4

Number Theory

INTEGERS ARE CENTRAL to the discrete mathematics we are emphasizing in this book. Therefore we want to explore the *theory of numbers*, an important branch of mathematics concerned with the properties of integers.

We tested the number theory waters in the previous chapter, by introducing binary operations called 'mod' and 'gcd'. Now let's plunge in and really immerse ourselves in the subject.

In other words, be prepared to drown.

4.1 DIVISIBILITY

We say that m divides n (or n is divisible by m) if $m > 0$ and the ratio n/m is an integer. This property underlies all of number theory, so it's convenient to have a special notation for it. We therefore write

$$m \backslash n \quad \Longleftrightarrow \quad m > 0 \text{ and } n = mk \text{ for some integer } k. \tag{4.1}$$

(The notation '$m|n$' is actually much more common than '$m\backslash n$' in current mathematics literature. But vertical lines are overused—for absolute values, set delimiters, conditional probabilities, etc.—and backward slashes are underused. Moreover, '$m\backslash n$' gives an impression that m is the denominator of an implied ratio. So we shall boldly let our divisibility symbol lean leftward.)

If m does not divide n we write '$m \not\backslash n$'.

There's a similar relation, "n is a multiple of m," which means almost the same thing except that m doesn't have to be positive. In this case we simply mean that $n = mk$ for some integer k. Thus, for example, there's only one multiple of 0 (namely 0), but nothing is divisible by 0. Every integer is a multiple of -1, but no integer is divisible by -1 (strictly speaking). These definitions apply when m and n are any real numbers; for example, 2π is divisible by π. But we'll almost always be using them when m and n are integers. After all, this is number theory.

"... no integer is divisible by -1 (strictly speaking)."
—Graham, Knuth, and Patashnik [161]

In Britain we call this 'hcf' (highest common factor).

The *greatest common divisor* of two integers m and n is the largest integer that divides them both:

$$\gcd(m,n) \;=\; \max\{\,k \mid k\backslash m \quad\text{and}\quad k\backslash n\,\}. \tag{4.2}$$

For example, $\gcd(12,18) = 6$. This is a familiar notion, because it's the common factor that fourth graders learn to take out of a fraction m/n when reducing it to lowest terms: $12/18 = (12/6)/(18/6) = 2/3$. Notice that if $n > 0$ we have $\gcd(0,n) = n$, because any positive number divides 0, and because n is the largest divisor of itself. The value of $\gcd(0,0)$ is undefined.

Not to be confused with the greatest common multiple.

Another familiar notion is the *least common multiple*,

$$\operatorname{lcm}(m,n) \;=\; \min\{\,k \mid k>0, \quad m\backslash k \quad\text{and}\quad n\backslash k\,\}; \tag{4.3}$$

this is undefined if $m \leqslant 0$ or $n \leqslant 0$. Students of arithmetic recognize this as the least common denominator, which is used when adding fractions with denominators m and n. For example, $\operatorname{lcm}(12,18) = 36$, and fourth graders know that $\frac{7}{12} + \frac{1}{18} = \frac{21}{36} + \frac{2}{36} = \frac{23}{36}$. The lcm is somewhat analogous to the gcd, but we don't give it equal time because the gcd has nicer properties.

One of the nicest properties of the gcd is that it is easy to compute, using a 2300-year-old method called *Euclid's algorithm*. To calculate $\gcd(m,n)$, for given values $0 \leqslant m < n$, Euclid's algorithm uses the recurrence

$$\begin{aligned} \gcd(0,n) \;&=\; n\,; \\ \gcd(m,n) \;&=\; \gcd(n \bmod m,\, m)\,, \qquad \text{for } m>0. \end{aligned} \tag{4.4}$$

Thus, for example, $\gcd(12,18) = \gcd(6,12) = \gcd(0,6) = 6$. The stated recurrence is valid, because any common divisor of m and n must also be a common divisor of both m and the number $n \bmod m$, which is $n - \lfloor n/m \rfloor m$. There doesn't seem to be any recurrence for $\operatorname{lcm}(m,n)$ that's anywhere near as simple as this. (See exercise 2.)

Euclid's algorithm also gives us more: We can extend it so that it will compute integers m' and n' satisfying

$$m'm + n'n \;=\; \gcd(m,n)\,. \tag{4.5}$$

(Remember that m' or n' can be negative.)

Here's how. If $m = 0$, we simply take $m' = 0$ and $n' = 1$. Otherwise we let $r = n \bmod m$ and apply the method recursively with r and m in place of m and n, computing $\bar{r}$ and $\bar{m}$ such that

$$\bar{r}\,r + \bar{m}\,m \;=\; \gcd(r,m)\,.$$

Since $r = n - \lfloor n/m \rfloor m$ and $\gcd(r,m) = \gcd(m,n)$, this equation tells us that

$$\bar{r}\,(n - \lfloor n/m \rfloor m) + \bar{m}\,m \;=\; \gcd(m,n)\,.$$

The left side can be rewritten to show its dependency on m and n:

$$(\overline{m} - \lfloor n/m \rfloor \overline{r})\, m + \overline{r}\, n \;=\; \gcd(m,n)\,;$$

hence $m' = \overline{m} - \lfloor n/m \rfloor \overline{r}$ and $n' = \overline{r}$ are the integers we need in (4.5). For example, in our favorite case $m = 12$, $n = 18$, this method gives $6 = 0{\cdot}0 + 1{\cdot}6 = 1\cdot 6 + 0\cdot 12 = (-1)\cdot 12 + 1\cdot 18$.

But why is (4.5) such a neat result? The main reason is that there's a sense in which the numbers m' and n' actually *prove* that Euclid's algorithm has produced the correct answer in any particular case. Let's suppose that our computer has told us after a lengthy calculation that $\gcd(m,n) = d$ and that $m'm + n'n = d$; but we're skeptical and think that there's really a greater common divisor, which the machine has somehow overlooked. This cannot be, however, because any common divisor of m and n has to divide $m'm + n'n$; so it has to divide d; so it has to be $\leqslant d$. Furthermore we can easily check that d does divide both m and n. (Algorithms that output their own proofs of correctness are called *self-certifying*.)

We'll be using (4.5) a lot in the rest of this chapter. One of its important consequences is the following mini-theorem:

$$k \backslash m \quad \text{and} \quad k \backslash n \qquad \Longleftrightarrow \qquad k \backslash \gcd(m,n)\,. \tag{4.6}$$

(Proof: If k divides both m and n, it divides $m'm + n'n$, so it divides $\gcd(m,n)$. Conversely, if k divides $\gcd(m,n)$, it divides a divisor of m and a divisor of n, so it divides both m and n.) We always knew that any common divisor of m and n must be *less than or equal to* their gcd; that's the definition of greatest common divisor. But now we know that any common divisor is, in fact, *a divisor of* their gcd.

Sometimes we need to do sums over all divisors of n. In this case it's often useful to use the handy rule

$$\sum_{m \backslash n} a_m \;=\; \sum_{m \backslash n} a_{n/m}\,, \qquad \text{integer } n > 0, \tag{4.7}$$

which holds since n/m runs through all divisors of n when m does. For example, when $n = 12$ this says that $a_1 + a_2 + a_3 + a_4 + a_6 + a_{12} = a_{12} + a_6 + a_4 + a_3 + a_2 + a_1$.

There's also a slightly more general identity,

$$\sum_{m \backslash n} a_m \;=\; \sum_{k} \sum_{m>0} a_m [n = mk]\,, \tag{4.8}$$

which is an immediate consequence of the definition (4.1). If n is positive, the right-hand side of (4.8) is $\sum_{k \backslash n} a_{n/k}$; hence (4.8) implies (4.7). And equation

(4.8) works also when n is negative. (In such cases, the nonzero terms on the right occur when k is the negative of a divisor of n.)

Moreover, a double sum over divisors can be "interchanged" by the law

$$\sum_{m\backslash n}\sum_{k\backslash m} a_{k,m} = \sum_{k\backslash n}\sum_{l\backslash(n/k)} a_{k,kl}\,. \tag{4.9}$$

For example, this law takes the following form when $n = 12$:

$$\begin{aligned} &a_{1,1} + (a_{1,2}+a_{2,2}) + (a_{1,3}+a_{3,3}) \\ &\qquad + (a_{1,4}+a_{2,4}+a_{4,4}) + (a_{1,6}+a_{2,6}+a_{3,6}+a_{6,6}) \\ &\qquad + (a_{1,12}+a_{2,12}+a_{3,12}+a_{4,12}+a_{6,12}+a_{12,12}) \\ &= (a_{1,1}+a_{1,2}+a_{1,3}+a_{1,4}+a_{1,6}+a_{1,12}) \\ &\qquad + (a_{2,2}+a_{2,4}+a_{2,6}+a_{2,12}) + (a_{3,3}+a_{3,6}+a_{3,12}) \\ &\qquad + (a_{4,4}+a_{4,12}) + (a_{6,6}+a_{6,12}) + a_{12,12}\,. \end{aligned}$$

We can prove (4.9) with Iversonian manipulation. The left-hand side is

$$\sum_{j,l}\sum_{k,m>0} a_{k,m}[n=jm][m=kl] = \sum_{j}\sum_{k,l>0} a_{k,kl}[n=jkl]\,;$$

the right-hand side is

$$\sum_{j,m}\sum_{k,l>0} a_{k,kl}[n=jk][n/k=ml] = \sum_{m}\sum_{k,l>0} a_{k,kl}[n=mlk]\,,$$

which is the same except for renaming the indices. This example indicates that the techniques we've learned in Chapter 2 will come in handy as we study number theory.

4.2 PRIMES

A positive integer p is called *prime* if it has just two divisors, namely 1 and p. *Throughout the rest of this chapter, the letter* p *will always stand for a prime number, even when we don't say so explicitly.* By convention, 1 isn't prime, so the sequence of primes starts out like this:

How about the p in 'explicitly'?

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \ldots\,.$$

Some numbers look prime but aren't, like 91 ($= 7\cdot13$) and 161 ($= 7\cdot23$). These numbers and others that have three or more divisors are called *composite*. Every integer greater than 1 is either prime or composite, but not both.

Primes are of great importance, because they're the fundamental building blocks of all the positive integers. Any positive integer n can be written as a

product of primes,

$$n = p_1 \dots p_m = \prod_{k=1}^{m} p_k, \qquad p_1 \leqslant \cdots \leqslant p_m. \tag{4.10}$$

For example, $12 = 2\cdot 2\cdot 3$; $11011 = 7\cdot 11\cdot 11\cdot 13$; $11111 = 41\cdot 271$. (Products denoted by $\prod$ are analogous to sums denoted by $\sum$, as explained in exercise 2.25. If $m = 0$, we consider this to be an empty product, whose value is 1 by definition; that's the way $n = 1$ gets represented by (4.10).) Such a factorization is always possible because if $n > 1$ is not prime it has a divisor n_1 such that $1 < n_1 < n$; thus we can write $n = n_1 \cdot n_2$, and (by induction) we know that n_1 and n_2 can be written as products of primes.

Moreover, the expansion in (4.10) is *unique*: There's only one way to write n as a product of primes in nondecreasing order. This statement is called the Fundamental Theorem of Arithmetic, and it seems so obvious that we might wonder why it needs to be proved. How could there be two different sets of primes with the same product? Well, there can't, but the reason *isn't* simply "by definition of prime numbers." For example, if we consider the set of all real numbers of the form $m + n\sqrt{10}$ when m and n are integers, the product of any two such numbers is again of the same form, and we can call such a number "prime" if it can't be factored in a nontrivial way. The number 6 has two representations, $2\cdot 3 = (4+\sqrt{10})(4-\sqrt{10})$; yet exercise 36 shows that 2, 3, $4+\sqrt{10}$, and $4-\sqrt{10}$ are all "prime" in this system.

Therefore we should prove rigorously that (4.10) is unique. There is certainly only one possibility when $n = 1$, since the product must be empty in that case; so let's suppose that $n > 1$ and that all smaller numbers factor uniquely. Suppose we have two factorizations

$$n = p_1 \dots p_m = q_1 \dots q_k, \qquad p_1 \leqslant \cdots \leqslant p_m \quad \text{and} \quad q_1 \leqslant \cdots \leqslant q_k,$$

where the p's and q's are all prime. We will prove that $p_1 = q_1$. If not, we can assume that $p_1 < q_1$, making p_1 smaller than all the q's. Since p_1 and q_1 are prime, their gcd must be 1; hence Euclid's self-certifying algorithm gives us integers a and b such that $ap_1 + bq_1 = 1$. Therefore

$$ap_1q_2 \dots q_k + bq_1q_2 \dots q_k = q_2 \dots q_k.$$

Now p_1 divides both terms on the left, since $q_1q_2 \dots q_k = n$; hence p_1 divides the right-hand side, $q_2 \dots q_k$. Thus $q_2 \dots q_k/p_1$ is an integer, and $q_2 \dots q_k$ has a prime factorization in which p_1 appears. But $q_2 \dots q_k < n$, so it has a unique factorization (by induction). This contradiction shows that p_1 must be equal to q_1 after all. Therefore we can divide both of n's factorizations by p_1, obtaining $p_2 \dots p_m = q_2 \dots q_k < n$. The other factors must likewise be equal (by induction), so our proof of uniqueness is complete.

It's the factorization, not the theorem, that's unique.

Sometimes it's more useful to state the Fundamental Theorem in another way: *Every positive integer can be written uniquely in the form*

$$n = \prod_p p^{n_p}, \qquad \text{where each } n_p \geqslant 0. \tag{4.11}$$

The right-hand side is a product over infinitely many primes; but for any particular n all but a few exponents are zero, so the corresponding factors are 1. Therefore it's really a finite product, just as many "infinite" sums are really finite because their terms are mostly zero.

Formula (4.11) represents n uniquely, so we can think of the sequence $\langle n_2, n_3, n_5, \ldots \rangle$ as a *number system* for positive integers. For example, the prime-exponent representation of 12 is $\langle 2, 1, 0, 0, \ldots \rangle$ and the prime-exponent representation of 18 is $\langle 1, 2, 0, 0, \ldots \rangle$. To multiply two numbers, we simply add their representations. In other words,

$$k = mn \qquad \Longleftrightarrow \qquad k_p = m_p + n_p \quad \text{for all } p. \tag{4.12}$$

This implies that

$$m \backslash n \qquad \Longleftrightarrow \qquad m_p \leqslant n_p \quad \text{for all } p, \tag{4.13}$$

and it follows immediately that

$$k = \gcd(m, n) \qquad \Longleftrightarrow \qquad k_p = \min(m_p, n_p) \quad \text{for all } p; \tag{4.14}$$

$$k = \operatorname{lcm}(m, n) \qquad \Longleftrightarrow \qquad k_p = \max(m_p, n_p) \quad \text{for all } p. \tag{4.15}$$

For example, since $12 = 2^2 \cdot 3^1$ and $18 = 2^1 \cdot 3^2$, we can get their gcd and lcm by taking the min and max of common exponents:

$$\gcd(12, 18) = 2^{\min(2,1)} \cdot 3^{\min(1,2)} = 2^1 \cdot 3^1 = 6;$$
$$\operatorname{lcm}(12, 18) = 2^{\max(2,1)} \cdot 3^{\max(1,2)} = 2^2 \cdot 3^2 = 36.$$

If the prime p divides a product mn then it divides either m or n, perhaps both, because of the unique factorization theorem. But composite numbers do not have this property. For example, the nonprime 4 divides $60 = 6 \cdot 10$, but it divides neither 6 nor 10. The reason is simple: In the factorization $60 = 6 \cdot 10 = (2 \cdot 3)(2 \cdot 5)$, the two prime factors of $4 = 2 \cdot 2$ have been split into two parts, hence 4 divides neither part. But a prime is unsplittable, so it must divide one of the original factors.

4.3 PRIME EXAMPLES

How many primes are there? A lot. In fact, infinitely many. Euclid proved this long ago in his Theorem 9 : 20, as follows. Suppose there were only

finitely many primes, say k of them—$2, 3, 5, \ldots, P_k$. Then, said Euclid, we should consider the number

$$M = 2\cdot 3\cdot 5\cdot \ldots \cdot P_k + 1.$$

None of the k primes can divide M, because each divides $M-1$. Thus there must be some other prime that divides M; perhaps M itself is prime. This contradicts our assumption that $2, 3, \ldots, P_k$ are the only primes, so there must indeed be infinitely many.

"Οἱ πρῶτοι ἀριθμοὶ πλείους εἰσὶ παντὸς τοῦ προτεθέντος πλήθους πρώτων ἀριθμῶν."
—*Euclid [98]*
[Translation: "There are more primes than in any given set of primes."]

