Introduction to number theory and logic

اسم المقرر: مقدمة في نظرية العدد والمنطق	المقرر: إجباري	✓	اختياري		عدد الساعات المعتمدة: ٢
Reports and truth tables, Reporting tools (negation, conjunction, conditional). The greatest common divisor (GCD) and least common multiples (LCM). The Division algorithm, The Euclidean algorithm, The Diophantine Equation. Prime numbers and Prime Factorizations. Congruence and its properties. Pythagorean Triples, sum of two squares and sum of four squares.					

Ref:

Elliott Mendelson, Introduction to Mathematical Logic, Sixth Edition, © 2015 by Taylor & Francis group, llc

CHAPTER 1

PRELIMINARIES

Number was born in superstition and reared in mystery, ... numbers were once made the foundation of religion and philosophy, and the tricks of figures have had a marvellous effect on a credulous people.
F. W. PARKER

1.1 MATHEMATICAL INDUCTION

The theory of numbers is concerned, at least in its elementary aspects, with properties of the integers and more particularly with the positive integers 1, 2, 3, ... (also known as the *natural numbers*). The origin of this misnomer harks back to the early Greeks for whom the word *number* meant positive integer, and nothing else. The natural numbers have been known to us for so long that the mathematician Leopold Kronecker once remarked, "God created the natural numbers, and all the rest is the work of man." Far from being a gift from Heaven, number theory has had a long and sometimes painful evolution, a story that is told in the ensuing pages.

We shall make no attempt to construct the integers axiomatically, assuming instead that they are already given and that any reader of this book is familiar with many elementary facts about them. Among these is the Well-Ordering Principle, stated here to refresh the memory.

Well-Ordering Principle. Every nonempty set S of nonnegative integers contains a least element; that is, there is some integer a in S such that $a \leq b$ for all b's belonging to S.

Because this principle plays a critical role in the proofs here and in subsequent chapters, let us use it to show that the set of positive integers has what is known as the Archimedean property.

Theorem 1.1 Archimedean property. If a and b are any positive integers, then there exists a positive integer n such that $na \geq b$.

Proof. Assume that the statement of the theorem is not true, so that for some a and b, $na < b$ for every positive integer n. Then the set

$$S = \{b - na \mid n \text{ a positive integer}\}$$

consists entirely of positive integers. By the Well-Ordering Principle, S will possess a least element, say, $b - ma$. Notice that $b - (m + 1)a$ also lies in S, because S contains all integers of this form. Furthermore, we have

$$b - (m + 1)a = (b - ma) - a < b - ma$$

contrary to the choice of $b - ma$ as the smallest integer in S. This contradiction arose out of our original assumption that the Archimedean property did not hold; hence, this property is proven true.

With the Well-Ordering Principle available, it is an easy matter to derive the First Principle of Finite Induction, which provides a basis for a method of proof called *mathematical induction*. Loosely speaking, the First Principle of Finite Induction asserts that if a set of positive integers has two specific properties, then it is the set of all positive integers. To be less cryptic, we state this principle in Theorem 1.2.

Theorem 1.2 First Principle of Finite Induction. Let S be a set of positive integers with the following properties:

(a) The integer 1 belongs to S.
(b) Whenever the integer k is in S, the next integer $k + 1$ must also be in S.

Then S is the set of all positive integers.

Proof. Let T be the set of all positive integers not in S, and assume that T is nonempty. The Well-Ordering Principle tells us that T possesses a least element, which we denote by a. Because 1 is in S, certainly $a > 1$, and so $0 < a - 1 < a$. The choice of a as the smallest positive integer in T implies that $a - 1$ is not a member of T, or equivalently that $a - 1$ belongs to S. By hypothesis, S must also contain $(a - 1) + 1 = a$, which contradicts the fact that a lies in T. We conclude that the set T is empty and in consequence that S contains all the positive integers.

Here is a typical formula that can be established by mathematical induction:

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(2n + 1)(n + 1)}{6} \tag{1}$$

for $n = 1, 2, 3, \ldots$. In anticipation of using Theorem 1.2, let S denote the set of all positive integers n for which Eq. (1) is true. We observe that when $n = 1$, the

formula becomes

$$1^2 = \frac{1(2+1)(1+1)}{6} = 1$$

This means that 1 is in S. Next, assume that k belongs to S (where k is a fixed but unspecified integer) so that

$$1^2 + 2^2 + 3^2 + \cdots + k^2 = \frac{k(2k+1)(k+1)}{6} \tag{2}$$

To obtain the sum of the first $k+1$ squares, we merely add the next one, $(k+1)^2$, to both sides of Eq. (2). This gives

$$1^2 + 2^2 + \cdots + k^2 + (k+1)^2 = \frac{k(2k+1)(k+1)}{6} + (k+1)^2$$

After some algebraic manipulation, the right-hand side becomes

$$(k+1)\left[\frac{k(2k+1)+6(k+1)}{6}\right] = (k+1)\left[\frac{2k^2+7k+6}{6}\right]$$

$$= \frac{(k+1)(2k+3)(k+2)}{6}$$

which is precisely the right-hand member of Eq. (1) when $n = k+1$. Our reasoning shows that the set S contains the integer $k+1$ whenever it contains the integer k. By Theorem 1.2, S must be all the positive integers; that is, the given formula is true for $n = 1, 2, 3, \ldots$.

Although mathematical induction provides a standard technique for attempting to prove a statement about the positive integers, one disadvantage is that it gives no aid in formulating such statements. Of course, if we can make an "educated guess" at a property that we believe might hold in general, then its validity can often be tested by the induction principle. Consider, for instance, the list of equalities

$$1 = 1$$
$$1 + 2 = 3$$
$$1 + 2 + 2^2 = 7$$
$$1 + 2 + 2^2 + 2^3 = 15$$
$$1 + 2 + 2^2 + 2^3 + 2^4 = 31$$
$$1 + 2 + 2^2 + 2^3 + 2^4 + 2^5 = 63$$

We seek a rule that gives the integers on the right-hand side. After a little reflection, the reader might notice that

$$1 = 2 - 1 \qquad 3 = 2^2 - 1 \qquad 7 = 2^3 - 1$$
$$15 = 2^4 - 1 \qquad 31 = 2^5 - 1 \qquad 63 = 2^6 - 1$$

(How one arrives at this observation is hard to say, but experience helps.) The pattern emerging from these few cases suggests a formula for obtaining the value of the

expression $1 + 2 + 2^2 + 2^3 + \cdots + 2^{n-1}$; namely,

$$1 + 2 + 2^2 + 2^3 + \cdots + 2^{n-1} = 2^n - 1 \tag{3}$$

for every positive integer n.

To confirm that our guess is correct, let S be the set of positive integers n for which Eq. (3) holds. For $n = 1$, Eq. (3) is certainly true, whence 1 belongs to the set S. We assume that Eq. (3) is true for a fixed integer k, so that for this k

$$1 + 2 + 2^2 + \cdots + 2^{k-1} = 2^k - 1$$

and we attempt to prove the validity of the formula for $k + 1$. Addition of the term 2^k to both sides of the last-written equation leads to

$$\begin{aligned} 1 + 2 + 2^2 + \cdots + 2^{k-1} + 2^k &= 2^k - 1 + 2^k \\ &= 2 \cdot 2^k - 1 = 2^{k+1} - 1 \end{aligned}$$

But this says that Eq. (3) holds when $n = k + 1$, putting the integer $k + 1$ in S so that $k + 1$ is in S whenever k is in S. According to the induction principle, S must be the set of all positive integers.

> **Remark.** When giving induction proofs, we shall usually shorten the argument by eliminating all reference to the set S, and proceed to show simply that the result in question is true for the integer 1, and if true for the integer k is then also true for $k + 1$.

We should inject a word of caution at this point, to wit, that one must be careful to establish both conditions of Theorem 1.2 before drawing any conclusions; neither is sufficient alone. The proof of condition (a) is usually called the *basis for the induction*, and the proof of (b) is called the *induction step*. The assumptions made in carrying out the induction step are known as the *induction hypotheses*. The induction situation has been likened to an infinite row of dominoes all standing on edge and arranged in such a way that when one falls it knocks down the next in line. If either no domino is pushed over (that is, there is no basis for the induction) or if the spacing is too large (that is, the induction step fails), then the complete line will not fall.

The validity of the induction step does not necessarily depend on the truth of the statement that one is endeavoring to prove. Let us look at the false formula

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2 + 3 \tag{4}$$

Assume that this holds for $n = k$; in other words,

$$1 + 3 + 5 + \cdots + (2k - 1) = k^2 + 3$$

Knowing this, we then obtain

$$\begin{aligned} 1 + 3 + 5 + \cdots + (2k - 1) + (2k + 1) &= k^2 + 3 + 2k + 1 \\ &= (k + 1)^2 + 3 \end{aligned}$$

which is precisely the form that Eq. (4) should take when $n = k + 1$. Thus, if Eq. (4) holds for a given integer, then it also holds for the succeeding integer. It is not possible, however, to find a value of n for which the formula is true.

There is a variant of the induction principle that is often used when Theorem 1.2 alone seems ineffective. As with the first version, this Second Principle of Finite Induction gives two conditions that guarantee a certain set of positive integers actually consists of all positive integers. This is what happens: We retain requirement (a), but (b) is replaced by

(b′) If k is a positive integer such that $1, 2, \ldots, k$ belong to S, then $k + 1$ must also be in S.

The proof that S consists of all positive integers has the same flavor as that of Theorem 1.2. Again, let T represent the set of positive integers not in S. Assuming that T is nonempty, we choose n to be the smallest integer in T. Then $n > 1$, by supposition (a). The minimal nature of n allows us to conclude that none of the integers $1, 2, \ldots, n - 1$ lies in T, or, if we prefer a positive assertion, $1, 2, \ldots, n - 1$ all belong to S. Property (b′) then puts $n = (n - 1) + 1$ in S, which is an obvious contradiction. The result of all this is to make T empty.

The First Principle of Finite Induction is used more often than is the Second; however, there are occasions when the Second is favored and the reader should be familiar with both versions. It sometimes happens that in attempting to show that $k + 1$ is a member of S, we require proof of the fact that not only k, but all positive integers that precede k, lie in S. Our formulation of these induction principles has been for the case in which the induction begins with 1. Each form can be generalized to start with any positive integer n_0. In this circumstance, the conclusion reads as "Then S is the set of all positive integers $n \geq n_0$."

Mathematical induction is often used as a method of definition as well as a method of proof. For example, a common way of introducing the symbol $n!$ (pronounced "n factorial") is by means of the inductive definition

(a) $1! = 1$,

(b) $n! = n \cdot (n - 1)!$ for $n > 1$.

This pair of conditions provides a rule whereby the meaning of $n!$ is specified for each positive integer n. Thus, by (a), $1! = 1$; (a) and (b) yield

$$2! = 2 \cdot 1! = 2 \cdot 1$$

while by (b), again,

$$3! = 3 \cdot 2! = 3 \cdot 2 \cdot 1$$

Continuing in this manner, using condition (b) repeatedly, the numbers $1!, 2!, 3!, \ldots, n!$ are defined in succession up to any chosen n. In fact,

$$n! = n \cdot (n - 1) \cdots 3 \cdot 2 \cdot 1$$

Induction enters in showing that $n!$, as a function on the positive integers, exists and is unique; however, we shall make no attempt to give the argument.

It will be convenient to extend the definition of $n!$ to the case in which $n = 0$ by stipulating that $0! = 1$.

Example 1.1. To illustrate a proof that requires the Second Principle of Finite Induction, consider the so-called *Lucas sequence*:

$$1, 3, 4, 7, 11, 18, 29, 47, 76, \ldots$$

Except for the first two terms, each term of this sequence is the sum of the preceding two, so that the sequence may be defined inductively by

$$\begin{aligned} a_1 &= 1 \\ a_2 &= 3 \\ a_n &= a_{n-1} + a_{n-2} \qquad \text{for all } n \geq 3 \end{aligned}$$

We contend that the inequality

$$a_n < (7/4)^n$$

holds for every positive integer n. The argument used is interesting because in the inductive step, it is necessary to know the truth of this inequality for two successive values of n to establish its truth for the following value.

First of all, for $n = 1$ and 2, we have

$$a_1 = 1 < (7/4)^1 = 7/4 \qquad \text{and} \qquad a_2 = 3 < (7/4)^2 = 49/16$$

whence the inequality in question holds in these two cases. This provides a basis for the induction. For the induction step, choose an integer $k \geq 3$ and assume that the inequality is valid for $n = 1, 2, \ldots, k-1$. Then, in particular,

$$a_{k-1} < (7/4)^{k-1} \qquad \text{and} \qquad a_{k-2} < (7/4)^{k-2}$$

By the way in which the Lucas sequence is formed, it follows that

$$\begin{aligned} a_k = a_{k-1} + a_{k-2} &< (7/4)^{k-1} + (7/4)^{k-2} \\ &= (7/4)^{k-2}(7/4 + 1) \\ &= (7/4)^{k-2}(11/4) \\ &< (7/4)^{k-2}(7/4)^2 = (7/4)^k \end{aligned}$$

Because the inequality is true for $n = k$ whenever it is true for the integers $1, 2, \ldots, k-1$, we conclude by the second induction principle that $a_n < (7/4)^n$ for all $n \geq 1$.

Among other things, this example suggests that if objects are defined inductively, then mathematical induction is an important tool for establishing the properties of these objects.

PROBLEMS 1.1

1. Establish the formulas below by mathematical induction:

(a) $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$ for all $n \geq 1$.

(b) $1 + 3 + 5 + \cdots + (2n-1) = n^2$ for all $n \geq 1$.

(c) $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \dfrac{n(n+1)(n+2)}{3}$ for all $n \geq 1$.

(d) $1^2 + 3^2 + 5^2 + \cdots + (2n-1)^2 = \dfrac{n(2n-1)(2n+1)}{3}$ for all $n \geq 1$.

(e) $1^3 + 2^3 + 3^3 + \cdots + n^3 = \left[\dfrac{n(n+1)}{2}\right]^2$ for all $n \geq 1$.

2. If $r \neq 1$, show that for any positive integer n,

$$a + ar + ar^2 + \cdots + ar^n = \frac{a(r^{n+1} - 1)}{r - 1}$$

3. Use the Second Principle of Finite Induction to establish that for all $n \geq 1$,

$$a^n - 1 = (a-1)(a^{n-1} + a^{n-2} + a^{n-3} + \cdots + a + 1)$$

[*Hint:* $a^{n+1} - 1 = (a+1)(a^n - 1) - a(a^{n-1} - 1)$.]

4. Prove that the cube of any integer can be written as the difference of two squares. [*Hint:* Notice that

$$n^3 = (1^3 + 2^3 + \cdots + n^3) - (1^3 + 2^3 + \cdots + (n-1)^3).]$$

5. (a) Find the values of $n \leq 7$ for which $n! + 1$ is a perfect square (it is unknown whether $n! + 1$ is a square for any $n > 7$).

(b) True or false? For positive integers m and n, $(mn)! = m!n!$ and $(m+n)! = m! + n!$.

6. Prove that $n! > n^2$ for every integer $n \geq 4$, whereas $n! > n^3$ for every integer $n \geq 6$.

7. Use mathematical induction to derive the following formula for all $n \geq 1$:

$$1(1!) + 2(2!) + 3(3!) + \cdots + n(n!) = (n+1)! - 1$$

8. (a) Verify that for all $n \geq 1$,

$$2 \cdot 6 \cdot 10 \cdot 14 \cdot \cdots \cdot (4n-2) = \frac{(2n)!}{n!}$$

(b) Use part (a) to obtain the inequality $2^n(n!)^2 \leq (2n)!$ for all $n \geq 1$.

9. Establish the Bernoulli inequality: If $1 + a > 0$, then

$$(1+a)^n \geq 1 + na$$

for all $n \geq 1$.

10. For all $n \geq 1$, prove the following by mathematical induction:

(a) $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \dfrac{1}{3^2} + \cdots + \dfrac{1}{n^2} \leq 2 - \dfrac{1}{n}$.

(b) $\dfrac{1}{2} + \dfrac{2}{2^2} + \dfrac{3}{2^3} + \cdots + \dfrac{n}{2^n} = 2 - \dfrac{n+2}{2^n}$.

11. Show that the expression $(2n)!/2^n n!$ is an integer for all $n \geq 0$.

12. Consider the function defined by

$$T(n) = \begin{cases} \dfrac{3n+1}{2} & \text{for } n \text{ odd} \\ \dfrac{n}{2} & \text{for } n \text{ even} \end{cases}$$

The $3n + 1$ conjecture is the claim that starting from any integer $n > 1$, the sequence of iterates $T(n)$, $T(T(n))$, $T(T(T(n)))$, ..., eventually reaches the integer 1 and subsequently runs through the values 1 and 2. This has been verified for all $n \leq 10^{16}$. Confirm the conjecture in the cases $n = 21$ and $n = 23$.

13. Suppose that the numbers a_n are defined inductively by $a_1 = 1, a_2 = 2, a_3 = 3$, and $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ for all $n \geq 4$. Use the Second Principle of Finite Induction to show that $a_n < 2^n$ for every positive integer n.

