, then use it to do the above example and some other examples.

```
sage: def ecm(N, B=10^3, trials=10):
...       m = prod([p^int(math.log(B)/math.log(p))
...                 for p in prime_range(B+1)])
...       R = Integers(N)
...       # Make Sage think that R is a field:
...       R.is_field = lambda : True
...       for _ in range(trials):
...           while True:
...               a = R.random_element()
...               if gcd(4*a.lift()^3 + 27, N) == 1: break
...           try:
...               m * EllipticCurve([a, 1])([0,1])
...           except ZeroDivisionError, msg:
...               # msg: "Inverse of <int> does not exist"
...               return gcd(Integer(str(msg).split()[2]), N)
...       return 1
sage: set_random_seed(2)
sage: ecm(5959, B=20)
101
sage: ecm(next_prime(10^20)*next_prime(10^7), B=10^3)
10000019
```

### *6.3.5 A Heuristic Explanation*

Let $N$ be a positive integer and, for simplicity of exposition, assume that $N = p_1 \cdots p_r$ with the $p_i$ distinct primes. It follows from Lemma 2.2.5 that there is a natural isomorphism

$$f : (\mathbf{Z}/N\mathbf{Z})^* \longrightarrow (\mathbf{Z}/p_1\mathbf{Z})^* \times \cdots \times (\mathbf{Z}/p_r\mathbf{Z})^*.$$

When using Pollard's method, we choose an $a \in (\mathbf{Z}/N\mathbf{Z})^*$, compute $a^m$, then compute $\gcd(a^m - 1, N)$. This gcd is divisible exactly by the primes $p_i$ such that $a^m \equiv 1 \pmod{p_i}$. To reinterpret Pollard's method using the above isomorphism, let $(a_1, \ldots, a_r) = f(a)$. Then $(a_1^m, \ldots, a_r^m) = f(a^m)$, and the $p_i$ that divide $\gcd(a^m - 1, N)$ are exactly the $p_i$ such that $a_i^m = 1$. By Theorem 2.1.20, these $p_i$ include the primes $p_j$ such that $p_j - 1$ is $B$-power smooth, where $m = \operatorname{lcm}(1, \ldots, m)$.

We will not define $E(\mathbf{Z}/N\mathbf{Z})$ when $N$ is composite, since this is not needed for the algorithm (where we assume that $N$ is prime and hope for a contradiction). However, for the remainder of this paragraph, we pretend that $E(\mathbf{Z}/N\mathbf{Z})$ is meaningful and describe a heuristic connection between Lenstra and Pollard's methods. The significant difference between Pollard's method and the elliptic curve method is that the isomorphism $f$ is replaced by an isomorphism (in quotes)

$$\text{“}g : E(\mathbf{Z}/N\mathbf{Z}) \to E(\mathbf{Z}/p_1\mathbf{Z}) \times \cdots \times E(\mathbf{Z}/p_r\mathbf{Z})\text{”}$$

where $E$ is $y^2 = x^3 + ax + 1$, and the $a$ of Pollard's method is replaced by $P = (0, 1)$. We put the isomorphism in quotes to emphasize that we have not defined $E(\mathbf{Z}/N\mathbf{Z})$. When carrying out the elliptic curve factorization algorithm, we attempt to compute $mP$, and if some components of $f(Q)$ are $\mathcal{O}$, for some point $Q$ that appears during the computation, but others are nonzero, we find a nontrivial factor of $N$.

## 6.4 Elliptic Curve Cryptography

The idea to use elliptic curves in cryptography was independently proposed by Neil Koblitz and Victor Miller in the mid 1980s. In this section, we discuss an analog of Diffie-Hellman that uses an elliptic curve instead of $(\mathbf{Z}/p\mathbf{Z})^*$. We then discuss the ElGamal elliptic curve cryptosystem.

### *6.4.1 Elliptic Curve Analogs of Diffie-Hellman*

The Diffie-Hellman key exchange from Section 3.2 works well on an elliptic curve with no serious modification. Michael and Nikita agree on a secret key as follows:

1. Michael and Nikita agree on a prime $p$, an elliptic curve $E$ over $\mathbf{Z}/p\mathbf{Z}$, and a point $P \in E(\mathbf{Z}/p\mathbf{Z})$.

2. Michael secretly chooses a random $m$ and sends $mP$.

3. Nikita secretly chooses a random $n$ and sends $nP$.

4. The secret key is $nmP$, which both Michael and Nikita can compute.

Presumably, an adversary can not compute $nmP$ without solving the discrete logarithm problem (see Problem 3.2.2 and Section 6.4.3 below) in $E(\mathbf{Z}/p\mathbf{Z})$. For well-chosen $E$, $P$, and $p$, experience suggests that the discrete logarithm problem in $E(\mathbf{Z}/p\mathbf{Z})$ is much more difficult than the discrete logarithm problem in $(\mathbf{Z}/p\mathbf{Z})^*$ (see Section 6.4.3 for more on the elliptic curve discrete log problem).

### *6.4.2 The ElGamal Cryptosystem and Digital Rights Management*

This section is about the ElGamal cryptosystem, which works well on an elliptic curve. This section draws on a paper by a computer hacker named Beale Screamer who cracked a "Digital Rights Management" (DRM) system.

The elliptic curve used in the DRM is an elliptic curve over the finite field $k = \mathbf{Z}/p\mathbf{Z}$, where

$$p = 7859631023794288223766947894468973962074985689 51.$$

The number $p$ in base 16 is

$$89ABCDEF012345672718281831415926141424F7,$$

which includes counting in hexadecimal, and digits of $e$, $\pi$, and $\sqrt{2}$. The elliptic curve $E$ is

$$y^2 = x^3 + 3176890812513255034763174764138276932727469559 27x + 7905289660787875871812057202571853543210065193 4.$$

We have

$$\#E(k) = 7859631023794288223766930248817149576126861574 29,$$

and the group $E(k)$ is cyclic with generator

$$B = (7715072162626498261706482685655798899077692541 76, 3901575102465566285252794592665149955625331966 55).$$

Our heroes Nikita and Michael share digital music when they are not out fighting terrorists. When Nikita installed the DRM software on her computer, it generated a private key

$$n = 67080503113991051351752720769306045630021705447 3,$$

which it hides in bits and pieces of files. In order for Nikita to play Juno Reactor's latest hit `juno.wma`, her web browser contacts a website that sells music. After Nikita sends her credit card number, that website allows Nikita to download a license file that allows her audio player to unlock and play `juno.wma`.

As we will see below, the license file was created using the ElGamal public-key cryptosystem in the group $E(k)$. Nikita can now use her license file to unlock `juno.wma`. However, when she shares both `juno.wma` and the license file with Michael, he is frustrated because even with the license, his computer still does not play `juno.wma`. This is because Michael's computer does not know Nikita's computer's private key (the integer $n$ above), so Michael's computer can not decrypt the license file.

