Chapter 2

Prime Numbers

The term *factoring* or *factorization* refers to the process of expressing an integer as the product of two or more integers in a nontrivial way, e.g., $42 = 6 \times 7$. Prime numbers are those for which this process cannot be done. We will soon see that prime numbers are the building blocks of the integers. Together with the theory of divisibility, the properties of primes are foundational elements of number theory.

2.1 Primes and Composites

Definition. We call an integer $p \geq 2$ a *prime* if it has no divisors strictly between 1 and p. An integer $n \geq 2$ which is not a prime is called *composite*. The dichotomy between primes and composites thus takes place:

primes:	2,	3,		5,		7,				11,		13,				17,	...
composites:			4,		6,		8,	9,	10,		12,		14,	15,	16,		...

The words prime and composite are also used as adjectives, as in a *prime* number,[1] or a *composite* integer. *Primality* and *compositeness* are the nouns associated with the two, respectively. Throughout this book, from now on, we shall designate the notation p to always represent a prime, lest we forget to remind the reader.

Proposition 2.1. The following statements hold.

1) Other than 2, all primes are odd numbers.[2]

[1] Do not mistake *prime* numbers (lacking nontrivial factors) for *relatively prime* numbers (lacking common factors). It is true, two distinct primes are always relatively prime.

[2] Being the only even prime, 2 is the odd one out!

2) Every composite has a prime divisor.

3) The number n is composite if and only if n has a prime divisor $p \leq \sqrt{n}$.

Proof. 1) By definition, even numbers are multiples of 2, hence they are all composite, except 2 itself is prime.

2) Suppose, by induction, the statement is true up to $n-1$. Either n is prime, and nothing to prove, or else n has a divisor d, with $1 < d < n$. It follows that d has a prime divisor which is also a divisor of n, by Proposition 1.1(3).

3) A prime has no prime divisor less than itself. For composite $n = ab$, where $a > 1$ and $b > 1$, either $a \leq \sqrt{n}$ or $b \leq \sqrt{n}$ must hold. Whichever is true, by (2) a or b has a prime divisor p, where $p \leq \sqrt{n}$ and $p \mid n$. $\triangledown$

Propositions 2.1(3) implies that in order to test the primality of a number n, it suffices to check divisibility by the primes 2, 3, 5, ..., up to $\sqrt{n}$. This testing algorithm is called the *trial division*. For example, $\sqrt{113} \approx 10.63$, and the only primes up to 10 are 2, 3, 5, 7—none of which divides 113. Hence, 113 is a prime.

Exercise 2.1. Determine prime or composite, using trial division.

a) 383
b) 447
c) 799
d) 811
e) 1763

Exercise 2.2. Investigate true or false.

a) The number $n + 99!$ is composite for each $2 \leq n \leq 100$.
b) The number $n^4 + 4$ is composite for each $n \geq 2$.
c) The number $n^2 + n + 41$ is prime for each $n \geq 0$.
d) The number $n^2 - 81n + 1681$ is prime for each $n \geq 1$.

Finding multiples is obviously much easier than finding divisors. Inspired by this fact, the trial division is turned into the *sieve of Eratosthenes*, a relatively efficient algorithm to identify the prime numbers up to a predetermined bound N. It works as follows.

We just ignore the even numbers, except the prime 2, then we list the odd numbers from 3 to N. The smallest in this list, 3, is prime. Now label all its multiples by this prime 3. Among the *unlabeled* remnant in the list, 5 is now smallest. Label all its multiples by this prime 5. Repeat this procedure until all up to $\sqrt{N}$ have been labeled. The unlabeled numbers which remain are all primes.

Example (The Sieve of Eratosthenes). With $N = 101$, it takes only 3 iterations, with primes 3, 5, 7, and the output is given below. The label following each number n, separated by a colon (:), also indicates the smallest prime divisor of n. The odd primes, surviving the sieve, are displayed in bold.