Euclid's proof suggests that we define *Euclid numbers* by the recurrence

$$e_n = e_1 e_2 \ldots e_{n-1} + 1, \qquad \text{when } n \geq 1. \tag{4.16}$$

The sequence starts out

$$e_1 = 1+1 = 2;$$
$$e_2 = 2+1 = 3;$$
$$e_3 = 2\cdot 3+1 = 7;$$
$$e_4 = 2\cdot 3\cdot 7+1 = 43;$$

these are all prime. But the next case, e_5, is $1807 = 13\cdot 139$. It turns out that $e_6 = 3263443$ is prime, while

$$e_7 = 547\cdot 607\cdot 1033\cdot 31051;$$
$$e_8 = 29881\cdot 67003\cdot 9119521\cdot 6212157481.$$

It is known that $e_9, \ldots, e_{17}$ are composite, and the remaining e_n are probably composite as well. However, the Euclid numbers are all *relatively prime* to each other; that is,

$$\gcd(e_m, e_n) = 1, \qquad \text{when } m \neq n.$$

Euclid's algorithm (what else?) tells us this in three short steps, because $e_n \bmod e_m = 1$ when $n > m$:

$$\gcd(e_m, e_n) = \gcd(1, e_m) = \gcd(0, 1) = 1.$$

Therefore, if we let q_j be the smallest factor of e_j for all $j \geq 1$, the primes q_1, q_2, q_3, ... are all different. This is a sequence of infinitely many primes.

Let's pause to consider the Euclid numbers from the standpoint of Chapter 1. Can we express e_n in closed form? Recurrence (4.16) can be simplified by removing the three dots: If $n > 1$ we have

$$e_n = e_1 \ldots e_{n-2} e_{n-1} + 1 = (e_{n-1} - 1)e_{n-1} + 1 = e_{n-1}^2 - e_{n-1} + 1.$$

Thus e_n has about twice as many decimal digits as e_{n-1}. Exercise 37 proves that there's a constant $E \approx 1.264$ such that

$$e_n = \lfloor E^{2^n} + \tfrac{1}{2} \rfloor . \tag{4.17}$$

And exercise 60 provides a similar formula that gives nothing but primes:

$$p_n = \lfloor P^{3^n} \rfloor , \tag{4.18}$$

for some constant P. But equations like (4.17) and (4.18) cannot really be considered to be in closed form, because the constants E and P are computed from the numbers e_n and p_n in a sort of sneaky way. No independent relation is known (or likely) that would connect them with other constants of mathematical interest.

Indeed, nobody knows *any* useful formula that gives arbitrarily large primes but only primes. Computer scientists at Chevron Geosciences did, however, strike mathematical oil in 1984. Using a program developed by David Slowinski, they discovered the largest prime known at that time,

$$2^{216091} - 1 ,$$

while testing a new Cray X-MP supercomputer. It's easy to compute this number in a few milliseconds on a personal computer, because modern computers work in binary notation and this number is simply $(11\ldots1)_2$. All 216,091 of its bits are '1'. But it's much harder to prove that this number is prime. In fact, just about any computation with it takes a lot of time, because it's so large. For example, even a sophisticated algorithm requires several minutes just to convert $2^{216091}-1$ to radix 10 on a PC. When printed out, its 65,050 decimal digits require 75 cents U.S. postage to mail first class.

Or probably more, by the time you read this.

Incidentally, $2^{216091} - 1$ is the number of moves necessary to solve the Tower of Hanoi problem when there are 216,091 disks. Numbers of the form

$$2^p - 1$$

(where p is prime, as always in this chapter) are called *Mersenne numbers*, after Father Marin Mersenne who investigated some of their properties in the seventeenth century [269]. . The Mersenne primes known to date occur for $p = 2$, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497, 86243, 110503, 132049, 216091, and 756839.

The number $2^n - 1$ can't possibly be prime if n is composite, because $2^{km} - 1$ has $2^m - 1$ as a factor:

$$2^{km} - 1 = (2^m - 1)(2^{m(k-1)} + 2^{m(k-2)} + \cdots + 1) .$$

But $2^p - 1$ isn't always prime when p is prime; $2^{11} - 1 = 2047 = 23 \cdot 89$ is the smallest such nonprime. (Mersenne knew this.)

Factoring and primality testing of large numbers are hot topics nowadays. A summary of what was known up to 1981 appears in Section 4.5.4 of [208], and many new results continue to be discovered. Pages 391–394 of that book explain a special way to test Mersenne numbers for primality.

For most of the last two hundred years, the largest known prime has been a Mersenne prime, although only 31 Mersenne primes are known. Many people are trying to find larger ones, but it's getting tough. So those really interested in fame (if not fortune) and a spot in *The Guinness Book of World Records* might instead try numbers of the form $2^n k + 1$, for small values of k like 3 or 5. These numbers can be tested for primality almost as quickly as Mersenne numbers can; exercise 4.5.4–27 of [208] gives the details.

We haven't fully answered our original question about how many primes there are. There are infinitely many, but some infinite sets are "denser" than others. For instance, among the positive integers there are infinitely many even numbers and infinitely many perfect squares, yet in several important senses there are more even numbers than perfect squares. One such sense looks at the size of the nth value. The nth even integer is $2n$ and the nth perfect square is n^2; since $2n$ is much less than n^2 for large n, the nth even integer occurs much sooner than the nth perfect square, so we can say there are many more even integers than perfect squares. A similar sense looks at the number of values not exceeding x. There are $\lfloor x/2 \rfloor$ such even integers and $\lfloor \sqrt{x} \rfloor$ perfect squares; since $x/2$ is much larger than $\sqrt{x}$ for large x, again we can say there are many more even integers.

Weird. I thought there were the same number of even integers as perfect squares, since there's a one-to-one correspondence between them.

What can we say about the primes in these two senses? It turns out that the nth prime, P_n, is about n times the natural log of n:

$$P_n \sim n \ln n .$$

(The symbol '$\sim$' can be read "is asymptotic to"; it means that the limit of the ratio $P_n / n \ln n$ is 1 as n goes to infinity.) Similarly, for the number of primes $\pi(x)$ not exceeding x we have what's known as the prime number theorem:

$$\pi(x) \sim \frac{x}{\ln x} .$$

Proving these two facts is beyond the scope of this book, although we can show easily that each of them implies the other. In Chapter 9 we will discuss the rates at which functions approach infinity, and we'll see that the function $n \ln n$, our approximation to P_n, lies between $2n$ and n^2 asymptotically. Hence there are fewer primes than even integers, but there are more primes than perfect squares.

These formulas, which hold only in the limit as n or $x \to \infty$, can be replaced by more exact estimates. For example, Rosser and Schoenfeld [312] have established the handy bounds

$$\ln x - \tfrac{3}{2} \;<\; \tfrac{x}{\pi(x)} \;<\; \ln x - \tfrac{1}{2}\,, \qquad \text{for } x \geqslant 67; \quad (4.19)$$

$$n\big(\ln n + \ln\ln n - \tfrac{3}{2}\big) < P_n < n\big(\ln n + \ln\ln n - \tfrac{1}{2}\big), \quad \text{for } n \geqslant 20. \quad (4.20)$$

If we look at a "random" integer n, the chances of its being prime are about one in $\ln n$. For example, if we look at numbers near 10^{16}, we'll have to examine about $16\ln 10 \approx 36.8$ of them before finding a prime. (It turns out that there are exactly 10 primes between $10^{16} - 370$ and $10^{16} - 1$.) Yet the distribution of primes has many irregularities. For example, all the numbers between $P_1P_2\ldots P_n + 2$ and $P_1P_2\ldots P_n + P_{n+1} - 1$ inclusive are composite. Many examples of "twin primes" p and $p+2$ are known (5 and 7, 11 and 13, 17 and 19, 29 and 31, ..., 9999999999999641 and 9999999999999643, ...), yet nobody knows whether or not there are infinitely many pairs of twin primes. (See Hardy and Wright [181, §1.4 and §2.8].)

One simple way to calculate all $\pi(x)$ primes $\leqslant x$ is to form the so-called sieve of Eratosthenes: First write down all integers from 2 through x. Next circle 2, marking it prime, and cross out all other multiples of 2. Then repeatedly circle the smallest uncircled, uncrossed number and cross out its other multiples. When everything has been circled or crossed out, the circled numbers are the primes. For example when $x = 10$ we write down 2 through 10, circle 2, then cross out its multiples 4, 6, 8, and 10. Next 3 is the smallest uncircled, uncrossed number, so we circle it and cross out 6 and 9. Now 5 is smallest, so we circle it and cross out 10. Finally we circle 7. The circled numbers are 2, 3, 5, and 7; so these are the $\pi(10) = 4$ primes not exceeding 10.

"Je me sers de la notation très simple $n!$ *pour désigner le produit de nombres décroissans depuis* n *jusqu'à l'unité, savoir* $n(n-1)(n-2)\ldots.3.2.1$. *L'emploi continuel de l'analyse combinatoire que je fais dans la plupart de mes démonstrations, a rendu cette notation indispensable."*
— Ch. Kramp [228]

4.4 FACTORIAL FACTORS

Now let's take a look at the factorization of some interesting highly composite numbers, the factorials:

$$n! \;=\; 1\cdot 2\cdot \ldots \cdot n \;=\; \prod_{k=1}^{n} k\,, \qquad \text{integer } n \geqslant 0. \quad (4.21)$$

According to our convention for an empty product, this defines 0! to be 1. Thus $n! = (n-1)!\,n$ for every positive integer n. This is the number of permutations of n distinct objects. That is, it's the number of ways to arrange n things in a row: There are n choices for the first thing; for each choice of first thing, there are $n-1$ choices for the second; for each of these $n(n-1)$ choices, there are $n-2$ for the third; and so on, giving $n(n-1)(n-2)\ldots(1)$

arrangements in all. Here are the first few values of the factorial function.

n	0	1	2	3	4	5	6	7	8	9	10
$n!$	1	1	2	6	24	120	720	5040	40320	362880	3628800

It's useful to know a few factorial facts, like the first six or so values, and the fact that 10! is about $3\frac{1}{2}$ million plus change; another interesting fact is that the number of digits in $n!$ exceeds n when $n \geqslant 25$.

We can prove that $n!$ is plenty big by using something like Gauss's trick of Chapter 1:

$$n!^2 \;=\; (1\cdot 2\cdot \ldots \cdot n)(n\cdot \ldots \cdot 2\cdot 1) \;=\; \prod_{k=1}^{n} k(n+1-k)\,.$$

We have $n \leqslant k(n+1-k) \leqslant \frac{1}{4}(n+1)^2$, since the quadratic polynomial $k(n+1-k) = \frac{1}{4}(n+1)^2 - \big(k-\frac{1}{2}(n+1)\big)^2$ has its smallest value at $k=1$ and its largest value at $k=\frac{1}{2}(n+1)$. Therefore

$$\prod_{k=1}^{n} n \;\leqslant\; n!^2 \;\leqslant\; \prod_{k=1}^{n} \frac{(n+1)^2}{4}\,;$$

that is,

$$n^{n/2} \;\leqslant\; n! \;\leqslant\; \frac{(n+1)^n}{2^n}\,. \tag{4.22}$$

This relation tells us that the factorial function grows exponentially!!

To approximate $n!$ more accurately for large n we can use Stirling's formula, which we will derive in Chapter 9:

$$n! \;\sim\; \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\,. \tag{4.23}$$

And a still more precise approximation tells us the asymptotic relative error: Stirling's formula undershoots $n!$ by a factor of about $1/(12n)$. Even for fairly small n this more precise estimate is pretty good. For example, Stirling's approximation (4.23) gives a value near 3598696 when $n=10$, and this is about $0.83\% \approx 1/120$ too small. Good stuff, asymptotics.

But let's get back to primes. We'd like to determine, for any given prime p, the largest power of p that divides $n!$; that is, we want the exponent of p in $n!$'s unique factorization. We denote this number by $\epsilon_p(n!)$, and we start our investigations with the small case $p=2$ and $n=10$. Since 10! is the product of ten numbers, $\epsilon_2(10!)$ can be found by summing the powers-of-2

contributions of those ten numbers; this calculation corresponds to summing the columns of the following array:

	1 2 3 4 5 6 7 8 9 10	powers of 2
divisible by 2	x x x x x	$5 = \lfloor 10/2 \rfloor$
divisible by 4	x x	$2 = \lfloor 10/4 \rfloor$
divisible by 8	x	$1 = \lfloor 10/8 \rfloor$
powers of 2	0 1 0 2 0 1 0 3 0 1	8

A powerful ruler.

(The column sums form what's sometimes called the *ruler function* $\rho(k)$, because of their similarity to '|'''|'''|'''|'''|''''|''''|''''|''''|''''|', the lengths of lines marking fractions of an inch.) The sum of these ten sums is 8; hence 2^8 divides 10! but 2^9 doesn't.

There's also another way: We can sum the contributions of the rows. The first row marks the numbers that contribute a power of 2 (and thus are divisible by 2); there are $\lfloor 10/2 \rfloor = 5$ of them. The second row marks those that contribute an additional power of 2; there are $\lfloor 10/4 \rfloor = 2$ of them. And the third row marks those that contribute yet another; there are $\lfloor 10/8 \rfloor = 1$ of them. These account for all contributions, so we have $\epsilon_2(10!) = 5+2+1 = 8$.

For general n this method gives

$$\epsilon_2(n!) \;=\; \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{n}{4} \right\rfloor + \left\lfloor \frac{n}{8} \right\rfloor + \cdots \;=\; \sum_{k \geqslant 1} \left\lfloor \frac{n}{2^k} \right\rfloor .$$

This sum is actually finite, since the summand is zero when $2^k > n$. Therefore it has only $\lfloor \lg n \rfloor$ nonzero terms, and it's computationally quite easy. For instance, when $n = 100$ we have

$$\epsilon_2(100!) \;=\; 50 + 25 + 12 + 6 + 3 + 1 \;=\; 97 .$$

Each term is just the floor of half the previous term. This is true for all n, because as a special case of (3.11) we have $\lfloor n/2^{k+1} \rfloor = \lfloor \lfloor n/2^k \rfloor / 2 \rfloor$. It's especially easy to see what's going on here when we write the numbers in binary:

$$\begin{aligned}
100 &= (1100100)_2 = 100 \\
\lfloor 100/2 \rfloor &= (110010)_2 = 50 \\
\lfloor 100/4 \rfloor &= (11001)_2 = 25 \\
\lfloor 100/8 \rfloor &= (1100)_2 = 12 \\
\lfloor 100/16 \rfloor &= (110)_2 = 6 \\
\lfloor 100/32 \rfloor &= (11)_2 = 3 \\
\lfloor 100/64 \rfloor &= (1)_2 = 1
\end{aligned}$$

We merely drop the least significant bit from one term to get the next.

The binary representation also shows us how to derive another formula,

$$\epsilon_2(n!) \;=\; n - \nu_2(n)\,, \tag{4.24}$$

where $\nu_2(n)$ is the number of 1's in the binary representation of n. This simplification works because each 1 that contributes 2^m to the value of n contributes $2^{m-1} + 2^{m-2} + \cdots + 2^0 = 2^m - 1$ to the value of $\epsilon_2(n!)$.

Generalizing our findings to an arbitrary prime p, we have

$$\epsilon_p(n!) \;=\; \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \;=\; \sum_{k \geqslant 1} \left\lfloor \frac{n}{p^k} \right\rfloor \tag{4.25}$$

by the same reasoning as before.

About how large is $\epsilon_p(n!)$? We get an easy (but good) upper bound by simply removing the floor from the summand and then summing an infinite geometric progression:

$$\begin{aligned} \epsilon_p(n!) \;&<\; \frac{n}{p} + \frac{n}{p^2} + \frac{n}{p^3} + \cdots \\ &=\; \frac{n}{p}\left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) \\ &=\; \frac{n}{p}\left(\frac{p}{p-1}\right) \\ &=\; \frac{n}{p-1}\,. \end{aligned}$$

For $p = 2$ and $n = 100$ this inequality says that $97 < 100$. Thus the upper bound 100 is not only correct, it's also close to the true value 97. In fact, the true value $n - \nu_2(n)$ is $\sim n$ in general, because $\nu_2(n) \leqslant \lceil \lg n \rceil$ is asymptotically much smaller than n.

When $p = 2$ and 3 our formulas give $\epsilon_2(n!) \sim n$ and $\epsilon_3(n!) \sim n/2$, so it seems reasonable that every once in awhile $\epsilon_3(n!)$ should be exactly half as big as $\epsilon_2(n!)$. For example, this happens when $n = 6$ and $n = 7$, because $6! = 2^4 \cdot 3^2 \cdot 5 = 7!/7$. But nobody has yet proved that such coincidences happen infinitely often.

The bound on $\epsilon_p(n!)$ in turn gives us a bound on $p^{\epsilon_p(n!)}$, which is p's contribution to $n!$:

$$p^{\epsilon_p(n!)} \;<\; p^{n/(p-1)}\,.$$

And we can simplify this formula (at the risk of greatly loosening the upper bound) by noting that $p \leqslant 2^{p-1}$; hence $p^{n/(p-1)} \leqslant (2^{p-1})^{n/(p-1)} = 2^n$. In other words, the contribution that any prime makes to $n!$ is less than 2^n.

We can use this observation to get another proof that there are infinitely many primes. For if there were only the k primes 2, 3, ..., P_k, then we'd have $n! < (2^n)^k = 2^{nk}$ for all $n > 1$, since each prime can contribute at most a factor of $2^n - 1$. But we can easily contradict the inequality $n! < 2^{nk}$ by choosing n large enough, say $n = 2^{2k}$. Then

$$n! \;<\; 2^{nk} \;=\; 2^{2^{2k}k} \;=\; n^{n/2},$$

contradicting the inequality $n! \geqslant n^{n/2}$ that we derived in (4.22). There are infinitely many primes, still.