We now describe the ElGamal cryptosystem, which lends itself well to implementation in the group $E(\mathbf{Z}/p\mathbf{Z})$. To illustrate ElGamal, we describe how Nikita would set up an ElGamal cryptosystem that anyone could use to encrypt messages for her. Nikita chooses a prime $p$, an elliptic curve $E$ over $\mathbf{Z}/p\mathbf{Z}$, and a point $B \in E(\mathbf{Z}/p\mathbf{Z})$, and publishes $p$, $E$, and $B$. She also chooses a random integer $n$, which she keeps secret, and publishes $nB$. Her public key is the four-tuple $(p, E, B, nB)$.

Suppose Michael wishes to encrypt a message for Nikita. If the message is encoded as an element $P \in E(\mathbf{Z}/p\mathbf{Z})$, Michael computes a random integer $r$ and the points $rB$ and $P + r(nB)$ on $E(\mathbf{Z}/p\mathbf{Z})$. Then $P$ is encrypted as the pair $(rB, P + r(nB))$. To decrypt the encrypted message, Nikita multiplies $rB$ by her secret key $n$ to find $n(rB) = r(nB)$, then subtracts this from $P + r(nB)$ to obtain

$$P = P + r(nB) - r(nB).$$

*Remark* 6.4.1. It also make sense to construct an ElGamal cryptosystem in the group $(\mathbf{Z}/p\mathbf{Z})^*$.

Returning to our story, Nikita's license file is an encrypted message to her. It contains the pair of points $(rB, P + r(nB))$, where

$$rB = (17967100321831574638502665573308604498219442466 0,$$
$$69783438535968636824930128267514183093517631471 8)$$

and

$$P + r(nB) = (13785103854826446737264515809300400034363911891 5,$$
$$11084858922867622405722923022358081502422487569 9).$$

When Nikita's computer plays `juno.wma`, it loads the secret key

$$n = 67080503113991051351752720769306045630021705447 3$$

into memory and computes

$$n(rB) = (328901393518732637577115650601768681044040715701,$$
$$5869478380878159936013505654887888462038879881 62).$$

It then subtracts this from $P + r(nB)$ to obtain

$$P = (14489646124220757767,$$
$$669337780373284096274895136618194604469696830074).$$

The $x$-coordinate 14489646124220757767 is the key that unlocks `juno.wma`.

If Nikita knew the private key $n$ that her computer generated, she could compute $P$ herself and unlock `juno.wma` and share her music with Michael. Beale Screamer found a weakness in the implementation of this system that allows Nikita to detetermine $n$, which is not a huge surprise since $n$ is stored on her computer after all.

*SAGE Example* 6.4.2. We do the above examples in Sage:

```
sage: p = 78596310237942882237669478944689739620749856895 1
sage: E = EllipticCurve(GF(p), \
...    [31768908125132550347631747641382769327274695592 7,
...     79052896607878758718120572025718535432100651934])
sage: E.cardinality()
78596310237942882237669302488171495761268615742 9
sage: E.cardinality().is_prime()
True
sage: B = E([
...    77150721626264982617064826856557988990776925417 6,
...    39015751024655662852527945926651499556253319665 5])
sage: n=67080503113991051351752720769306045630021705447 3
sage: r=70674630913457179596452846564371866229568459543
sage: P = E([14489646124220757767,
...    669337780373284096274895136618194604469696830074])
sage: encrypt = (r*B, P + r*(n*B))
sage: encrypt[1] - n*encrypt[0] == P   # decrypting works
True
```

### *6.4.3 The Elliptic Curve Discrete Logarithm Problem*

**Problem 6.4.3** (Elliptic Curve Discrete Log Problem)**.** Suppose $E$ is an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$ and $P \in E(\mathbf{Z}/p\mathbf{Z})$. Given a multiple $Q$ of $P$, the *elliptic curve discrete log problem* is to find $n \in \mathbf{Z}$ such that $nP = Q$.

For example, let $E$ be the elliptic curve given by $y^2 = x^3 + x + 1$ over the field $\mathbf{Z}/7\mathbf{Z}$. We have

$$E(\mathbf{Z}/7\mathbf{Z}) = \{\mathcal{O}, (2,2), (0,1), (0,6), (2,5)\}.$$

If $P = (2,2)$ and $Q = (0,6)$, then $3P = Q$, so $n = 3$ is a solution to the discrete logarithm problem.

If $E(\mathbf{Z}/p\mathbf{Z})$ has order $p$ or $p \pm 1$, or is a product of reasonably small primes, then there are some methods for attacking the discrete log problem on $E$, which are beyond the scope of this book. It is therefore important to be able to compute $\#E(\mathbf{Z}/p\mathbf{Z})$ efficiently, in order to verify that the elliptic curve one wishes to use for a cryptosystem doesn't have any obvious vulnerabilities. The naive algorithm to compute $\#E(\mathbf{Z}/p\mathbf{Z})$ is to try each value of $x \in \mathbf{Z}/p\mathbf{Z}$ and count how often $x^3+ax+b$ is a perfect square mod $p$, but this is of no use when $p$ is large enough to be useful for cryptography. Fortunately, there is an algorithm due to Schoof, Elkies, and Atkin for computing $\#E(\mathbf{Z}/p\mathbf{Z})$ efficiently (polynomial time in the number of digits of $p$), but this algorithm is beyond the scope of this book.

In Section 3.2.1, we discussed the discrete log problem in $(\mathbf{Z}/p\mathbf{Z})^*$. There are general attacks called "index calculus attacks" on the discrete log problem in $(\mathbf{Z}/p\mathbf{Z})^*$ that are slow, but still faster than the known algorithms for solving the discrete log in a "general" group (one with no extra structure). For most elliptic curves, there is no known analog of index calculus attacks on the discrete log problem. At present, it appears that given $p$, the discrete log problem in $E(\mathbf{Z}/p\mathbf{Z})$ is much harder than the discrete log problem in the multiplicative group $(\mathbf{Z}/p\mathbf{Z})^*$. This suggests that by using an elliptic curve-based cryptosystem instead of one based on $(\mathbf{Z}/p\mathbf{Z})^*$, one gets equivalent security with much smaller numbers, which is one reason why building cryptosystems using elliptic curves is attractive to some cryptographers. For example, Certicom, a company that strongly supports elliptic curve cryptography, claims:

> "[Elliptic curve crypto] devices require less storage, less power, less memory, and less bandwidth than other systems. This allows you to implement cryptography in platforms that are constrained, such as wireless devices, handheld computers, smart cards, and thin-clients. It also provides a big win in situations where efficiency is important."