14. If the numbers a_n are defined by $a_1 = 11, a_2 = 21$, and $a_n = 3a_{n-1} - 2a_{n-2}$ for $n \geq 3$, prove that

$$a_n = 5 \cdot 2^n + 1 \qquad n \geq 1$$

1.2 THE BINOMIAL THEOREM

Closely connected with the factorial notation are the *binomial coefficients* $\binom{n}{k}$. For any positive integer n and any integer k satisfying $0 \leq k \leq n$, these are defined by

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

By canceling out either $k!$ or $(n-k)!$, $\binom{n}{k}$ can be written as

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

For example, with $n = 8$ and $k = 3$, we have

$$\binom{8}{3} = \frac{8!}{3!5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{5!} = \frac{8 \cdot 7 \cdot 6}{3!} = 56$$

Also observe that if $k = 0$ or $k = n$, the quantity $0!$ appears on the right-hand side of the definition of $\binom{n}{k}$; because we have taken $0!$ as 1, these special values of k give

$$\binom{n}{0} = \binom{n}{n} = 1$$

There are numerous useful identities connecting binomial coefficients. One that we require here is *Pascal's rule*:

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \qquad 1 \leq k \leq n$$

Its proof consists of multiplying the identity

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

by $n!/(k-1)!(n-k)!$ to obtain

$$\frac{n!}{k(k-1)!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)(n-k)!}$$
$$= \frac{(n+1)n!}{k(k-1)!(n-k+1)(n-k)!}$$

Falling back on the definition of the factorial function, this says that

$$\frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} = \frac{(n+1)!}{k!(n+1-k)!}$$

from which Pascal's rule follows.

This relation gives rise to a configuration, known as *Pascal's triangle*, in which the binomial coefficient $\binom{n}{k}$ appears as the $(k+1)$th number in the nth row:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & 1 & & & \\
 & & & & 1 & & 2 & & 1 & & \\
 & & & 1 & & 3 & & 3 & & 1 & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & & & & & \cdots & & & & & &
\end{array}$$

The rule of formation should be clear. The borders of the triangle are composed of 1's; a number not on the border is the sum of the two numbers nearest it in the row immediately above.

The so-called *binomial theorem* is in reality a formula for the complete expansion of $(a+b)^n$, $n \geq 1$, into a sum of powers of a and b. This expression appears with great frequency in all phases of number theory, and it is well worth our time to look at it now. By direct multiplication, it is easy to verify that

$$\begin{aligned}
(a+b)^1 &= a+b \\
(a+b)^2 &= a^2 + 2ab + b^2 \\
(a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3 \\
(a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4, \text{ etc.}
\end{aligned}$$

The question is how to predict the coefficients. A clue lies in the observation that the coefficients of these first few expansions form the successive rows of Pascal's triangle. This leads us to suspect that the general binomial expansion takes the form

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \cdots + \binom{n}{n-1} ab^{n-1} + \binom{n}{n} b^n$$

or, written more compactly,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k$$

Mathematical induction provides the best means for confirming this guess. When $n = 1$, the conjectured formula reduces to

$$(a+b)^1 = \sum_{k=0}^{1} \binom{1}{k} a^{1-k}b^k = \binom{1}{0} a^1b^0 + \binom{1}{1} a^0b^1 = a+b$$

which is certainly correct. Assuming that the formula holds for some fixed integer m, we go on to show that it also must hold for $m + 1$. The starting point is to notice that

$$(a + b)^{m+1} = a(a + b)^m + b(a + b)^m$$

Under the induction hypothesis,

$$\begin{aligned} a(a + b)^m &= \sum_{k=0}^{m} \binom{m}{k} a^{m-k+1} b^k \\ &= a^{m+1} + \sum_{k=1}^{m} \binom{m}{k} a^{m+1-k} b^k \end{aligned}$$

and

$$\begin{aligned} b(a + b)^m &= \sum_{j=0}^{m} \binom{m}{j} a^{m-j} b^{j+1} \\ &= \sum_{k=1}^{m} \binom{m}{k-1} a^{m+1-k} b^k + b^{m+1} \end{aligned}$$

Upon adding these expressions, we obtain

$$\begin{aligned} (a + b)^{m+1} &= a^{m+1} + \sum_{k=1}^{m} \left[\binom{m}{k} + \binom{m}{k-1} \right] a^{m+1-k} b^k + b^{m+1} \\ &= \sum_{k=0}^{m+1} \binom{m+1}{k} a^{m+1-k} b^k \end{aligned}$$

which is the formula in the case $n = m + 1$. This establishes the binomial theorem by induction.

Before abandoning these ideas, we might remark that the first acceptable formulation of the method of mathematical induction appears in the treatise *Traité du Triangle Arithmetiqué*, by the 17th century French mathematician and philosopher Blaise Pascal. This short work was written in 1653, but not printed until 1665 because Pascal had withdrawn from mathematics (at the age of 25) to dedicate his talents to religion. His careful analysis of the properties of the binomial coefficients helped lay the foundations of probability theory.

PROBLEMS 1.2

1. (a) Derive Newton's identity

$$\binom{n}{k}\binom{k}{r} = \binom{n}{r}\binom{n-r}{k-r} \qquad n \geq k \geq r \geq 0$$

(b) Use part (a) to express $\binom{n}{k}$ in terms of its predecessor:

$$\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1} \qquad n \geq k \geq 1$$

2. If $2 \leq k \leq n-2$, show that

$$\binom{n}{k} = \binom{n-2}{k-2} + 2\binom{n-2}{k-1} + \binom{n-2}{k} \qquad n \geq 4$$

3. For $n \geq 1$, derive each of the identities below:

(a) $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$

[*Hint:* Let $a = b = 1$ in the binomial theorem.]

(b) $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n\binom{n}{n} = 0.$

(c) $\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n2^{n-1}.$

[*Hint:* After expanding $n(1+b)^{n-1}$ by the binomial theorem, let $b = 1$; note also that

$$n\binom{n-1}{k} = (k+1)\binom{n}{k+1}.]$$

(d) $\binom{n}{0} + 2\binom{n}{1} + 2^2\binom{n}{2} + \cdots + 2^n\binom{n}{n} = 3^n.$

(e) $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \binom{n}{6} + \cdots$

$$= \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots = 2^{n-1}.$$

[*Hint:* Use parts (a) and (b).]

(f) $\binom{n}{0} - \frac{1}{2}\binom{n}{1} + \frac{1}{3}\binom{n}{2} - \cdots + \frac{(-1)^n}{n+1}\binom{n}{n} = \frac{1}{n+1}.$

[*Hint:* The left-hand side equals

$$\frac{1}{n+1}\left[\binom{n+1}{1} - \binom{n+1}{2} + \binom{n+1}{3} - \cdots + (-1)^n\binom{n+1}{n+1}\right].]$$

4. Prove the following for $n \geq 1$:

(a) $\binom{n}{r} < \binom{n}{r+1}$ if and only if $0 \leq r < \frac{1}{2}(n-1)$.

(b) $\binom{n}{r} > \binom{n}{r+1}$ if and only if $n-1 \geq r > \frac{1}{2}(n-1)$.

(c) $\binom{n}{r} = \binom{n}{r+1}$ if and only if n is an odd integer, and $r = \frac{1}{2}(n-1)$.

Pythagoras divided those who attended his lectures into two groups: the Probationers (or listeners) and the Pythagoreans. After three years in the first class, a listener could be initiated into the second class, to whom were confided the main discoveries of the school. The Pythagoreans were a closely knit brotherhood, holding all worldly goods in common and bound by an oath not to reveal the founder's secrets. Legend has it that a talkative Pythagorean was drowned in a shipwreck as the gods' punishment for publicly boasting that he had added the dodecahedron to the number of regular solids enumerated by Pythagoras. For a time, the autocratic Pythagoreans succeeded in dominating the local government in Croton, but a popular revolt in 501 B.C. led to the murder of many of its prominent members, and Pythagoras himself was killed shortly thereafter. Although the political influence of the Pythagoreans thus was destroyed, they continued to exist for at least two centuries more as a philosophical and mathematical society. To the end, they remained a secret order, publishing nothing and, with noble self-denial, ascribing all their discoveries to the Master.

The Pythagoreans believed that the key to an explanation of the universe lay in number and form, their general thesis being that "Everything is Number." (By number, they meant, of course, a positive integer.) For a rational understanding of nature, they considered it sufficient to analyze the properties of certain numbers. Pythagoras himself, we are told "seems to have attached supreme importance to the study of arithmetic, which he advanced and took out of the realm of commercial utility."

The Pythagorean doctrine is a curious mixture of cosmic philosophy and number mysticism, a sort of supernumerology that assigned to everything material or spiritual a definite integer. Among their writings, we find that 1 represented reason, for reason could produce only one consistent body of truth; 2 stood for man and 3 for woman; 4 was the Pythagorean symbol for justice, being the first number that is the product of equals; 5 was identified with marriage, because it is formed by the union of 2 and 3; and so forth. All the even numbers, after the first one, were capable of separation into other numbers; hence, they were prolific and were considered as feminine and earthy—and somewhat less highly regarded in general. Being a predominantly male society, the Pythagoreans classified the odd numbers, after the first two, as masculine and divine.

Although these speculations about numbers as models of "things" appear frivolous today, it must be borne in mind that the intellectuals of the classical Greek period were largely absorbed in philosophy and that these same men, because they had such intellectual interests, were the very ones who were engaged in laying the foundations for mathematics as a system of thought. To Pythagoras and his followers, mathematics was largely a means to an end, the end being philosophy. Only with the founding of the School of Alexandria do we enter a new phase in which the cultivation of mathematics was pursued for its own sake.

It was at Alexandria, not Athens, that a science of numbers divorced from mystic philosophy first began to develop. For nearly a thousand years, until its destruction by the Arabs in 641 A.D., Alexandria stood at the cultural and commercial center of the Hellenistic world. (After the fall of Alexandria, most of its scholars migrated to Constantinople. During the next 800 years, while formal learning in the West all but disappeared, this enclave at Constantinople preserved for us the mathematical works

of the various Greek schools.) The so-called Alexandrian Museum, a forerunner of the modern university, brought together the leading poets and scholars of the day; adjacent to it there was established an enormous library, reputed to hold over 700,000 volumes—hand-copied—at its height. Of all the distinguished names connected with the Museum, that of Euclid (fl. c.300 B.C.), founder of the School of Mathematics, is in a special class. Posterity has come to know him as the author of the *Elements*, the oldest Greek treatise on mathematics to reach us in its entirety. The *Elements* is a compilation of much of the mathematical knowledge available at that time, organized into 13 parts or Books, as they are called. The name of Euclid is so often associated with geometry that one tends to forget that three of the Books, VII, VIII, and IX, are devoted to number theory.

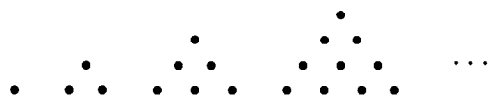
Euclid's *Elements* constitutes one of the great success stories of world literature. Scarcely any other book save the Bible has been more widely circulated or studied. Over a thousand editions of it have appeared since the first printed version in 1482, and before its printing, manuscript copies dominated much of the teaching of mathematics in Western Europe. Unfortunately, no copy of the work has been found that actually dates from Euclid's own time; the modern editions are descendants of a revision prepared by Theon of Alexandria, a commentator of the 4th century A.D.

PROBLEMS 2.1

1. Each of the numbers

$$1 = 1, 3 = 1 + 2, 6 = 1 + 2 + 3, 10 = 1 + 2 + 3 + 4, \ldots$$

represents the number of dots that can be arranged evenly in an equilateral triangle:



This led the ancient Greeks to call a number *triangular* if it is the sum of consecutive integers, beginning with 1. Prove the following facts concerning triangular numbers:

(a) A number is triangular if and only if it is of the form $n(n + 1)/2$ for some $n \geq 1$. (Pythagoras, circa 550 B.C.)

(b) The integer n is a triangular number if and only if $8n + 1$ is a perfect square. (Plutarch, circa 100 A.D.)

(c) The sum of any two consecutive triangular numbers is a perfect square. (Nicomachus, circa 100 A.D.)

(d) If n is a triangular number, then so are $9n + 1$, $25n + 3$, and $49n + 6$. (Euler, 1775)

2. If t_n denotes the nth triangular number, prove that in terms of the binomial coefficients,

$$t_n = \binom{n+1}{2} \qquad n \geq 1$$

3. Derive the following formula for the sum of triangular numbers, attributed to the Hindu mathematician Aryabhata (circa 500 A.D.):

$$t_1 + t_2 + t_3 + \cdots + t_n = \frac{n(n+1)(n+2)}{6} \qquad n \geq 1$$

[*Hint:* Group the terms on the left-hand side in pairs, noting the identity $t_{k-1} + t_k = k^2$.]

5. (a) For $n \geq 2$, prove that

$$\binom{2}{2} + \binom{3}{2} + \binom{4}{2} + \cdots + \binom{n}{2} = \binom{n+1}{3}$$

[*Hint:* Use induction, and Pascal's rule.]

(b) From part (a), and the relation $m^2 = 2\binom{m}{2} + m$ for $m \geq 2$, deduce the formula

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

(c) Apply the formula in part (a) to obtain a proof that

$$1 \cdot 2 + 2 \cdot 3 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}$$

[*Hint:* Observe that $(m-1)m = 2\binom{m}{2}$.]

6. Derive the binomial identity

$$\binom{2}{2} + \binom{4}{2} + \binom{6}{2} + \cdots + \binom{2n}{2} = \frac{n(n+1)(4n-1)}{6} \qquad n \geq 2$$

[*Hint:* For $m \geq 2$, $\binom{2m}{2} = 2\binom{m}{2} + m^2$.]

7. For $n \geq 1$, verify that

$$1^2 + 3^2 + 5^2 + \cdots + (2n-1)^2 = \binom{2n+1}{3}$$

8. Show that, for $n \geq 1$,

$$\binom{2n}{n} = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n} 2^{2n}$$

9. Establish the inequality $2^n < \binom{2n}{n} < 2^{2n}$, for $n > 1$.

[*Hint:* Put $x = 2 \cdot 4 \cdot 6 \cdots (2n)$, $y = 1 \cdot 3 \cdot 5 \cdots (2n-1)$, and $z = 1 \cdot 2 \cdot 3 \cdots n$; show that $x > y > z$, hence $x^2 > xy > xz$.]

10. The *Catalan numbers*, defined by

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!} \qquad n = 0, 1, 2, \ldots$$

form the sequence 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, ... They first appeared in 1838 when Eugène Catalan (1814–1894) showed that there are C_n ways of parenthesizing a nonassociative product of $n+1$ factors. [For instance, when $n = 3$ there are five ways: $((ab)c)d$, $(a(bc))d$, $a((bc)d)$, $a(b(cd))$, $(ab)(ac)$.] For $n \geq 1$, prove that C_n can be given inductively by

$$C_n = \frac{2(2n-1)}{n+1} C_{n-1}$$

CHAPTER 2

DIVISIBILITY THEORY IN THE INTEGERS

Integral numbers are the fountainhead of all mathematics.
H. MINKOWSKI

2.1 EARLY NUMBER THEORY

Before becoming weighted down with detail, we should say a few words about the origin of number theory. The theory of numbers is one of the oldest branches of mathematics; an enthusiast, by stretching a point here and there, could extend its roots back to a surprisingly remote date. Although it seems probable that the Greeks were largely indebted to the Babylonians and ancient Egyptians for a core of information about the properties of the natural numbers, the first rudiments of an actual theory are generally credited to Pythagoras and his disciples.

Our knowledge of the life of Pythagoras is scanty, and little can be said with any certainty. According to the best estimates, he was born between 580 and 562 B.C. on the Aegean island of Samos. It seems that he studied not only in Egypt, but may even have extended his journeys as far east as Babylonia. When Pythagoras reappeared after years of wandering, he sought out a favorable place for a school and finally settled upon Croton, a prosperous Greek settlement on the heel of the Italian boot. The school concentrated on four *mathemata*, or subjects of study: *arithmetica* (arithmetic, in the sense of number theory, rather than the art of calculating), *harmonia* (music), *geometria* (geometry), and *astrologia* (astronomy). This fourfold division of knowledge became known in the Middle Ages as the *quadrivium*, to which was added the *trivium* of logic, grammar, and rhetoric. These seven liberal arts came to be looked upon as the necessary course of study for an educated person.

4. Prove that the square of any odd multiple of 3 is the difference of two triangular numbers; specifically, that

$$9(2n+1)^2 = t_{9n+4} - t_{3n+1}$$

5. In the sequence of triangular numbers, find the following:
(a) Two triangular numbers whose sum and difference are also triangular numbers.
(b) Three successive triangular numbers whose product is a perfect square.
(c) Three successive triangular numbers whose sum is a perfect square.

6. (a) If the triangular number t_n is a perfect square, prove that $t_{4n(n+1)}$ is also a square.
(b) Use part (a) to find three examples of squares that are also triangular numbers.

7. Show that the difference between the squares of two consecutive triangular numbers is always a cube.

8. Prove that the sum of the reciprocals of the first n triangular numbers is less than 2; that is,

$$\frac{1}{1} + \frac{1}{3} + \frac{1}{6} + \frac{1}{10} + \cdots + \frac{1}{t_n} < 2$$

[*Hint:* Observe that $\frac{2}{n(n+1)} = 2(\frac{1}{n} - \frac{1}{n+1})$.]

9. (a) Establish the identity $t_x = t_y + t_z$, where

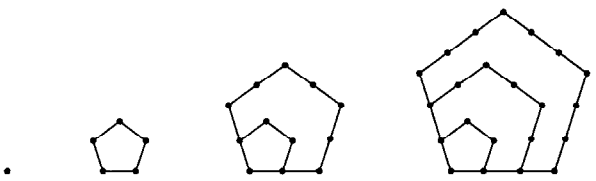
$$x = \frac{n(n+3)}{2} + 1 \qquad y = n+1 \qquad z = \frac{n(n+3)}{2}$$

and $n \geq 1$, thereby proving that there are infinitely many triangular numbers that are the sum of two other such numbers.
(b) Find three examples of triangular numbers that are sums of two other triangular numbers.