We can even beef up this argument to get a crude bound on $\pi(n)$, the number of primes not exceeding n. Every such prime contributes a factor of less than 2^n to $n!$; so, as before,

$$n! \;<\; 2^{n\pi(n)}.$$

If we replace $n!$ here by Stirling's approximation (4.23), which is a lower bound, and take logarithms, we get

$$n\pi(n) \;>\; n\lg(n/e) + \tfrac{1}{2}\lg(2\pi n)\,;$$

hence

$$\pi(n) \;>\; \lg(n/e)\,.$$

This lower bound is quite weak, compared with the actual value $\pi(n) \sim n/\ln n$, because $\log n$ is much smaller than $n/\log n$ when n is large. But we didn't have to work very hard to get it, and a bound is a bound.

4.5 RELATIVE PRIMALITY

When $\gcd(m, n) = 1$, the integers m and n have no prime factors in common and we say that they're *relatively prime*.

This concept is so important in practice, we ought to have a special notation for it; but alas, number theorists haven't agreed on a very good one yet. Therefore we cry: HEAR US, O MATHEMATICIANS OF THE WORLD! LET US NOT WAIT ANY LONGER! WE CAN MAKE MANY FORMULAS CLEARER BY ADOPTING A NEW NOTATION NOW! LET US AGREE TO WRITE '$m \perp n$', AND TO SAY "m IS PRIME TO n," IF m AND n ARE RELATIVELY PRIME. In other words, let us declare that

Like perpendicular lines don't have a common direction, perpendicular numbers don't have common factors.

$$m \perp n \quad\Longleftrightarrow\quad m, n \text{ are integers and } \gcd(m, n) = 1. \tag{4.26}$$

A fraction m/n is in lowest terms if and only if $m \perp n$. Since we reduce fractions to lowest terms by casting out the largest common factor of numerator and denominator, we suspect that, in general,

$$m/\gcd(m,n) \quad \perp \quad n/\gcd(m,n)\,; \tag{4.27}$$

and indeed this is true. It follows from a more general law, $\gcd(km,kn) = k\gcd(m,n)$, proved in exercise 14.

The $\perp$ relation has a simple formulation when we work with the prime-exponent representations of numbers, because of the gcd rule (4.14):

$$m \perp n \qquad \Longleftrightarrow \qquad \min(m_p, n_p) = 0 \quad \text{for all } p. \tag{4.28}$$

Furthermore, since m_p and n_p are nonnegative, we can rewrite this as

The dot product is zero, like orthogonal vectors.

$$m \perp n \qquad \Longleftrightarrow \qquad m_p n_p = 0 \quad \text{for all } p. \tag{4.29}$$

And now we can prove an important law by which we can split and combine two $\perp$ relations with the same left-hand side:

$$k \perp m \quad \text{and} \quad k \perp n \qquad \Longleftrightarrow \qquad k \perp mn\,. \tag{4.30}$$

In view of (4.29), this law is another way of saying that $k_p m_p = 0$ and $k_p n_p = 0$ if and only if $k_p(m_p + n_p) = 0$, when m_p and n_p are nonnegative.

There's a beautiful way to construct the set of all nonnegative fractions m/n with $m \perp n$, called the *Stern–Brocot tree* because it was discovered independently by Moriz Stern [339], a German mathematician, and Achille Brocot [40], a French clockmaker. The idea is to start with the two fractions $(\frac{0}{1}, \frac{1}{0})$ and then to repeat the following operation as many times as desired:

Interesting how mathematicians will say "discovered" when absolutely anyone else would have said "invented."

$$\text{Insert } \frac{m+m'}{n+n'} \text{ between two adjacent fractions } \frac{m}{n} \text{ and } \frac{m'}{n'}\,.$$

The new fraction $(m+m')/(n+n')$ is called the *mediant* of m/n and m'/n'. For example, the first step gives us one new entry between $\frac{0}{1}$ and $\frac{1}{0}$,

$$\tfrac{0}{1},\ \tfrac{1}{1},\ \tfrac{1}{0}\,;$$

and the next gives two more:

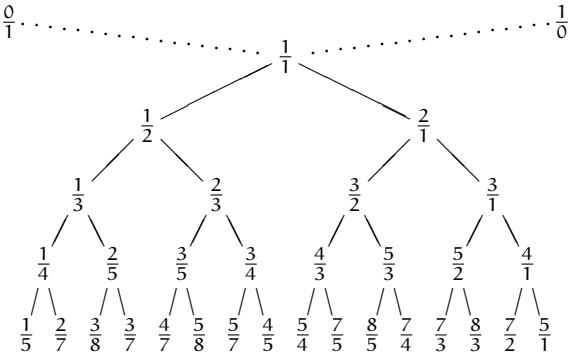
$$\tfrac{0}{1},\ \tfrac{1}{2},\ \tfrac{1}{1},\ \tfrac{2}{1},\ \tfrac{1}{0}\,.$$

The next gives four more,

$$\tfrac{0}{1},\ \tfrac{1}{3},\ \tfrac{1}{2},\ \tfrac{2}{3},\ \tfrac{1}{1},\ \tfrac{3}{2},\ \tfrac{2}{1},\ \tfrac{3}{1},\ \tfrac{1}{0}\,;$$

and then we'll get 8, 16, and so on. The entire array can be regarded as an infinite binary tree structure whose top levels look like this:

I guess $1/0$ *is infinity, "in lowest terms."*



Each fraction is $\frac{m+m'}{n+n'}$, where $\frac{m}{n}$ is the nearest ancestor above and to the left, and $\frac{m'}{n'}$ is the nearest ancestor above and to the right. (An "ancestor" is a fraction that's reachable by following the branches upward.) Many patterns can be observed in this tree.

Why does this construction work? Why, for example, does each mediant fraction $(m+m')/(n+n')$ turn out to be in lowest terms when it appears in this tree? (If m, m', n, and n' were all odd, we'd get even/even; somehow the construction guarantees that fractions with odd numerators and denominators never appear next to each other.) And why do all possible fractions m/n occur exactly once? Why can't a particular fraction occur twice, or not at all?

Conserve parody.

All of these questions have amazingly simple answers, based on the following fundamental fact: *If* m/n *and* m'/n' *are consecutive fractions at any stage of the construction, we have*

$$m'n - mn' = 1. \tag{4.31}$$

This relation is true initially ($1\cdot 1 - 0\cdot 0 = 1$); and when we insert a new mediant $(m+m')/(n+n')$, the new cases that need to be checked are

$$(m+m')n - m(n+n') = 1;$$
$$m'(n+n') - (m+m')n' = 1.$$

Both of these equations are equivalent to the original condition (4.31) that they replace. Therefore (4.31) is invariant at all stages of the construction.

Furthermore, if $m/n < m'/n'$ and if all values are nonnegative, it's easy to verify that

$$m/n < (m+m')/(n+n') < m'/n'.$$

A mediant fraction isn't halfway between its progenitors, but it does lie somewhere in between. Therefore the construction preserves order, and we couldn't possibly get the same fraction in two different places.

True, but if you get a compound fracture you'd better go see a doctor.

One question still remains. Can any positive fraction a/b with $a \perp b$ possibly be omitted? The answer is no, because we can confine the construction to the immediate neighborhood of a/b, and in this region the behavior is easy to analyze: Initially we have

$$\frac{m}{n} = \frac{0}{1} < \left(\frac{a}{b}\right) < \frac{1}{0} = \frac{m'}{n'},$$

where we put parentheses around $\frac{a}{b}$ to indicate that it's not really present yet. Then if at some stage we have

$$\frac{m}{n} < \left(\frac{a}{b}\right) < \frac{m'}{n'},$$

the construction forms $(m+m')/(n+n')$ and there are three cases. Either $(m+m')/(n+n') = a/b$ and we win; or $(m+m')/(n+n') < a/b$ and we can set $m \leftarrow m+m'$, $n \leftarrow n+n'$; or $(m+m')/(n+n') > a/b$ and we can set $m' \leftarrow m+m'$, $n' \leftarrow n+n'$. This process cannot go on indefinitely, because the conditions

$$\frac{a}{b} - \frac{m}{n} > 0 \qquad \text{and} \qquad \frac{m'}{n'} - \frac{a}{b} > 0$$

imply that

$$an - bm \geqslant 1 \qquad \text{and} \qquad bm' - an' \geqslant 1;$$

hence

$$(m'+n')(an-bm) + (m+n)(bm'-an') \geqslant m'+n'+m+n;$$

and this is the same as $a+b \geqslant m'+n'+m+n$ by (4.31). Either m or n or m' or n' increases at each step, so we must win after at most $a+b$ steps.

The *Farey series* of order N, denoted by $\mathcal{F}_N$, is the set of all reduced fractions between 0 and 1 whose denominators are N or less, arranged in increasing order. For example, if $N = 6$ we have

$$\mathcal{F}_6 = \frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1}.$$

We can obtain $\mathcal{F}_N$ in general by starting with $\mathcal{F}_1 = \frac{0}{1}, \frac{1}{1}$ and then inserting mediants whenever it's possible to do so without getting a denominator that is too large. We don't miss any fractions in this way, because we know that the Stern–Brocot construction doesn't miss any, and because a mediant with denominator $\leqslant N$ is never formed from a fraction whose denominator is $> N$. (In other words, $\mathcal{F}_N$ defines a *subtree* of the Stern–Brocot tree, obtained by

pruning off unwanted branches.) It follows that $m'n - mn' = 1$ whenever m/n and m'/n' are consecutive elements of a Farey series.

This method of construction reveals that $\mathcal{F}_N$ can be obtained in a simple way from $\mathcal{F}_{N-1}$: We simply insert the fraction $(m + m')/N$ between consecutive fractions m/n, m'/n' of $\mathcal{F}_{N-1}$ whose denominators sum to N. For example, it's easy to obtain $\mathcal{F}_7$ from the elements of $\mathcal{F}_6$, by inserting $\frac{1}{7}$, $\frac{2}{7}$, $\ldots$, $\frac{6}{7}$ according to the stated rule:

$$\mathcal{F}_7 = \tfrac{0}{1}, \tfrac{1}{7}, \tfrac{1}{6}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{2}{7}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{3}{7}, \tfrac{1}{2}, \tfrac{4}{7}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{5}{7}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{5}{6}, \tfrac{6}{7}, \tfrac{1}{1}.$$

When N is prime, $N - 1$ new fractions will appear; but otherwise we'll have fewer than $N - 1$, because this process generates only numerators that are relatively prime to N.

Long ago in (4.5) we proved — in different words — that whenever $m \perp n$ and $0 < m \leqslant n$ we can find integers a and b such that

$$ma - nb = 1. \tag{4.32}$$

(Actually we said $m'm + n'n = \gcd(m, n)$, but we can write 1 for $\gcd(m, n)$, a for m', and b for $-n'$.) The Farey series gives us another proof of (4.32), because we can let b/a be the fraction that precedes m/n in $\mathcal{F}_n$. Thus (4.5) is just (4.31) again. For example, one solution to $3a - 7b = 1$ is $a = 5$, $b = 2$, since $\frac{2}{5}$ precedes $\frac{3}{7}$ in $\mathcal{F}_7$. This construction implies that we can always find a solution to (4.32) with $0 \leqslant b < a < n$, if $0 < m \leqslant n$. Similarly, if $0 \leqslant n < m$ and $m \perp n$, we can solve (4.32) with $0 < a \leqslant b \leqslant m$ by letting a/b be the fraction that *follows* n/m in $\mathcal{F}_m$.

Sequences of three consecutive terms in a Farey series have an amazing property that is proved in exercise 61. But we had better not discuss the Farey series any further, because the entire Stern–Brocot tree turns out to be even more interesting.

Farey 'nough.

We can, in fact, regard the Stern–Brocot tree as a *number system* for representing rational numbers, because each positive, reduced fraction occurs exactly once. Let's use the letters L and R to stand for going down to the left or right branch as we proceed from the root of the tree to a particular fraction; then a string of L's and R's uniquely identifies a place in the tree. For example, LRRL means that we go left from $\frac{1}{1}$ down to $\frac{1}{2}$, then right to $\frac{2}{3}$, then right to $\frac{3}{4}$, then left to $\frac{5}{7}$. We can consider LRRL to be a representation of $\frac{5}{7}$. Every positive fraction gets represented in this way as a unique string of L's and R's.

Well, actually there's a slight problem: The fraction $\frac{1}{1}$ corresponds to the *empty* string, and we need a notation for that. Let's agree to call it I, because that looks something like 1 and it stands for "identity."

This representation raises two natural questions: (1) Given positive integers m and n with $m \perp n$, what is the string of L's and R's that corresponds to m/n? (2) Given a string of L's and R's, what fraction corresponds to it? Question 2 seems easier, so let's work on it first. We define

$$f(S) \;=\; \text{fraction corresponding to } S$$

when S is a string of L's and R's. For example, $f(\mathrm{LRRL}) = \frac{5}{7}$.

According to the construction, $f(S) = (m+m')/(n+n')$ if m/n and m'/n' are the closest fractions preceding and following S in the upper levels of the tree. Initially $m/n = 0/1$ and $m'/n' = 1/0$; then we successively replace either m/n or m'/n' by the mediant $(m+m')/(n+n')$ as we move right or left in the tree, respectively.

How can we capture this behavior in mathematical formulas that are easy to deal with? A bit of experimentation suggests that the best way is to maintain a 2×2 matrix

$$M(S) \;=\; \begin{pmatrix} n & n' \\ m & m' \end{pmatrix}$$

that holds the four quantities involved in the ancestral fractions m/n and m'/n' enclosing $f(S)$. We could put the m's on top and the n's on the bottom, fractionwise; but this upside-down arrangement works out more nicely because we have $M(I) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ when the process starts, and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is traditionally called the identity matrix I.

A step to the left replaces n' by $n+n'$ and m' by $m+m'$; hence

$$M(SL) \;=\; \begin{pmatrix} n & n+n' \\ m & m+m' \end{pmatrix} \;=\; \begin{pmatrix} n & n' \\ m & m' \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \;=\; M(S) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

(This is a special case of the general rule

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} w & x \\ y & z \end{pmatrix} \;=\; \begin{pmatrix} aw+by & ax+bz \\ cw+dy & cx+dz \end{pmatrix}$$

for multiplying 2×2 matrices.) Similarly it turns out that

If you're clueless about matrices, don't panic; this book uses them only here.

$$M(SR) \;=\; \begin{pmatrix} n+n' & n' \\ m+m' & m' \end{pmatrix} \;=\; M(S) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Therefore if we define L and R as 2×2 matrices,

$$L \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad R \;=\; \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \tag{4·33}$$

we get the simple formula $M(S) = S$, by induction on the length of S. Isn't that nice? (The letters L and R serve dual roles, as matrices and as letters in the string representation.) For example,

$$M(\mathrm{LRRL}) = \mathrm{LRRL} = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right) = \left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&1\\1&2\end{smallmatrix}\right) = \left(\begin{smallmatrix}3&4\\2&3\end{smallmatrix}\right);$$

the ancestral fractions that enclose $\mathrm{LRRL} = \frac{5}{7}$ are $\frac{2}{3}$ and $\frac{3}{4}$. And this construction gives us the answer to Question 2:

$$f(S) = f\left(\begin{pmatrix} n & n' \\ m & m' \end{pmatrix}\right) = \frac{m+m'}{n+n'}. \tag{4·34}$$

How about Question 1? That's easy, now that we understand the fundamental connection between tree nodes and 2×2 matrices. Given a pair of positive integers m and n, with $m \perp n$, we can find the position of m/n in the Stern–Brocot tree by "binary search" as follows:

$S := I;$
while $m/n \neq f(S)$ **do**
  if $m/n < f(S)$ **then** $\big(\mathrm{output}(L);\ S := SL\big)$
  else $\big(\mathrm{output}(R);\ S := SR\big).$

This outputs the desired string of L's and R's.

There's also another way to do the same job, by changing m and n instead of maintaining the state S. If S is any 2×2 matrix, we have

$$f(RS) = f(S) + 1$$

because RS is like S but with the top row added to the bottom row. (Let's look at it in slow motion:

$$S = \begin{pmatrix} n & n' \\ m & m' \end{pmatrix}; \qquad RS = \begin{pmatrix} n & n' \\ m+n & m'+n' \end{pmatrix};$$

hence $f(S) = (m+m')/(n+n')$ and $f(RS) = \big((m+n)+(m'+n')\big)/(n+n').$) If we carry out the binary search algorithm on a fraction m/n with $m > n$, the first output will be R; hence the subsequent behavior of the algorithm will have $f(S)$ exactly 1 greater than if we had begun with $(m-n)/n$ instead of m/n. A similar property holds for L, and we have

$$\frac{m}{n} = f(RS) \iff \frac{m-n}{n} = f(S), \quad \text{when } m > n;$$
$$\frac{m}{n} = f(LS) \iff \frac{m}{n-m} = f(S), \quad \text{when } m < n.$$

This means that we can transform the binary search algorithm to the following matrix-free procedure:

$$\begin{aligned}&\textbf{while}\ \ m \neq n\ \ \textbf{do}\\ &\quad \textbf{if}\ \ m < n\ \ \textbf{then}\ \big(\text{output(L)};\ \ n := n - m\big)\\ &\qquad\quad \textbf{else}\ \ \big(\text{output(R)};\ \ m := m - n\big)\,.\end{aligned}$$

For example, given $m/n = 5/7$, we have successively

$$\begin{array}{rccccc} m = & 5 & 5 & 3 & 1 & 1\\ n = & 7 & 2 & 2 & 2 & 1\\ \text{output} & \text{L} & \text{R} & \text{R} & \text{L} & \end{array}$$

in the simplified algorithm.