For an up-to-date list of elliptic curve discrete log challenge problems that Certicom sponsors, see [Cer]. For example, in April 2004, a specific cryptosystem was cracked that was based on an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$, where $p$ has 109 bits. The first unsolved challenge problem involves an elliptic curve over $\mathbf{Z}/p\mathbf{Z}$, where $p$ has 131 bits, and the next challenge after that is one in which $p$ has 163 bits. Certicom claims at [Cer] that the 163-bit challenge problem is computationally infeasible.

FIGURE 6.4. Louis J. Mordell

## 6.5 Elliptic Curves Over the Rational Numbers

Let $E$ be an elliptic curve defined over $\mathbf{Q}$. The following is a deep theorem about the group $E(\mathbf{Q})$.

**Theorem 6.5.1** (Mordell)**.** *The group $E(\mathbf{Q})$ is finitely generated. That is, there are points $P_1, \ldots, P_s \in E(\mathbf{Q})$ such that every element of $E(\mathbf{Q})$ is of the form $n_1P_1 + \cdots + n_sP_s$ for integers $n_1, \ldots n_s \in \mathbf{Z}$.*

Mordell's theorem implies that it makes sense to ask whether or not we can compute $E(\mathbf{Q})$, where by "compute" we mean find a finite set $P_1, \ldots, P_s$ of points on $E$ that generate $E(\mathbf{Q})$ as an abelian group. There is a systematic approach to computing $E(\mathbf{Q})$ called "descent" (see, for example, [Cre97, Cre, Sil86]). It is widely believed that the method of descent will always succeed, but nobody has yet proved that it will. Proving that descent works for all curves is one of the central open problems in number theory, and is closely related to the Birch and Swinnerton-Dyer conjecture (one of the Clay Math Institute's million dollar prize problems). The crucial difficulty amounts to deciding whether or not certain explicitly given curves have any rational points on them or not (these are curves that have points over $\mathbf{R}$ and modulo $n$ for all $n$).

The details of using descent to compute $E(\mathbf{Q})$ are beyond the scope of this book. In several places below, we will simply assert that $E(\mathbf{Q})$ has a certain structure or is generated by certain elements. In each case, we computed $E(\mathbf{Q})$ using a computer implementation of this method.

### *6.5.1 The Torsion Subgroup of $E(\mathbf{Q})$*

For any abelian group $G$, let $G_{\text{tor}}$ be the subgroup of elements of finite order. If $E$ is an elliptic curve over $\mathbf{Q}$, then $E(\mathbf{Q})_{\text{tor}}$ is a subgroup of $E(\mathbf{Q})$, which must be finite because of Theorem 6.5.1 (see Exercise 6.6).

One can also prove that $E(\mathbf{Q})_{\text{tor}}$ is finite by showing that there is a prime $p$ and an injective reduction homomorphism $E(\mathbf{Q})_{\text{tor}} \hookrightarrow E(\mathbf{Z}/p\mathbf{Z})$, then noting that $E(\mathbf{Z}/p\mathbf{Z})$ is finite. For example, if $E$ is $y^2 = x^3 - 5x + 4$, then $E(\mathbf{Q})_{\text{tor}} = \{\mathcal{O}, (1,0)\} \cong \mathbf{Z}/2\mathbf{Z}$.

The possibilities for $E(\mathbf{Q})_{\text{tor}}$ are known.

**Theorem 6.5.2** (Mazur, 1976)**.** *Let $E$ be an elliptic curve over $\mathbf{Q}$. Then $E(\mathbf{Q})_{\text{tor}}$ is isomorphic to one of the following 15 groups:*

$$\begin{aligned} \mathbf{Z}/n\mathbf{Z} \qquad & \text{for } n \leq 10 \text{ or } n = 12, \\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2n \qquad & \text{for } n \leq 4. \end{aligned}$$

*SAGE Example* 6.5.3. We compute the structure of the torsion subgroups of some elliptic curves. In each case, the output of the function $T(a,b)$ below is a pair $c, d \in \mathbf{Z}$ (or integer $c$) such that the torsion subgroup of $y^3 = x^3 + ax + b$ is $\mathbf{Z}/c\mathbf{Z} \times \mathbf{Z}/d\mathbf{Z}$.

```
sage: T = lambda v: EllipticCurve(v
...           ).torsion_subgroup().invariants()
sage: T([-5,4])
[2]
sage: T([-43,166])
[7]
sage: T([-4,0])
[2, 2]
sage: T([-1386747, 368636886])
[8, 2]
```

### *6.5.2 The Rank of $E(\mathbf{Q})$*

The quotient $E(\mathbf{Q})/E(\mathbf{Q})_{\text{tor}}$ is a finitely generated free abelian group, so it is isomorphism to $\mathbf{Z}^r$ for some integer $r$, called the *rank* of $E(\mathbf{Q})$. For example, one can prove that if $E$ is $y^2 = x^3 - 5x + 4$, then $E(\mathbf{Q})/E(\mathbf{Q})_{\text{tor}}$ is generated by the point $(0, 2)$.

*SAGE Example* 6.5.4. We use Sage to compute the ranks of some elliptic curves $y^2 = x^3 + ax + b$. The function $r(a,b)$ below returns the rank of this curve over $\mathbf{Q}$.

```
sage: r = lambda v: EllipticCurve(v).rank()
sage: r([-5,4])
1
sage: r([0,1])
0
sage: r([-3024, 46224])
2
sage: r([-112, 400])
```

```
3
sage: r([-102627, 12560670])
4
```

The following is a folklore conjecture, not associated with any particular mathematician:

**Conjecture 6.5.5.** *There are elliptic curves over* $\mathbf{Q}$ *of arbitrarily large rank.*

The world record is the following curve, whose rank is at least 28:

$$\begin{aligned} y^2+xy+y = x^3 - x^2 - \\ 20067762415575526585033208209338542750930230312178956502x + \\ 3448161179503055646703298569039072037485594435931918036 12\ldots \\ \ldots 66008296291939448732243429 \end{aligned}$$

It was discovered in May 2006 by Noam Elkies of Harvard University.

### *6.5.3 The Congruent Number Problem*

**Definition 6.5.6** (Congruent Number)**.** We call a nonzero rational number $n$ a *congruent number* if $\pm n$ is the area of a right triangle with rational side lengths. Equivalently, $n$ is a *congruent number* if the system of two equations

$$\begin{aligned} a^2 + b^2 &= c^2 \\ \frac{1}{2}ab &= n \end{aligned}$$

has a solution with $a, b, c \in \mathbf{Q}$.

For example, 6 is the area of the right triangle with side lengths 3, 4, and 5, so 6 is a congruent number. Less obvious is that 5 is also a congruent number; it is the area of the right triangle with side lengths 3/2, 20/3, and 41/6. It is nontrivial to prove that 1, 2, 3, and 4 are not congruent numbers. Here is a list of the integer congruent numbers up to 50:

$$5, 6, 7, 13, 14, 15, 20, 21, 22, 23, 24, 28, 29, 30, 31, 34, 37, 38, 39, 41, 45, 46, 47.$$

Every congruence class modulo 8 except 3 is represented in this list, which incorrectly suggests that if $n \equiv 3 \pmod 8$ then $n$ is not a congruent number. Though no $n \leq 218$ with $n \equiv 3 \pmod 8$ is a congruent number, $n = 219$ is a congruent number and $219 \equiv 3 \pmod 8$.