10. Each of the numbers

$$1, 5 = 1+4, 12 = 1+4+7, 22 = 1+4+7+10, \ldots$$

represents the number of dots that can be arranged evenly in a pentagon:



The ancient Greeks called these *pentagonal* numbers. If p_n denotes the nth pentagonal number, where $p_1 = 1$ and $p_n = p_{n-1} + (3n-2)$ for $n \geq 2$, prove that

$$p_n = \frac{n(3n-1)}{2}, \qquad n \geq 1$$

11. For $n \geq 2$, verify the following relations between the pentagonal, square, and triangular numbers:
(a) $p_n = t_{n-1} + n^2$
(b) $p_n = 3t_{n-1} + n = 2t_{n-1} + t_n$

2.2 THE DIVISION ALGORITHM

We have been exposed to relationships between integers for several pages and, as yet, not a single divisibility property has been derived. It is time to remedy this situation. One theorem, the Division Algorithm, acts as the foundation stone upon which our whole development rests. The result is familiar to most of us; roughly, it asserts that an integer a can be "divided" by a positive integer b in such a way that the remainder is smaller than is b. The exact statement of this fact is Theorem 2.1.

Theorem 2.1 Division Algorithm. Given integers a and b, with $b > 0$, there exist unique integers q and r satisfying

$$a = qb + r \qquad 0 \leq r < b$$

The integers q and r are called, respectively, the *quotient* and *remainder* in the division of a by b.

Proof. We begin by proving that the set

$$S = \{a - xb \mid x \text{ an integer}; a - xb \geq 0\}$$

is nonempty. To do this, it suffices to exhibit a value of x making $a - xb$ nonnegative. Because the integer $b \geq 1$, we have $|a| b \geq |a|$, and so

$$a - (-|a|)b = a + |a| b \geq a + |a| \geq 0$$

For the choice $x = -|a|$, then, $a - xb$ lies in S. This paves the way for an application of the Well-Ordering Principle (Chapter 1), from which we infer that the set S contains a smallest integer; call it r. By the definition of S, there exists an integer q satisfying

$$r = a - qb \qquad 0 \leq r$$

We argue that $r < b$. If this were not the case, then $r \geq b$ and

$$a - (q + 1)b = (a - qb) - b = r - b \geq 0$$

The implication is that the integer $a - (q + 1)b$ has the proper form to belong to the set S. But $a - (q + 1)b = r - b < r$, leading to a contradiction of the choice of r as the smallest member of S. Hence, $r < b$.

Next we turn to the task of showing the uniqueness of q and r. Suppose that a has two representations of the desired form, say,

$$a = qb + r = q'b + r'$$

where $0 \leq r < b, 0 \leq r' < b$. Then $r' - r = b(q - q')$ and, owing to the fact that the absolute value of a product is equal to the product of the absolute values,

$$|r' - r| = b|q - q'|$$

Upon adding the two inequalities $-b < -r \leq 0$ and $0 \leq r' < b$, we obtain $-b < r' - r < b$ or, in equivalent terms, $|r' - r| < b$. Thus, $b|q - q'| < b$, which yields

$$0 \leq |q - q'| < 1$$

Because $|q - q'|$ is a nonnegative integer, the only possibility is that $|q - q'| = 0$, whence $q = q'$; this, in turn, gives $r = r'$, ending the proof.

A more general version of the Division Algorithm is obtained on replacing the restriction that b must be positive by the simple requirement that $b \neq 0$.

Corollary. If a and b are integers, with $b \neq 0$, then there exist unique integers q and r such that

$$a = qb + r \qquad 0 \leq r < |b|$$

Proof. It is enough to consider the case in which b is negative. Then $|b| > 0$, and Theorem 2.1 produces unique integers q' and r for which

$$a = q'|b| + r \qquad 0 \leq r < |b|$$

Noting that $|b| = -b$, we may take $q = -q'$ to arrive at $a = qb + r$, with $0 \leq r < |b|$.

To illustrate the Division Algorithm when $b < 0$, let us take $b = -7$. Then, for the choices of $a = 1, -2, 61$, and -59, we obtain the expressions

$$\begin{aligned} 1 &= 0(-7) + 1 \\ -2 &= 1(-7) + 5 \\ 61 &= (-8)(-7) + 5 \\ -59 &= 9(-7) + 4 \end{aligned}$$

We wish to focus our attention on the applications of the Division Algorithm, and not so much on the algorithm itself. As a first illustration, note that with $b = 2$ the possible remainders are $r = 0$ and $r = 1$. When $r = 0$, the integer a has the form $a = 2q$ and is called *even*; when $r = 1$, the integer a has the form $a = 2q + 1$ and is called *odd*. Now a^2 is either of the form $(2q)^2 = 4k$ or $(2q + 1)^2 = 4(q^2 + q) + 1 = 4k + 1$. The point to be made is that the square of an integer leaves the remainder 0 or 1 upon division by 4.

We also can show the following: The square of any odd integer is of the form $8k + 1$. For, by the Division Algorithm, any integer is representable as one of the four forms: $4q, 4q + 1, 4q + 2, 4q + 3$. In this classification, only those integers of the forms $4q + 1$ and $4q + 3$ are odd. When the latter are squared, we find that

$$(4q + 1)^2 = 8(2q^2 + q) + 1 = 8k + 1$$

and similarly

$$(4q + 3)^2 = 8(2q^2 + 3q + 1) + 1 = 8k + 1$$

As examples, the square of the odd integer 7 is $7^2 = 49 = 8 \cdot 6 + 1$, and the square of 13 is $13^2 = 169 = 8 \cdot 21 + 1$.

As these remarks indicate, the advantage of the Division Algorithm is that it allows us to prove assertions about all the integers by considering only a finite number of cases. Let us illustrate this with one final example.

Example 2.1. We propose to show that the expression $a(a^2 + 2)/3$ is an integer for all $a \geq 1$. According to the Division Algorithm, every a is of the form $3q, 3q + 1$, or

$3q + 2$. Assume the first of these cases. Then

$$\frac{a(a^2 + 2)}{3} = q(9q^2 + 2)$$

which clearly is an integer. Similarly, if $a = 3q + 1$, then

$$\frac{(3q + 1)((3q + 1)^2 + 2)}{3} = (3q + 1)(3q^2 + 2q + 1)$$

and $a(a^2 + 2)/3$ is an integer in this instance also. Finally, for $a = 3q + 2$, we obtain

$$\frac{(3q + 2)((3q + 2)^2 + 2)}{3} = (3q + 2)(3q^2 + 4q + 2)$$

an integer once more. Consequently, our result is established in all cases.

PROBLEMS 2.2

1. Prove that if a and b are integers, with $b > 0$, then there exist unique integers q and r satisfying $a = qb + r$, where $2b \leq r < 3b$.
2. Show that any integer of the form $6k + 5$ is also of the form $3j + 2$, but not conversely.
3. Use the Division Algorithm to establish the following:
   (a) The square of any integer is either of the form $3k$ or $3k + 1$.
   (b) The cube of any integer has one of the forms: $9k$, $9k + 1$, or $9k + 8$.
   (c) The fourth power of any integer is either of the form $5k$ or $5k + 1$.
4. Prove that $3a^2 - 1$ is never a perfect square.
   [*Hint:* Problem 3(a).]
5. For $n \geq 1$, prove that $n(n + 1)(2n + 1)/6$ is an integer.
   [*Hint:* By the Division Algorithm, n has one of the forms $6k, 6k + 1, \ldots, 6k + 5$; establish the result in each of these six cases.]
6. Show that the cube of any integer is of the form $7k$ or $7k \pm 1$.
7. Obtain the following version of the Division Algorithm: For integers a and b, with $b \neq 0$, there exist unique integers q and r that satisfy $a = qb + r$, where $-\frac{1}{2}|b| < r \leq \frac{1}{2}|b|$.
   [*Hint:* First write $a = q'b + r'$, where $0 \leq r' < |b|$. When $0 \leq r' \leq \frac{1}{2}|b|$, let $r = r'$ and $q = q'$; when $\frac{1}{2}|b| < r' < |b|$, let $r = r' - |b|$ and $q = q' + 1$ if $b > 0$ or $q = q' - 1$ if $b < 0$.]
8. Prove that no integer in the following sequence is a perfect square:

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

   [*Hint:* A typical term $111 \cdots 111$ can be written as

   $$111 \cdots 111 = 111 \cdots 108 + 3 = 4k + 3.]$$

9. Verify that if an integer is simultaneously a square and a cube (as is the case with $64 = 8^2 = 4^3$), then it must be either of the form $7k$ or $7k + 1$.
10. For $n \geq 1$, establish that the integer $n(7n^2 + 5)$ is of the form $6k$.
11. If n is an odd integer, show that $n^4 + 4n^2 + 11$ is of the form $16k$.

2.3 THE GREATEST COMMON DIVISOR

Of special significance is the case in which the remainder in the Division Algorithm turns out to be zero. Let us look into this situation now.

Definition 2.1. An integer b is said to be *divisible* by an integer $a \neq 0$, in symbols $a \mid b$, if there exists some integer c such that $b = ac$. We write $a \nmid b$ to indicate that b is not divisible by a.

Thus, for example, -12 is divisible by 4, because $-12 = 4(-3)$. However, 10 is not divisible by 3; for there is no integer c that makes the statement $10 = 3c$ true.

There is other language for expressing the divisibility relation $a \mid b$. We could say that a is a *divisor* of b, that a is a *factor* of b, or that b is a *multiple* of a. Notice that in Definition 2.1 there is a restriction on the divisor a: Whenever the notation $a \mid b$ is employed, it is understood that a is different from zero.

If a is a divisor of b, then b is also divisible by $-a$ (indeed, $b = ac$ implies that $b = (-a)(-c)$), so that the divisors of an integer always occur in pairs. To find all the divisors of a given integer, it is sufficient to obtain the positive divisors and then adjoin to them the corresponding negative integers. For this reason, we shall usually limit ourselves to a consideration of positive divisors.

It will be helpful to list some immediate consequences of Definition 2.1. (The reader is again reminded that, although not stated, divisors are assumed to be nonzero.)

Theorem 2.2. For integers a, b, c, the following hold:

(a) $a \mid 0$, $1 \mid a$, $a \mid a$.
(b) $a \mid 1$ if and only if $a = \pm 1$.
(c) If $a \mid b$ and $c \mid d$, then $ac \mid bd$.
(d) If $a \mid b$ and $b \mid c$, then $a \mid c$.
(e) $a \mid b$ and $b \mid a$ if and only if $a = \pm b$.
(f) If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.
(g) If $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for arbitrary integers x and y.

Proof. We shall prove assertions (f) and (g), leaving the other parts as an exercise. If $a \mid b$, then there exists an integer c such that $b = ac$; also, $b \neq 0$ implies that $c \neq 0$. Upon taking absolute values, we get $|b| = |ac| = |a||c|$. Because $c \neq 0$, it follows that $|c| \geq 1$, whence $|b| = |a||c| \geq |a|$.

As regards (g), the relations $a \mid b$ and $a \mid c$ ensure that $b = ar$ and $c = as$ for suitable integers r and s. But then whatever the choice of x and y,

$$bx + cy = arx + asy = a(rx + sy)$$

Because $rx + sy$ is an integer, this says that $a \mid (bx + cy)$, as desired.

It is worth pointing out that property (g) of Theorem 2.2 extends by induction to sums of more than two terms. That is, if $a \mid b_k$ for $k = 1, 2, \ldots, n$, then

$$a \mid (b_1x_1 + b_2x_2 + \cdots + b_nx_n)$$

for all integers $x_1, x_2, \ldots, x_n$. The few details needed for the proof are so straightforward that we omit them.

If a and b are arbitrary integers, then an integer d is said to be a *common divisor* of a and b if both $d \mid a$ and $d \mid b$. Because 1 is a divisor of every integer,

1 is a common divisor of a and b; hence, their set of positive common divisors is nonempty. Now every integer divides zero, so that if $a = b = 0$, then every integer serves as a common divisor of a and b. In this instance, the set of positive common divisors of a and b is infinite. However, when at least one of a or b is different from zero, there are only a finite number of positive common divisors. Among these, there is a largest one, called the greatest common divisor of a and b. We frame this as Definition 2.2.

Definition 2.2. Let a and b be given integers, with at least one of them different from zero. The *greatest common divisor* of a and b, denoted by $\gcd(a, b)$, is the positive integer d satisfying the following:

(a) $d \mid a$ and $d \mid b$.
(b) If $c \mid a$ and $c \mid b$, then $c \leq d$.

Example 2.2. The positive divisors of -12 are 1, 2, 3, 4, 6, 12, whereas those of 30 are 1, 2, 3, 5, 6, 10, 15, 30; hence, the positive common divisors of -12 and 30 are 1, 2, 3, 6. Because 6 is the largest of these integers, it follows that $\gcd(-12, 30) = 6$. In the same way, we can show that

$$\gcd(-5, 5) = 5 \qquad \gcd(8, 17) = 1 \qquad \gcd(-8, -36) = 4$$

The next theorem indicates that $\gcd(a, b)$ can be represented as a linear combination of a and b. (By a *linear combination* of a and b, we mean an expression of the form $ax + by$, where x and y are integers.) This is illustrated by, say,

$$\gcd(-12, 30) = 6 = (-12)2 + 30 \cdot 1$$

or

$$\gcd(-8, -36) = 4 = (-8)4 + (-36)(-1)$$

Now for the theorem.

Theorem 2.3. Given integers a and b, not both of which are zero, there exist integers x and y such that

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

Proof. Consider the set S of all positive linear combinations of a and b:

$$S = \{au + bv \mid au + bv > 0; u, v \text{ integers}\}$$

Notice first that S is not empty. For example, if $a \neq 0$, then the integer $|a| = au + b \cdot 0$ lies in S, where we choose $u = 1$ or $u = -1$ according as a is positive or negative. By virtue of the Well-Ordering Principle, S must contain a smallest element d. Thus, from the very definition of S, there exist integers x and y for which $d = ax + by$. We claim that $d = \gcd(a, b)$.

Taking stock of the Division Algorithm, we can obtain integers q and r such that $a = qd + r$, where $0 \leq r < d$. Then r can be written in the form

$$\begin{aligned} r &= a - qd = a - q(ax + by) \\ &= a(1 - qx) + b(-qy) \end{aligned}$$

If r were positive, then this representation would imply that r is a member of S, contradicting the fact that d is the least integer in S (recall that $r < d$). Therefore, $r = 0$, and so $a = qd$, or equivalently $d \mid a$. By similar reasoning, $d \mid b$, the effect of which is to make d a common divisor of a and b.

Now if c is an arbitrary positive common divisor of the integers a and b, then part (g) of Theorem 2.2 allows us to conclude that $c \mid (ax + by)$; that is, $c \mid d$. By part (f) of the same theorem, $c = |c| \leq |d| = d$, so that d is greater than every positive common divisor of a and b. Piecing the bits of information together, we see that $d = \gcd(a, b)$.

It should be noted that the foregoing argument is merely an "existence" proof and does not provide a practical method for finding the values of x and y. This will come later.

A perusal of the proof of Theorem 2.3 reveals that the greatest common divisor of a and b may be described as the smallest positive integer of the form $ax + by$. Consider the case in which $a = 6$ and $b = 15$. Here, the set S becomes

$$\begin{aligned} S &= \{6(-2) + 15 \cdot 1, 6(-1) + 15 \cdot 1, 6 \cdot 1 + 15 \cdot 0, \ldots\} \\ &= \{3, 9, 6, \ldots\} \end{aligned}$$

We observe that 3 is the smallest integer in S, whence $3 = \gcd(6, 15)$.

The nature of the members of S appearing in this illustration suggests another result, which we give in the next corollary.

Corollary. If a and b are given integers, not both zero, then the set

$$T = \{ax + by \mid x, y \text{ are integers}\}$$

is precisely the set of all multiples of $d = \gcd(a, b)$.

Proof. Because $d \mid a$ and $d \mid b$, we know that $d \mid (ax + by)$ for all integers x, y. Thus, every member of T is a multiple of d. Conversely, d may be written as $d = ax_0 + by_0$ for suitable integers x_0 and y_0, so that any multiple nd of d is of the form

$$nd = n(ax_0 + by_0) = a(nx_0) + b(ny_0)$$

Hence, nd is a linear combination of a and b, and, by definition, lies in T.

It may happen that 1 and -1 are the only common divisors of a given pair of integers a and b, whence $\gcd(a, b) = 1$. For example:

$$\gcd(2, 5) = \gcd(-9, 16) = \gcd(-27, -35) = 1$$

This situation occurs often enough to prompt a definition.

Definition 2.3. Two integers a and b, not both of which are zero, are said to be *relatively prime* whenever $\gcd(a, b) = 1$.

The following theorem characterizes relatively prime integers in terms of linear combinations.

Theorem 2.4. Let a and b be integers, not both zero. Then a and b are relatively prime if and only if there exist integers x and y such that $1 = ax + by$.

Proof. If a and b are relatively prime so that $\gcd(a, b) = 1$, then Theorem 2.3 guarantees the existence of integers x and y satisfying $1 = ax + by$. As for the converse, suppose that $1 = ax + by$ for some choice of x and y, and that $d = \gcd(a, b)$. Because $d \mid a$ and $d \mid b$, Theorem 2.2 yields $d \mid (ax + by)$, or $d \mid 1$. Inasmuch as d is a positive integer, this last divisibility condition forces d to equal 1 (part (b) of Theorem 2.2 plays a role here), and the desired conclusion follows.

This result leads to an observation that is useful in certain situations; namely,

Corollary 1. If $\gcd(a, b) = d$, then $\gcd(a/d, b/d) = 1$.