Irrational numbers don't appear in the Stern–Brocot tree, but all the rational numbers that are "close" to them do. For example, if we try the binary search algorithm with the number $e = 2.71828\ldots$, instead of with a fraction m/n, we'll get an infinite string of L's and R's that begins

$$\text{RRLRRLRLLLLRLRRRRRRLRLLLLLLLLLRLR}\ldots\,.$$

We can consider this infinite string to be the representation of e in the Stern–Brocot number system, just as we can represent e as an infinite decimal $2.718281828459\ldots$ or as an infinite binary fraction $(10.101101111110\ldots)_2$. Incidentally, it turns out that e's representation has a regular pattern in the Stern–Brocot system:

$$e \;=\; \text{RL}^0\text{RLR}^2\text{LRL}^4\text{RLR}^6\text{LRL}^8\text{RLR}^{10}\text{LRL}^{12}\text{RL}\ldots\,;$$

this is equivalent to a special case of something that Euler [105] discovered when he was 24 years old.

From this representation we can deduce that the fractions

$$\begin{array}{cccccccccccccccccc} & \text{R} & \text{R} & \text{L} & \text{R} & \text{R} & \text{L} & \text{R} & \text{L} & \text{L} & \text{L} & \text{L} & \text{R} & \text{L} & \text{R} & \text{R} & \text{R} & \text{R} & \\ \frac{1}{1}, & \frac{2}{1}, & \frac{3}{1}, & \frac{5}{2}, & \frac{8}{3}, & \frac{11}{4}, & \frac{19}{7}, & \frac{30}{11}, & \frac{49}{18}, & \frac{68}{25}, & \frac{87}{32}, & \frac{106}{39}, & \frac{193}{71}, & \frac{299}{110}, & \frac{492}{181}, & \frac{685}{252}, & \frac{878}{323}, & \cdots \end{array}$$

are the simplest rational upper and lower approximations to e. For if m/n does not appear in this list, then some fraction in this list whose numerator is $\leqslant m$ and whose denominator is $\leqslant n$ lies between m/n and e. For example, $\frac{27}{10}$ is not as simple an approximation as $\frac{19}{7} = 2.714\ldots$, which appears in the list and is closer to e. We can see this because the Stern–Brocot tree not only includes all rationals, it includes them in order, and because all fractions with small numerator and denominator appear above all less simple ones. Thus, $\frac{27}{10} = \text{RRLRRLL}$ is less than $\frac{19}{7} = \text{RRLRRL}$, which is less than

$e = \mathrm{RRLRRLR}\ldots$. Excellent approximations can be found in this way. For example, $\frac{1264}{465} \approx 2.718280$ agrees with e to six decimal places; we obtained this fraction from the first 19 letters of e's Stern–Brocot representation, and the accuracy is about what we would get with 19 bits of e's binary representation.

We can find the infinite representation of an irrational number α by a simple modification of the matrix-free binary search procedure:

$$\textbf{if } \alpha < 1 \textbf{ then } \big(\mathrm{output}(\mathrm{L});\ \alpha := \alpha/(1-\alpha)\big) \\ \textbf{else } \big(\mathrm{output}(\mathrm{R});\ \alpha := \alpha - 1\big)\,.$$

(These steps are to be repeated infinitely many times, or until we get tired.) If α is rational, the infinite representation obtained in this way is the same as before but with RL^∞ appended at the right of α's (finite) representation. For example, if $\alpha = 1$, we get $\mathrm{RLLL}\ldots$, corresponding to the infinite sequence of fractions $\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots$, which approach 1 in the limit. This situation is exactly analogous to ordinary binary notation, if we think of L as 0 and R as 1: Just as every real number x in $[0\,..\,1)$ has an infinite binary representation $(.b_1b_2b_3\ldots)_2$ not ending with all 1's, every real number α in $[0\,..\,\infty)$ has an infinite Stern–Brocot representation $B_1B_2B_3\ldots$ not ending with all R's. Thus we have a one-to-one order-preserving correspondence between $[0\,..\,1)$ and $[0\,..\,\infty)$ if we let $0 \leftrightarrow \mathrm{L}$ and $1 \leftrightarrow \mathrm{R}$.

There's an intimate relationship between Euclid's algorithm and the Stern–Brocot representations of rationals. Given $\alpha = m/n$, we get $\lfloor m/n \rfloor$ R's, then $\lfloor n/(m \bmod n) \rfloor$ L's, then $\lfloor (m \bmod n)/(n \bmod (m \bmod n)) \rfloor$ R's, and so on. These numbers $m \bmod n$, $n \bmod (m \bmod n)$, ... are just the values examined in Euclid's algorithm. (A little fudging is needed at the end to make sure that there aren't infinitely many R's.) We will explore this relationship further in Chapter 6.

4.6 'MOD': THE CONGRUENCE RELATION

"Numerorum congruentiam hoc signo, $\equiv$, in posterum denotabimus, modulum ubi opus erit in clausulis adiungentes, $-16 \equiv 9 \pmod 5$, $-7 \equiv 15 \pmod{11}$."
— C. F. Gauss [142]

Modular arithmetic is one of the main tools provided by number theory. We got a glimpse of it in Chapter 3 when we used the binary operation 'mod', usually as one operation amidst others in an expression. In this chapter we will use 'mod' also with entire equations, for which a slightly different notation is more convenient:

$$a \equiv b \pmod m \qquad \Longleftrightarrow \qquad a \bmod m = b \bmod m\,. \tag{4.35}$$

For example, $9 \equiv -16 \pmod 5$, because $9 \bmod 5 = 4 = (-16) \bmod 5$. The formula '$a \equiv b \pmod m$' can be read "$a$ is congruent to b modulo m." The definition makes sense when a, b, and m are arbitrary real numbers, but we almost always use it with integers only.

Since $x \bmod m$ differs from x by a multiple of m, we can understand congruences in another way:

$$a \equiv b \pmod m \qquad \Longleftrightarrow \qquad a - b \text{ is a multiple of } m. \tag{4.36}$$

For if $a \bmod m = b \bmod m$, then the definition of 'mod' in (3.21) tells us that $a - b = a \bmod m + km - (b \bmod m + lm) = (k - l)m$ for some integers k and l. Conversely if $a - b = km$, then $a = b$ if $m = 0$; otherwise

$$\begin{aligned} a \bmod m \;=\; a - \lfloor a/m \rfloor m \;&=\; b + km - \lfloor (b + km)/m \rfloor m \\ &=\; b - \lfloor b/m \rfloor m \;=\; b \bmod m\,. \end{aligned}$$

The characterization of $\equiv$ in (4.36) is often easier to apply than (4.35). For example, we have $8 \equiv 23 \pmod 5$ because $8 - 23 = -15$ is a multiple of 5; we don't have to compute both $8 \bmod 5$ and $23 \bmod 5$.

"I feel fine today modulo a slight headache."
— The Hacker's Dictionary [337]

The congruence sign '$\equiv$' looks conveniently like '$=$', because congruences are almost like equations. For example, congruence is an *equivalence relation*; that is, it satisfies the reflexive law '$a \equiv a$', the symmetric law '$a \equiv b \;\Rightarrow\; b \equiv a$', and the transitive law '$a \equiv b \equiv c \;\Rightarrow\; a \equiv c$'. All these properties are easy to prove, because any relation '$\equiv$' that satisfies '$a \equiv b \iff f(a) = f(b)$' for some function f is an equivalence relation. (In our case, $f(x) = x \bmod m$.) Moreover, we can add and subtract congruent elements without losing congruence:

$$\begin{aligned} &a \equiv b \quad\text{and}\quad c \equiv d \qquad \Longrightarrow \qquad a + c \equiv b + d \qquad (\bmod m)\,; \\ &a \equiv b \quad\text{and}\quad c \equiv d \qquad \Longrightarrow \qquad a - c \equiv b - d \qquad (\bmod m)\,. \end{aligned}$$

For if $a - b$ and $c - d$ are both multiples of m, so are $(a + c) - (b + d) = (a - b) + (c - d)$ and $(a - c) - (b - d) = (a - b) - (c - d)$. Incidentally, it isn't necessary to write '$(\bmod m)$' once for every appearance of '$\equiv$'; if the modulus is constant, we need to name it only once in order to establish the context. This is one of the great conveniences of congruence notation.

Multiplication works too, provided that we are dealing with integers:

$$a \equiv b \quad\text{and}\quad c \equiv d \qquad \Longrightarrow \qquad ac \equiv bd \qquad (\bmod m)\,, \quad \text{integers } b, c.$$

Proof: $ac - bd = (a - b)c + b(c - d)$. Repeated application of this multiplication property now allows us to take powers:

$$a \equiv b \qquad \Longrightarrow \qquad a^n \equiv b^n \qquad (\bmod m)\,, \quad \text{integers } a, b;\ \text{integer } n \geqslant 0.$$

For example, since $2 \equiv -1 \pmod 3$, we have $2^n \equiv (-1)^n \pmod 3$; this means that $2^n - 1$ is a multiple of 3 if and only if n is even.

Thus, most of the algebraic operations that we customarily do with equations can also be done with congruences. Most, but not all. The operation of division is conspicuously absent. If $ad \equiv bd \pmod m$, we can't always conclude that $a \equiv b$. For example, $3 \cdot 2 \equiv 5 \cdot 2 \pmod 4$, but $3 \not\equiv 5$.

We can salvage the cancellation property for congruences, however, in the common case that d and m are relatively prime:

$$ad \equiv bd \iff a \equiv b \pmod m, \quad \text{integers } a, b, d, m \text{ and } d \perp m. \tag{4.37}$$

For example, it's legit to conclude from $15 \equiv 35 \pmod m$ that $3 \equiv 7 \pmod m$, unless the modulus m is a multiple of 5.

To prove this property, we use the extended gcd law (4.5) again, finding d' and m' such that $d'd + m'm = 1$. Then if $ad \equiv bd$ we can multiply both sides of the congruence by d', obtaining $ad'd \equiv bd'd$. Since $d'd \equiv 1$, we have $ad'd \equiv a$ and $bd'd \equiv b$; hence $a \equiv b$. This proof shows that the number d' acts almost like $1/d$ when congruences are considered $\pmod m$; therefore we call it the "inverse of d modulo m."

Another way to apply division to congruences is to divide the modulus as well as the other numbers:

$$ad \equiv bd \pmod{md} \iff a \equiv b \pmod m, \quad \text{for } d \neq 0. \tag{4.38}$$

This law holds for all real a, b, d, and m, because it depends only on the distributive law $(a \bmod m)d = ad \bmod md$: We have $a \bmod m = b \bmod m \iff (a \bmod m)d = (b \bmod m)d \iff ad \bmod md = bd \bmod md$. Thus, for example, from $3 \cdot 2 \equiv 5 \cdot 2 \pmod 4$ we conclude that $3 \equiv 5 \pmod 2$.

We can combine (4.37) and (4.38) to get a general law that changes the modulus as little as possible:

$$ad \equiv bd \pmod m \iff a \equiv b \left(\bmod \frac{m}{\gcd(d, m)}\right), \quad \text{integers } a, b, d, m. \tag{4.39}$$

For we can multiply $ad \equiv bd$ by d', where $d'd + m'm = \gcd(d, m)$; this gives the congruence $a \cdot \gcd(d, m) \equiv b \cdot \gcd(d, m) \pmod m$, which can be divided by $\gcd(d, m)$.

Let's look a bit further into this idea of changing the modulus. If we know that $a \equiv b \pmod{100}$, then we also must have $a \equiv b \pmod{10}$, or modulo any divisor of 100. It's stronger to say that $a - b$ is a multiple of 100

than to say that it's a multiple of 10. In general,

$$a \equiv b \pmod{md} \quad \Longrightarrow \quad a \equiv b \pmod{m}, \quad \text{integer } d, \qquad (4.40)$$

because any multiple of md is a multiple of m.

Conversely, if we know that $a \equiv b$ with respect to two small moduli, can we conclude that $a \equiv b$ with respect to a larger one? Yes; the rule is *Modulitos?*

$$a \equiv b \pmod{m} \quad \text{and} \quad a \equiv b \pmod{n} \\ \iff \quad a \equiv b \pmod{\operatorname{lcm}(m,n)}, \qquad \text{integers } m, n > 0. \qquad (4.41)$$

For example, if we know that $a \equiv b$ modulo 12 and 18, we can safely conclude that $a \equiv b \pmod{36}$. The reason is that if $a - b$ is a common multiple of m and n, it is a multiple of $\operatorname{lcm}(m,n)$. This follows from the principle of unique factorization.

The special case $m \perp n$ of this law is extremely important, because $\operatorname{lcm}(m,n) = mn$ when m and n are relatively prime. Therefore we will state it explicitly:

$$a \equiv b \pmod{mn} \\ \iff \quad a \equiv b \pmod{m} \text{ and } a \equiv b \pmod{n}, \quad \text{if } m \perp n. \qquad (4.42)$$

For example, $a \equiv b \pmod{100}$ if and only if $a \equiv b \pmod{25}$ and $a \equiv b \pmod{4}$. Saying this another way, if we know $x \bmod 25$ and $x \bmod 4$, then we have enough facts to determine $x \bmod 100$. This is a special case of the *Chinese Remainder Theorem* (see exercise 30), so called because it was discovered by Sun Tsŭ in China, about A.D. 350.

The moduli m and n in (4.42) can be further decomposed into relatively prime factors until every distinct prime has been isolated. Therefore

$$a \equiv b \pmod{m} \qquad \iff \qquad a \equiv b \pmod{p^{m_p}} \text{ for all } p,$$

if the prime factorization (4.11) of m is $\prod_p p^{m_p}$. Congruences modulo powers of primes are the building blocks for all congruences modulo integers.

4.7 INDEPENDENT RESIDUES

One of the important applications of congruences is a *residue number system*, in which an integer x is represented as a sequence of residues (or remainders) with respect to moduli that are prime to each other:

$$\operatorname{Res}(x) = (x \bmod m_1, \ldots, x \bmod m_r), \quad \text{if } m_j \perp m_k \text{ for } 1 \leqslant j < k \leqslant r.$$

Knowing $x \bmod m_1, \ldots, x \bmod m_r$ doesn't tell us everything about x. But it does allow us to determine $x \bmod m$, where m is the product $m_1 \ldots m_r$.

In practical applications we'll often know that x lies in a certain range; then we'll know everything about x if we know $x \bmod m$ and if m is large enough.

For example, let's look at a small case of a residue number system that has only two moduli, 3 and 5:

$x \bmod 15$	$x \bmod 3$	$x \bmod 5$
0	0	0
1	1	1
2	2	2
3	0	3
4	1	4
5	2	0
6	0	1
7	1	2
8	2	3
9	0	4
10	1	0
11	2	1
12	0	2
13	1	3
14	2	4

Each ordered pair $(x \bmod 3, x \bmod 5)$ is different, because $x \bmod 3 = y \bmod 3$ and $x \bmod 5 = y \bmod 5$ if and only if $x \bmod 15 = y \bmod 15$.

We can perform addition, subtraction, and multiplication on the two components *independently*, because of the rules of congruences. For example, if we want to multiply $7 = (1,2)$ by $13 = (1,3)$ modulo 15, we calculate $1 \cdot 1 \bmod 3 = 1$ and $2 \cdot 3 \bmod 5 = 1$. The answer is $(1,1) = 1$; hence $7 \cdot 13 \bmod 15$ must equal 1. Sure enough, it does.

This independence principle is useful in computer applications, because different components can be worked on separately (for example, by different computers). If each modulus m_k is a distinct prime p_k, chosen to be slightly less than 2^{31}, then a computer whose basic arithmetic operations handle integers in the range $[-2^{31} \mathinner{..} 2^{31})$ can easily compute sums, differences, and products modulo p_k. A set of r such primes makes it possible to add, subtract, and multiply "multiple-precision numbers" of up to almost $31r$ bits, and the residue system makes it possible to do this faster than if such large numbers were added, subtracted, or multiplied in other ways.

For example, the Mersenne prime $2^{31} - 1$ works well.

We can even do division, in appropriate circumstances. For example, suppose we want to compute the exact value of a large determinant of integers. The result will be an integer D, and bounds on $|D|$ can be given based on the size of its entries. But the only fast ways known for calculating determinants

require division, and this leads to fractions (and loss of accuracy, if we resort to binary approximations). The remedy is to evaluate $D \bmod p_k = D_k$, for various large primes p_k. We can safely divide modulo p_k unless the divisor happens to be a multiple of p_k. That's very unlikely, but if it does happen we can choose another prime. Finally, knowing D_k for sufficiently many primes, we'll have enough information to determine D.