Deciding whether an integer $n$ is a congruent number can be subtle, since the simplest triangle with area $n$ can be very complicated. For example,

as Zagier pointed out, the number 157 is a congruent number, and the "simplest" rational right triangle with area 157 has side lengths

$$a = \frac{6803298487826435051217540}{411340519227716149383203} \text{ and } b = \frac{411340519227716149383203}{21666555693714761309610}.$$

This solution would be difficult to find by a brute force search.

We call congruent numbers "congruent" because of the following proposition, which asserts that any congruent number is the common "congruence" between three perfect squares.

**Proposition 6.5.7.** *Suppose $n$ is the area of a right triangle with rational side lengths $a, b, c$, with $a \leq b < c$. Let $A = (c/2)^2$. Then*

$$A - n, \quad A, \text{ and } A + n$$

*are all perfect squares of rational numbers.*

*Proof.* We have

$$\begin{aligned} a^2 + b^2 &= c^2 \\ \frac{1}{2}ab &= n \end{aligned}$$

Add or subtract 4 times the second equation to the first to get

$$\begin{aligned} a^2 \pm 2ab + b^2 &= c^2 \pm 4n \\ (a \pm b)^2 &= c^2 \pm 4n \\ \left(\frac{a \pm b}{2}\right)^2 &= \left(\frac{c}{2}\right)^2 \pm n \\ &= A \pm n \end{aligned}$$

□

The main motivating open problem related to congruent numbers is to give a systematic way to recognize them.

**Open Problem 6.5.8.** *Give an algorithm which, given $n$, outputs whether or not $n$ is a congruent number.*

Fortunately, the vast theory developed about elliptic curves has something to say about the above problem. In order to understand this connection, we begin with an elementary algebraic proposition that establishes a link between elliptic curves and the congruent number problem.

**Proposition 6.5.9** (Congruent numbers and elliptic curves)**.** *Let $n$ be a rational number. There is a bijection between*

$$A = \left\{(a, b, c) \in \mathbf{Q}^3 \,:\, \frac{ab}{2} = n,\, a^2 + b^2 = c^2\right\}$$

*and*

$$B = \left\{(x, y) \in \mathbf{Q}^2 \,:\, y^2 = x^3 - n^2 x, \text{ with } y \neq 0\right\}$$

*given explicitly by the maps*

$$f(a, b, c) = \left(-\frac{nb}{a+c},\ \frac{2n^2}{a+c}\right)$$

*and*

$$g(x, y) = \left(\frac{n^2 - x^2}{y},\ -\frac{2xn}{y},\ \frac{n^2 + x^2}{y}\right).$$

The proof of this proposition is not deep, but involves substantial (elementary) algebra and we will not prove it in this book.

For $n \neq 0$, let $E_n$ be the elliptic curve $y^2 = x^3 - n^2 x$.

**Proposition 6.5.10** (Congruent number criterion)**.** *The rational number $n$ is a congruent number if and only if there is a point $P = (x, y) \in E_n(\mathbf{Q})$ with $y \neq 0$.*

*Proof.* The number $n$ is a congruent number if and only if the set $A$ from Proposition 6.5.9 is nonempty. By the proposition $A$ is nonempty if and only if $B$ is nonempty. □

*Example* 6.5.11. Let $n = 5$. Then $E_n$ is $y^2 = x^3 - 25x$, and we notice that $(-4, -6) \in E_n(\mathbf{Q})$. We next use the bijection of Proposition 6.5.9 to find the corresponding right triangle:

$$g(-4, -6) = \left(\frac{25 - 16}{-6}, -\frac{-40}{-6}, \frac{25 + 16}{-6}\right) = \left(-\frac{3}{2}, -\frac{20}{3}, -\frac{41}{6}\right).$$

Multiplying through by $-1$ yields the side lengths of a rational right triangle with area 5. *Are there any others?*

Observe that we can apply $g$ to any point in $E_n(\mathbf{Q})$ with $y \neq 0$. Using the group law, we find that $2(-4, -6) = (1681/144, 62279/1728)$ and

$$g(2(-4, -6)) = \left(-\frac{1519}{492}, -\frac{4920}{1519}, \frac{3344161}{747348}\right).$$

This example foreshadows Theorem 6.5.14.

*Example* 6.5.12. Let $n = 1$, so $E_1$ is defined by $y^2 = x^3 - x$. Since 1 is not a congruent number, the elliptic curve $E_1$ has no point with $y \neq 0$. See Exercise 6.11.

*SAGE Example* 6.5.13. We implement the `cong` function in Sage, which returns a triple $(a, b, c)$ whose entries are the sides of a rational right triangle of area $n$ if one exists, and returns False if there are no such triangles.

```
sage: def cong(n):
...       G = EllipticCurve([-n^2,0]).gens()
...       if len(G) == 0: return False
...       x,y,_ = G[0]
...       return ((n^2-x^2)/y,-2*x*n/y,(n^2+x^2)/y)
sage: cong(6)
(3, 4, 5)
sage: cong(5)
(3/2, 20/3, 41/6)
sage: cong(1)
False
sage: cong(13)
(323/30, 780/323, 106921/9690)
sage: (323/30 * 780/323)/2
13
sage: (323/30)^2 + (780/323)^2 == (106921/9690)^2
True
```

**Theorem 6.5.14** (Infinitely Many Triangles)**.** *If $n$ is a congruent number, then there are infinitely many distinct right triangles with rational side lengths and area $n$.*

We will not prove this theorem, except to note that one proves it by showing that $E_n(\mathbf{Q})_{\text{tor}} = \{\mathcal{O}, (0,0), (n,0), (-n,0)\}$, so the elements of the set $B$ in Proposition 6.5.9 all have infinite order. Hence, $B$ is infinite so $A$ is infinite.

Tunnell has proved that the Birch and Swinnerton-Dyer conjecture (alluded to above), implies the existence of an elementary way to decide whether or not an integer $n$ is a congruent number. We state Tunnell's elementary way in the form of a conjecture.