Proof. Before starting with the proof proper, we should observe that although a/d and b/d have the appearance of fractions, in fact, they are integers because d is a divisor both of a and of b. Now, knowing that $\gcd(a, b) = d$, it is possible to find integers x and y such that $d = ax + by$. Upon dividing each side of this equation by d, we obtain the expression

$$1 = \left(\frac{a}{d}\right)x + \left(\frac{b}{d}\right)y$$

Because a/d and b/d are integers, an appeal to the theorem is legitimate. The conclusion is that a/d and b/d are relatively prime.

For an illustration of the last corollary, let us observe that $\gcd(-12, 30) = 6$ and

$$\gcd(-12/6, 30/6) = \gcd(-2, 5) = 1$$

as it should be.

It is not true, without adding an extra condition, that $a \mid c$ and $b \mid c$ together give $ab \mid c$. For instance, $6 \mid 24$ and $8 \mid 24$, but $6 \cdot 8 \nmid 24$. If 6 and 8 were relatively prime, of course, this situation would not arise. This brings us to Corollary 2.

Corollary 2. If $a \mid c$ and $b \mid c$, with $\gcd(a, b) = 1$, then $ab \mid c$.

Proof. Inasmuch as $a \mid c$ and $b \mid c$, integers r and s can be found such that $c = ar = bs$. Now the relation $\gcd(a, b) = 1$ allows us to write $1 = ax + by$ for some choice of integers x and y. Multiplying the last equation by c, it appears that

$$c = c \cdot 1 = c(ax + by) = acx + bcy$$

If the appropriate substitutions are now made on the right-hand side, then

$$c = a(bs)x + b(ar)y = ab(sx + ry)$$

or, as a divisibility statement, $ab \mid c$.

Our next result seems mild enough, but is of fundamental importance.

Theorem 2.5 Euclid's lemma. If $a \mid bc$, with $\gcd(a, b) = 1$, then $a \mid c$.

Proof. We start again from Theorem 2.3, writing $1 = ax + by$, where x and y are integers. Multiplication of this equation by c produces

$$c = 1 \cdot c = (ax + by)c = acx + bcy$$

Because $a \mid ac$ and $a \mid bc$, it follows that $a \mid (acx + bcy)$, which can be recast as $a \mid c$.

If a and b are not relatively prime, then the conclusion of Euclid's lemma may fail to hold. Here is a specific example: $12 \mid 9 \cdot 8$, but $12 \nmid 9$ and $12 \nmid 8$.

The subsequent theorem often serves as a definition of $\gcd(a, b)$. The advantage of using it as a definition is that order relationship is not involved. Thus, it may be used in algebraic systems having no order relation.

Theorem 2.6. Let a, b be integers, not both zero. For a positive integer d, $d = \gcd(a, b)$ if and only if

(a) $d \mid a$ and $d \mid b$.
(b) Whenever $c \mid a$ and $c \mid b$, then $c \mid d$.

Proof. To begin, suppose that $d = \gcd(a, b)$. Certainly, $d \mid a$ and $d \mid b$, so that (a) holds. In light of Theorem 2.3, d is expressible as $d = ax + by$ for some integers x, y. Thus, if $c \mid a$ and $c \mid b$, then $c \mid (ax + by)$, or rather $c \mid d$. In short, condition (b) holds. Conversely, let d be any positive integer satisfying the stated conditions. Given any common divisor c of a and b, we have $c \mid d$ from hypothesis (b). The implication is that $d \geq c$, and consequently d is the greatest common divisor of a and b.

PROBLEMS 2.3

1. If $a \mid b$, show that $(-a) \mid b$, $a \mid (-b)$, and $(-a) \mid (-b)$.
2. Given integers a, b, c, d, verify the following:
   (a) If $a \mid b$, then $a \mid bc$.
   (b) If $a \mid b$ and $a \mid c$, then $a^2 \mid bc$.
   (c) $a \mid b$ if and only if $ac \mid bc$, where $c \neq 0$.
   (d) If $a \mid b$ and $c \mid d$, then $ac \mid bd$.
3. Prove or disprove: If $a \mid (b + c)$, then either $a \mid b$ or $a \mid c$.
4. For $n \geq 1$, use mathematical induction to establish each of the following divisibility statements:
   (a) $8 \mid 5^{2n} + 7$.
   [*Hint:* $5^{2(k+1)} + 7 = 5^2(5^{2k} + 7) + (7 - 5^2 \cdot 7)$.]
   (b) $15 \mid 2^{4n} - 1$.
   (c) $5 \mid 3^{3n+1} + 2^{n+1}$.
   (d) $21 \mid 4^{n+1} + 5^{2n-1}$.
   (e) $24 \mid 2 \cdot 7^n + 3 \cdot 5^n - 5$.
5. Prove that for any integer a, one of the integers a, $a + 2$, $a + 4$ is divisible by 3.

6. For an arbitrary integer a, verify the following:
   (a) $2 \mid a(a+1)$, and $3 \mid a(a+1)(a+2)$.
   (b) $3 \mid a(2a^2+7)$.
   (c) If a is odd, then $32 \mid (a^2+3)(a^2+7)$.

7. Prove that if a and b are both odd integers, then $16 \mid a^4 + b^4 - 2$.

8. Prove the following:
   (a) The sum of the squares of two odd integers cannot be a perfect square.
   (b) The product of four consecutive integers is 1 less than a perfect square.

9. Establish that the difference of two consecutive cubes is never divisible by 2.

10. For a nonzero integer a, show that $\gcd(a, 0) = |a|$, $\gcd(a, a) = |a|$, and $\gcd(a, 1) = 1$.

11. If a and b are integers, not both of which are zero, verify that

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

12. Prove that, for a positive integer n and any integer a, $\gcd(a, a+n)$ divides n; hence, $\gcd(a, a+1) = 1$.

13. Given integers a and b, prove the following:
   (a) There exist integers x and y for which $c = ax + by$ if and only if $\gcd(a, b) \mid c$.
   (b) If there exist integers x and y for which $ax + by = \gcd(a, b)$, then $\gcd(x, y) = 1$.

14. For any integer a, show the following:
   (a) $\gcd(2a+1, 9a+4) = 1$.
   (b) $\gcd(5a+2, 7a+3) = 1$.
   (c) If a is odd, then $\gcd(3a, 3a+2) = 1$.

15. If a and b are integers, not both of which are zero, prove that $\gcd(2a - 3b, 4a - 5b)$ divides b; hence, $\gcd(2a+3, 4a+5) = 1$.

16. Given an odd integer a, establish that

$$a^2 + (a+2)^2 + (a+4)^2 + 1$$

is divisible by 12.

17. Prove that the expression $(3n)!/(3!)^n$ is an integer for all $n \geq 0$.

18. Prove: The product of any three consecutive integers is divisible by 6; the product of any four consecutive integers is divisible by 24; the product of any five consecutive integers is divisible by 120.
   [*Hint:* See Corollary 2 to Theorem 2.4.]

19. Establish each of the assertions below:
   (a) If a is an arbitrary integer, then $6 \mid a(a^2+11)$.
   (b) If a is an odd integer, then $24 \mid a(a^2-1)$.
   [*Hint:* The square of an odd integer is of the form $8k+1$.]
   (c) If a and b are odd integers, then $8 \mid (a^2 - b^2)$.
   (d) If a is an integer not divisible by 2 or 3, then $24 \mid (a^2+23)$.
   (e) If a is an arbitrary integer, then $360 \mid a^2(a^2-1)(a^2-4)$.

20. Confirm the following properties of the greatest common divisor:
   (a) If $\gcd(a, b) = 1$, and $\gcd(a, c) = 1$, then $\gcd(a, bc) = 1$.
   [*Hint:* Because $1 = ax + by = au + cv$ for some x, y, u, v,
   $1 = (ax+by)(au+cv) = a(aux + cvx + byu) + bc(yv)$.]
   (b) If $\gcd(a, b) = 1$, and $c \mid a$, then $\gcd(b, c) = 1$.
   (c) If $\gcd(a, b) = 1$, then $\gcd(ac, b) = \gcd(c, b)$.
   (d) If $\gcd(a, b) = 1$, and $c \mid a + b$, then $\gcd(a, c) = \gcd(b, c) = 1$.
   [*Hint:* Let $d = \gcd(a, c)$. Then $d \mid a$, $d \mid c$ implies that $d \mid (a+b) - a$, or $d \mid b$.]
   (e) If $\gcd(a, b) = 1$, $d \mid ac$, and $d \mid bc$, then $d \mid c$.
   (f) If $\gcd(a, b) = 1$, then $\gcd(a^2, b^2) = 1$.
   [*Hint:* First show that $\gcd(a, b^2) = \gcd(a^2, b) = 1$.]

21. (a) Prove that if $d \mid n$, then $2^d - 1 \mid 2^n - 1$.
[*Hint:* Use the identity

$$x^k - 1 = (x - 1)(x^{k-1} + x^{k-2} + \cdots + x + 1).]$$

(b) Verify that $2^{35} - 1$ is divisible by 31 and 127.

22. Let t_n denote the nth triangular number. For what values of n does t_n divide the sum $t_1 + t_2 + \cdots + t_n$?
[*Hint:* See Problem 1(c), Section 1.1.]

23. If $a \mid bc$, show that $a \mid \gcd(a, b) \gcd(a, c)$.

2.4 THE EUCLIDEAN ALGORITHM

The greatest common divisor of two integers can, of course, be found by listing all their positive divisors and choosing the largest one common to each; but this is cumbersome for large numbers. A more efficient process, involving repeated application of the Division Algorithm, is given in the seventh Book of the *Elements*. Although there is historical evidence that this method predates Euclid, today it is referred to as the *Euclidean Algorithm*.

The Euclidean Algorithm may be described as follows: Let a and b be two integers whose greatest common divisor is desired. Because $\gcd(|a|, |b|) = \gcd(a, b)$, there is no harm in assuming that $a \geq b > 0$. The first step is to apply the Division Algorithm to a and b to get

$$a = q_1 b + r_1 \qquad 0 \leq r_1 < b$$

If it happens that $r_1 = 0$, then $b \mid a$ and $\gcd(a, b) = b$. When $r_1 \neq 0$, divide b by r_1 to produce integers q_2 and r_2 satisfying

$$b = q_2 r_1 + r_2 \qquad 0 \leq r_2 < r_1$$

If $r_2 = 0$, then we stop; otherwise, proceed as before to obtain

$$r_1 = q_3 r_2 + r_3 \qquad 0 \leq r_3 < r_2$$

This division process continues until some zero remainder appears, say, at the $(n + 1)$th stage where r_{n-1} is divided by r_n (a zero remainder occurs sooner or later because the decreasing sequence $b > r_1 > r_2 > \cdots \geq 0$ cannot contain more than b integers).

The result is the following system of equations:

$$\begin{aligned}
a &= q_1 b + r_1 & 0 < r_1 < b \\
b &= q_2 r_1 + r_2 & 0 < r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3 & 0 < r_3 < r_2 \\
&\vdots \\
r_{n-2} &= q_n r_{n-1} + r_n & 0 < r_n < r_{n-1} \\
r_{n-1} &= q_{n+1} r_n + 0
\end{aligned}$$

We argue that r_n, the last nonzero remainder that appears in this manner, is equal to $\gcd(a, b)$. Our proof is based on the lemma below.

Lemma. If $a = qb + r$, then $\gcd(a, b) = \gcd(b, r)$.

Proof. If $d = \gcd(a, b)$, then the relations $d \mid a$ and $d \mid b$ together imply that $d \mid (a - qb)$, or $d \mid r$. Thus, d is a common divisor of both b and r. On the other hand, if c is an arbitrary common divisor of b and r, then $c \mid (qb + r)$, whence $c \mid a$. This makes c a common divisor of a and b, so that $c \leq d$. It now follows from the definition of $\gcd(b, r)$ that $d = \gcd(b, r)$.

Using the result of this lemma, we simply work down the displayed system of equations, obtaining

$$\gcd(a, b) = \gcd(b, r_1) = \cdots = \gcd(r_{n-1}, r_n) = \gcd(r_n, 0) = r_n$$

as claimed.

Theorem 2.3 asserts that $\gcd(a, b)$ can be expressed in the form $ax + by$, but the proof of the theorem gives no hint as to how to determine the integers x and y. For this, we fall back on the Euclidean Algorithm. Starting with the next-to-last equation arising from the algorithm, we write

$$r_n = r_{n-2} - q_n r_{n-1}$$

Now solve the preceding equation in the algorithm for r_{n-1} and substitute to obtain

$$\begin{aligned} r_n &= r_{n-2} - q_n(r_{n-3} - q_{n-1}r_{n-2}) \\ &= (1 + q_n q_{n-1})r_{n-2} + (-q_n)r_{n-3} \end{aligned}$$

This represents r_n as a linear combination of r_{n-2} and r_{n-3}. Continuing backward through the system of equations, we successively eliminate the remainders r_{n-1}, $r_{n-2}, \ldots, r_2, r_1$ until a stage is reached where $r_n = \gcd(a, b)$ is expressed as a linear combination of a and b.

Example 2.3. Let us see how the Euclidean Algorithm works in a concrete case by calculating, say, $\gcd(12378, 3054)$. The appropriate applications of the Division Algorithm produce the equations

$$\begin{aligned} 12378 &= 4 \cdot 3054 + 162 \\ 3054 &= 18 \cdot 162 + 138 \\ 162 &= 1 \cdot 138 + 24 \\ 138 &= 5 \cdot 24 + 18 \\ 24 &= 1 \cdot 18 + 6 \\ 18 &= 3 \cdot 6 + 0 \end{aligned}$$

Our previous discussion tells us that the last nonzero remainder appearing in these equations, namely, the integer 6, is the greatest common divisor of 12378 and 3054:

$$6 = \gcd(12378, 3054)$$

To represent 6 as a linear combination of the integers 12378 and 3054, we start with the next-to-last of the displayed equations and successively eliminate the remainders

18, 24, 138, and 162:

$$\begin{aligned} 6 &= 24 - 18 \\ &= 24 - (138 - 5 \cdot 24) \\ &= 6 \cdot 24 - 138 \\ &= 6(162 - 138) - 138 \\ &= 6 \cdot 162 - 7 \cdot 138 \\ &= 6 \cdot 162 - 7(3054 - 18 \cdot 162) \\ &= 132 \cdot 162 - 7 \cdot 3054 \\ &= 132(12378 - 4 \cdot 3054) - 7 \cdot 3054 \\ &= 132 \cdot 12378 + (-535)3054 \end{aligned}$$

Thus, we have

$$6 = \gcd(12378\,,\,3054) = 12378x + 3054y$$

where $x = 132$ and $y = -535$. Note that this is not the only way to express the integer 6 as a linear combination of 12378 and 3054; among other possibilities, we could add and subtract $3054 \cdot 12378$ to get

$$\begin{aligned} 6 &= (132 + 3054)12378 + (-535 - 12378)3054 \\ &= 3186 \cdot 12378 + (-12913)3054 \end{aligned}$$

The French mathematician Gabriel Lamé (1795–1870) proved that the number of steps required in the Euclidean Algorithm is at most five times the number of digits in the smaller integer. In Example 2.3, the smaller integer (namely, 3054) has four digits, so that the total number of divisions cannot be greater than 20; in actuality only six divisions were needed. Another observation of interest is that for each $n > 0$, it is possible to find integers a_n and b_n such that exactly n divisions are required to compute $\gcd(a_n\,,\,b_n)$ by the Euclidean Algorithm. We shall prove this fact in Chapter 14.

One more remark is necessary. The number of steps in the Euclidean Algorithm usually can be reduced by selecting remainders r_{k+1} such that $|\,r_{k+1}\,| < r_k/2$, that is, by working with least absolute remainders in the divisions. Thus, repeating Example 2.3, it is more efficient to write

$$\begin{aligned} 12378 &= 4 \cdot 3054 + 162 \\ 3054 &= 19 \cdot 162 - 24 \\ 162 &= 7 \cdot 24 - 6 \\ 24 &= (-4)(-6) + 0 \end{aligned}$$

As evidenced by this set of equations, this scheme is apt to produce the negative of the value of the greatest common divisor of two integers (the last nonzero remainder being -6), rather than the greatest common divisor itself.

An important consequence of the Euclidean Algorithm is the following theorem.

Theorem 2.7. If $k > 0$, then $\gcd(ka\,,\,kb) = k\,\gcd(a\,,\,b)$.

Proof. If each of the equations appearing in the Euclidean Algorithm for a and b (see page 28) is multiplied by k, we obtain

$$\begin{aligned} ak &= q_1(bk) + r_1k & 0 < r_1k < bk \\ bk &= q_2(r_1k) + r_2k & 0 < r_2k < r_1k \\ &\vdots \\ r_{n-2}k &= q_n(r_{n-1}k) + r_nk & 0 < r_nk < r_{n-1}k \\ r_{n-1}k &= q_{n+1}(r_nk) + 0 \end{aligned}$$

But this is clearly the Euclidean Algorithm applied to the integers ak and bk, so that their greatest common divisor is the last nonzero remainder r_nk; that is,

$$\gcd(ka\,,\,kb) = r_nk = k\,\gcd(a\,,\,b)$$

as stated in the theorem.

Corollary. For any integer $k \neq 0$, $\gcd(ka\,,\,kb) = |\,k\,|\,\gcd(a\,,\,b)$.

Proof. It suffices to consider the case in which $k < 0$. Then $-k = |\,k\,| > 0$ and, by Theorem 2.7,

$$\begin{aligned} \gcd(ak\,,\,bk) &= \gcd(-ak\,,\,-bk) \\ &= \gcd(a\,|\,k|\,,\,b\,|\,k|) \\ &= |\,k\,|\,\gcd(a\,,\,b) \end{aligned}$$

An alternate proof of Theorem 2.7 runs very quickly as follows: $\gcd(ak\,,\,bk)$ is the smallest positive integer of the form $(ak)x + (bk)y$, which, in turn, is equal to k times the smallest positive integer of the form $ax + by$; the latter value is equal to $k\,\gcd(a\,,\,b)$.