But we haven't explained how to get from a given sequence of residues $(x \bmod m_1, \ldots, x \bmod m_r)$ back to $x \bmod m$. We've shown that this conversion can be done in principle, but the calculations might be so formidable that they might rule out the idea in practice. Fortunately, there is a reasonably simple way to do the job, and we can illustrate it in the situation $(x \bmod 3, x \bmod 5)$ shown in our little table. The key idea is to solve the problem in the two cases $(1,0)$ and $(0,1)$; for if $(1,0) = a$ and $(0,1) = b$, then $(x,y) = (ax + by) \bmod 15$, since congruences can be multiplied and added.

In our case $a = 10$ and $b = 6$, by inspection of the table; but how could we find a and b when the moduli are huge? In other words, if $m \perp n$, what is a good way to find numbers a and b such that the equations

$$a \bmod m = 1, \quad a \bmod n = 0, \quad b \bmod m = 0, \quad b \bmod n = 1$$

all hold? Once again, (4.5) comes to the rescue: With Euclid's algorithm, we can find m' and n' such that

$$m'm + n'n = 1.$$

Therefore we can take $a = n'n$ and $b = m'm$, reducing them both mod mn if desired.

Further tricks are needed in order to minimize the calculations when the moduli are large; the details are beyond the scope of this book, but they can be found in [208, page 274]. Conversion from residues to the corresponding original numbers is feasible, but it is sufficiently slow that we save total time only if a sequence of operations can all be done in the residue number system before converting back.

Let's firm up these congruence ideas by trying to solve a little problem: How many solutions are there to the congruence

$$x^2 \equiv 1 \pmod m, \tag{4.43}$$

if we consider two solutions x and x' to be the same when $x \equiv x'$?

According to the general principles explained earlier, we should consider first the case that m is a prime power, p^k, where $k > 0$. Then the congruence $x^2 \equiv 1$ can be written

$$(x-1)(x+1) \equiv 0 \pmod{p^k},$$

so p must divide either $x-1$ or $x+1$, or both. But p can't divide both $x-1$ and $x+1$ unless $p=2$; we'll leave that case for later. If $p>2$, then $p^k\backslash(x-1)(x+1) \iff p^k\backslash(x-1)$ or $p^k\backslash(x+1)$; so there are exactly two solutions, $x\equiv+1$ and $x\equiv-1$.

The case $p=2$ is a little different. If $2^k\backslash(x-1)(x+1)$ then either $x-1$ or $x+1$ is divisible by 2 but not by 4, so the other one must be divisible by 2^{k-1}. This means that we have four solutions when $k\geqslant 3$, namely $x\equiv\pm1$ and $x\equiv 2^{k-1}\pm 1$. (For example, when $p^k=8$ the four solutions are $x\equiv 1$, 3, 5, 7 (mod 8); it's often useful to know that *the square of any odd integer has the form* $8n+1$.)

Now $x^2\equiv 1 \pmod m$ if and only if $x^2\equiv 1 \pmod{p^{m_p}}$ for all primes p with $m_p>0$ in the complete factorization of m. Each prime is independent of the others, and there are exactly two possibilities for $x \bmod p^{m_p}$ except when $p=2$. Therefore if m has exactly r different prime divisors, the total number of solutions to $x^2\equiv 1$ is 2^r, except for a correction when m is even. The exact number in general is

All primes are odd except 2, which is the oddest of all.

$$2^{r+[8\backslash m]+[4\backslash m]-[2\backslash m]}. \tag{4.44}$$

For example, there are four "square roots of unity modulo 12," namely 1, 5, 7, and 11. When $m=15$ the four are those whose residues mod 3 and mod 5 are ±1, namely $(1,1)$, $(1,4)$, $(2,1)$, and $(2,4)$ in the residue number system. These solutions are 1, 4, 11, and 14 in the ordinary (decimal) number system.

4.8 ADDITIONAL APPLICATIONS

There's some unfinished business left over from Chapter 3: We wish to prove that the m numbers

$$0 \bmod m,\quad n \bmod m,\quad 2n \bmod m,\quad \ldots,\quad (m-1)n \bmod m \tag{4.45}$$

consist of precisely d copies of the m/d numbers

$$0,\quad d,\quad 2d,\quad \ldots,\quad m-d$$

in some order, where $d=\gcd(m,n)$. For example, when $m=12$ and $n=8$ we have $d=4$, and the numbers are 0, 8, 4, 0, 8, 4, 0, 8, 4, 0, 8, 4.

The first part of the proof—to show that we get d copies of the first m/d values—is now trivial. We have

Mathematicians love to say that things are trivial.

$$jn\equiv kn \pmod m \qquad\Longleftrightarrow\qquad j(n/d)\equiv k(n/d) \pmod{m/d}$$

by (4.38); hence we get d copies of the values that occur when $0\leqslant k<m/d$.

Now we must show that those m/d numbers are $\{0, d, 2d, \ldots, m-d\}$ in some order. Let's write $m = m'd$ and $n = n'd$. Then $kn \bmod m = d(kn' \bmod m')$, by the distributive law (3.23); so the values that occur when $0 \leqslant k < m'$ are d times the numbers

$$0 \bmod m',\ \ n' \bmod m',\ \ 2n' \bmod m',\ \ \ldots,\ \ (m'-1)n' \bmod m'.$$

But we know that $m' \perp n'$ by (4.27); we've divided out their gcd. Therefore we need only consider the case $d = 1$, namely the case that m and n are relatively prime.

So let's assume that $m \perp n$. In this case it's easy to see that the numbers (4.45) are just $\{0, 1, \ldots, m-1\}$ in some order, by using the "pigeonhole principle." This principle states that if m pigeons are put into m pigeonholes, there is an empty hole if and only if there's a hole with more than one pigeon. (Dirichlet's box principle, proved in exercise 3.8, is similar.) We know that the numbers (4.45) are distinct, because

$$jn \equiv kn \pmod m \qquad \Longleftrightarrow \qquad j \equiv k \pmod m$$

when $m \perp n$; this is (4.37). Therefore the m different numbers must fill all the pigeonholes $0, 1, \ldots, m-1$. Therefore the unfinished business of Chapter 3 is finished.

The proof is complete, but we can prove even more if we use a direct method instead of relying on the indirect pigeonhole argument. If $m \perp n$ and if a value $j \in [0\,..\,m)$ is given, we can explicitly compute $k \in [0\,..\,m)$ such that $kn \bmod m = j$ by solving the congruence

$$kn \ \equiv\ j \pmod m$$

for k. We simply multiply both sides by n', where $m'm + n'n = 1$, to get

$$k \ \equiv\ jn' \pmod m;$$

hence $k = jn' \bmod m$.

We can use the facts just proved to establish an important result discovered by Pierre de Fermat in 1640. Fermat was a great mathematician who contributed to the discovery of calculus and many other parts of mathematics. He left notebooks containing dozens of theorems stated without proof, and each of those theorems has subsequently been verified — with the possible exception of one that became the most famous of all, because it baffled the world's best mathematicians for 350 years. The famous one, called "Fermat's Last Theorem," states that

$$a^n + b^n \ \neq\ c^n \tag{4.46}$$

NEWS FLASH

Euler [115] conjectured that $a^4+b^4+c^4 \neq d^4$, *but Noam Elkies [92] found infinitely many solutions in August, 1987.*

Now Roger Frye has done an exhaustive computer search, proving (after about 110 hours on a Connection Machine) that the only solution with $d < 1000000$ *is:*

$$95800^4 + 217519^4 + 414560^4 = 422481^4 .$$

for all positive integers a, b, c, and n, when $n > 2$. (Of course there are lots of solutions to the equations $a + b = c$ and $a^2 + b^2 = c^2$.) Andrew Wiles culminated many years of research by announcing a proof of (4.46) in 1993; his proof is currently being subjected to intense scrutiny.

Fermat's theorem of 1640 is much easier to verify. It's now called Fermat's Little Theorem (or just Fermat's theorem, for short), and it states that

$$n^{p-1} \equiv 1 \pmod p, \qquad \text{if } n \perp p. \tag{4.47}$$

Proof: As usual, we assume that p denotes a prime. We know that the $p-1$ numbers $n \bmod p$, $2n \bmod p$, ..., $(p-1)n \bmod p$ are the numbers 1, 2, ..., $p-1$ in some order. Therefore if we multiply them together we get

$$\begin{aligned} &n \cdot (2n) \cdot \ldots \cdot \big((p-1)n\big) \\ &\qquad \equiv (n \bmod p) \cdot (2n \bmod p) \cdot \ldots \cdot \big((p-1)n \bmod p\big) \\ &\qquad \equiv (p-1)!\,, \end{aligned}$$

where the congruence is modulo p. This means that

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

and we can cancel the $(p-1)!$ since it's not divisible by p. QED.

An alternative form of Fermat's theorem is sometimes more convenient:

$$n^p \equiv n \pmod p, \qquad \text{integer } n. \tag{4.48}$$

This congruence holds for all integers n. The proof is easy: If $n \perp p$ we simply multiply (4.47) by n. If not, $p \backslash n$, so $n^p \equiv 0 \equiv n$.

In the same year that he discovered (4.47), Fermat wrote a letter to Mersenne, saying he suspected that the number

$$f_n = 2^{2^n} + 1$$

"... laquelle proposition, si elle est vraie, est de très grand usage."
—P. de Fermat [121]

would turn out to be prime for all $n \geqslant 0$. He knew that the first five cases gave primes:

$$2^1+1 = 3;\ 2^2+1 = 5;\ 2^4+1 = 17;\ 2^8+1 = 257;\ 2^{16}+1 = 65537;$$

but he couldn't see how to prove that the next case, $2^{32} + 1 = 4294967297$, would be prime.

It's interesting to note that Fermat could have proved that $2^{32}+1$ is *not* prime, using his own recently discovered theorem, if he had taken time to perform a few dozen multiplications: We can set $n = 3$ in (4.47), deducing that

$$3^{2^{32}} \equiv 1 \pmod{2^{32}+1}, \qquad \text{if } 2^{32}+1 \text{ is prime.}$$

And it's possible to test this relation by hand, beginning with 3 and squaring 32 times, keeping only the remainders mod $2^{32}+1$. First we have $3^2 = 9$, then $3^{2^2} = 81$, then $3^{2^3} = 6561$, and so on until we reach

If this is Fermat's Little Theorem, the other one was last but not least.

$$3^{2^{32}} \equiv 3029026160 \pmod{2^{32}+1}.$$

The result isn't 1, so $2^{32}+1$ isn't prime. This method of disproof gives us no clue about what the factors might be, but it does prove that factors exist. (They are 641 and 6700417, first found by Euler in 1732 [102].)

If $3^{2^{32}}$ had turned out to be 1, modulo $2^{32}+1$, the calculation wouldn't have proved that $2^{32}+1$ is prime; it just wouldn't have disproved it. But exercise 47 discusses a converse to Fermat's theorem by which we *can* prove that large prime numbers are prime, without doing an enormous amount of laborious arithmetic.

We proved Fermat's theorem by cancelling $(p-1)!$ from both sides of a congruence. It turns out that $(p-1)!$ is always congruent to -1, modulo p; this is part of a classical result known as Wilson's theorem:

$$(n-1)! \equiv -1 \pmod{n} \iff n \text{ is prime}, \quad \text{if } n > 1. \tag{4.49}$$

One half of this theorem is trivial: If $n > 1$ is not prime, it has a prime divisor p that appears as a factor of $(n-1)!$, so $(n-1)!$ cannot be congruent to -1. (If $(n-1)!$ were congruent to -1 modulo n, it would also be congruent to -1 modulo p, but it isn't.)

The other half of Wilson's theorem states that $(p-1)! \equiv -1 \pmod{p}$. We can prove this half by pairing up numbers with their inverses mod p. If $n \perp p$, we know that there exists n' such that

$$n'n \equiv 1 \pmod{p};$$

here n' is the inverse of n, and n is also the inverse of n'. Any two inverses of n must be congruent to each other, since $nn' \equiv nn''$ implies $n' \equiv n''$.

If p is prime, is p' prime prime?

Now suppose we pair up each number between 1 and $p-1$ with its inverse. Since the product of a number and its inverse is congruent to 1, the product of all the numbers in all pairs of inverses is also congruent to 1; so it seems that $(p-1)!$ is congruent to 1. Let's check, say for $p = 5$. We get $4! = 24$; but this is congruent to 4, not 1, modulo 5. Oops — what went wrong? Let's take a closer look at the inverses:

$$1' = 1, \quad 2' = 3, \quad 3' = 2, \quad 4' = 4.$$

Ah so; 2 and 3 pair up but 1 and 4 don't — they're their own inverses.

To resurrect our analysis we must determine which numbers are their own inverses. If x is its own inverse, then $x^2 \equiv 1 \pmod{p}$; and we have

already proved that this congruence has exactly two roots when $p > 2$. (If $p = 2$ it's obvious that $(p-1)! \equiv -1$, so we needn't worry about that case.) The roots are 1 and $p-1$, and the other numbers (between 1 and $p-1$) pair up; hence

$$(p-1)! \;\equiv\; 1 \cdot (p-1) \;\equiv\; -1\,,$$

as desired.

Unfortunately, we can't compute factorials efficiently, so Wilson's theorem is of no use as a practical test for primality. It's just a theorem.

4.9 PHI AND MU

How many of the integers $\{0, 1, \ldots, m-1\}$ are relatively prime to m? This is an important quantity called $\varphi(m)$, the "totient" of m (so named by J. J. Sylvester [347], a British mathematician who liked to invent new words). We have $\varphi(1) = 1$, $\varphi(p) = p-1$, and $\varphi(m) < m-1$ for all composite numbers m.

The φ function is called *Euler's totient function*, because Euler was the first person to study it. Euler discovered, for example, that Fermat's theorem (4.47) can be generalized to nonprime moduli in the following way:

"Si fuerit N *ad* x *numerus primus et* n *numerus partium ad* N *primarum, tum potestas* x^n *unitate minuta semper per numerum* N *erit divisibilis."*
—L. Euler [111]

$$n^{\varphi(m)} \;\equiv\; 1 \pmod m\,, \qquad \text{if } n \perp m. \tag{4.50}$$

(Exercise 32 asks for a proof of Euler's theorem.)

If m is a prime power p^k, it's easy to compute $\varphi(m)$, because $n \perp p^k \iff p \nmid n$. The multiples of p in $\{0, 1, \ldots, p^k-1\}$ are $\{0, p, 2p, \ldots, p^k-p\}$; hence there are p^{k-1} of them, and $\varphi(p^k)$ counts what is left:

$$\varphi(p^k) \;=\; p^k - p^{k-1}\,.$$

Notice that this formula properly gives $\varphi(p) = p-1$ when $k = 1$.

If $m > 1$ is not a prime power, we can write $m = m_1 m_2$ where $m_1 \perp m_2$. Then the numbers $0 \leqslant n < m$ can be represented in a residue number system as $(n \bmod m_1, n \bmod m_2)$. We have

$$n \;\perp\; m \qquad \Longleftrightarrow \qquad n \bmod m_1 \;\perp\; m_1 \quad \text{and} \quad n \bmod m_2 \;\perp\; m_2$$

by (4.30) and (4.4). Hence, $n \bmod m$ is "good" if and only if $n \bmod m_1$ and $n \bmod m_2$ are both "good," if we consider relative primality to be a virtue. The total number of good values modulo m can now be computed, recursively: It is $\varphi(m_1)\varphi(m_2)$, because there are $\varphi(m_1)$ good ways to choose the first component $n \bmod m_1$ and $\varphi(m_2)$ good ways to choose the second component $n \bmod m_2$ in the residue representation.

For example, $\varphi(12) = \varphi(4)\varphi(3) = 2\cdot 2 = 4$, because n is prime to 12 if and only if $n \bmod 4 = (1 \text{ or } 3)$ and $n \bmod 3 = (1 \text{ or } 2)$. The four values prime to 12 are $(1,1)$, $(1,2)$, $(3,1)$, $(3,2)$ in the residue number system; they are 1, 5, 7, 11 in ordinary decimal notation. Euler's theorem states that $n^4 \equiv 1 \pmod{12}$ whenever $n \perp 12$.

"Si sint A *et* B *numeri inter se primi et numerus partium ad* A *primarum sit* $= a$, *numerus vero partium ad* B *primarum sit* $= b$, *tum numerus partium ad productum* AB *primarum erit* $= ab$."
—*L. Euler [111]*

A function $f(m)$ of positive integers is called *multiplicative* if $f(1) = 1$ and

$$f(m_1 m_2) \;=\; f(m_1)f(m_2) \qquad \text{whenever } m_1 \perp m_2. \tag{4.51}$$

We have just proved that $\varphi(m)$ is multiplicative. We've also seen another instance of a multiplicative function earlier in this chapter: The number of incongruent solutions to $x^2 \equiv 1 \pmod m$ is multiplicative. Still another example is $f(m) = m^\alpha$ for any power α.

A multiplicative function is defined completely by its values at prime powers, because we can decompose any positive integer m into its prime-power factors, which are relatively prime to each other. The general formula

$$f(m) \;=\; \prod_p f(p^{m_p}), \qquad \text{if } m = \prod_p p^{m_p} \tag{4.52}$$

holds if and only if f is multiplicative.