3:3	**5**:5	**7**:7	9:3	**11**	**13**	15:3	**17**	**19**	21:3
23	25:5	27:3	**29**	**31**	33:3	35:5	**37**	39:3	**41**
43	45:3	**47**	49:7	51:3	**53**	55:5	57:3	**59**	**61**
63:3	65:5	**67**	69:3	**71**	**73**	75:3	77:7	**79**	81:3
83	85:5	87:3	**89**	91:7	93:3	95:5	**97**	99:3	**101**

Table 2.1: The sieve of Eratosthenes for odd primes up to 101.

EXERCISE 2.3. Using the sieve of Eratosthenes, find all primes up to 1,000. Alternately, implement this algorithm using a programming language of your choice, say with $N = 10,000$.

Some divisibility properties involving primes will now be presented. The simplest result is perhaps this next lemma, to be followed by a very useful theorem, a consequence of Euclid's lemma.

Lemma 2.2. Let p be a prime. For any integer n, we have $\gcd(p, n) = p$ if $p \mid n$, otherwise $\gcd(p, n) = 1$.

Proof. The claim is justified since 1 and p are the only divisors of p. ▽

Theorem 2.3. If $p \mid mn$, then either $p \mid m$ or $p \mid n$. More generally, if a prime p divides the product of integers, then p divides one of them.

Proof. If $p \nmid n$ then by Lemma 2.2, $\gcd(p, n) = 1$, and by Theorem 1.8(1) (Euclid's lemma) we then have $p \mid m$. Repeated use of this argument establishes the general claim. ▽

EXERCISE 2.4. Show that if $p \mid n^2$ then $p^2 \mid n^2$, where p is prime.

2.2 Factoring Composites into Primes

Intuitively, by factoring a given composite n and further factoring its factors if necessary, we should be able to write n as a product of only prime numbers. The next theorem, which is of greatest importance in the theory of numbers, assures not only that *factorization into primes* can always be done, but also that the end collection of prime factors is uniquely determined by n. For example, in factoring the number 1998, one may obtain $1998 = 2 \times 3 \times 3 \times$

3×37, or $1998 = 3 \times 37 \times 3 \times 2 \times 3$, but it would be impossible to find another prime factor outside the collection $\{2, 3, 3, 3, 37\}$. Here, a *collection* is like a set in which repetition of elements is taken into account.

Theorem 2.4 (The Fundamental Theorem of Arithmetic)**.** Every composite is the product of a unique collection of prime numbers.

Proof. Claim first that every composite is a product of primes. Suppose this is true up to $n-1$. If n is prime, there is nothing to prove. Else by Proposition 2.1(2), $n = pn'$ for a prime p and $n' < n$. Either n' is prime or, by the induction hypothesis, a product of primes. Thus the claim is true.

To prove uniqueness, we proceed by contradiction. Suppose we have two different collections of primes, p's and q's, whose products both equal n. Equating these products, after canceling out all common terms, will result in $p_1 p_2 \cdots p_j = q_1 q_2 \cdots q_k$, where none of the p's equals any of the q's. By Theorem 2.3, p_1 must divide one of the q's. But it is impossible for a prime to divide another prime other than itself. ▽

Exercise 2.5. Factor these numbers into primes.

a) 123
b) 400
c) 720
d) 7575
e) 19392

We will become familiar with the notation $n = \prod p_i^{e_i}$, where n has been factored into powers of distinct primes, $e_i \geq 1$. We call this the *factorization into prime powers*—distinct is implicitly assumed, always. Occasionally it will be more convenient to have the product $n = \prod p_i^{e_i}$ span over all prime numbers by allowing $e_i \geq 0$. Of course, $e_i = 0$ for all but finitely many of them. With this assumption, if $d \mid n$ then it is a consequence of Theorem 2.4 that $d = \prod p_i^{d_i}$, where each $d_i \leq e_i$.

Exercise 2.6. Prove that if $d^2 \mid n^2$ then $d \mid n$.

Exercise 2.7. Count how many positive divisors each number has.

a) 323
b) 720
c) 1024
d) 2310
e) 19392

Corollary 2.5. Suppose that m and n have been factored into prime powers, $m = \prod p_i^{f_i}$ and $n = \prod p_i^{e_i}$, with $e_i, f_i \geq 0$. Then $\gcd(m, n) = \prod p_i^{d_i}$, where $d_i = \min(e_i, f_i)$, the lesser of the two.