By way of illustrating Theorem 2.7, we see that

$$\gcd(12\,,\,30) = 3\,\gcd(4\,,\,10) = 3 \cdot 2\,\gcd(2\,,\,5) = 6 \cdot 1 = 6$$

There is a concept parallel to that of the greatest common divisor of two integers, known as their least common multiple; but we shall not have much occasion to make use of it. An integer c is said to be a *common multiple* of two nonzero integers a and b whenever $a \mid c$ and $b \mid c$. Evidently, zero is a common multiple of a and b. To see there exist common multiples that are not trivial, just note that the products ab and $-(ab)$ are both common multiples of a and b, and one of these is positive. By the Well-Ordering Principle, the set of positive common multiples of a and b must contain a smallest integer; we call it the least common multiple of a and b.

For the record, here is the official definition.

Definition 2.4. The *least common multiple* of two nonzero integers a and b, denoted by $\operatorname{lcm}(a\,,\,b)$, is the positive integer m satisfying the following:

(a) $a \mid m$ and $b \mid m$.
(b) If $a \mid c$ and $b \mid c$, with $c > 0$, then $m \leq c$.

As an example, the positive common multiples of the integers -12 and 30 are 60, 120, 180, ...; hence, $\text{lcm}(-12, 30) = 60$.

The following remark is clear from our discussion: Given nonzero integers a and b, $\text{lcm}(a, b)$ always exists and $\text{lcm}(a, b) \leq |ab|$.

We lack a relationship between the ideas of greatest common divisor and least common multiple. This gap is filled by Theorem 2.8.

Theorem 2.8. For positive integers a and b

$$\gcd(a, b)\,\text{lcm}(a, b) = ab$$

Proof. To begin, put $d = \gcd(a, b)$ and write $a = dr$, $b = ds$ for integers r and s. If $m = ab/d$, then $m = as = rb$, the effect of which is to make m a (positive) common multiple of a and b.

Now let c be any positive integer that is a common multiple of a and b; say, for definiteness, $c = au = bv$. As we know, there exist integers x and y satisfying $d = ax + by$. In consequence,

$$\frac{c}{m} = \frac{cd}{ab} = \frac{c(ax + by)}{ab} = \left(\frac{c}{b}\right)x + \left(\frac{c}{a}\right)y = vx + uy$$

This equation states that $m \mid c$, allowing us to conclude that $m \leq c$. Thus, in accordance with Definition 2.4, $m = \text{lcm}(a, b)$; that is,

$$\text{lcm}(a, b) = \frac{ab}{d} = \frac{ab}{\gcd(a, b)}$$

which is what we started out to prove.

Theorem 2.8 has a corollary that is worth a separate statement.

Corollary. For any choice of positive integers a and b, $\text{lcm}(a, b) = ab$ if and only if $\gcd(a, b) = 1$.

Perhaps the chief virtue of Theorem 2.8 is that it makes the calculation of the least common multiple of two integers dependent on the value of their greatest common divisor—which, in turn, can be calculated from the Euclidean Algorithm. When considering the positive integers 3054 and 12378, for instance, we found that $\gcd(3054, 12378) = 6$; whence,

$$\text{lcm}(3054, 12378) = \frac{3054 \cdot 12378}{6} = 6300402$$

Before moving on to other matters, let us observe that the notion of greatest common divisor can be extended to more than two integers in an obvious way. In the case of three integers, a, b, c, not all zero, $\gcd(a, b, c)$ is defined to be the positive integer d having the following properties:

(a) d is a divisor of each of a, b, c.

(b) If e divides the integers a, b, c, then $e \leq d$.

We cite two examples:

$$\gcd(39, 42, 54) = 3 \qquad \text{and} \qquad \gcd(49, 210, 350) = 7$$

The reader is cautioned that it is possible for three integers to be relatively prime as a triple (in other words, $\gcd(a, b, c) = 1$), yet not relatively prime in pairs; this is brought out by the integers 6, 10, and 15.

PROBLEMS 2.4

1. Find $\gcd(143, 227)$, $\gcd(306, 657)$, and $\gcd(272, 1479)$.
2. Use the Euclidean Algorithm to obtain integers x and y satisfying the following:
   (a) $\gcd(56, 72) = 56x + 72y$.
   (b) $\gcd(24, 138) = 24x + 138y$.
   (c) $\gcd(119, 272) = 119x + 272y$.
   (d) $\gcd(1769, 2378) = 1769x + 2378y$.
3. Prove that if d is a common divisor of a and b, then $d = \gcd(a, b)$ if and only if $\gcd(a/d, b/d) = 1$.
   [*Hint:* Use Theorem 2.7.]
4. Assuming that $\gcd(a, b) = 1$, prove the following:
   (a) $\gcd(a + b, a - b) = 1$ or 2.
   [*Hint:* Let $d = \gcd(a + b, a - b)$ and show that $d \mid 2a$, $d \mid 2b$, and thus that $d \leq \gcd(2a, 2b) = 2\gcd(a, b)$.]
   (b) $\gcd(2a + b, a + 2b) = 1$ or 3.
   (c) $\gcd(a + b, a^2 + b^2) = 1$ or 2.
   [*Hint:* $a^2 + b^2 = (a + b)(a - b) + 2b^2$.]
   (d) $\gcd(a + b, a^2 - ab + b^2) = 1$ or 3.
   [*Hint:* $a^2 - ab + b^2 = (a + b)^2 - 3ab$.]
5. For $n \geq 1$, and positive integers a, b, show the following:
   (a) If $\gcd(a, b) = 1$, then $\gcd(a^n, b^n) = 1$.
   [*Hint:* See Problem 20(a), Section 2.2.]
   (b) The relation $a^n \mid b^n$ implies that $a \mid b$.
   [*Hint:* Put $d = \gcd(a, b)$ and write $a = rd$, $b = sd$, where $\gcd(r, s) = 1$. By part (a), $\gcd(r^n, s^n) = 1$. Show that $r = 1$, whence $a = d$.]
6. Prove that if $\gcd(a, b) = 1$, then $\gcd(a + b, ab) = 1$.
7. For nonzero integers a and b, verify that the following conditions are equivalent:
   (a) $a \mid b$.
   (b) $\gcd(a, b) = |a|$.
   (c) $\operatorname{lcm}(a, b) = |b|$.
8. Find $\operatorname{lcm}(143, 227)$, $\operatorname{lcm}(306, 657)$, and $\operatorname{lcm}(272, 1479)$.
9. Prove that the greatest common divisor of two positive integers divides their least common multiple.
10. Given nonzero integers a and b, establish the following facts concerning $\operatorname{lcm}(a, b)$:
    (a) $\gcd(a, b) = \operatorname{lcm}(a, b)$ if and only if $a = \pm b$.
    (b) If $k > 0$, then $\operatorname{lcm}(ka, kb) = k \operatorname{lcm}(a, b)$.
    (c) If m is any common multiple of a and b, then $\operatorname{lcm}(a, b) \mid m$.
    [*Hint:* Put $t = \operatorname{lcm}(a, b)$ and use the Division Algorithm to write $m = qt + r$, where $0 \leq r < t$. Show that r is a common multiple of a and b.]
11. Let a, b, c be integers, no two of which are zero, and $d = \gcd(a, b, c)$. Show that

$$d = \gcd(\gcd(a, b), c) = \gcd(a, \gcd(b, c)) = \gcd(\gcd(a, c), b)$$

12. Find integers x, y, z satisfying

$$\gcd(198, 288, 512) = 198x + 288y + 512z$$

[*Hint:* Put $d = \gcd(198, 288)$. Because $\gcd(198, 288, 512) = \gcd(d, 512)$, first find integers u and v for which $\gcd(d, 512) = du + 512v$.]

2.5 THE DIOPHANTINE EQUATION $ax + by = c$

We now change focus somewhat and take up the study of Diophantine equations. The name honors the mathematician Diophantus, who initiated the study of such equations. Practically nothing is known of Diophantus as an individual, save that he lived in Alexandria sometime around 250 A.D. The only positive evidence as to the date of his activity is that the Bishop of Laodicea, who began his episcopate in 270, dedicated a book on Egyptian computation to his friend Diophantus. Although Diophantus' works were written in Greek and he displayed the Greek genius for theoretical abstraction, he was most likely a Hellenized Babylonian. The only personal particulars we have of his career come from the wording of an epigram-problem (apparently dating from the 4th century): His boyhood lasted 1/6 of his life; his beard grew after 1/12 more; after 1/7 more he married, and his son was born 5 years later; the son lived to half his father's age and the father died 4 years after his son. If x was the age at which Diophantus died, these data lead to the equation

$$\frac{1}{6}x + \frac{1}{12}x + \frac{1}{7}x + 5 + \frac{1}{2}x + 4 = x$$

with solution $x = 84$. Thus, he must have reached an age of 84, but in what year or even in what century is not certain.

The great work upon which the reputation of Diophantus rests is his *Arithmetica*, which may be described as the earliest treatise on algebra. Only six Books of the original thirteen have been preserved. It is in the *Arithmetica* that we find the first systematic use of mathematical notation, although the signs employed are of the nature of abbreviations for words rather than algebraic symbols in the sense with which we use them today. Special symbols are introduced to represent frequently occurring concepts, such as the unknown quantity in an equation and the different powers of the unknown up to the sixth power; Diophantus also had a symbol to express subtraction, and another for equality.

It is customary to apply the term *Diophantine equation* to any equation in one or more unknowns that is to be solved in the integers. The simplest type of Diophantine equation that we shall consider is the linear Diophantine equation in two unknowns:

$$ax + by = c$$

where a, b, c are given integers and a, b are not both zero. A solution of this equation is a pair of integers x_0, y_0 that, when substituted into the equation, satisfy it; that is, we ask that $ax_0 + by_0 = c$. Curiously enough, the linear equation does not appear in the extant works of Diophantus (the theory required for its solution is to be found in Euclid's *Elements*), possibly because he viewed it as trivial; most of his problems deal with finding squares or cubes with certain properties.

A given linear Diophantine equation can have a number of solutions, as is the case with $3x + 6y = 18$, where

$$3 \cdot 4 + 6 \cdot 1 = 18$$
$$3(-6) + 6 \cdot 6 = 18$$
$$3 \cdot 10 + 6(-2) = 18$$

By contrast, there is no solution to the equation $2x + 10y = 17$. Indeed, the left-hand side is an even integer whatever the choice of x and y, whereas the right-hand side is not. Faced with this, it is reasonable to enquire about the circumstances under which a solution is possible and, when a solution does exist, whether we can determine all solutions explicitly.

The condition for solvability is easy to state: the linear Diophantine equation $ax + by = c$ admits a solution if and only if $d \mid c$, where $d = \gcd(a, b)$. We know that there are integers r and s for which $a = dr$ and $b = ds$. If a solution of $ax + by = c$ exists, so that $ax_0 + by_0 = c$ for suitable x_0 and y_0, then

$$c = ax_0 + by_0 = drx_0 + dsy_0 = d(rx_0 + sy_0)$$

which simply says that $d \mid c$. Conversely, assume that $d \mid c$, say $c = dt$. Using Theorem 2.3, integers x_0 and y_0 can be found satisfying $d = ax_0 + by_0$. When this relation is multiplied by t, we get

$$c = dt = (ax_0 + by_0)t = a(tx_0) + b(ty_0)$$

Hence, the Diophantine equation $ax + by = c$ has $x = tx_0$ and $y = ty_0$ as a particular solution. This proves part of our next theorem.

Theorem 2.9. The linear Diophantine equation $ax + by = c$ has a solution if and only if $d \mid c$, where $d = \gcd(a, b)$. If x_0, y_0 is any particular solution of this equation, then all other solutions are given by

$$x = x_0 + \left(\frac{b}{d}\right)t \qquad y = y_0 - \left(\frac{a}{d}\right)t$$

where t is an arbitrary integer.

Proof. To establish the second assertion of the theorem, let us suppose that a solution x_0, y_0 of the given equation is known. If x', y' is any other solution, then

$$ax_0 + by_0 = c = ax' + by'$$

which is equivalent to

$$a(x' - x_0) = b(y_0 - y')$$

By the corollary to Theorem 2.4, there exist relatively prime integers r and s such that $a = dr$, $b = ds$. Substituting these values into the last-written equation and canceling the common factor d, we find that

$$r(x' - x_0) = s(y_0 - y')$$

The situation is now this: $r \mid s(y_0 - y')$, with $\gcd(r, s) = 1$. Using Euclid's lemma, it must be the case that $r \mid (y_0 - y')$; or, in other words, $y_0 - y' = rt$ for some integer t.

Substituting, we obtain

$$x' - x_0 = st$$

This leads us to the formulas

$$x' = x_0 + st = x_0 + \left(\frac{b}{d}\right)t$$

$$y' = y_0 - rt = y_0 - \left(\frac{a}{d}\right)t$$

It is easy to see that these values satisfy the Diophantine equation, regardless of the choice of the integer t; for

$$\begin{aligned} ax' + by' &= a\left[x_0 + \left(\frac{b}{d}\right)t\right] + b\left[y_0 - \left(\frac{a}{d}\right)t\right] \\ &= (ax_0 + by_0) + \left(\frac{ab}{d} - \frac{ab}{d}\right)t \\ &= c + 0 \cdot t \\ &= c \end{aligned}$$

Thus, there are an infinite number of solutions of the given equation, one for each value of t.

Example 2.4. Consider the linear Diophantine equation

$$172x + 20y = 1000$$

Applying the Euclidean's Algorithm to the evaluation of gcd(172 , 20), we find that

$$\begin{aligned} 172 &= 8 \cdot 20 + 12 \\ 20 &= 1 \cdot 12 + 8 \\ 12 &= 1 \cdot 8 + 4 \\ 8 &= 2 \cdot 4 \end{aligned}$$

whence gcd(172 , 20) = 4. Because 4 | 1000, a solution to this equation exists. To obtain the integer 4 as a linear combination of 172 and 20, we work backward through the previous calculations, as follows:

$$\begin{aligned} 4 &= 12 - 8 \\ &= 12 - (20 - 12) \\ &= 2 \cdot 12 - 20 \\ &= 2(172 - 8 \cdot 20) - 20 \\ &= 2 \cdot 172 + (-17)20 \end{aligned}$$

Upon multiplying this relation by 250, we arrive at

$$\begin{aligned} 1000 = 250 \cdot 4 &= 250[2 \cdot 172 + (-17)20] \\ &= 500 \cdot 172 + (-4250)20 \end{aligned}$$

so that $x = 500$ and $y = -4250$ provide one solution to the Diophantine equation in question. All other solutions are expressed by

$$x = 500 + (20/4)t = 500 + 5t$$
$$y = -4250 - (172/4)t = -4250 - 43t$$

for some integer t.

A little further effort produces the solutions in the positive integers, if any happen to exist. For this, t must be chosen to satisfy simultaneously the inequalities

$$5t + 500 > 0 \qquad -43t - 4250 > 0$$

or, what amounts to the same thing,

$$-98\frac{36}{43} > t > -100$$

Because t must be an integer, we are forced to conclude that $t = -99$. Thus, our Diophantine equation has a unique positive solution $x = 5$, $y = 7$ corresponding to the value $t = -99$.

It might be helpful to record the form that Theorem 2.9 takes when the coefficients are relatively prime integers.

Corollary. If $\gcd(a, b) = 1$ and if x_0, y_0 is a particular solution of the linear Diophantine equation $ax + by = c$, then all solutions are given by

$$x = x_0 + bt \qquad y = y_0 - at$$

for integral values of t.

Here is an example. The equation $5x + 22y = 18$ has $x_0 = 8$, $y_0 = -1$ as one solution; from the corollary, a complete solution is given by $x = 8 + 22t$, $y = -1 - 5t$ for arbitrary t.

Diophantine equations frequently arise when solving certain types of traditional word problems, as evidenced by Example 2.5.

Example 2.5. A customer bought a dozen pieces of fruit, apples and oranges, for \$1.32. If an apple costs 3 cents more than an orange and more apples than oranges were purchased, how many pieces of each kind were bought?

To set up this problem as a Diophantine equation, let x be the number of apples and y be the number of oranges purchased; in addition, let z represent the cost (in cents) of an orange. Then the conditions of the problem lead to

$$(z + 3)x + zy = 132$$

or equivalently

$$3x + (x + y)z = 132$$

Because $x + y = 12$, the previous equation may be replaced by

$$3x + 12z = 132$$

which, in turn, simplifies to $x + 4z = 44$.

Stripped of inessentials, the object is to find integers x and z satisfying the Diophantine equation

$$x + 4z = 44 \tag{1}$$

Inasmuch as gcd $(1, 4) = 1$ is a divisor of 44, there is a solution to this equation. Upon multiplying the relation $1 = 1(-3) + 4 \cdot 1$ by 44 to get

$$44 = 1(-132) + 4 \cdot 44$$

it follows that $x_0 = -132$, $z_0 = 44$ serves as one solution. All other solutions of Eq. (1) are of the form

$$x = -132 + 4t \qquad z = 44 - t$$

where t is an integer.

Not all of the choices for t furnish solutions to the original problem. Only values of t that ensure $12 \geq x > 6$ should be considered. This requires obtaining those values of t such that

$$12 \geq -132 + 4t > 6$$

Now, $12 \geq -132 + 4t$ implies that $t \leq 36$, whereas $-132 + 4t > 6$ gives $t > 34\frac{1}{2}$. The only integral values of t to satisfy both inequalities are $t = 35$ and $t = 36$. Thus, there are two possible purchases: a dozen apples costing 11 cents apiece (the case where $t = 36$), or 8 apples at 12 cents each and 4 oranges at 9 cents each (the case where $t = 35$).