**Conjecture 6.5.15.** *Let $a, b, c$ denote integers. If $n$ is an even square-free integer, then $n$ is a congruent number if and only if*

$$\#\left\{(a,b,c) \in \mathbf{Z}^3 : 4a^2 + b^2 + 8c^2 = \frac{n}{2} : c \text{ is even}\right\}$$
$$= \#\left\{(a,b,c) : 4a^2 + b^2 + 8c^2 = \frac{n}{2} : c \text{ is odd}\right\}.$$

*If $n$ is odd and square free then $n$ is a congruent number if and only if*

$$\#\left\{(a,b,c) : 2a^2 + b^2 + 8c^2 = n : c \text{ is even}\right\}$$
$$= \#\left\{(a,b,c) : 2a^2 + b^2 + 8c^2 = n : c \text{ is odd}\right\}.$$

Enough of the Birch and Swinnerton-Dyer conjecture is known to prove one direction of Conjecture 6.5.15. In particular, it is a very deep theorem that if we do not have equality of the displayed cardinalities, then $n$ is not a congruent number.

The even more difficult (and still open!) part of Conjecture 6.5.15 is the converse: If one has equality of the displayed cardinalities, prove that $n$ is a congruent number. The difficulty in this direction, which appears to be very deep, is that we must somehow construct (or prove the existence of) elements of $E_n(\mathbf{Q})$. This has been accomplished in some cases due to the groundbreaking work of Gross and Zagier ([GZ86]) but much work remains to be done.

The excellent book [Kob84] is about congruent numbers and Conjecture 6.5.15, and we encourage the reader to consult it. The Birch and Swinnerton-Dyer conjecture is a Clay Math Institute million dollar millennium prize problem (see [Cla, Wil00]).

## 6.6 Exercises

6.1 Write down an equation $y^2 = x^3 + ax + b$ over a field $K$ such that $-16(4a^3+27b^2) = 0$. Precisely what goes wrong when trying to endow the set $E(K) = \{(x, y) \in K \times K : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\}$ with a group structure?

6.2 One rational solution to the equation $y^2 = x^3 - 2$ is $(3, 5)$. Find a rational solution with $x \neq 3$ by drawing the tangent line to $(3, 5)$ and computing the second point of intersection.

6.3 Let $E$ be the elliptic curve over the finite field $K = \mathbf{Z}/5\mathbf{Z}$ defined by the equation

$$y^2 = x^3 + x + 1.$$

(a) List all 9 elements of $E(K)$.

(b) What is the structure of $E(K)$, as a product of cyclic groups?

6.4 Let $E$ be the elliptic curve defined by the equation $y^2 = x^3 + 1$. For each prime $p \geq 5$, let $N_p$ be the cardinality of the group $E(\mathbf{Z}/p\mathbf{Z})$ of points on this curve having coordinates in $\mathbf{Z}/p\mathbf{Z}$. For example, we have that $N_5 = 6, N_7 = 12, N_{11} = 12, N_{13} = 12, N_{17} = 18, N_{19} = 12, , N_{23} = 24,$ and $N_{29} = 30$ (you do not have to prove this).

(a) For the set of primes satisfying $p \equiv 2 \pmod 3$, can you see a pattern for the values of $N_p$? Make a general conjecture for the value of $N_p$ when $p \equiv 2 \pmod 3$.

(b) (*) Prove your conjecture.

6.5 Let $E$ be an elliptic curve over the real numbers $\mathbf{R}$. Prove that $E(\mathbf{R})$ is not a finitely generated abelian group.

6.6 (*) Suppose $G$ is a finitely generated abelian group. Prove that the subgroup $G_{\text{tor}}$ of elements of finite order in $G$ is finite.

6.7 Suppose $y^2 = x^3 + ax + b$ with $a, b \in \mathbf{Q}$ defines an elliptic curve. Show that there is another equation $Y^2 = X^3 + AX + B$ with $A, B \in \mathbf{Z}$ whose solutions are in bijection with the solutions to $y^2 = x^3 + ax + b$.

6.8 Suppose $a$, $b$, $c$ are relatively prime integers with $a^2 + b^2 = c^2$. Then there exist integers $x$ and $y$ with $x > y$ such that $c = x^2 + y^2$ and either $a = x^2 - y^2$, $b = 2xy$ or $a = 2xy$, $b = x^2 - y^2$.

6.9 (*) Fermat's Last Theorem for exponent 4 asserts that any solution to the equation $x^4 + y^4 = z^4$ with $x, y, z \in \mathbf{Z}$ satisfies $xyz = 0$. Prove Fermat's Last Theorem for exponent 4, as follows.

(a) Show that if the equation $x^2 + y^4 = z^4$ has no integer solutions with $xyz \neq 0$, then Fermat's Last Theorem for exponent 4 is true.

(b) Prove that $x^2 + y^4 = z^4$ has no integer solutions with $xyz \neq 0$ as follows. Suppose $n^2 + k^4 = m^4$ is a solution with $m > 0$ minimal among all solutions. Show that there exists a solution with $m$ smaller using Exercise 6.8 (consider two cases).

6.10 This problem requires a computer.

(a) Show that the set of numbers $59 + 1 \pm s$ for $s \leq 15$ contains 14 numbers that are $B$-power smooth for $B = 20$.

(b) Find the proportion of primes $p$ in the interval from $10^{12}$ and $10^{12} + 1000$ such that $p - 1$ is $B = 10^5$ power smooth.

6.11 (*) Prove that 1 is not a congruent number by showing that the elliptic curve $y^2 = x^3 - x$ has no rational solutions except $(0, \pm 1)$ and $(0, 0)$, as follows:

(a) Write $y = \frac{p}{q}$ and $x = \frac{r}{s}$, where $p, q, r, s$ are all positive integers and $\gcd(p, q) = \gcd(r, s) = 1$. Prove that $s \mid q$, so $q = sk$ for some $k \in \mathbf{Z}$.

(b) Prove that $s = k^2$, and substitute to see that $p^2 = r^3 - rk^4$.

(c) Prove that $r$ is a perfect square by supposing that there is a prime $\ell$ such that $\mathrm{ord}_\ell(r)$ is odd, and analyzing $\mathrm{ord}_\ell$ of both sides of $p^2 = r^3 - rk^4$.

(d) Write $r = m^2$, and substitute to see that $p^2 = m^6 - m^2k^4$. Prove that $m \mid p$.

(e) Divide through by $m^2$ and deduce a contradiction to Exercise 6.9.

# Answers and Hints

- **Chapter 1. Prime Numbers**

  2. They are $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97$.
  3. Emulate the proof of Proposition 1.2.5.

- **Chapter 2. The Ring of Integers Modulo $n$**

  2. They are 5, 13, 3, and 8.
  3. For example, $x = 22$, $y = -39$.
  4. Hint: Use the binomial theorem and prove that if $r \geq 1$, then $p$ divides $\binom{p}{r}$.
  7. For example, $S_1 = \{0, 1, 2, 3, 4, 5, 6\}$, $S_2 = \{1, 3, 5, 7, 9, 11, 13\}$, $S_3 = \{0, 2, 4, 6, 8, 10, 12\}$, and $S_4 = \{2, 3, 5, 7, 11, 13, 29\}$. In each we find $S_i$ by listing the first seven numbers satisfying the $i$th condition, then adjust the last number if necessary so that the reductions will be distinct modulo 7.
  8. An integer is divisible by 5 if and only if the last digits is 0 or 5. An integer is divisible by 9 if and only if the sum of the digits is divisible by 9. An integer is divisible by 11 if and only if the alternating sum of the digits is divisible by 11.
  9. Hint for part (a): Use the divisibility rule you found in Exercise 1.8.