In particular, this formula gives us the value of Euler's totient function for general m:

$$\varphi(m) \;=\; \prod_{p\backslash m} (p^{m_p} - p^{m_p-1}) \;=\; m \prod_{p\backslash m} \Big(1 - \frac{1}{p}\Big). \tag{4.53}$$

For example, $\varphi(12) = (4-2)(3-1) = 12(1-\frac12)(1-\frac13)$.

Now let's look at an application of the φ function to the study of rational numbers mod 1. We say that the fraction m/n is *basic* if $0 \leqslant m < n$. Therefore $\varphi(n)$ is the number of reduced basic fractions with denominator n; and the Farey series $\mathcal{F}_n$ contains all the reduced basic fractions with denominator n or less, as well as the non-basic fraction $\frac11$.

The set of *all* basic fractions with denominator 12, before reduction to lowest terms, is

$$\tfrac{0}{12},\ \tfrac{1}{12},\ \tfrac{2}{12},\ \tfrac{3}{12},\ \tfrac{4}{12},\ \tfrac{5}{12},\ \tfrac{6}{12},\ \tfrac{7}{12},\ \tfrac{8}{12},\ \tfrac{9}{12},\ \tfrac{10}{12},\ \tfrac{11}{12}\,.$$

Reduction yields

$$\tfrac{0}{1},\ \tfrac{1}{12},\ \tfrac{1}{6},\ \tfrac{1}{4},\ \tfrac{1}{3},\ \tfrac{5}{12},\ \tfrac{1}{2},\ \tfrac{7}{12},\ \tfrac{2}{3},\ \tfrac{3}{4},\ \tfrac{5}{6},\ \tfrac{11}{12}\,,$$

and we can group these fractions by their denominators:

$$\frac{0}{1};\quad \frac{1}{2};\quad \frac{1}{3},\frac{2}{3};\quad \frac{1}{4},\frac{3}{4};\quad \frac{1}{6},\frac{5}{6};\quad \frac{1}{12},\frac{5}{12},\frac{7}{12},\frac{11}{12}.$$

What can we make of this? Well, every divisor d of 12 occurs as a denominator, together with all $\varphi(d)$ of its numerators. The only denominators that occur are divisors of 12. Thus

$$\varphi(1)+\varphi(2)+\varphi(3)+\varphi(4)+\varphi(6)+\varphi(12) \;=\; 12\,.$$

A similar thing will obviously happen if we begin with the unreduced fractions $\frac{0}{m}, \frac{1}{m}, \ldots, \frac{m-1}{m}$ for any m, hence

$$\sum_{d\backslash m}\varphi(d) \;=\; m\,. \tag{4.54}$$

We said near the beginning of this chapter that problems in number theory often require sums over the divisors of a number. Well, (4.54) is one such sum, so our claim is vindicated. (We will see other examples.)

Now here's a curious fact: If f is any function such that the sum

$$g(m) \;=\; \sum_{d\backslash m} f(d)$$

is multiplicative, then f itself is multiplicative. (This result, together with (4.54) and the fact that $g(m)=m$ is obviously multiplicative, gives another reason why $\varphi(m)$ is multiplicative.) We can prove this curious fact by induction on m: The basis is easy because $f(1)=g(1)=1$. Let $m>1$, and assume that $f(m_1m_2)=f(m_1)f(m_2)$ whenever $m_1\perp m_2$ and $m_1m_2<m$. If $m=m_1m_2$ and $m_1\perp m_2$, we have

$$g(m_1m_2) \;=\; \sum_{d\backslash m_1m_2} f(d) \;=\; \sum_{d_1\backslash m_1}\;\sum_{d_2\backslash m_2} f(d_1d_2)\,,$$

and $d_1\perp d_2$ since all divisors of m_1 are relatively prime to all divisors of m_2. By the induction hypothesis, $f(d_1d_2)=f(d_1)f(d_2)$ except possibly when $d_1=m_1$ and $d_2=m_2$; hence we obtain

$$\begin{aligned}&\Bigl(\sum_{d_1\backslash m_1} f(d_1)\sum_{d_2\backslash m_2} f(d_2)\Bigr) - f(m_1)f(m_2)+f(m_1m_2)\\ &\qquad = \; g(m_1)g(m_2)-f(m_1)f(m_2)+f(m_1m_2)\,.\end{aligned}$$

But this equals $g(m_1m_2)=g(m_1)g(m_2)$, so $f(m_1m_2)=f(m_1)f(m_2)$.

Conversely, if $f(m)$ is multiplicative, the corresponding sum-over-divisors function $g(m) = \sum_{d\backslash m} f(d)$ is always multiplicative. In fact, exercise 33 shows that even more is true. Hence the curious fact and its converse are both facts.

The *Möbius function* $\mu(m)$, named after the nineteenth-century mathematician August Möbius who also had a famous band, can be defined for all integers $m \geq 1$ by the equation

$$\sum_{d\backslash m} \mu(d) = [m=1]. \tag{4.55}$$

This equation is actually a recurrence, since the left-hand side is a sum consisting of $\mu(m)$ and certain values of $\mu(d)$ with $d < m$. For example, if we plug in $m = 1, 2, \ldots, 12$ successively we can compute the first twelve values:

m	1	2	3	4	5	6	7	8	9	10	11	12
$\mu(m)$	1	−1	−1	0	−1	1	−1	0	0	1	−1	0

Richard Dedekind [77] and Joseph Liouville [251] noticed the following important "inversion principle" in 1857:

$$g(m) = \sum_{d\backslash m} f(d) \qquad \Longleftrightarrow \qquad f(m) = \sum_{d\backslash m} \mu(d)\, g(\frac{m}{d}). \tag{4.56}$$

According to this principle, the μ function gives us a new way to understand any function $f(m)$ for which we know $\sum_{d\backslash m} f(d)$.

Now is a good time to try warmup exercise 11.

The proof of (4.56) uses two tricks (4.7) and (4.9) that we described near the beginning of this chapter: If $g(m) = \sum_{d\backslash m} f(d)$ then

$$\begin{aligned}
\sum_{d\backslash m} \mu(d)\, g(\frac{m}{d}) &= \sum_{d\backslash m} \mu(\frac{m}{d})\, g(d) \\
&= \sum_{d\backslash m} \mu(\frac{m}{d}) \sum_{k\backslash d} f(k) \\
&= \sum_{k\backslash m} \sum_{d\backslash (m/k)} \mu(\frac{m}{kd})\, f(k) \\
&= \sum_{k\backslash m} \sum_{d\backslash (m/k)} \mu(d)\, f(k) \\
&= \sum_{k\backslash m} [m/k=1]\, f(k) = f(m).
\end{aligned}$$

The other half of (4.56) is proved similarly (see exercise 12).

Relation (4.56) gives us a useful property of the Möbius function, and we have tabulated the first twelve values; but what is the value of $\mu(m)$ when m is large? How can we solve the recurrence (4.55)? Well, the function $g(m) = [m=1]$ is obviously multiplicative — after all, it's zero except when $m = 1$. So the Möbius function defined by (4.55) must be multiplicative, by the curious fact we proved a minute or two ago. Therefore we can figure out what $\mu(m)$ is if we compute $\mu(p^k)$.

Depending on how fast you read.

When $m = p^k$, (4.55) says that

$$\mu(1) + \mu(p) + \mu(p^2) + \cdots + \mu(p^k) \;=\; 0$$

for all $k \geqslant 1$, since the divisors of p^k are $1, \ldots, p^k$. It follows that

$$\mu(p) \;=\; -1\,; \qquad \mu(p^k) \;=\; 0 \quad \text{for } k > 1.$$

Therefore by (4.52), we have the general formula

$$\mu(m) \;=\; \prod_{p \backslash m} \mu(p^{m_p}) \;=\; \begin{cases} (-1)^r, & \text{if } m = p_1 p_2 \ldots p_r; \\ 0, & \text{if } m \text{ is divisible by some } p^2. \end{cases} \tag{4.57}$$

That's μ.

If we regard (4.54) as a recurrence for the function $\varphi(m)$, we can solve that recurrence by applying the Dedekind-Liouville rule (4.56). We get

$$\varphi(m) \;=\; \sum_{d \backslash m} \mu(d)\, \frac{m}{d}\,. \tag{4.58}$$

For example,

$$\begin{aligned} \varphi(12) \;&=\; \mu(1)\cdot 12 + \mu(2)\cdot 6 + \mu(3)\cdot 4 + \mu(4)\cdot 3 + \mu(6)\cdot 2 + \mu(12)\cdot 1 \\ &=\; 12 - 6 - 4 + 0 + 2 + 0 \;=\; 4\,. \end{aligned}$$

If m is divisible by r different primes, say $\{p_1, \ldots, p_r\}$, the sum (4.58) has only 2^r nonzero terms, because the μ function is often zero. Thus we can see that (4.58) checks with formula (4.53), which reads

$$\varphi(m) \;=\; m\Big(1 - \frac{1}{p_1}\Big) \ldots \Big(1 - \frac{1}{p_r}\Big)\,;$$

if we multiply out the r factors $(1 - 1/p_j)$, we get precisely the 2^r nonzero terms of (4.58). The advantage of the Möbius function is that it applies in many situations besides this one.

For example, let's try to figure out how many fractions are in the Farey series $\mathcal{F}_n$. This is the number of reduced fractions in $[0\,..\,1]$ whose denominators do not exceed n, so it is 1 greater than $\Phi(n)$ where we define

$$\Phi(x) = \sum_{1\leqslant k\leqslant x} \varphi(k)\,. \tag{4.59}$$

(We must add 1 to $\Phi(n)$ because of the final fraction $\frac{1}{1}$.) The sum in (4.59) looks difficult, but we can determine $\Phi(x)$ indirectly by observing that

$$\sum_{d\geqslant 1} \Phi\Big(\frac{x}{d}\Big) = \frac{1}{2}\lfloor x\rfloor\lfloor 1+x\rfloor \tag{4.60}$$

for all real $x \geqslant 0$. Why does this identity hold? Well, it's a bit awesome yet not really beyond our ken. There are $\frac{1}{2}\lfloor x\rfloor\lfloor 1+x\rfloor$ basic fractions m/n with $0 \leqslant m < n \leqslant x$, counting both reduced and unreduced fractions; that gives us the right-hand side. The number of such fractions with $\gcd(m,n) = d$ is $\Phi(x/d)$, because such fractions are m'/n' with $0 \leqslant m' < n' \leqslant x/d$ after replacing m by $m'd$ and n by $n'd$. So the left-hand side counts the same fractions in a different way, and the identity must be true.

Let's look more closely at the situation, so that equations (4.59) and (4.60) become clearer. The definition of $\Phi(x)$ implies that $\Phi(x) = \Phi\big(\lfloor x\rfloor\big)$; but it turns out to be convenient to define $\Phi(x)$ for arbitrary real values, not just for integers. At integer values we have the table

(This extension to real values is a useful trick for many recurrences that arise in the analysis of algorithms.)

n	0	1	2	3	4	5	6	7	8	9	10	11	12
$\varphi(n)$	–	1	1	2	2	4	2	6	4	6	4	10	4
$\Phi(n)$	0	1	2	4	6	10	12	18	22	28	32	42	46

and we can check (4.60) when $x = 12$:

$$\begin{aligned}&\Phi(12)+\Phi(6)+\Phi(4)+\Phi(3)+\Phi(2)+\Phi(2)+6\cdot\Phi(1)\\ &\quad = 46+12+6+4+2+2+6 \;=\; 78 \;=\; \tfrac{1}{2}\cdot 12\cdot 13\,.\end{aligned}$$

Amazing.

Identity (4.60) can be regarded as an implicit recurrence for $\Phi(x)$; for example, we've just seen that we could have used it to calculate $\Phi(12)$ from certain values of $\Phi(m)$ with $m < 12$. And we can solve such recurrences by using another beautiful property of the Möbius function:

In fact, Möbius [273] invented his function because of (4.61), not (4.56).

$$g(x) = \sum_{d\geqslant 1} f(x/d) \qquad\Longleftrightarrow\qquad f(x) = \sum_{d\geqslant 1} \mu(d)\,g(x/d)\,. \tag{4.61}$$

This inversion law holds for all functions f such that $\sum_{k,d\geq 1}|f(x/kd)| < \infty$; we can prove it as follows. Suppose $g(x) = \sum_{d\geq 1} f(x/d)$. Then

$$\begin{aligned}\sum_{d\geq 1}\mu(d)g(x/d) &= \sum_{d\geq 1}\mu(d)\sum_{k\geq 1}f(x/kd)\\ &= \sum_{m\geq 1}f(x/m)\sum_{d,k\geq 1}\mu(d)[m=kd]\\ &= \sum_{m\geq 1}f(x/m)\sum_{d\backslash m}\mu(d) \;=\; \sum_{m\geq 1}f(x/m)[m=1] \;=\; f(x)\,.\end{aligned}$$

The proof in the other direction is essentially the same.

So now we can solve the recurrence (4.60) for $\Phi(x)$:

$$\Phi(x) \;=\; \frac{1}{2}\sum_{d\geq 1}\mu(d)\lfloor x/d\rfloor\lfloor 1+x/d\rfloor\,. \tag{4.62}$$

This is always a finite sum. For example,

$$\begin{aligned}\Phi(12) &= \tfrac{1}{2}(12\cdot 13-6\cdot 7-4\cdot 5+0-2\cdot 3+2\cdot 3\\ &\qquad -1\cdot 2+0+0+1\cdot 2-1\cdot 2+0)\\ &= 78-21-10-3+3-1+1-1 \;=\; 46\,.\end{aligned}$$

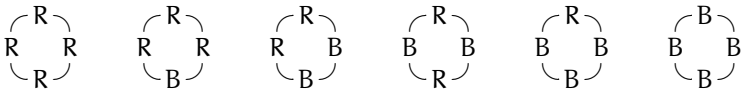
In Chapter 9 we'll see how to use (4.62) to get a good approximation to $\Phi(x)$; in fact, we'll prove a result due to Mertens in 1874 [270],

$$\Phi(x) \;=\; \frac{3}{\pi^2}x^2 + O(x\log x)\,.$$

Therefore the function $\Phi(x)$ grows "smoothly"; it averages out the erratic behavior of $\varphi(k)$.

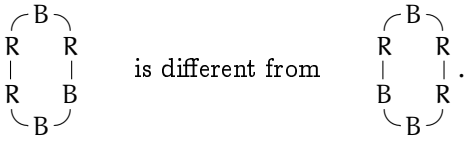
In keeping with the tradition established last chapter, let's conclude this chapter with a problem that illustrates much of what we've just seen and that also points ahead to the next chapter. Suppose we have beads of n different colors; our goal is to count how many different ways there are to string them into circular necklaces of length m. We can try to "name and conquer" this problem by calling the number of possible necklaces $N(m,n)$.

For example, with two colors of beads R and B, we can make necklaces of length 4 in $N(4,2) = 6$ different ways:



All other ways are equivalent to one of these, because rotations of a necklace do not change it. However, reflections are considered to be different; in the

case $m = 6$, for example,



The problem of counting these configurations was first solved by P. A. MacMahon in 1892 [264].

There's no obvious recurrence for $N(m,n)$, but we can count the necklaces by breaking them each into linear strings in m ways and considering the resulting fragments. For example, when $m = 4$ and $n = 2$ we get

RRRR	RRRR	RRRR	RRRR
RRBR	RRRB	BRRR	RBRR
RBBR	RRBB	BRRB	BBRR
RBRB	BRBR	RBRB	BRBR
RBBB	BRBB	BBRB	BBBR
BBBB	BBBB	BBBB	BBBB

Each of the n^m possible patterns appears at least once in this array of $mN(m,n)$ strings, and some patterns appear more than once. How many times does a pattern $a_0 \dots a_{m-1}$ appear? That's easy: It's the number of cyclic shifts $a_k \dots a_{m-1} a_0 \dots a_{k-1}$ that produce the same pattern as the original $a_0 \dots a_{m-1}$. For example, BRBR occurs twice, because the four ways to cut the necklace formed from BRBR produce four cyclic shifts (BRBR, RBRB, BRBR, RBRB); two of these coincide with BRBR itself. This argument shows that

$$\begin{aligned} mN(m,n) &= \sum_{a_0,\dots,a_{m-1}\in S_n} \sum_{0\leqslant k<m} \left[a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}\right] \\ &= \sum_{0\leqslant k<m} \sum_{a_0,\dots,a_{m-1}\in S_n} \left[a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}\right]. \end{aligned}$$

Here S_n is a set of n different colors.