Linear indeterminate problems such as these have a long history, occurring as early as the 1st century in the Chinese mathematical literature. Owing to a lack of algebraic symbolism, they often appeared in the guise of rhetorical puzzles or riddles. The contents of the *Mathematical Classic* of Chang Ch′ iu-chien (6th century) attest to the algebraic abilities of the Chinese scholars. This elaborate treatise contains one of the most famous problems in indeterminate equations, in the sense of transmission to other societies—the problem of the "hundred fowls." The problem states:

> If a cock is worth 5 coins, a hen 3 coins, and three chicks together 1 coin, how many cocks, hens, and chicks, totaling 100, can be bought for 100 coins?

In terms of equations, the problem would be written (if x equals the number of cocks, y the number of hens, z the number of chicks):

$$5x + 3y + \frac{1}{3}z = 100 \qquad x + y + z = 100$$

Eliminating one of the unknowns, we are left with a linear Diophantine equation in the two other unknowns. Specifically, because the quantity $z = 100 - x - y$, we have $5x + 3y + \frac{1}{3}(100 - x - y) = 100$, or

$$7x + 4y = 100$$

This equation has the general solution $x = 4t$, $y = 25 - 7t$, so that $z = 75 + 3t$, where t is an arbitrary integer. Chang himself gave several answers:

$$\begin{array}{lll} x = 4 & y = 18 & z = 78 \\ x = 8 & y = 11 & z = 81 \\ x = 12 & y = 4 & z = 84 \end{array}$$

A little further effort produces all solutions in the positive integers. For this, t must be chosen to satisfy simultaneously the inequalities

$$4t > 0 \qquad 25 - 7t > 0 \qquad 75 + 3t > 0$$

The last two of these are equivalent to the requirement $-25 < t < 3\frac{4}{7}$. Because t must have a positive value, we conclude that $t = 1, 2, 3$, leading to precisely the values Chang obtained.

PROBLEMS 2.5

1. Which of the following Diophantine equations cannot be solved?
   (a) $6x + 51y = 22$.
   (b) $33x + 14y = 115$.
   (c) $14x + 35y = 93$.
2. Determine all solutions in the integers of the following Diophantine equations:
   (a) $56x + 72y = 40$.
   (b) $24x + 138y = 18$.
   (c) $221x + 35y = 11$.
3. Determine all solutions in the positive integers of the following Diophantine equations:
   (a) $18x + 5y = 48$.
   (b) $54x + 21y = 906$.
   (c) $123x + 360y = 99$.
   (d) $158x - 57y = 7$.
4. If a and b are relatively prime positive integers, prove that the Diophantine equation $ax - by = c$ has infinitely many solutions in the positive integers.
   [*Hint:* There exist integers x_0 and y_0 such that $ax_0 + by_0 = c$. For any integer t, which is larger than both $|x_0|/b$ and $|y_0|/a$, a positive solution of the given equation is $x = x_0 + bt$, $y = -(y_0 - at)$.]
5. (a) A man has \$4.55 in change composed entirely of dimes and quarters. What are the maximum and minimum number of coins that he can have? Is it possible for the number of dimes to equal the number of quarters?
   (b) The neighborhood theater charges \$1.80 for adult admissions and \$.75 for children. On a particular evening the total receipts were \$90. Assuming that more adults than children were present, how many people attended?
   (c) A certain number of sixes and nines is added to give a sum of 126; if the number of sixes and nines is interchanged, the new sum is 114. How many of each were there originally?
6. A farmer purchased 100 head of livestock for a total cost of \$4000. Prices were as follow: calves, \$120 each; lambs, \$50 each; piglets, \$25 each. If the farmer obtained at least one animal of each type, how many of each did he buy?
7. When Mr. Smith cashed a check at his bank, the teller mistook the number of cents for the number of dollars and vice versa. Unaware of this, Mr. Smith spent 68 cents and then

noticed to his surprise that he had twice the amount of the original check. Determine the smallest value for which the check could have been written.
[*Hint:* If x denotes the number of dollars and y the number of cents in the check, then $100y + x - 68 = 2(100x + y)$.]

8. Solve each of the puzzle-problems below:

(a) Alcuin of York, 775. One hundred bushels of grain are distributed among 100 persons in such a way that each man receives 3 bushels, each woman 2 bushels, and each child $\frac{1}{2}$ bushel. How many men, women, and children are there?

(b) Mahaviracarya, 850. There were 63 equal piles of plantain fruit put together and 7 single fruits. They were divided evenly among 23 travelers. What is the number of fruits in each pile?
[*Hint:* Consider the Diophantine equation $63x + 7 = 23y$.]

(c) Yen Kung, 1372. We have an unknown number of coins. If you make 77 strings of them, you are 50 coins short; but if you make 78 strings, it is exact. How many coins are there?
[*Hint:* If N is the number of coins, then $N = 77x + 27 = 78y$ for integers x and y.]

(d) Christoff Rudolff, 1526. Find the number of men, women, and children in a company of 20 persons if together they pay 20 coins, each man paying 3, each woman 2, and each child $\frac{1}{2}$.

(e) Euler, 1770. Divide 100 into two summands such that one is divisible by 7 and the other by 11.

CHAPTER 3

PRIMES AND THEIR DISTRIBUTION

Mighty are numbers, joined with art resistless.
EURIPIDES

3.1 THE FUNDAMENTAL THEOREM OF ARITHMETIC

Essential to everything discussed herein—in fact, essential to every aspect of number theory—is the notion of a prime number. We have previously observed that any integer $a > 1$ is divisible by ± 1 and $\pm a$; if these exhaust the divisors of a, then it is said to be a prime number. In Definition 3.1 we state this somewhat differently.

> **Definition 3.1.** An integer $p > 1$ is called a *prime number*, or simply a *prime*, if its only positive divisors are 1 and p. An integer greater than 1 that is not a prime is termed *composite*.

Among the first ten positive integers, 2, 3, 5, 7 are primes and 4, 6, 8, 9, 10 are composite numbers. Note that the integer 2 is the only even prime, and according to our definition the integer 1 plays a special role, being neither prime nor composite.

In the rest of this book, the letters p and q will be reserved, so far as is possible, for primes.

Proposition 14 of Book IX of Euclid's *Elements* embodies the result that later became known as the Fundamental Theorem of Arithmetic, namely, that every integer greater than 1 can, except for the order of the factors, be represented as a product of primes in one and only one way. To quote the proposition itself: "If a number be the least that is measured by prime numbers, it will not be measured by any other

prime except those originally measuring it." Because every number $a > 1$ is either a prime or, by the Fundamental Theorem, can be broken down into unique prime factors and no further, the primes serve as the building blocks from which all other integers can be made. Accordingly, the prime numbers have intrigued mathematicians through the ages, and although a number of remarkable theorems relating to their distribution in the sequence of positive integers have been proved, even more remarkable is what remains unproved. The open questions can be counted among the outstanding unsolved problems in all of mathematics.

To begin on a simpler note, we observe that the prime 3 divides the integer 36, where 36 may be written as any one of the products

$$6 \cdot 6 = 9 \cdot 4 = 12 \cdot 3 = 18 \cdot 2$$

In each instance, 3 divides at least one of the factors involved in the product. This is typical of the general situation, the precise result being Theorem 3.1.

Theorem 3.1. If p is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof. If $p \mid a$, then we need go no further, so let us assume that $p \nmid a$. Because the only positive divisors of p are 1 and p itself, this implies that $\gcd(p, a) = 1$. (In general, $\gcd(p, a) = p$ or $\gcd(p, a) = 1$ according as $p \mid a$ or $p \nmid a$.) Hence, citing Euclid's lemma, we get $p \mid b$.

This theorem easily extends to products of more than two terms.

Corollary 1. If p is a prime and $p \mid a_1 a_2 \cdots a_n$, then $p \mid a_k$ for some k, where $1 \leq k \leq n$.

Proof. We proceed by induction on n, the number of factors. When $n = 1$, the stated conclusion obviously holds; whereas when $n = 2$, the result is the content of Theorem 3.1. Suppose, as the induction hypothesis, that $n > 2$ and that whenever p divides a product of less than n factors, it divides at least one of the factors. Now let $p \mid a_1 a_2 \cdots a_n$. From Theorem 3.1, either $p \mid a_n$ or $p \mid a_1 a_2 \cdots a_{n-1}$. If $p \mid a_n$, then we are through. As regards the case where $p \mid a_1 a_2 \cdots a_{n-1}$, the induction hypothesis ensures that $p \mid a_k$ for some choice of k, with $1 \leq k \leq n - 1$. In any event, p divides one of the integers $a_1, a_2, \ldots, a_n$.

Corollary 2. If $p, q_1, q_2, \ldots, q_n$ are all primes and $p \mid q_1 q_2 \cdots q_n$, then $p = q_k$ for some k, where $1 \leq k \leq n$.

Proof. By virtue of Corollary 1, we know that $p \mid q_k$ for some k, with $1 \leq k \leq n$. Being a prime, q_k is not divisible by any positive integer other than 1 or q_k itself. Because $p > 1$, we are forced to conclude that $p = q_k$.

With this preparation out of the way, we arrive at one of the cornerstones of our development, the Fundamental Theorem of Arithmetic. As indicated earlier, this theorem asserts that every integer greater than 1 can be factored into primes in essentially one way; the linguistic ambiguity *essentially* means that $2 \cdot 3 \cdot 2$ is not considered as being a different factorization of 12 from $2 \cdot 2 \cdot 3$. We state this precisely in Theorem 3.2.

Theorem 3.2 Fundamental Theorem of Arithmetic. Every positive integer $n > 1$ can be expressed as a product of primes; this representation is unique, apart from the order in which the factors occur.

Proof. Either n is a prime or it is composite; in the former case, there is nothing more to prove. If n is composite, then there exists an integer d satisfying $d \mid n$ and $1 < d < n$. Among all such integers d, choose p_1 to be the smallest (this is possible by the Well-Ordering Principle). Then p_1 must be a prime number. Otherwise it too would have a divisor q with $1 < q < p_1$; but then $q \mid p_1$ and $p_1 \mid n$ imply that $q \mid n$, which contradicts the choice of p_1 as the smallest positive divisor, not equal to 1, of n.

We therefore may write $n = p_1 n_1$, where p_1 is prime and $1 < n_1 < n$. If n_1 happens to be a prime, then we have our representation. In the contrary case, the argument is repeated to produce a second prime number p_2 such that $n_1 = p_2 n_2$; that is,

$$n = p_1 p_2 n_2 \qquad 1 < n_2 < n_1$$

If n_2 is a prime, then it is not necessary to go further. Otherwise, write $n_2 = p_3 n_3$, with p_3 a prime:

$$n = p_1 p_2 p_3 n_3 \qquad 1 < n_3 < n_2$$

The decreasing sequence

$$n > n_1 > n_2 > \cdots > 1$$

cannot continue indefinitely, so that after a finite number of steps n_{k-1} is a prime, call it, p_k. This leads to the prime factorization

$$n = p_1 p_2 \cdots p_k$$

To establish the second part of the proof—the uniqueness of the prime factorization—let us suppose that the integer n can be represented as a product of primes in two ways; say,

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s \qquad r \leq s$$

where the p_i and q_j are all primes, written in increasing magnitude so that

$$p_1 \leq p_2 \leq \cdots \leq p_r \qquad q_1 \leq q_2 \leq \cdots \leq q_s$$

Because $p_1 \mid q_1 q_2 \cdots q_s$, Corollary 2 of Theorem 3.1 tells us that $p_1 = q_k$ for some k; but then $p_1 \geq q_1$. Similar reasoning gives $q_1 \geq p_1$, whence $p_1 = q_1$. We may cancel this common factor and obtain

$$p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$$

Now repeat the process to get $p_2 = q_2$ and, in turn,

$$p_3 p_4 \cdots p_r = q_3 q_4 \cdots q_s$$

Continue in this fashion. If the inequality $r < s$ were to hold, we would eventually arrive at

$$1 = q_{r+1} q_{r+2} \cdots q_s$$

which is absurd, because each $q_j > 1$. Hence, $r = s$ and

$$p_1 = q_1 \qquad p_2 = q_2, \ldots, p_r = q_r$$

making the two factorizations of n identical. The proof is now complete.

Of course, several of the primes that appear in the factorization of a given positive integer may be repeated, as is the case with $360 = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 5$. By collecting like primes and replacing them by a single factor, we can rephrase Theorem 3.2 as a corollary.

Corollary. Any positive integer $n > 1$ can be written uniquely in a *canonical form*

$$n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

where, for $i = 1, 2, \ldots, r$, each k_i is a positive integer and each p_i is a prime, with $p_1 < p_2 < \cdots < p_r$.

To illustrate, the canonical form of the integer 360 is $360 = 2^3 \cdot 3^2 \cdot 5$. As further examples we cite

$$4725 = 3^3 \cdot 5^2 \cdot 7 \qquad \text{and} \qquad 17460 = 2^3 \cdot 3^2 \cdot 5 \cdot 7^2$$

Theorem 3.2 should not be taken lightly because number systems do exist in which the factorization into "primes" is not unique. Perhaps the most elemental example is the set E of all positive even integers. Let us agree to call an even integer an e-prime if it is not the product of two other even integers. Thus, 2, 6, 10, 14, ... all are e-primes, whereas 4, 8, 12, 16, ... are not. It is not difficult to see that the integer 60 can be factored into e-primes in two distinct ways; namely,

$$60 = 2 \cdot 30 = 6 \cdot 10$$

Part of the difficulty arises from the fact that Theorem 3.1 is lacking in the set E; that is, $6 \mid 2 \cdot 30$, but $6 \nmid 2$ and $6 \nmid 30$.

This is an opportune moment to insert a famous result of Pythagoras. Mathematics as a science began with Pythagoras (569–500 B.C.), and much of the content of Euclid's *Elements* is due to Pythagoras and his School. The Pythagoreans deserve the credit for being the first to classify numbers into odd and even, prime and composite.

Theorem 3.3 Pythagoras. The number $\sqrt{2}$ is irrational.

Proof. Suppose, to the contrary, that $\sqrt{2}$ is a rational number, say, $\sqrt{2} = a/b$, where a and b are both integers with $\gcd(a, b) = 1$. Squaring, we get $a^2 = 2b^2$, so that $b \mid a^2$. If $b > 1$, then the Fundamental Theorem of Arithmetic guarantees the existence of a prime p such that $p \mid b$. It follows that $p \mid a^2$ and, by Theorem 3.1, that $p \mid a$; hence, $\gcd(a, b) \geq p$. We therefore arrive at a contradiction, unless $b = 1$. But if this happens, then $a^2 = 2$, which is impossible (we assume that the reader is willing to grant that no integer can be multiplied by itself to give 2). Our supposition that $\sqrt{2}$ is a rational number is untenable, and so $\sqrt{2}$ must be irrational.

There is an interesting variation on the proof of Theorem 3.3. If $\sqrt{2} = a/b$ with $\gcd(a, b) = 1$, there must exist integers r and s satisfying $ar + bs = 1$. As a result,

$$\sqrt{2} = \sqrt{2}(ar + bs) = (\sqrt{2}a)r + (\sqrt{2}b)s = 2br + as$$

This representation of $\sqrt{2}$ leads us to conclude that $\sqrt{2}$ is an integer, an obvious impossibility.

PROBLEMS 3.1

1. It has been conjectured that there are infinitely many primes of the form $n^2 - 2$. Exhibit five such primes.
2. Give an example to show that the following conjecture is not true: Every positive integer can be written in the form $p + a^2$, where p is either a prime or 1, and $a \geq 0$.
3. Prove each of the assertions below:
   (a) Any prime of the form $3n + 1$ is also of the form $6m + 1$.
   (b) Each integer of the form $3n + 2$ has a prime factor of this form.
   (c) The only prime of the form $n^3 - 1$ is 7.
   [*Hint:* Write $n^3 - 1$ as $(n - 1)(n^2 + n + 1)$.]
   (d) The only prime p for which $3p + 1$ is a perfect square is $p = 5$.
   (e) The only prime of the form $n^2 - 4$ is 5.
4. If $p \geq 5$ is a prime number, show that $p^2 + 2$ is composite.
   [*Hint:* p takes one of the forms $6k + 1$ or $6k + 5$.]
5. (a) Given that p is a prime and $p \mid a^n$, prove that $p^n \mid a^n$.
   (b) If $\gcd(a, b) = p$, a prime, what are the possible values of $\gcd(a^2, b^2)$, $\gcd(a^2, b)$ and $\gcd(a^3, b^2)$?
6. Establish each of the following statements:
   (a) Every integer of the form $n^4 + 4$, with $n > 1$, is composite.
   [*Hint:* Write $n^4 + 4$ as a product of two quadratic factors.]
   (b) If $n > 4$ is composite, then n divides $(n - 1)!$.
   (c) Any integer of the form $8^n + 1$, where $n \geq 1$, is composite.
   [*Hint:* $2^n + 1 \mid 2^{3n} + 1$.]
   (d) Each integer $n > 11$ can be written as the sum of two composite numbers.
   [*Hint:* If n is even, say $n = 2k$, then $n - 6 = 2(k - 3)$; for n odd, consider the integer $n - 9$.]
7. Find all prime numbers that divide 50!.
8. If $p \geq q \geq 5$ and p and q are both primes, prove that $24 \mid p^2 - q^2$.
9. (a) An unanswered question is whether there are infinitely many primes that are 1 more than a power of 2, such as $5 = 2^2 + 1$. Find two more of these primes.
   (b) A more general conjecture is that there exist infinitely many primes of the form $n^2 + 1$; for example, $257 = 16^2 + 1$. Exhibit five more primes of this type.
10. If $p \neq 5$ is an odd prime, prove that either $p^2 - 1$ or $p^2 + 1$ is divisible by 10.
11. Another unproven conjecture is that there are an infinitude of primes that are 1 less than a power of 2, such as $3 = 2^2 - 1$.
   (a) Find four more of these primes.
   (b) If $p = 2^k - 1$ is prime, show that k is an odd integer, except when $k = 2$.
   [*Hint:* $3 \mid 4^n - 1$ for all $n \geq 1$.]
12. Find the prime factorization of the integers 1234, 10140, and 36000.
13. If $n > 1$ is an integer not of the form $6k + 3$, prove that $n^2 + 2^n$ is composite.
   [*Hint:* Show that either 2 or 3 divides $n^2 + 2^n$.]
14. It has been conjectured that every even integer can be written as the difference of two consecutive primes in infinitely many ways. For example,

$$6 = 29 - 23 = 137 - 131 = 599 - 593 = 1019 - 1013 = \cdots$$

   Express the integer 10 as the difference of two consecutive primes in 15 ways.
15. Prove that a positive integer $a > 1$ is a square if and only if in the canonical form of a all the exponents of the primes are even integers.