10. 71

11. 8

12. As explained on page 23, we know that $\mathbf{Z}/n\mathbf{Z}$ is a ring for any $n$. Thus to show that $\mathbf{Z}/p\mathbf{Z}$ is a field it suffices to show that every nonzero element $\overline{a} \in \mathbf{Z}/p\mathbf{Z}$ has an inverse. Lift $a$ to an element $a \in \mathbf{Z}$, and set $b = p$ in Proposition 2.3.1. Because $p$ is prime, $\gcd(a, p) = 1$, so there exists $x, y$ such that $ax+py = 1$. Reducing this equality modulo $p$ proves that $\overline{a}$ has an inverse $x \pmod{p}$. Alternatively, one could argue just like after Definition 2.1.16 that $\overline{a}^m = 1$ for some $m$, so some power of $\overline{a}$ is the inverse of $\overline{a}$.

13. 302

15. Only for $n = 1, 2$. If $n > 2$, then $n$ is either divisible by an odd prime $p$ or 4. If $4 \mid n$, then $2^e - 2^{e-1}$ divides $\varphi(n)$ for some $e \geq 2$, so $\varphi(n)$ is even. If an odd $p$ divides $n$, then the even number $p^e - p^{e-1}$ divides $\varphi(n)$ for some $e \geq 1$.

16. The map $\psi$ is a homomorphism since both reduction maps

$$\mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \quad \text{and} \quad \mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$$

are homomorphisms. It is injective because if $a \in \mathbf{Z}$ is such that $\psi(a) = 0$, then $m \mid a$ and $n \mid a$, so $mn \mid a$ (since $m$ and $n$ are coprime), so $a \equiv 0 \pmod{mn}$. The cardinality of $\mathbf{Z}/mn\mathbf{Z}$ is $mn$ and the cardinality of the product $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/n\mathbf{Z}$ is also $mn$, so $\psi$ must be an isomorphism. The units $(\mathbf{Z}/mn\mathbf{Z})^*$ are thus in bijection with the units $(\mathbf{Z}/m\mathbf{Z})^* \times (\mathbf{Z}/n\mathbf{Z})^*$.

For the second part of the exercise, let $g = \gcd(m, n)$ and set $a = mn/g$. Then $a \not\equiv 0 \pmod{mn}$, but $m \mid a$ and $n \mid a$, so $a \ker(\psi)$.

17. We express the question as a system of linear equations modulo various numbers, and use the Chinese remainder theorem. Let $x$ be the number of books. The problem asserts that

$$\begin{aligned} x &\equiv 6 \pmod{7} \\ x &\equiv 2 \pmod{6} \\ x &\equiv 1 \pmod{5} \\ x &\equiv 0 \pmod{4} \end{aligned}$$

Applying CRT to the first pair of equations, we find that $x \equiv 20 \pmod{42}$. Applying CRT to this equation and the third, we find that $x \equiv 146 \pmod{210}$. Since 146 is not divisible by 4, we add multiples of 210 to 146 until we find the first $x$ that is divisible by 4. The first multiple works, and we find that the aspiring mathematicians have 356 math books.

18. Note that $p = 3$ works, since $11 = 3^2 + 2$ is prime. Now suppose $p \neq 3$ is any prime such that $p$ and $p^2 + 2$ are both prime. We must have $p \equiv 1 \pmod 3$ or $p \equiv 2 \pmod 3$. Then $p^2 \equiv 1 \pmod 3$, so $p^2 + 2 \equiv 0 \pmod 3$. Since $p^2 + 2$ is prime, we must have $p^2 + 2 = 3$, so $p = 1$, a contradiction as $p$ is assumed prime.

19. For (a) $n = 1, 2$, see solution to Exercise 2.15. For (b), yes there are many such examples. For example, $m = 2$, $n = 4$.

20. By repeated application of multiplicativity and Equation (2.2.2) on page 31, we see that if $n = \prod_i p_i^{e_i}$ is the prime factorization of $n$, then

$$\varphi(n) = \prod_i (p_i^{e_i} - p_i^{e_i - 1}) = \prod_i p_i^{e_i - 1} \cdot \prod_i (p_i - 1).$$

23. 1, 6, 29, 34

24. Let $g = \gcd(12n+1, 30n+2)$. Then $g \mid 30n+2-2\cdot(12n+1) = 6n$. For the same reason, $g$ also divides $12n + 1 - 2 \cdot (6n) = 1$, so $g = 1$, as claimed.

27. There is no primitive root modulo 8, since $(\mathbf{Z}/8\mathbf{Z})^*$ has order 4, but every element of $(\mathbf{Z}/8\mathbf{Z})^*$ has order 2. Prove that if $\zeta$ is a primitive root modulo $2^n$, for $n \geq 3$, then the reduction of $\zeta$ mod 8 is a primitive root, a contradiction.

28. 2 is a primitive root modulo 125.

29. Let $\prod_{i=1}^m p_i^{e_i}$ be the prime factorization of $n$. Slightly generalizing Exercise 16, we see that

$$(\mathbf{Z}/n\mathbf{Z})^* \cong \prod (\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*.$$

Thus $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if the product $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ is cyclic. If $8 \mid n$, then there is no chance $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic, so assume $8 \nmid n$. Then by Exercise 2.28, each group $(\mathbf{Z}/p_i^{e_i}\mathbf{Z})^*$ is itself cyclic. A product of cyclic groups is cyclic if and only the orders of the factors in the product are coprime (this follows from Exercise 2.16). Thus $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if the numbers $p_i(p_i - 1)$, for $i = 1, \ldots, m$ are pairwise coprime. Since $p_i - 1$ is even, there can be at most one odd prime in the factorization of $n$, and we see that $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic if and only if $n$ is an odd prime power, twice an odd prime power, or $n = 4$.

## • Chapter 3. Public-Key Cryptography

1. The best case is that each letter is A. Then the question is to find the largest $n$ such that $1 + 27 + \cdots + 27^n \leq 10^{20}$. By computing

$\log_{27}(10^{20})$, we see that $27^{13} < 10^{20}$ and $27^{14} > 10^{20}$. Thus $n \leq 13$, and since $1+27+\cdots+27^{n-1} < 27^n$, and $2 \cdot 27^{13} < 10^{20}$, it follows that $n = 13$.

2. This is not secure, since it is just equivalent to a "Ceaser Cipher," that is a permutation of the letters of the alphabet, which is well-known to be easily broken using a frequency analysis.