Proof. By Theorem 2.4, a divisor of m must be of the form $d = \prod p_i^{d_i}$ with $d_i \leq f_i$. Similarly, if $d \mid n$ then $d_i \leq e_i$. So the greatest possible value for such d is when $d_i = \min(e_i, f_i)$. ▽

Example. We evaluate $\gcd(27720, 61152)$ using prime factorization,

$$\begin{aligned} 27720 &= 2^3 \times 3^2 \times 5^1 \times 7^1 \times 11^1 \times 13^0 \\ 61152 &= 2^5 \times 3^1 \times 5^0 \times 7^2 \times 11^0 \times 13^1 \end{aligned}$$

and get the result, $\gcd(27720, 61152) = 2^3 \times 3 \times 7 = 168$.

Exercise 2.8. Evaluate $\gcd(m, n)$ by factoring m and n.

a) $\gcd(400, 720)$
b) $\gcd(19392, 29391)$
c) $\gcd(2^3 \times 3^8 \times 5^4 \times 7^5,\ 3^7 \times 5^2 \times 7^2)$
d) $\gcd(2^5 \times 5^7 \times 11^3,\ 3^7 \times 7^2 \times 13^9)$
e) $\gcd(2^4 \times 5^2 \times 7 \times 11^3,\ 2^7 \times 3^2 \times 5^2 \times 11)$

Exercise 2.9. Show that $\gcd(m^2, n^2) = \gcd(m, n)^2$.

Exercise 2.10. The *least common multiple* of two nonzero integers is the smallest positive integer which is divisible by both. For example, $\operatorname{lcm}(4, 6) = 12$ because it is the smallest positive integer such that $4 \mid 12$ and $6 \mid 12$.

a) Suppose $m = \prod p_i^{f_i}$ and $n = \prod p_i^{e_i}$, where $e_i \geq 0$ and $f_i \geq 0$. Prove that $\operatorname{lcm}(m, n) = \prod p_i^{d_i}$, where $d_i = \max(e_i, f_i)$, the greater of the two.
b) Show that if $m \mid k$ and $n \mid k$, then $\operatorname{lcm}(m, n) \mid k$.
c) Find an equation relating $\gcd(m, n)$ to $\operatorname{lcm}(m, n)$.
d) Illustrate your answer in (c) using $m = 600$ and $n = 630$.

Thus we have now another method for evaluating $\gcd(m, n)$, totally independent from the Euclidean algorithm. In contrast, however, factoring is slow and the computation time grows exponentially with the size of the integer.

More about factorization will be discussed in Chapter 7; in the meantime, we demonstrate next a factorization technique due to Fermat. Although the method is old, many modern and powerful factoring algorithms are actually based on this principle.

If $n = x^2 - y^2$, then n factors as $n = (x + y)(x - y)$. This fact is the simple idea behind the method of *Fermat factorization*. We seek a factor of n by calculating the numbers $y^2 = x^2 - n$ for each integer $x \geq \sqrt{n}$ until we find a *perfect square*, i.e., the square of an integer.

Example (Fermat Factorization). Let $n = 4277$. We have $\sqrt{4277} \approx 65.39$, so let us start with $x = 66$.

$$66^2 - 4277 = 79$$
$$67^2 - 4277 = 212$$
$$68^2 - 4277 = 347$$
$$69^2 - 4277 = 484 = 22^2$$

The result is, $4277 = 69^2 - 22^2 = (69 + 22)(69 - 22) = 91 \times 47$.

EXERCISE 2.11. Illustrate Fermat factorization using the following numbers.
a) 2117
b) 16781
c) 17933
d) 70027

Fermat factorization works for all odd numbers, for if $n = ab$ with both a and b odd, then $n = x^2 - y^2$, where $x = (a + b)/2$ and $y = (a - b)/2$. Moreover, this shows that we should terminate the algorithm when we reach $x = (n + 1)/2$, in which case the result is trivial, i.e., $n = n \times 1$, and n is prime. For large n, however, the iterations from $\sqrt{n}$ to $(n + 1)/2$ make this algorithm too slow to be practical.