16. An integer is said to be *square-free* if it is not divisible by the square of any integer greater than 1. Prove the following:
   (a) An integer $n > 1$ is square-free if and only if n can be factored into a product of distinct primes.
   (b) Every integer $n > 1$ is the product of a square-free integer and a perfect square. [*Hint:* If $n = p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$ is the canonical factorization of n, then write $k_i = 2q_i + r_i$ where $r_i = 0$ or 1 according as k_i is even or odd.]

17. Verify that any integer n can be expressed as $n = 2^k m$, where $k \geq 0$ and m is an odd integer.

18. Numerical evidence makes it plausible that there are infinitely many primes p such that $p + 50$ is also prime. List 15 of these primes.

19. A positive integer n is called *square-full*, or *powerful*, if $p^2 \mid n$ for every prime factor p of n (there are 992 square-full numbers less than 250,000). If n is square-full, show that it can be written in the form $n = a^2 b^3$, with a and b positive integers.

3.2 THE SIEVE OF ERATOSTHENES

Given a particular integer, how can we determine whether it is prime or composite and, in the latter case, how can we actually find a nontrivial divisor? The most obvious approach consists of successively dividing the integer in question by each of the numbers preceding it; if none of them (except 1) serves as a divisor, then the integer must be prime. Although this method is very simple to describe, it cannot be regarded as useful in practice. For even if one is undaunted by large calculations, the amount of time and work involved may be prohibitive.

There is a property of composite numbers that allows us to reduce materially the necessary computations—but still the process remains cumbersome. If an integer $a > 1$ is composite, then it may be written as $a = bc$, where $1 < b < a$ and $1 < c < a$. Assuming that $b \leq c$, we get $b^2 \leq bc = a$, and so $b \leq \sqrt{a}$. Because $b > 1$, Theorem 3.2 ensures that b has at least one prime factor p. Then $p \leq b \leq \sqrt{a}$; furthermore, because $p \mid b$ and $b \mid a$, it follows that $p \mid a$. The point is simply this: A composite number a will always possess a prime divisor p satisfying $p \leq \sqrt{a}$.

In testing the primality of a specific integer $a > 1$, it therefore suffices to divide a by those primes not exceeding $\sqrt{a}$ (presuming, of course, the availability of a list of primes up to $\sqrt{a}$). This may be clarified by considering the integer $a = 509$. Inasmuch as $22 < \sqrt{509} < 23$, we need only try out the primes that are not larger than 22 as possible divisors, namely, the primes 2, 3, 5, 7, 11, 13, 17, 19. Dividing 509 by each of these, in turn, we find that none serves as a divisor of 509. The conclusion is that 509 must be a prime number.

Example 3.1. The foregoing technique provides a practical means for determining the canonical form of an integer, say $a = 2093$. Because $45 < \sqrt{2093} < 46$, it is enough to examine the primes 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43. By trial, the first of these to divide 2093 is 7, and $2093 = 7 \cdot 299$. As regards the integer 299, the seven primes that are less than 18 (note that $17 < \sqrt{299} < 18$) are 2, 3, 5, 7, 11, 13, 17. The first prime divisor of 299 is 13 and, carrying out the required division, we obtain $299 = 13 \cdot 23$. But 23 is itself a prime, whence 2093 has exactly three prime factors, 7, 13, and 23:

$$2093 = 7 \cdot 13 \cdot 23$$

Another Greek mathematician whose work in number theory remains significant is Eratosthenes of Cyrene (276–194 B.C.). Although posterity remembers him mainly as the director of the world-famous library at Alexandria, Eratosthenes was gifted in all branches of learning, if not of first rank in any; in his own day, he was nicknamed "Beta" because, it was said, he stood at least second in every field. Perhaps the most impressive feat of Eratosthenes was the accurate measurement of the earth's circumference by a simple application of Euclidean geometry.

We have seen that if an integer $a > 1$ is not divisible by any prime $p \leq \sqrt{a}$, then a is of necessity a prime. Eratosthenes used this fact as the basis of a clever technique, called the *Sieve of Eratosthenes*, for finding all primes below a given integer n. The scheme calls for writing down the integers from 2 to n in their natural order and then systematically eliminating all the composite numbers by striking out all multiples $2p, 3p, 4p, 5p, \ldots$ of the primes $p \leq \sqrt{n}$. The integers that are left on the list—those that do not fall through the "sieve"—are primes.

To see an example of how this works, suppose that we wish to find all primes not exceeding 100. Consider the sequence of consecutive integers $2, 3, 4, \ldots, 100$. Recognizing that 2 is a prime, we begin by crossing out all even integers from our listing, except 2 itself. The first of the remaining integers is 3, which must be a prime. We keep 3, but strike out all higher multiples of 3, so that 9, 15, 21, ... are now removed (the even multiples of 3 having been removed in the previous step). The smallest integer after 3 that has not yet been deleted is 5. It is not divisible by either 2 or 3—otherwise it would have been crossed out—hence, it is also a prime. All proper multiples of 5 being composite numbers, we next remove 10, 15, 20, ... (some of these are, of course, already missing), while retaining 5 itself. The first surviving integer 7 is a prime, for it is not divisible by 2, 3, or 5, the only primes that precede it. After eliminating the proper multiples of 7, the largest prime less than $\sqrt{100} = 10$, all composite integers in the sequence $2, 3, 4, \ldots, 100$ have fallen through the sieve. The positive integers that remain, to wit, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, are all of the primes less than 100.

The following table represents the result of the completed sieve. The multiples of 2 are crossed out by \; the multiples of 3 are crossed out by /; the multiples of 5 are crossed out by —; the multiples of 7 are crossed out by ~.

	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	83	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~

By this point, an obvious question must have occurred to the reader. Is there a largest prime number, or do the primes go on forever? The answer is to be found in a remarkably simple proof given by Euclid in Book IX of his *Elements*. Euclid's argument is universally regarded as a model of mathematical elegance. Loosely

speaking, it goes like this: Given any finite list of prime numbers, one can always find a prime not on the list; hence, the number of primes is infinite. The actual details appear below.

Theorem 3.4 Euclid. There is an infinite number of primes.

Proof. Euclid's proof is by contradiction. Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7, \ldots$ be the primes in ascending order, and suppose that there is a last prime, called p_n. Now consider the positive integer

$$P = p_1 p_2 \cdots p_n + 1$$

Because $P > 1$, we may put Theorem 3.2 to work once again and conclude that P is divisible by some prime p. But $p_1, p_2, \ldots, p_n$ are the only prime numbers, so that p must be equal to one of $p_1, p_2, \ldots, p_n$. Combining the divisibility relation $p \mid p_1 p_2 \cdots p_n$ with $p \mid P$, we arrive at $p \mid P - p_1 p_2 \cdots p_n$ or, equivalently, $p \mid 1$. The only positive divisor of the integer 1 is 1 itself and, because $p > 1$, a contradiction arises. Thus, no finite list of primes is complete, whence the number of primes is infinite.

For a prime p, define $p^{\#}$ to be the product of all primes that are less than or equal to p. Numbers of the form $p^{\#} + 1$ might be termed *Euclidean numbers*, because they appear in Euclid's scheme for proving the infinitude of primes. It is interesting to note that in forming these integers, the first five, namely,

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

are all prime numbers. However,

$$\begin{aligned} 13^{\#} + 1 &= 59 \cdot 509 \\ 17^{\#} + 1 &= 19 \cdot 97 \cdot 277 \\ 19^{\#} + 1 &= 347 \cdot 27953 \end{aligned}$$

are not prime. A question whose answer is not known is whether there are infinitely many primes p for which $p^{\#} + 1$ is also prime. For that matter, are there infinitely many composite $p^{\#} + 1$?

At present, 19 primes of the form $p^{\#} + 1$ have been identified. These correspond to the values $p = 2, 3, 5, 7, 11, 31, 379, 1019, 1021, 2657, 3229, 4547, 4787, 11549, 13649, 18523, 23801, 24029$, and 42209; the largest of these, a number consisting of 18241 digits, was discovered in 2000. The integer $p^{\#} + 1$ is composite for all other $p \leq 120000$.

Euclid's theorem is too important for us to be content with a single proof. Here is a variation in the reasoning: Form the infinite sequence of positive integers

$$\begin{aligned} n_1 &= 2 \\ n_2 &= n_1 + 1 \\ n_3 &= n_1 n_2 + 1 \\ n_4 &= n_1 n_2 n_3 + 1 \\ &\vdots \\ n_k &= n_1 n_2 \cdots n_{k-1} + 1 \\ &\vdots \end{aligned}$$

Because each $n_k > 1$, each of these integers is divisible by a prime. But no two n_k can have the same prime divisor. To see this, let $d = \gcd(n_i, n_k)$ and suppose that $i < k$. Then d divides n_i and, hence, must divide $n_1 n_2 \cdots n_{k-1}$. Because $d \mid n_k$, Theorem 2.2 (g) tells us that $d \mid n_k - n_1 n_2 \cdots n_{k-1}$ or $d \mid 1$. The implication is that $d = 1$, and so the integers $n_k (k = 1, 2, \ldots)$ are pairwise relatively prime. The point we wish to make is that there are as many distinct primes as there are integers n_k, namely, infinitely many of them.

Let p_n denote the nth of the prime numbers in their natural order. Euclid's proof shows that the expression $p_1 p_2 \cdots p_n + 1$ is divisible by at least one prime. If there are several such prime divisors, then p_{n+1} cannot exceed the smallest of these so that $p_{n+1} \leq p_1 p_2 \cdots p_n + 1$ for $n \geq 1$. Another way of saying the same thing is that

$$p_n \leq p_1 p_2 \cdots p_{n-1} + 1 \qquad n \geq 2$$

With a slight modification of Euclid's reasoning, this inequality can be improved to give

$$p_n \leq p_1 p_2 \cdots p_{n-1} - 1 \qquad n \geq 3$$

For instance, when $n = 5$, this tells us that

$$11 = p_5 \leq 2 \cdot 3 \cdot 5 \cdot 7 - 1 = 209$$

We can see that the estimate is rather extravagant. A sharper limitation on the size of p_n is given by *Bonse's inequality*, which states that

$$p_n^2 < p_1 p_2 \cdots p_{n-1} \qquad n \geq 5$$

This inequality yields $p_5^2 < 210$, or $p_5 \leq 14$. A somewhat better size-estimate for p_5 comes from the inequality

$$p_{2n} \leq p_2 p_3 \cdots p_n - 2 \qquad n \geq 3$$

Here, we obtain

$$p_5 < p_6 \leq p_2 p_3 - 2 = 3 \cdot 5 - 2 = 13$$

To approximate the size of p_n from these formulas, it is necessary to know the values of $p_1, p_2, \ldots, p_{n-1}$. For a bound in which the preceding primes do not enter the picture, we have the following theorem.

Theorem 3.5. If p_n is the nth prime number, then $p_n \leq 2^{2^{n-1}}$.

Proof. Let us proceed by induction on n, the asserted inequality being clearly true when $n = 1$. As the hypothesis of the induction, we assume that $n > 1$ and that the result holds for all integers up to n. Then

$$\begin{aligned} p_{n+1} &\leq p_1 p_2 \cdots p_n + 1 \\ &\leq 2 \cdot 2^2 \cdots 2^{2^{n-1}} + 1 = 2^{1+2+2^2+\cdots+2^{n-1}} + 1 \end{aligned}$$

Recalling the identity $1 + 2 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$, we obtain

$$p_{n+1} \leq 2^{2^n-1} + 1$$

However, $1 \leq 2^{2^n-1}$ for all n; whence

$$\begin{aligned} p_{n+1} &\leq 2^{2^n-1} + 2^{2^n-1} \\ &= 2 \cdot 2^{2^n-1} = 2^{2^n} \end{aligned}$$

completing the induction step, and the argument.

There is a corollary to Theorem 3.5 that is of interest.

Corollary. For $n \geq 1$, there are at least $n + 1$ primes less than 2^{2^n}.

Proof. From the theorem, we know that $p_1, p_2, \ldots, p_{n+1}$ are all less than 2^{2^n}.

We can do considerably better than is indicated by Theorem 3.5. In 1845, Joseph Bertrand conjectured that the prime numbers are well-distributed in the sense that between $n \geq 2$ and $2n$ there is at least one prime. He was unable to establish his conjecture, but verified it for all $n \leq 3{,}000{,}000$. (One way of achieving this is to consider a sequence of primes 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259, 2503, 5003, 9973, 19937, 39869, 79699, 159389, ... each of which is less than twice the preceding.) Because it takes some real effort to substantiate this famous conjecture, let us content ourselves with saying that the first proof was carried out by the Russian mathematician P. L. Tchebycheff in 1852. Granting the result, it is not difficult to show that

$$p_n < 2^n \qquad n \geq 2$$

and as a direct consequence, $p_{n+1} < 2p_n$ for $n \geq 2$. In particular,

$$11 = p_5 < 2 \cdot p_4 = 14$$

To see that $p_n < 2^n$, we argue by induction on n. Clearly, $p_2 = 3 < 2^2$, so that the inequality is true here. Now assume that the inequality holds for an integer n, whence $p_n < 2^n$. Invoking Bertrand's conjecture, there exists a prime number p satisfying $2^n < p < 2^{n+1}$; that is, $p_n < p$. This immediately leads to the conclusion that $p_{n+1} \leq p < 2^{n+1}$, which completes the induction and the proof.

Primes of special form have been of perennial interest. Among these, the repunit primes are outstanding in their simplicity. A *repunit* is an integer written (in decimal notation) as a string of 1's, such as 11, 111, or 1111. Each such integer must have the form $(10^n - 1)/9$. We use the symbol R_n to denote the repunit consisting of n consecutive 1's. A peculiar feature of these numbers is the apparent scarcity of primes among them. So far, only R_2, R_{19}, R_{23}, R_{317}, R_{1031}, R_{49081}, and R_{86453}

have been identified as primes (the last one in 2001). It is known that the only possible repunit primes R_n for all $n \leq 45000$ are the seven numbers just indicated. No conjecture has been made as to the existence of any others. For a repunit R_n to be prime, the subscript n must be a prime; that this is not a sufficient condition is shown by

$$R_5 = 11111 = 41 \cdot 271 \qquad R_7 = 1111111 = 239 \cdot 4649$$

PROBLEMS 3.2

1. Determine whether the integer 701 is prime by testing all primes $p \leq \sqrt{701}$ as possible divisors. Do the same for the integer 1009.
2. Employing the Sieve of Eratosthenes, obtain all the primes between 100 and 200.
3. Given that $p \nmid n$ for all primes $p \leq \sqrt[3]{n}$, show that $n > 1$ is either a prime or the product of two primes.
   [*Hint:* Assume to the contrary that n contains at least three prime factors.]
4. Establish the following facts:
   (a) $\sqrt{p}$ is irrational for any prime p.
   (b) If $a > 0$ and $\sqrt[n]{a}$ is rational, then $\sqrt[n]{a}$ must be an integer.
   (c) For $n \geq 2$, $\sqrt[n]{n}$ is irrational.
   [*Hint:* Use the fact that $2^n > n$.]
5. Show that any composite three-digit number must have a prime factor less than or equal to 31.
6. Fill in any missing details in this sketch of a proof of the infinitude of primes: Assume that there are only finitely many primes, say $p_1, p_2, \ldots, p_n$. Let A be the product of any r of these primes and put $B = p_1 p_2 \cdots p_n / A$. Then each p_k divides either A or B, but not both. Because $A + B > 1$, $A + B$ has a prime divisor different from any of the p_k, which is a contradiction.
7. Modify Euclid's proof that there are infinitely many primes by assuming the existence of a largest prime p and using the integer $N = p! + 1$ to arrive at a contradiction.
8. Give another proof of the infinitude of primes by assuming that there are only finitely many primes, say $p_1, p_2, \ldots, p_n$, and using the following integer to arrive at a contradiction:

$$N = p_2 p_3 \cdots p_n + p_1 p_3 \cdots p_n + \cdots + p_1 p_2 \cdots p_{n-1}$$

9. (a) Prove that if $n > 2$, then there exists a prime p satisfying $n < p < n!$.
   [*Hint:* If $n! - 1$ is not prime, then it has a prime divisor p; and $p \leq n$ implies $p \mid n!$, leading to a contradiction.]
   (b) For $n > 1$, show that every prime divisor of $n! + 1$ is an odd integer that is greater than n.
10. Let q_n be the smallest prime that is strictly greater than $P_n = p_1 p_2 \cdots p_n + 1$. It has been conjectured that the difference $q_n - (p_1 p_2 \cdots p_n)$ is always a prime. Confirm this for the first five values of n.
11. If p_n denotes the nth prime number, put $d_n = p_{n+1} - p_n$. An open question is whether the equation $d_n = d_{n+1}$ has infinitely many solutions. Give five solutions.
12. Assuming that p_n is the nth prime number, establish each of the following statements:
   (a) $p_n > 2n - 1$ for $n \geq 5$.
   (b) None of the integers $P_n = p_1 p_2 \cdots p_n + 1$ is a perfect square.
   [*Hint:* Each P_n is of the form $4k + 3$ for $n > 1$.]