3. If we can compute the polynomial

$$f = (x-p)(x-q)(x-r) = x^3-(p+q+r)x^2+(pq+pr+qr)x-pqr,$$

then we can factor $n$ by finding the roots of $f$, for example, using Newton's method (or Cardona's formula for the roots of a cubic). Because $p$, $q$, $r$, are distinct odd primes, we have

$$\varphi(n) = (p-1)(q-1)(r-1) = pqr - (pq+pr+qr) + p + q + r,$$

and

$$\sigma(n) = 1 + (p+q+r) + (pq+pr+qr) + pqr.$$

Since we know $n$, $\varphi(n)$, and $\sigma(n)$, we know

$$\begin{aligned} \sigma(n) - 1 - n &= (p+q+r) + (pq+pr+qr), \quad \text{and} \\ \varphi(n) - n &= (p+q+r) - (pq+pr+qr). \end{aligned}$$

We can thus compute both $p+q+r$ and $pq+pr+qr$, hence deduce $f$ and find $p, q, r$.

## • Chapter 4. Quadratic Reciprocity

1. They are all 1, $-1$, 0, and 1.

3. By Proposition 4.3.4, the value of $\left(\frac{3}{p}\right)$ depends only on the reduction $\pm p \pmod{12}$. List enough primes $p$ such that $\pm p$ reduce to $1, 5, 7, 11$ modulo 12 and verify that the asserted formula holds for each of them.

7. Since $p = 2^{13} - 1$ is prime, there are either two solutions or no solutions to $x^2 \equiv 5 \pmod{p}$, and we can decide which using quadratic reciprocity. We have

$$\left(\frac{5}{p}\right) = (-1)^{(p-1)/2\cdot(5-1)/2}\left(\frac{p}{5}\right) = \left(\frac{p}{5}\right),$$

so there are two solutions if and only if $p = 2^{13}-1$ is $\pm 1$ mod 5. In fact, $p \equiv 1 \pmod{5}$, so there are two solutions.

8. We have $4^{48} = 2^{96}$. By Euler's Theorem, $2^{96} = 1$, so $x = 1$.

9. For (a), take $a = 19$ and $n = 20$. We found this example using the Chinese remainder theorem applied to $4 \pmod 5$ and $3 \pmod 4$, and used that $\left(\frac{19}{20}\right) = \left(\frac{19}{5}\right) \cdot \left(\frac{19}{4}\right) = (-1)(-1) = 1$, yet 19 is not a square modulo either 5 or 4, so is certainly not a square modulo 20.
10. Hint: First reduce to the case that $6k - 1$ is prime, by using that if $p$ and $q$ are primes not of the form $6k - 1$, then neither is their product. If $p = 6k - 1$ divides $n^2 + n + 1$, it divides $4n^2 + 4n + 4 = (2n + 1)^2 + 3$, so $-3$ is a quadratic residue modulo $p$. Now use quadratic reciprocity to show that $-3$ is not a quadratic residue modulo $p$.

## • Chapter 5. Continued Fractions

9. Suppose $n = x^2 + y^2$, with $x, y \in \mathbf{Q}$. Let $d$ be such that $dx, dy \in \mathbf{Z}$. Then $d^2n = (dx)^2 + (dy)^2$ is a sum of two integer squares, so by Theorem 5.7.1, if $p \mid d^2n$ and $p \equiv 3 \pmod 4$, then $\mathrm{ord}_p(d^2n)$ is even. We have $\mathrm{ord}_p(d^2n)$ is even if and only if $\mathrm{ord}_p(n)$ is even, so Theorem 5.7.1 implies that $n$ is also a sum of two squares.
11. The squares modulo 8 are $0, 1, 4$, so a sum of two squares reduces modulo 8 to one of $0, 1, 2, 4$, or 5. Four consecutive integers that are sums of squares would reduce to four consecutive integers in the set $\{0, 1, 2, 4, 5\}$, which is impossible.

## • Chapter 6. Elliptic Curves

2. The second point of intersection is $(129/100, 383/1000)$.
3. The group is cyclic of order 9, generated by $(4, 2)$. The elements of $E(K)$ are

$$\{\mathcal{O}, (4, 2), (3, 4), (2, 4), (0, 4), (0, 1), (2, 1), (3, 1), (4, 3)\}.$$

4. In part (a), the pattern is that $N_p = p + 1$. For part (b), a hint is that when $p \equiv 2 \pmod 3$, the map $x \mapsto x^3$ on $(\mathbf{Z}/p\mathbf{Z})^*$ is an automorphism, so $x \mapsto x^3 + 1$ is a bijection. Now use what you learned about squares in $\mathbf{Z}/p\mathbf{Z}$ from Chapter 4.
5. For all sufficiently large real $x$, the equation $y^2 = x^3 + ax + b$ has a real solution $y$. Thus, the group $E(\mathbf{R})$ is not countable, since $\mathbf{R}$ is not countable. But any finitely generated group is countable.
6. In a course on abstract algebra, one often proves the nontrivial fact that every subgroup of a finitely generated abelian group is finitely generated. In particular, the torsion subgroup $G_{\mathrm{tor}}$ is

finitely generated. However, a finitely generated abelian torsion group is finite.

7. Hint: Multiply both sides of $y^2 = x^3 + ax + b$ by a power of a common denominator, and "absorb" powers into $x$ and $y$.
8. Hint: see Exercise 4.6.

# References

[ACD+99] K. Aardal, S. Cavallar, B. Dodson, A. Lenstra, W. Lioen, P. L. Montgomery, B. Murphy, J. Gilchrist, G. Guillerm, P. Leyland, J. Marchand, F. Morain, A. Muffett, C.&C. Putnam, and P. Zimmermann, *Factorization of a 512-bit RSA key using the Number Field Sieve*, `http://www.loria.fr/~zimmerma/records/RSA155` (1999).

[AGP94] W. R. Alford, Andrew Granville, and Carl Pomerance, *There are infinitely many Carmichael numbers*, Ann. of Math. (2) **139** (1994), no. 3, 703–722. MR 95k:11114

[AKS02] M. Agrawal, N. Kayal, and N. Saxena, *PRIMES is in P*, to appear in Annals of Math., `http://www.cse.iitk.ac.in/users/manindra/primality.ps` (2002).

[BS76] Leonard E. Baum and Melvin M. Sweet, *Continued fractions of algebraic power series in characteristic* 2, Ann. of Math. (2) **103** (1976), no. 3, 593–610. MR 53 #13127

[Bur89] D. M. Burton, *Elementary Number Theory*, second ed., W. C. Brown Publishers, Dubuque, IA, 1989. MR 90e:11001

[Cal] C. Caldwell, *The Largest Known Primes,* `http://www.utm.edu/research/primes/largest.html`.