Let's see how many patterns satisfy $a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}$, when k is given. For example, if $m = 12$ and $k = 8$, we want to count the number of solutions to

$$a_0 a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} \;=\; a_8 a_9 a_{10} a_{11} a_0 a_1 a_2 a_3 a_4 a_5 a_6 a_7 \,.$$

This means $a_0 = a_8 = a_4$; $a_1 = a_9 = a_5$; $a_2 = a_{10} = a_6$; and $a_3 = a_{11} = a_7$. So the values of a_0, a_1, a_2, and a_3 can be chosen in n^4 ways, and the

remaining a's depend on them. Does this look familiar? In general, the solution to

$$a_j \;=\; a_{(j+k) \bmod m}\,, \qquad \text{for } 0 \leqslant j < m$$

makes us equate a_j with $a_{(j+kl) \bmod m}$ for $l = 1, 2, \ldots$; and we know that the multiples of k modulo m are $\{0, d, 2d, \ldots, m-d\}$, where $d = \gcd(k, m)$. Therefore the general solution is to choose $a_0, \ldots, a_{d-1}$ independently and then to set $a_j = a_{j-d}$ for $d \leqslant j < m$. There are n^d solutions.

We have just proved that

$$mN(m,n) \;=\; \sum_{0 \leqslant k < m} n^{\gcd(k,m)}\,.$$

This sum can be simplified, since it includes only terms n^d where $d \backslash m$. Substituting $d = \gcd(k, m)$ yields

$$\begin{aligned} N(m,n) \;&=\; \frac{1}{m} \sum_{d \backslash m} n^d \sum_{0 \leqslant k < m} \big[d = \gcd(k,m)\big] \\ &= \frac{1}{m} \sum_{d \backslash m} n^d \sum_{0 \leqslant k < m} \big[k/d \perp m/d\big] \\ &= \frac{1}{m} \sum_{d \backslash m} n^d \sum_{0 \leqslant k < m/d} \big[k \perp m/d\big]\,. \end{aligned}$$

(We are allowed to replace k/d by k because k must be a multiple of d.) Finally, we have $\sum_{0 \leqslant k < m/d} [k \perp m/d] = \varphi(m/d)$ by definition, so we obtain MacMahon's formula:

$$N(m,n) \;=\; \frac{1}{m} \sum_{d \backslash m} n^d\, \varphi\Big(\frac{m}{d}\Big) \;=\; \frac{1}{m} \sum_{d \backslash m} \varphi(d)\, n^{m/d}\,. \tag{4.63}$$

When $m = 4$ and $n = 2$, for example, the number of necklaces is $\frac{1}{4}(1 \cdot 2^4 + 1 \cdot 2^2 + 2 \cdot 2^1) = 6$, just as we suspected.

It's not immediately obvious that the value $N(m,n)$ defined by MacMahon's sum is an integer! Let's try to prove directly that

$$\sum_{d \backslash m} \varphi(d)\, n^{m/d} \;\equiv\; 0 \qquad (\mathrm{mod}\ m)\,, \tag{4.64}$$

without using the clue that this is related to necklaces. In the special case that m is prime, this congruence reduces to $n^p + (p-1)n \equiv 0 \pmod p$; that is, it reduces to $n^p \equiv n$. We've seen in (4.48) that this congruence is an alternative form of Fermat's theorem. Therefore (4.64) holds when $m = p$;

we can regard it as a generalization of Fermat's theorem to the case when the modulus is not prime. (Euler's generalization (4.50) is different.)

We've proved (4.64) for all prime moduli, so let's look at the smallest case left, $m = 4$. We must prove that

$$n^4 + n^2 + 2n \equiv 0 \pmod 4 .$$

The proof is easy if we consider even and odd cases separately. If n is even, all three terms on the left are congruent to 0 modulo 4, so their sum is too. If n is odd, n^4 and n^2 are each congruent to 1, and $2n$ is congruent to 2; hence the left side is congruent to $1+1+2$ and thus to 0 modulo 4, and we're done.

Next, let's be a bit daring and try $m = 12$. This value of m ought to be interesting because it has lots of factors, including the square of a prime, yet it is fairly small. (Also there's a good chance we'll be able to generalize a proof for 12 to a proof for general m.) The congruence we must prove is

$$n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \equiv 0 \pmod{12} .$$

Now what? By (4.42) this congruence holds if and only if it also holds modulo 3 and modulo 4. So let's prove that it holds modulo 3. Our congruence (4.64) holds for primes, so we have $n^3 + 2n \equiv 0 \pmod 3$. Careful scrutiny reveals that we can use this fact to group terms of the larger sum:

$$\begin{aligned} &n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \\ &\quad = (n^{12} + 2n^4) + (n^6 + 2n^2) + 2(n^3 + 2n) \\ &\quad \equiv 0 + 0 + 2\cdot 0 \equiv 0 \pmod 3 . \end{aligned}$$

So it works modulo 3.

We're half done. To prove congruence modulo 4 we use the same trick. We've proved that $n^4 + n^2 + 2n \equiv 0 \pmod 4$, so we use this pattern to group:

$$\begin{aligned} &n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \\ &\quad = (n^{12} + n^6 + 2n^3) + 2(n^4 + n^2 + 2n) \\ &\quad \equiv 0 + 2\cdot 0 \equiv 0 \pmod 4 . \end{aligned}$$

QED for the case $m = 12$.

QED: Quite Easily Done.

So far we've proved our congruence for prime m, for $m = 4$, and for $m = 12$. Now let's try to prove it for prime powers. For concreteness we may suppose that $m = p^3$ for some prime p. Then the left side of (4.64) is

$$\begin{aligned} &n^{p^3} + \varphi(p)n^{p^2} + \varphi(p^2)n^p + \varphi(p^3)n \\ &\quad = n^{p^3} + (p-1)n^{p^2} + (p^2 - p)n^p + (p^3 - p^2)n \\ &\quad = (n^{p^3} - n^{p^2}) + p(n^{p^2} - n^p) + p^2(n^p - n) + p^3 n . \end{aligned}$$

We can show that this is congruent to 0 modulo p^3 if we can prove that $n^{p^3} - n^{p^2}$ is divisible by p^3, that $n^{p^2} - n^p$ is divisible by p^2, and that $n^p - n$ is divisible by p, because the whole thing will then be divisible by p^3. By the alternative form of Fermat's theorem we have $n^p \equiv n \pmod p$, so p divides $n^p - n$; hence there is an integer q such that

$$n^p = n + pq\,.$$

Now we raise both sides to the pth power, expand the right side according to the binomial theorem (which we'll meet in Chapter 5), and regroup, giving

$$\begin{aligned} n^{p^2} &= (n+pq)^p = n^p + (pq)^1 n^{p-1}\binom{p}{1} + (pq)^2 n^{p-2}\binom{p}{2} + \cdots \\ &= n^p + p^2 Q \end{aligned}$$

for some other integer Q. We're able to pull out a factor of p^2 here because $\binom{p}{1} = p$ in the second term, and because a factor of $(pq)^2$ appears in all the terms that follow. So we find that p^2 divides $n^{p^2} - n^p$.

Again we raise both sides to the pth power, expand, and regroup, to get

$$\begin{aligned} n^{p^3} &= (n^p + p^2 Q)^p \\ &= n^{p^2} + (p^2 Q)^1 n^{p(p-1)}\binom{p}{1} + (p^2 Q)^2 n^{p(p-2)}\binom{p}{2} + \cdots \\ &= n^{p^2} + p^3 \mathbf{Q} \end{aligned}$$

for yet another integer $\mathbf{Q}$. So p^3 divides $n^{p^3} - n^{p^2}$. This finishes the proof for $m = p^3$, because we've shown that p^3 divides the left-hand side of (4.64).

Moreover we can prove by induction that

$$n^{p^k} = n^{p^{k-1}} + p^k \mathfrak{Q}$$

for some final integer $\mathfrak{Q}$ (final because we're running out of fonts); hence

$$n^{p^k} \equiv n^{p^{k-1}} \pmod{p^k}, \qquad \text{for } k > 0. \tag{4.65}$$

Thus the left side of (4.64), which is

$$(n^{p^k} - n^{p^{k-1}}) + p(n^{p^{k-1}} - n^{p^{k-2}}) + \cdots + p^{k-1}(n^p - n) + p^k n\,,$$

is divisible by p^k and so is congruent to 0 modulo p^k.

We're almost there. Now that we've proved (4.64) for prime powers, all that remains is to prove it when $m = m_1 m_2$, where $m_1 \perp m_2$, assuming that the congruence is true for m_1 and m_2. Our examination of the case $m = 12$, which factored into instances of $m = 3$ and $m = 4$, encourages us to think that this approach will work.

We know that the φ function is multiplicative, so we can write

$$\begin{aligned}\sum_{d\backslash m}\varphi(d)\,n^{m/d} &= \sum_{d_1\backslash m_1,\,d_2\backslash m_2}\varphi(d_1d_2)\,n^{m_1m_2/d_1d_2}\\ &= \sum_{d_1\backslash m_1}\varphi(d_1)\Bigg(\sum_{d_2\backslash m_2}\varphi(d_2)(n^{m_1/d_1})^{m_2/d_2}\Bigg).\end{aligned}$$

But the inner sum is congruent to 0 modulo m_2, because we've assumed that (4.64) holds for m_2; so the entire sum is congruent to 0 modulo m_2. By a symmetric argument, we find that the entire sum is congruent to 0 modulo m_1 as well. Thus by (4.42) it's congruent to 0 modulo m. QED.

Exercises

Warmups

1 What is the smallest positive integer that has exactly k divisors, for $1 \leqslant k \leqslant 6$?

2 Prove that $\gcd(m,n)\cdot\operatorname{lcm}(m,n) = m\cdot n$, and use this identity to express $\operatorname{lcm}(m,n)$ in terms of $\operatorname{lcm}(n \bmod m, m)$, when $n \bmod m \neq 0$. *Hint:* Use (4.12), (4.14), and (4.15).

3 Let $\pi(x)$ be the number of primes not exceeding x. Prove or disprove:

$$\pi(x) - \pi(x-1) \;=\; [x \text{ is prime}]\,.$$

4 What would happen if the Stern–Brocot construction started with the five fractions $(\frac{0}{1}, \frac{1}{0}, \frac{0}{-1}, \frac{-1}{0}, \frac{0}{1})$ instead of with $(\frac{0}{1}, \frac{1}{0})$?

5 Find simple formulas for L^k and R^k, when L and R are the 2×2 matrices of (4.33).

6 What does '$a \equiv b \pmod 0$' mean?

7 Ten people numbered 1 to 10 are lined up in a circle as in the Josephus problem, and every mth person is executed. (The value of m may be much larger than 10.) Prove that the first three people to go cannot be 10, k, and $k+1$ (in this order), for any k.

8 The residue number system $(x \bmod 3, x \bmod 5)$ considered in the text has the curious property that 13 corresponds to $(1,3)$, which looks almost the same. Explain how to find all instances of such a coincidence, without calculating all fifteen pairs of residues. In other words, find all solutions to the congruences

$$10x + y \;\equiv\; x \pmod 3\,, \qquad 10x + y \;\equiv\; y \pmod 5\,.$$

Hint: Use the facts that $10u+6v \equiv u \pmod 3$ and $10u+6v \equiv v \pmod 5$.

9 Show that $(3^{77}-1)/2$ is odd and composite. *Hint:* What is $3^{77} \bmod 4$?

10 Compute $\varphi(999)$.

11 Find a function $\sigma(n)$ with the property that

$$g(n) = \sum_{0\leqslant k\leqslant n} f(k) \qquad \Longleftrightarrow \qquad f(n) = \sum_{0\leqslant k\leqslant n} \sigma(k)\, g(n-k)\,.$$

(This is analogous to the Möbius function; see (4.56).)

12 Simplify the formula $\sum_{d\backslash m} \sum_{k\backslash d} \mu(k)\, g(d/k)$.

13 A positive integer n is called *squarefree* if it is not divisible by m^2 for any $m > 1$. Find a necessary and sufficient condition that n is squarefree,

a in terms of the prime-exponent representation (4.11) of n;

b in terms of $\mu(n)$.

Basics

14 Prove or disprove:

a $\gcd(km, kn) = k\gcd(m, n)$;

b $\operatorname{lcm}(km, kn) = k\operatorname{lcm}(m, n)$.

15 Does every prime occur as a factor of some Euclid number e_n?

16 What is the sum of the reciprocals of the first n Euclid numbers?

17 Let f_n be the "Fermat number" $2^{2^n}+1$. Prove that $f_m \perp f_n$ if $m < n$.

18 Show that if 2^n+1 is prime then n is a power of 2.

19 Prove the following identities when n is a positive integer:

$$\sum_{1\leqslant k<n} \left\lfloor \frac{\varphi(k+1)}{k} \right\rfloor = \sum_{1<m\leqslant n} \left\lfloor \left(\sum_{1\leqslant k<m} \lfloor (m/k)/\lceil m/k \rceil \rfloor \right)^{-1} \right\rfloor$$
$$= n - 1 - \sum_{k=1}^{n} \left\lceil \left\{ \frac{(k-1)!+1}{k} \right\} \right\rceil .$$

Hint: This is a trick question and the answer is pretty easy.

20 For every positive integer n there's a prime p such that $n < p \leqslant 2n$. (This is essentially "Bertrand's postulate," which Joseph Bertrand verified for $n < 3000000$ in 1845 and Chebyshev proved for all n in 1850.) Use Bertrand's postulate to prove that there's a constant $b \approx 1.25$ such that the numbers

$$\lfloor 2^b \rfloor,\ \lfloor 2^{2^b} \rfloor,\ \lfloor 2^{2^{2^b}} \rfloor,\ \ldots$$

are all prime.

21 Let P_n be the nth prime number. Find a constant K such that

$$\lfloor (10^{n^2} K) \bmod 10^n \rfloor = P_n .$$

22 The number 1111111111111111 is prime. Prove that, in any radix b, $(11\ldots1)_b$ can be prime only if the number of 1's is prime. *Is this a test for strabismus?*

23 State a recurrence for $\rho(k)$, the ruler function in the text's discussion of $\epsilon_2(n!)$. Show that there's a connection between $\rho(k)$ and the disk that's moved at step k when an n-disk Tower of Hanoi is being transferred in $2^n - 1$ moves, for $1 \leqslant k \leqslant 2^n - 1$.

24 Express $\epsilon_p(n!)$ in terms of $\nu_p(n)$, the sum of the digits in the radix p representation of n, thereby generalizing (4.24). *Look, ma, sideways addition.*

25 We say that m *exactly divides* n, written $m \backslash\backslash n$, if $m \backslash n$ and $m \perp n/m$. For example, in the text's discussion of factorial factors, $p^{\epsilon_p(n!)} \backslash\backslash n!$. Prove or disprove the following:

a $k \backslash\backslash n$ and $m \backslash\backslash n \iff km \backslash\backslash n$, if $k \perp m$.

b For all $m, n > 0$, either $\gcd(m,n) \backslash\backslash m$ or $\gcd(m,n) \backslash\backslash n$.

26 Consider the sequence $\mathcal{G}_N$ of all nonnegative reduced fractions m/n such that $mn \leqslant N$. For example,

$$\mathcal{G}_{10} = \tfrac{0}{1}, \tfrac{1}{10}, \tfrac{1}{9}, \tfrac{1}{8}, \tfrac{1}{7}, \tfrac{1}{6}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{1}{1}, \tfrac{3}{2}, \tfrac{2}{1}, \tfrac{5}{2}, \tfrac{3}{1}, \tfrac{4}{1}, \tfrac{5}{1}, \tfrac{6}{1}, \tfrac{7}{1}, \tfrac{8}{1}, \tfrac{9}{1}, \tfrac{10}{1}.$$

Is it true that $m'n - mn' = 1$ whenever m/n immediately precedes m'/n' in $\mathcal{G}_N$?

27 Give a simple rule for comparing rational numbers based on their representations as L's and R's in the Stern–Brocot number system.

28 The Stern–Brocot representation of π is

$$\pi = R^3L^7R^{15}LR^{292}LRLR^2LR^3LR^{14}L^2R\ldots ;$$

use it to find all the simplest rational approximations to π whose denominators are less than 50. Is $\frac{22}{7}$ one of them?

29 The text describes a correspondence between binary real numbers $x = (.b_1b_2b_3\ldots)_2$ in $[0\,..\,1)$ and Stern–Brocot real numbers $\alpha = B_1B_2B_3\ldots$ in $[0\,..\,\infty)$. If x corresponds to α and $x \neq 0$, what number corresponds to $1 - x$?

30 Prove the following statement (the Chinese Remainder Theorem): Let $m_1, \ldots, m_r$ be integers with $m_j \perp m_k$ for $1 \leqslant j < k \leqslant r$; let $m = m_1 \ldots m_r$; and let $a_1, \ldots, a_r, A$ be integers. Then there is exactly one integer a such that

$$a \equiv a_k \pmod{m_k} \text{ for } 1 \leqslant k \leqslant r \quad \text{and} \quad A \leqslant a < A + m .$$

31 A number in decimal notation is divisible by 3 if and only if the sum of its digits is divisible by 3. Prove this well-known rule, and generalize it.

Why is "Euler" pronounced "Oiler" when "Euclid" is "Yooklid"?

32 Prove Euler's theorem (4.50) by generalizing the proof of (4.47).

33 Show that if $f(m)$ and $g(m)$ are multiplicative functions, then so is $h(m) = \sum_{d \backslash m} f(d)\, g(m/d)$.

34 Prove that (4.56) is a special case of (4.61).

Homework exercises

35 Let $I(m, n)$ be a function that satisfies the relation

$$I(m,n)m + I(n,m)n \;=\; \gcd(m,n)\,,$$

when m and n are nonnegative integers with $m \neq n$. Thus, $I(m,n) = m'$ and $I(n,m) = n'$ in (4.5); the value of $I(m,n)$ is an *inverse* of m with respect to n. Find a recurrence that defines $I(m,n)$.