EXERCISE 2.12. Fermat factorization is efficient when n has at least one factor relatively close to $\sqrt{n}$. Why is this statement true?

2.3 The Infinitude of Primes

One relevant question concerning primes is whether or not there exist infinitely many primes of a special form, e.g., $4n + 3$ or $n^2 + 1$. This will turn to generate very difficult problems, many of which are still unsolved. But first, of course, we need to be convinced that prime numbers are indeed infinitely many—and this fact is not hard to demonstrate.

Theorem 2.6. There are infinitely many prime numbers.

Proof. If there were only finitely many primes, let n be the product of them all. Too large to be prime, the number $n + 1$ would be composite and divisible by p, which is one of the primes dividing n. Then p would divide $(n + 1) - n = 1$, by Proposition 1.1(4). This is absurd since $p \geq 2$. ▽

Furthermore, we actually have a way to estimate the distribution of primes among the positive integers in a given interval. To make precise the statement, we need the next definition.

Definition. The *prime counting function* $\pi(x)$ denotes the number of primes up to x, where x can be any real number.[3] For example, $\pi(13) = 6$ because there are exactly six primes in this range, i.e., 2, 3, 5, 7, 11, 13. Similarly, $\pi(100) = 25$, according to Table 2.1.

For large values of x, the function $\pi(x)$ behaves much like $x/\log x$, where $\log x$ denotes the *natural logarithm function*. We state this result as the next theorem, the proof of which requires advanced techniques from complex analysis and, unfortunately, will not be provided here.

Theorem 2.7 (The Prime Number Theorem)**.** We have

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

Moreover, it has been found that $x/(\log x - 1)$ is a slightly better function than $x/\log x$ in approximating $\pi(x)$ for large values of x.

Example. Up to 25 billion, the number of primes is estimated by

$$\frac{25,000,000,000}{\log 25,000,000,000 - 1} \approx 1,089,697,743$$

which is comparable to the actual count, $\pi(25 \times 10^9) = 1,091,987,405$.

Exercise 2.13. Estimate how many prime numbers there are,

a) up to one million.
b) up to ten million.
c) between 9 million and 10 million.
d) among the ten-digit integers.

Now back to primes of special forms. The first case we shall consider involves primes that come in the sequence $\{an+b\}$, where a and b are fixed. According to Proposition 1.6(1), every number of this form is a multiple of $\gcd(a, b)$. Hence if $\gcd(a, b) > 1$, then the sequence $\{an+b\}$ can contain only composites, except perhaps $\gcd(a, b)$ itself, if prime. So to avoid triviality, we assume that $\gcd(a, b) = 1$—a condition which, claimed below, is sufficient to ensure the infinitude of such primes.

Theorem 2.8 (Dirichlet's Theorem on Primes in Arithmetic Progressions)**.** Primes of the form $an + b$ are infinitely many if, and only if, $\gcd(a, b) = 1$.

This theorem is a very advanced general result whose proof lies in the domain of analytic number theory and, unfortunately again, cannot be given within the scope of this book. Instead we will supply, by way of illustration, a simple proof for the specific case $a = 4$ and $b = 3$.

[3]For example, $\pi(\pi) = 2$. Ha!

There are infinitely many primes of the form $4n + 3$. To see this, first note that every odd prime has the form either $4n+1$ or $4n+3$. Second, the product of two numbers of the form $4n + 1$ is again of the same form. Therefore, a number of the form $4n + 3$ must have a prime divisor of the form $4n + 3$.

If primes of the form $4n+3$ were finite in number, let n be the product of them all. As noted, one of these prime divisors of n must divide $4(n-1)+3$, hence it would also divide $4(n-1)+3-4n = -1$, a contradiction. $\bigtriangledown$

EXERCISE 2.14. Prove the infinitude of primes of the form $6n + 5$.

There remain many open problems today concerning the infinitude of primes of a curious type, such as $n^2 + 1$ or $n! + 1$. Among the most popular is the so-called *Mersenne primes*—of the form $2^p - 1$, e.g., the prime $31 = 2^5 - 1$. The exponent p must be a prime as a necessary, but not sufficient, condition for a Mersenne prime. This claim is in the next exercise.