[Cer] Certicom, *The certicom ECC challenge,*
`http://www.certicom.com/`
`index.php?action=res,ecc_challenge`.

[Cla] Clay Mathematics Institute, *Millennium prize problems,*
`http://www.claymath.org/millennium_prize_problems/`.

[Coh] H. Cohn, *A short proof of the continued fraction expansion of e,*
`http://research.microsoft.com/~cohn/publications.html`.

[Coh93] H. Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics, vol. 138, Springer-Verlag, Berlin, 1993. MR 94i:11105

[Con97] John H. Conway, *The Sensual (Quadratic) Form*, Carus Mathematical Monographs, vol. 26, Mathematical Association of America, Washington, DC, 1997, With the assistance of Francis Y. C. Fung. MR 98k:11035

[CP01] R. Crandall and C. Pomerance, *Prime Numbers*, Springer-Verlag, New York, 2001, A computational perspective. MR 2002a:11007

[Cre] J. E. Cremona, `mwrank` *(computer software)*,
`http://www.maths.nott.ac.uk/personal/jec/ftp/progs/`.

[Cre97] ———, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[Dav99] H. Davenport, *The Higher Arithmetic*, seventh ed., Cambridge University Press, Cambridge, 1999, An introduction to the theory of numbers, Chapter VIII by J. H. Davenport. MR 2000k:11002

[DH76] W. Diffie and M. E. Hellman, *New directions in cryptography*, IEEE Trans. Information Theory **IT-22** (1976), no. 6, 644–654. MR 55 #10141

[Eul85] Leonhard Euler, *An essay on continued fractions*, Math. Systems Theory **18** (1985), no. 4, 295–328, Translated from the Latin by B. F. Wyman and M. F. Wyman. MR 87d:01011b

[FT93] A. Fröhlich and M. J. Taylor, *Algebraic Number Theory*, Cambridge University Press, Cambridge, 1993. MR 94d:11078

[Guy94] R. K. Guy, *Unsolved Problems in Number Theory*, second ed., Springer-Verlag, New York, 1994, Unsolved Problems in Intuitive Mathematics, I. MR 96e:11002

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[Har77] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[Hoo67] C. Hooley, *On Artin's conjecture*, J. Reine Angew. Math. **225** (1967), 209–220. MR 34 #7445

[HW79] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, fifth ed., The Clarendon Press Oxford University Press, New York, 1979. MR 81i:10002

[IBM01] IBM, *IBM's Test-Tube Quantum Computer Makes History,* `http://www.research.ibm.com/resources/news/20011219_quantum.shtml`.

[IR90] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, second ed., Springer-Verlag, New York, 1990. MR 92e:11001

[Khi63] A. Ya. Khintchine, *Continued fractions*, Translated by Peter Wynn, P. Noordhoff Ltd., Groningen, 1963. MR 28 #5038

[Knu97] Donald E. Knuth, *The Art of Computer Programming*, third ed., Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1997, Volume 1: Fundamental algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[Knu98] ———, *The Art of Computer Programming. Vol. 2*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1998, Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing. MR 83i:68003

[Kob84] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Graduate Texts in Mathematics, vol. 97, Springer-Verlag, New York, 1984. MR 86c:11040

[Leh14] D. N. Lehmer, *List of Primes Numbers from* 1 *to* 10,006,721, Carnegie Institution Washington, D.C. (1914).

[Lem] F. Lemmermeyer, *Proofs of the Quadratic Reciprocity Law,* `http://www.rzuser.uni-heidelberg.de/~hb3/rchrono.html`.

[Len87] H. W. Lenstra, Jr., *Factoring integers with elliptic curves*, Ann. of Math. (2) **126** (1987), no. 3, 649–673. MR 89g:11125

[LL93] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The Development of the Number Field Sieve*, Lecture Notes in Mathematics, vol. 1554, Springer-Verlag, Berlin, 1993. MR 96m:11116

[LMG+01] Vandersypen L. M., Steffen M., Breyta G., Yannoni C. S., Sherwood M. H., and Chuang I. L., *Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance*, Nature **414** (2001), no. 6866, 883–887.

[LT72] S. Lang and H. Trotter, *Continued fractions for some algebraic numbers*, J. Reine Angew. Math. **255** (1972), 112–134; addendum, ibid. **267** (1974), 219–220; MR **50** #2086. MR 46 #5258

[LT74] ———, *Addendum to: Continued fractions for some algebraic numbers (J. Reine Angew. Math.* **255** *(1972), 112–134)*, J. Reine Angew. Math. **267** (1974), 219–220. MR 50 #2086

[Mor93] P. Moree, *A note on Artin's conjecture*, Simon Stevin **67** (1993), no. 3-4, 255–257. MR 95e:11106

[MS08] B. Mazur and W. Stein, *What is Riemann's Hypothesis?*, 2008, In preparation.

[NZM91] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory of Numbers*, fifth ed., John Wiley & Sons Inc., New York, 1991. MR 91i:11001

[Old70] C. D. Olds, *The Simple Continued Fraction Expression of e*, Amer. Math. Monthly **77** (1970), 968–974.

[Per57] O. Perron, *Die Lehre von den Kettenbrüchen. Dritte, verbesserte und erweiterte Aufl. Bd. II. Analytisch-funktionentheoretische Kettenbrüche*, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1957. MR 19,25c

[RSA] RSA, *The New RSA Factoring Challenge,* `http://www.rsasecurity.com/rsalabs/challenges/factoring`.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Comm. ACM **21** (1978), no. 2, 120–126. MR 83m:94003

[Sag08] Sage, *Free Open Source Mathematical Software (Version 3.0.4)*, 2008, `http://www.sagemath.org`.

[Sch85] R. Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*, Mathematics of Computation **44** (1985), no. 170, 483–494.

[Sho97] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput. **26** (1997), no. 5, 1484–1509. MR 98i:11108

[Sho05] V. Shoup, *A Computational Introduction to Number Theory and Algebra*, Cambridge University Press, Cambridge, 2005. MR MR2151586 (2006g:11003)

[Sil86] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, New York, 1986. MR 87g:11070

[Sin99] S. Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, Doubleday, 1999.

[Slo] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

[ST92] J. H. Silverman and J. Tate, *Rational points on elliptic curves*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1992. MR 93g:11003

[Wal48] H. S. Wall, *Analytic Theory of Continued Fractions*, D. Van Nostrand Company, Inc., New York, N. Y., 1948. MR 10,32d

[Wei03] E. W. Weisstein, *RSA-576 Factored,* `http://mathworld.wolfram.com/news/2003-12-05/rsa/`.

[Wil00] A. J. Wiles, *The Birch and Swinnerton-Dyer Conjecture*, `http://www.claymath.org/prize_problems/birchsd.htm`.

[Zag75] D. Zagier, *The first 50 million prime numbers*, `http://modular.fas.harvard.edu/scans/papers/zagier/`.

# Index