36 Consider the set $Z(\sqrt{10}) = \{\, m + n\sqrt{10} \mid \text{integer } m, n \,\}$. The number $m + n\sqrt{10}$ is called a *unit* if $m^2 - 10n^2 = \pm 1$, since it has an inverse (that is, since $(m+n\sqrt{10}\,)\cdot\pm(m-n\sqrt{10}\,) = 1$). For example, $3+\sqrt{10}$ is a unit, and so is $19-6\sqrt{10}$. Pairs of cancelling units can be inserted into any factorization, so we ignore them. Nonunit numbers of $Z(\sqrt{10})$ are called prime if they cannot be written as a product of two nonunits. Show that 2, 3, and $4 \pm \sqrt{10}$ are primes of $Z(\sqrt{10})$. *Hint:* If $2 = (k + l\sqrt{10}\,) \times (m + n\sqrt{10}\,)$ then $4 = (k^2 - 10l^2)(m^2 - 10n^2)$. Furthermore, the square of any integer mod 10 is 0, 1, 4, 5, 6, or 9.

37 Prove (4.17). *Hint:* Show that $e_n - \frac{1}{2} = (e_{n-1} - \frac{1}{2})^2 + \frac{1}{4}$, and consider $2^{-n} \log(e_n - \frac{1}{2})$.

38 Prove that if $a \perp b$ and $a > b$ then

$$\gcd(a^m - b^m,\, a^n - b^n) \;=\; a^{\gcd(m,n)} - b^{\gcd(m,n)}\,, \qquad 0 \leqslant m < n.$$

(All variables are integers.) *Hint:* Use Euclid's algorithm.

39 Let $S(m)$ be the smallest positive integer n for which there exists an increasing sequence of integers

$$m \;=\; a_1 \;<\; a_2 \;<\; \cdots \;<\; a_t \;=\; n$$

such that $a_1 a_2 \ldots a_t$ is a perfect square. (If m is a perfect square, we can let $t = 1$ and $n = m$.) For example, $S(2) = 6$ because the best such sequence is $a_1 = 2$, $a_2 = 3$, $a_3 = 6$. We have

n	1	2	3	4	5	6	7	8	9	10	11	12
$S(n)$	1	6	8	4	10	12	14	15	9	18	22	20

Prove that $S(m) \neq S(m')$ whenever $0 < m < m'$.

40 If the radix p representation of n is $(a_m \ldots a_1 a_0)_p$, prove that

$$n!/p^{\epsilon_p(n!)} \equiv (-1)^{\epsilon_p(n!)} a_m! \ldots a_1! \, a_0! \pmod{p}.$$

(The left side is simply $n!$ with all p factors removed. When $n = p$ this reduces to Wilson's theorem.)

Wilson's theorem: "Martha, that boy is a menace."

41 **a** Show that if $p \bmod 4 = 3$, there is no integer n such that p divides $n^2 + 1$. *Hint:* Use Fermat's theorem.

b But show that if $p \bmod 4 = 1$, there is such an integer. *Hint:* Write $(p-1)!$ as $\left(\prod_{k=1}^{(p-1)/2} k(p-k)\right)$ and think about Wilson's theorem.

42 Consider two fractions m/n and m'/n' in lowest terms. Prove that when the sum $m/n + m'/n'$ is reduced to lowest terms, the denominator will be nn' if and only if $n \perp n'$. (In other words, $(mn' + m'n)/nn'$ will already be in lowest terms if and only if n and n' have no common factor.)

43 There are 2^k nodes at level k of the Stern–Brocot tree, corresponding to the matrices L^k, $L^{k-1}R$, ..., R^k. Show that this sequence can be obtained by starting with L^k and then multiplying successively by

$$\begin{pmatrix} 0 & -1 \\ 1 & 2\rho(n)+1 \end{pmatrix}$$

for $1 \leqslant n < 2^k$, where $\rho(n)$ is the ruler function.

Radio announcer: "... pitcher Mark LeChiffre hits a two-run single! Mark, who was batting .080, gets his second hit of the year." Anything wrong?

44 Prove that a baseball player whose batting average is .316 must have batted at least 19 times. (If he has m hits in n times at bat, then $m/n \in [0.3155 \ldots 0.3165)$.)

45 The number 9376 has the peculiar self-reproducing property that

$$9376^2 = 87909376.$$

How many 4-digit numbers x satisfy the equation $x^2 \bmod 10000 = x$? How many n-digit numbers x satisfy the equation $x^2 \bmod 10^n = x$?

46 **a** Prove that if $n^j \equiv 1$ and $n^k \equiv 1 \pmod{m}$, then $n^{\gcd(j,k)} \equiv 1$.

b Show that $2^n \not\equiv 1 \pmod{n}$, if $n > 1$. *Hint:* Consider the least prime factor of n.

47 Show that if $n^{m-1} \equiv 1 \pmod{m}$ and if $n^{(m-1)/p} \not\equiv 1 \pmod{m}$ for all primes such that $p \backslash (m-1)$, then m is prime. *Hint:* Show that if this condition holds, the numbers $n^k \bmod m$ are distinct, for $1 \leqslant k < m$.

48 Generalize Wilson's theorem (4.49) by ascertaining the value of the expression $\left(\prod_{1 \leqslant n < m,\, n \perp m} n\right) \bmod m$, when $m > 1$.

49 Let $R(N)$ be the number of pairs of integers (m, n) such that $0 \leqslant m < N$, $0 \leqslant n < N$, and $m \perp n$.

a Express $R(N)$ in terms of the Φ function.

b Prove that $R(N) = \sum_{d \geqslant 1} \lfloor N/d \rfloor^2 \mu(d)$.

50 Let m be a positive integer and let

$$\omega \;=\; e^{2\pi i/m} \;=\; \cos(2\pi/m) + i \sin(2\pi/m)\,.$$

What are the roots of disunity?

We say that ω is an mth *root of unity*, since $\omega^m = e^{2\pi i} = 1$. In fact, each of the m complex numbers ω^0, ω^1, $\ldots$, ω^{m-1} is an mth root of unity, because $(\omega^k)^m = e^{2\pi k i} = 1$; therefore $z - \omega^k$ is a factor of the polynomial $z^m - 1$, for $0 \leqslant k < m$. Since these factors are distinct, the complete factorization of $z^m - 1$ over the complex numbers must be

$$z^m - 1 \;=\; \prod_{0 \leqslant k < m} (z - \omega^k)\,.$$

a Let $\Psi_m(z) = \prod_{0 \leqslant k < m,\, k \perp m} (z - \omega^k)$. (This polynomial of degree $\varphi(m)$ is called the *cyclotomic polynomial of order* m.) Prove that

$$z^m - 1 \;=\; \prod_{d \backslash m} \Psi_d(z)\,.$$

b Prove that $\Psi_m(z) = \prod_{d \backslash m} (z^d - 1)^{\mu(m/d)}$.

Exam problems

51 Prove Fermat's theorem (4.48) by expanding $(1 + 1 + \cdots + 1)^p$ via the multinomial theorem.

52 Let n and x be positive integers such that x has no divisors $\leqslant n$ (except 1), and let p be a prime number. Prove that at least $\lfloor n/p \rfloor$ of the numbers $\{x - 1, x^2 - 1, \ldots, x^{n-1} - 1\}$ are multiples of p.

53 Find all positive integers n such that $n \setminus \lfloor (n-1)!/(n+1) \rfloor$.

54 Determine the value of $1000! \bmod 10^{250}$ by hand calculation.

55 Let P_n be the product of the first n factorials, $\prod_{k=1}^{n} k!$. Prove that P_{2n}/P_n^4 is an integer, for all positive integers n.

56 Show that

$$\Bigg(\prod_{k=1}^{2n-1} k^{\min(k,\, 2n-k)} \Bigg) \Bigg/ \Bigg(\prod_{k=1}^{n-1} (2k+1)^{2n-2k-1} \Bigg)$$

is a power of 2.

57 Let $S(m,n)$ be the set of all integers k such that

$$m \bmod k + n \bmod k \;\geqslant\; k\,.$$

For example, $S(7,9) = \{2,4,5,8,10,11,12,13,14,15,16\}$. Prove that

$$\sum_{k\in S(m,n)} \varphi(k) \;=\; mn\,.$$

Hint: Prove first that $\sum_{1\leqslant m\leqslant n}\sum_{d\backslash m}\varphi(d) = \sum_{d\geqslant 1}\varphi(d)\lfloor n/d\rfloor$. Then consider $\lfloor (m+n)/d\rfloor - \lfloor m/d\rfloor - \lfloor n/d\rfloor$.

58 Let $f(m) = \sum_{d\backslash m} d$. Find a necessary and sufficient condition that $f(m)$ is a power of 2.

Bonus problems

59 Prove that if $x_1, \ldots, x_n$ are positive integers with $1/x_1 + \cdots + 1/x_n = 1$, then $\max(x_1,\ldots,x_n) < e_n$. *Hint:* Prove the following stronger result by induction: "If $1/x_1 + \cdots + 1/x_n + 1/\alpha = 1$, where $x_1, \ldots, x_n$ are positive integers and α is a rational number $\geqslant \max(x_1,\ldots,x_n)$, then $\alpha+1 \leqslant e_{n+1}$ and $x_1 \ldots x_n(\alpha+1) \leqslant e_1 \ldots e_n e_{n+1}$." (The proof is nontrivial.)

60 Prove that there's a constant P such that (4.18) gives only primes. You may use the following (highly nontrivial) fact: There is a prime between p and $p + p^\theta$, for all sufficiently large p, if $\theta > \frac{6}{11}$.

61 Prove that if m/n, m'/n', and m''/n'' are consecutive elements of $\mathcal{F}_N$, then

$$\begin{aligned} m'' &= \lfloor (n+N)/n' \rfloor m' - m\,, \\ n'' &= \lfloor (n+N)/n' \rfloor n' - n\,. \end{aligned}$$

(This recurrence allows us to compute the elements of $\mathcal{F}_N$ in order, starting with $\frac{0}{1}$ and $\frac{1}{N}$.)

62 What binary number corresponds to e, in the binary $\leftrightarrow$ Stern–Brocot correspondence? (Express your answer as an infinite sum; you need not evaluate it in closed form.)

63 Using only the methods of this chapter, show that if Fermat's Last Theorem (4.46) were false, the least n for which it fails would have to be prime. (You may assume that (4.46) holds when $n = 4$.) Furthermore, if $a^p + b^p = c^p$ is the smallest counterexample, show that

$$a + b \;=\; \begin{cases} m^p, & \text{if } p\not\backslash c, \\ p^{p-1}m^p, & \text{if } p\backslash c, \end{cases}$$

for some integer m. Thus $c \geqslant m^p/2$ must be really huge. *Hint:* Let $x = a + b$, and note that $\gcd\big(x, (a^p + (x-a)^p)/x\big) = \gcd(x, pa^{p-1})$.

64 The *Peirce sequence* $\mathcal{P}_N$ *of order* N is an infinite string of fractions separated by '$<$' or '$=$' signs, containing all the nonnegative fractions m/n with $m \geqslant 0$ and $n \leqslant N$ (including fractions that are not reduced). It is defined recursively by starting with

$$\mathcal{P}_1 \;=\; \tfrac{0}{1}<\tfrac{1}{1}<\tfrac{2}{1}<\tfrac{3}{1}<\tfrac{4}{1}<\tfrac{5}{1}<\tfrac{6}{1}<\tfrac{7}{1}<\tfrac{8}{1}<\tfrac{9}{1}<\tfrac{10}{1}<\cdots.$$

For $N \geqslant 1$, we form $\mathcal{P}_{N+1}$ by inserting two symbols just before the kNth symbol of $\mathcal{P}_N$, for all $k > 0$. The two inserted symbols are

$$\begin{array}{ll} \dfrac{k-1}{N+1} & = , \qquad \text{if } kN \text{ is odd;} \\ \mathcal{P}_{N,kN} & \dfrac{k-1}{N+1}\,, \qquad \text{if } kN \text{ is even.} \end{array}$$

Here $\mathcal{P}_{N,j}$ denotes the jth symbol of $\mathcal{P}_N$, which will be either '$<$' or '$=$' when j is even; it will be a fraction when j is odd. For example,

$$\begin{aligned}
\mathcal{P}_2 &= \tfrac{0}{2}=\tfrac{0}{1}<\tfrac{1}{2}<\tfrac{2}{2}=\tfrac{1}{1}<\tfrac{3}{2}<\tfrac{4}{2}=\tfrac{2}{1}<\tfrac{5}{2}<\tfrac{6}{2}=\tfrac{3}{1}<\tfrac{7}{2}<\tfrac{8}{2}=\tfrac{4}{1}<\tfrac{9}{2}<\tfrac{10}{2}=\tfrac{5}{1}<\cdots; \\
\mathcal{P}_3 &= \tfrac{0}{2}=\tfrac{0}{3}=\tfrac{0}{1}<\tfrac{1}{3}<\tfrac{1}{2}<\tfrac{2}{3}<\tfrac{2}{2}=\tfrac{3}{3}=\tfrac{1}{1}<\tfrac{4}{3}<\tfrac{3}{2}<\tfrac{5}{3}<\tfrac{4}{2}=\tfrac{6}{3}=\tfrac{2}{1}<\tfrac{7}{3}<\tfrac{5}{2}<\cdots; \\
\mathcal{P}_4 &= \tfrac{0}{2}=\tfrac{0}{4}=\tfrac{0}{3}=\tfrac{0}{1}<\tfrac{1}{4}<\tfrac{1}{3}<\tfrac{2}{4}=\tfrac{1}{2}<\tfrac{2}{3}<\tfrac{3}{4}<\tfrac{2}{2}=\tfrac{4}{4}=\tfrac{3}{3}=\tfrac{1}{1}<\tfrac{5}{4}<\tfrac{4}{3}<\tfrac{6}{4}=\cdots; \\
\mathcal{P}_5 &= \tfrac{0}{2}=\tfrac{0}{4}=\tfrac{0}{5}=\tfrac{0}{3}=\tfrac{0}{1}<\tfrac{1}{5}<\tfrac{1}{4}<\tfrac{1}{3}<\tfrac{2}{5}<\tfrac{2}{4}=\tfrac{1}{2}<\tfrac{2}{5}<\tfrac{2}{3}<\tfrac{3}{4}<\tfrac{4}{5}<\tfrac{2}{2}=\tfrac{4}{4}=\cdots; \\
\mathcal{P}_6 &= \tfrac{0}{2}=\tfrac{0}{4}=\tfrac{0}{6}=\tfrac{0}{5}=\tfrac{0}{3}=\tfrac{0}{1}<\tfrac{1}{6}<\tfrac{1}{5}<\tfrac{1}{4}<\tfrac{2}{6}=\tfrac{1}{3}<\tfrac{2}{5}<\tfrac{2}{4}=\tfrac{3}{6}=\tfrac{1}{2}<\tfrac{3}{5}<\tfrac{4}{6}=\cdots.
\end{aligned}$$

(Equal elements occur in a slightly peculiar order.) Prove that the '$<$' and '$=$' signs defined by the rules above correctly describe the relations between adjacent fractions in the Peirce sequence.

Research problems

65 Are the Euclid numbers e_n all squarefree?

66 Are the Mersenne numbers $2^p - 1$ all squarefree?

67 Prove or disprove that $\max_{1\leqslant j<k\leqslant n} a_k/\gcd(a_j, a_k) \geqslant n$, for all sequences of integers $0 < a_1 < \cdots < a_n$.

68 Is there a constant Q such that $\lfloor Q^{2^n} \rfloor$ is prime for all $n \geqslant 0$?

69 Let P_n denote the nth prime. Prove or disprove that $P_{n+1} - P_n = O(\log P_n)^2$.

70 Does $\epsilon_3(n!) = \epsilon_2(n!)/2$ for infinitely many n?

71 Prove or disprove: If $k \neq 1$ there exists $n > 1$ such that $2^n \equiv k \pmod n$. Are there infinitely many such n?

72 Prove or disprove: For all integers a, there exist infinitely many n such that $\varphi(n)\backslash(n + a)$.

73 If the $\Phi(n)+1$ terms of the Farey series

$$\mathcal{F}_n \;=\; \langle \mathcal{F}_n(0), \mathcal{F}_n(1), \ldots, \mathcal{F}_n\big(\Phi(n)\big)\rangle$$

were fairly evenly distributed, we would expect $\mathcal{F}_n(k) \approx k/\Phi(n)$. Therefore the sum $D(n) = \sum_{k=0}^{\Phi(n)} \big|\mathcal{F}_n(k) - k/\Phi(n)\big|$ measures the "deviation of $\mathcal{F}_n$ from uniformity." Is it true that $D(n) = O(n^{1/2+\epsilon})$ for all $\epsilon > 0$?

74 Approximately how many distinct values are there in the set $\{0! \bmod p, 1! \bmod p, \ldots, (p-1)! \bmod p\}$, as $p \to \infty$?