EXERCISE 2.15. Let m and n be two positive integers. Prove these facts.
a) $(2^m - 1) \,\%\, (2^n - 1) = 2^{m\,\%\,n} - 1$.
b) $2^m - 1$ is composite if m is.
c) $\gcd(2^m - 1, 2^n - 1) = 2^{\gcd(m,n)} - 1$.
d) $2^m - 1$ and $2^n - 1$ are relatively prime if and only if m and n are.

It is unsettled whether or not there are infinitely many Mersenne primes. Computational evidence suggests that there probably are. The largest finding, announced August 2008 at the Prime Pages [eCal], was the 12,978,189-digit Mersenne prime $2^p - 1$ corresponding to $p = 43112609$.

This is quite common where, along with the search for a settling proof, computational efforts are being done to find the largest prime of the specific type—here, largest means record breaking.

Another famous prime search is that after *twin primes,* i.e., a pair of primes differing by 2. At the writing of this revision, the record of twin primes at the Prime Pages shows the pair $65516468355 \times 2^{333333} \pm 1$, of 100,355 decimal digits each.

The search for numerical support, however, has not always been that fruitful—at least not with the next class of primes, to be found among the *Fermat numbers,* i.e.,

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

for $n \geq 0$. The first five Fermat numbers happen to be primes—such primes are named *Fermat primes*. But no other Fermat primes have been discovered until now, if any more exists.

The big challenge, since F_n grows quite rapidly, is not merely in evaluating a Fermat number or proving its primality, but even more in finding one of its factors, if composite. We will see more about this topic in Section 9.3.

EXERCISE 2.16. Let $F_n = 2^{2^n} + 1$ denote a Fermat number, $n \geq 0$.
a) Find the first five Fermat primes.
b) Verify the recurrence relation $F_n = F_0 F_1 F_2 \cdots F_{n-1} + 2$ for $n \geq 1$.
c) Prove that Fermat numbers are relatively prime one to another.
d) Use (c) to prove again that there are infinitely many prime numbers.

EXERCISE 2.17. Show that $2^k + 1$ is composite if k has an odd prime factor.

2.4 Euler's Factorization Method [Project 2]

While Fermat factorization is based on the difference of two squares, there is another factoring technique, due to Euler, which involves the *sum* of two squares.

Suppose that an odd number n can be expressed as a sum of two squares in two different ways, $n = x^2 + y^2 = z^2 + w^2$. Without loss of generality, assume that x and z are even, while y and w odd. Let $d = \gcd(x - z, w - y)$ and $c = \gcd(x + z, w + y)$. This will lead us to finding a factor of n, given by $(c/2)^2 + (d/2)^2$.

Example (Euler's Factorization Method). Given that $493 = 22^2 + 3^2 = 18^2 + 13^2$. We have $d = \gcd(22 - 18, 13 - 3) = 2$ and $c = \gcd(22 + 18, 13 + 3) = 8$. Then $(c/2)^2 + (d/2)^2 = 17$ divides 493. Indeed, $493 = 17 \times 29$.

PROJECT 2.4.1. Factor the following numbers by Euler's method.
a) 10049 $(100^2 + 7^2 = 32^2 + 95^2)$
b) 10081 $(100^2 + 9^2 = 84^2 + 55^2)$
c) 10121
d) 1000049

PROJECT 2.4.2. Justify the claim of Euler's method by showing that

$$n = \frac{c^2 + d^2}{4} \times \frac{(x - z)^2 + (w - y)^2}{d^2}$$

This method of factorization obviously has its limitations; for one thing, there is no reason to expect that every number can be represented as a sum of two squares. The aim of the next project (Section 3.4) is to identify the positive integers which are representable in this way. As a preview, the answer will be given by the following theorem, first discovered by Fermat.

Theorem 2.9. Let $n = \prod p_i^{e_i}$ be the usual factorization of n into prime powers. The diophantine equation $n = x^2 + y^2$ has a solution if and only if e_i is even whenever $p_i \,\%\, 4 = 3$.

In particular, a consequence of Euler's method, Theorem 2.9 implies that primes of the form $4k + 1$ can be *uniquely* written as a sum of two squares!