(c) The sum

$$\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n}$$

is never an integer.

13. For the repunits R_n, verify the assertions below:

(a) If $n \mid m$, then $R_n \mid R_m$.

[*Hint:* If $m = kn$, consider the identity

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

(b) If $d \mid R_n$ and $d \mid R_m$, then $d \mid R_{n+m}$.

[*Hint:* Show that $R_{m+n} = R_n 10^m + R_m$.]

(c) If $\gcd(n, m) = 1$, then $\gcd(R_n, R_m) = 1$.

14. Use the previous problem to obtain the prime factors of the repunit R_{10}.

3.3 THE GOLDBACH CONJECTURE

Although there is an infinitude of primes, their distribution within the positive integers is most mystifying. Repeatedly in their distribution we find hints or, as it were, shadows of a pattern; yet an actual pattern amenable to precise description remains elusive. The difference between consecutive primes can be small, as with the pairs 11 and 13, 17 and 19, or for that matter 1000000000061 and 1000000000063. At the same time there exist arbitrarily long intervals in the sequence of integers that are totally devoid of any primes.

It is an unanswered question whether there are infinitely many pairs of *twin primes*; that is, pairs of successive odd integers p and $p + 2$ that are both primes. Numerical evidence leads us to suspect an affirmative conclusion. Electronic computers have discovered 152892 pairs of twin primes less than 30000000 and 20 pairs between 10^{12} and $10^{12}+$ 10000, which hints at their growing scarcity as the positive integers increase in magnitude. Many examples of immense twins are known. The largest twins to date, each 51090 digits long,

$$33218925 \cdot 2^{169690} \pm 1$$

were discovered in 2002.

Consecutive primes cannot only be close together, but also can be far apart; that is, arbitrarily large gaps can occur between consecutive primes. Stated precisely: Given any positive integer n, there exist n consecutive integers, all of which are composite. To prove this, we simply need to consider the integers

$$(n + 1)! + 2, (n + 1)! + 3, \ldots, (n + 1)! + (n + 1)$$

where $(n + 1)! = (n + 1) \cdot n \cdots 3 \cdot 2 \cdot 1$. Clearly, there are n integers listed and they are consecutive. What is important is that each integer is composite. Indeed, $(n + 1)! + 2$ is divisible by 2, $(n + 1)! + 3$ is divisible by 3, and so on.

For instance, if a sequence of four consecutive composite integers is desired, then the previous argument produces 122, 123, 124, and 125:

$$5! + 2 = 122 = 2 \cdot 61$$
$$5! + 3 = 123 = 3 \cdot 41$$
$$5! + 4 = 124 = 4 \cdot 31$$
$$5! + 5 = 125 = 5 \cdot 25$$

Of course, we can find other sets of four consecutive composites, such as 24, 25, 26, 27 or 32, 33, 34, 35.

As this example suggests, our procedure for constructing gaps between two consecutive primes gives a gross overestimate of where they occur among the integers. The first occurrences of prime gaps of specific lengths, where all the intervening integers are composite, have been the subject of computer searches. For instance, there is a gap of length 778 (that is, $p_{n+1} - p_n = 778$) following the prime 42842283925351. No gap of this size exists between two smaller primes. The largest effectively calculated gap between consecutive prime numbers has length 1132, with a string of 1131 composites immediately after the prime

$$1693182318746371$$

Interestingly, computer researchers have not identified gaps of every possible width up to 1132. The smallest missing gap size is 796. The conjecture is that there is a prime gap (a string of $2k - 1$ consecutive composites between two primes) for every even integer $2k$.

This brings us to another unsolved problem concerning the primes, the Goldbach conjecture. In a letter to Leonhard Euler in the year 1742, Christian Goldbach hazarded the guess that every even integer is the sum of two numbers that are either primes or 1. A somewhat more general formulation is that every even integer greater than 4 can be written as a sum of two odd prime numbers. This is easy to confirm for the first few even integers:

$$\begin{aligned}
2 &= 1 + 1 \\
4 &= 2 + 2 = 1 + 3 \\
6 &= 3 + 3 = 1 + 5 \\
8 &= 3 + 5 = 1 + 7 \\
10 &= 3 + 7 = 5 + 5 \\
12 &= 5 + 7 = 1 + 11 \\
14 &= 3 + 11 = 7 + 7 = 1 + 13 \\
16 &= 3 + 13 = 5 + 11 \\
18 &= 5 + 13 = 7 + 11 = 1 + 17 \\
20 &= 3 + 17 = 7 + 13 = 1 + 19 \\
22 &= 3 + 19 = 5 + 17 = 11 + 11 \\
24 &= 5 + 19 = 7 + 17 = 11 + 13 = 1 + 23 \\
26 &= 3 + 23 = 7 + 19 = 13 + 13 \\
28 &= 5 + 23 = 11 + 17 \\
30 &= 7 + 23 = 11 + 19 = 13 + 17 = 1 + 29
\end{aligned}$$

Although it seems that Euler never tried to prove the result, upon writing to Goldbach at a later date, Euler countered with a conjecture of his own: Any even integer (≥ 6) of the form $4n + 2$ is a sum of two numbers each being either a prime of the form $4n + 1$ or 1.

The numerical data suggesting the truth of Goldbach's conjecture are overwhelming. It has been verified by computers for all even integers less than $4 \cdot 10^{14}$.

As the integers become larger, the number of different ways in which $2n$ can be expressed as the sum of two primes increases. For example, there are 219400 such representations for the even integer 100000000. Although this supports the feeling that Goldbach was correct in his conjecture, it is far from a mathematical proof, and all attempts to obtain a proof have been completely unsuccessful. One of the most famous number theorists of the last century, G. H. Hardy, in his address to the Mathematical Society of Copenhagen in 1921, stated that the Goldbach conjecture appeared "... probably as difficult as any of the unsolved problems in mathematics." It is currently known that every even integer is the sum of six or fewer primes.

We remark that if the conjecture of Goldbach is true, then each odd number larger than 7 must be the sum of three odd primes. To see this, take n to be an odd integer greater than 7, so that $n - 3$ is even and greater than 4; if $n - 3$ could be expressed as the sum of two odd primes, then n would be the sum of three.

The first real progress on the conjecture in nearly 200 years was made by Hardy and Littlewood in 1922. On the basis of a certain unproved hypothesis, the so-called generalized Riemann hypothesis, they showed that every sufficiently large odd number is the sum of three odd primes. In 1937, the Russian mathematician I. M. Vinogradov was able to remove the dependence on the generalized Riemann hypothesis, thereby giving an unconditional proof of this result; that is to say, he established that all odd integers greater than some effectively computable n_0 can be written as the sum of three odd primes.

$$n = p_1 + p_2 + p_3 \qquad (n \text{ odd}, n \text{ sufficiently large})$$

Vinogradov was unable to decide how large n_0 should be, but Borozdkin (1956) proved that $n_0 < 3^{3^{15}}$. In 2002, the bound on n_0 was reduced to 10^{1346}. It follows immediately that every even integer from some point on is the sum of either two or four primes. Thus, it is enough to answer the question for every odd integer n in the range $9 \leq n \leq n_0$, which, for a given integer, becomes a matter of tedious computation (unfortunately, n_0 is so large that this exceeds the capabilities of the most modern electronic computers).

Because of the strong evidence in favor of the famous Goldbach conjecture, we readily become convinced that it is true. Nevertheless, it might be false. Vinogradov showed that if $A(x)$ is the number of even integers $n \leq x$ that are not the sum of two primes, then

$$\lim_{x \to \infty} A(x)/x = 0$$

This allows us to say that "almost all" even integers satisfy the conjecture. As Edmund Landau so aptly put it, "The Goldbach conjecture is false for at most 0% of all even integers; this *at most* 0% does not exclude, of course, the possibility that there are infinitely many exceptions."

Having digressed somewhat, let us observe that according to the Division Algorithm, every positive integer can be written uniquely in one of the forms

$$4n \qquad 4n + 1 \qquad 4n + 2 \qquad 4n + 3$$

for some suitable $n \geq 0$. Clearly, the integers $4n$ and $4n + 2 = 2(2n + 1)$ are both even. Thus, all odd integers fall into two progressions: one containing integers of the form $4n + 1$, and the other containing integers of the form $4n + 3$.

The question arises as to how these two types of primes are distributed within the set of positive integers. Let us display the first few odd prime numbers in consecutive order, putting the $4n + 3$ primes in the top row and the $4n + 1$ primes under them:

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

At this point, one might have the general impression that primes of the form $4n + 3$ are more abundant than are those of the form $4n + 1$. To obtain more precise information, we require the help of the function $\pi_{a,b}(x)$, which counts the number of primes of the form $p = an + b$ not exceeding x. Our small table, for instance, indicates that $\pi_{4,1}(89) = 10$ and $\pi_{4,3}(89) = 13$.

In a famous letter written in 1853, Tchebycheff remarked that $\pi_{4,1}(x) \leq \pi_{4,3}(x)$ for small values of x. He also implied that he had a proof that the inequality always held. In 1914, J. E. Littlewood showed that the inequality fails infinitely often, but his method gave no indication of the value of x for which this first happens. It turned out to be quite difficult to find. Not until 1957 did a computer search reveal that $x = 26861$ is the smallest prime for which $\pi_{4,1}(x) > \pi_{4,3}(x)$; here, $\pi_{4,1}(x) = 1473$ and $\pi_{4,3}(x) = 1472$. This is an isolated situation, because the next prime at which a reversal occurs is $x = 616{,}841$. Remarkably, $\pi_{4,1}(x) > \pi_{4,3}(x)$ for the 410 million successive integers x lying between 18540000000 and 18950000000.

The behavior of primes of the form $3n \pm 1$ provided more of a computational challenge: the inequality $\pi_{3,1}(x) \leq \pi_{3,2}(x)$ holds for all x until one reaches $x = 608981813029$.

This furnishes a pleasant opportunity for a repeat performance of Euclid's method for proving the existence of an infinitude of primes. A slight modification of his argument reveals that there is an infinite number of primes of the form $4n + 3$. We approach the proof through a simple lemma.

Lemma. The product of two or more integers of the form $4n + 1$ is of the same form.

Proof. It is sufficient to consider the product of just two integers. Let us take $k = 4n + 1$ and $k' = 4m + 1$. Multiplying these together, we obtain

$$\begin{aligned} kk' &= (4n + 1)(4m + 1) \\ &= 16nm + 4n + 4m + 1 = 4(4nm + n + m) + 1 \end{aligned}$$

which is of the desired form.

This paves the way for Theorem 3.6.

Theorem 3.6. There are an infinite number of primes of the form $4n + 3$.

Proof. In anticipation of a contradiction, let us assume that there exist only finitely many primes of the form $4n + 3$; call them $q_1, q_2, \ldots, q_s$. Consider the positive integer

$$N = 4q_1q_2 \cdots q_s - 1 = 4(q_1q_2 \cdots q_s - 1) + 3$$

and let $N = r_1r_2 \cdots r_t$ be its prime factorization. Because N is an odd integer, we have $r_k \neq 2$ for all k, so that each r_k is either of the form $4n + 1$ or $4n + 3$. By the lemma, the product of any number of primes of the form $4n + 1$ is again an integer of this type. For N to take the form $4n + 3$, as it clearly does, N must contain at least one prime factor r_i of the form $4n + 3$. But r_i cannot be found among the listing $q_1, q_2, \ldots, q_s$, for this would lead to the contradiction that $r_i \mid 1$. The only possible conclusion is that there are infinitely many primes of the form $4n + 3$.

Having just seen that there are infinitely many primes of the form $4n + 3$, we might reasonably ask: Is the number of primes of the form $4n + 1$ also infinite? This answer is likewise in the affirmative, but a demonstration must await the development of the necessary mathematical machinery. Both these results are special cases of a remarkable theorem by P. G. L. Dirichlet on primes in arithmetic progressions, established in 1837. The proof is much too difficult for inclusion here, so that we must content ourselves with the mere statement.

Theorem 3.7 Dirichlet. If a and b are relatively prime positive integers, then the arithmetic progression

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

contains infinitely many primes.

Dirichlet's theorem tells us, for instance, that there are infinitely many prime numbers ending in 999, such as 1999, 100999, 1000999, ... for these appear in the arithmetic progression determined by $1000n + 999$, where $\gcd(1000, 999) = 1$.

There is no arithmetic progression $a, a + b, a + 2b, \ldots$ that consists solely of prime numbers. To see this, suppose that $a + nb = p$, where p is a prime. If we put $n_k = n + kp$ for $k = 1, 2, 3, \ldots$ then the n_kth term in the progression is

$$a + n_kb = a + (n + kp)b = (a + nb) + kpb = p + kpb$$

Because each term on the right-hand side is divisible by p, so is $a + n_kb$. In other words, the progression must contain infinitely many composite numbers.

It is an old, but still unsolved question of whether there exist arbitrarily long but finite arithmetic progressions consisting only of prime numbers (not necessarily consecutive primes). The longest progression found to date is composed of the 22 primes:

$$11410337850553 + 4609098694200n \qquad 0 \leq n \leq 21$$

The prime factorization of the common difference between the terms is

$$2^3 \cdot 3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 1033$$

which is divisible by 9699690, the product of the primes less than 22. This takes place according to Theorem 3.8.

Theorem 3.8. If all the $n > 2$ terms of the arithmetic progression

$$p, p + d, p + 2d, \ldots, p + (n - 1)d$$

are prime numbers, then the common difference d is divisible by every prime $q < n$.

Proof. Consider a prime number $q < n$ and assume to the contrary that $q \nmid d$. We claim that the first q terms of the progression

$$p, p + d, p + 2d, \ldots, p + (q - 1)d \tag{1}$$

will leave different remainders when divided by q. Otherwise there exist integers j and k, with $0 \leq j < k \leq q - 1$, such that the numbers $p + jd$ and $p + kd$ yield the same remainder upon division by q. Then q divides their difference $(k - j)d$. But $\gcd(q , d) = 1$, and so Euclid's lemma leads to $q \mid k - j$, which is nonsense in light of the inequality $k - j \leq q - 1$.

Because the q different remainders produced from Eq. (1) are drawn from the q integers $0, 1, \ldots, q - 1$, one of these remainders must be zero. This means that $q \mid p + td$ for some t satisfying $0 \leq t \leq q - 1$. Because of the inequality $q < n \leq p \leq p + td$, we are forced to conclude that $p + td$ is composite. (If p were less than n, one of the terms of the progression would be $p + pd = p(1 + d)$.) With this contradiction, the proof that $q \mid d$ is complete.

It has been conjectured that there exist arithmetic progressions of finite (but otherwise arbitrary) length, composed of consecutive prime numbers. Examples of such progressions consisting of three and four primes, respectively, are 47, 53, 59, and 251, 257, 263, 269.

Most recently a sequence of 10 consecutive primes was discovered in which each term exceeds its predecessor by just 210; the smallest of these primes has 93 digits. Finding an arithmetic progression consisting of 11 consecutive primes is likely to be out of reach for some time. Absent the restriction that the primes involved be consecutive, strings of 11-term arithmetic progressions are easily located. One such is

$$110437 + 13860n \qquad 0 \leq n \leq 10$$

In the interest of completeness, we might mention another famous problem that, so far, has resisted the most determined attack. For centuries, mathematicians have sought a simple formula that would yield every prime number or, failing this, a formula that would produce nothing but primes. At first glance, the request seems modest enough: Find a function $f(n)$ whose domain is, say, the nonnegative integers and whose range is some infinite subset of the set of all primes. It was widely believed years ago that the quadratic polynomial

$$f(n) = n^2 + n + 41$$

assumed only prime values. This was shown to be false by Euler, in 1772. As evidenced by the following table, the claim is a correct one for $n = 0, 1, 2, \ldots, 39$.

n	$f(n)$	n	$f(n)$	n	$f(n)$
0	41	14	251	28	853
1	43	15	281	29	911
2	47	16	313	30	971
3	53	17	347	31	1033
4	61	18	383	32	1097
5	71	19	421	33	1163
6	83	20	461	34	1231
7	97	21	503	35	1301
8	113	22	547	36	1373
9	131	23	593	37	1447
10	151	24	641	38	1523
11	173	25	691	39	1601
12	197	26	743		
13	223	27	797		

However, this provocative conjecture is shattered in the cases $n = 40$ and $n = 41$, where there is a factor of 41:

$$f(40) = 40 \cdot 41 + 41 = 41^2$$

and

$$f(41) = 41 \cdot 42 + 41 = 41 \cdot 43$$

The next value $f(42) = 1847$ turns out to be prime once again. In fact, for the first 100 integer values of n, the so-called Euler polynomial represents 86 primes. Although it starts off very well in the production of primes, there are other quadratics such as

$$g(n) = n^2 + n + 27941$$

that begin to best $f(n)$ as the values of n become larger. For example, $g(n)$ is prime for 286129 values of $0 \leq n \leq 10^6$, whereas its famous rival yields 261081 primes in this range.

It has been shown that no polynomial of the form $n^2 + n + q$, with q a prime, can do better than the Euler polynomial in giving primes for successive values of n. Indeed, until fairly recently no other quadratic polynomial of any kind was known to produce more than 40 successive prime values. The polynomial

$$h(n) = 103n^2 - 3945n + 34381$$

found in 1988, produces 43 distinct prime values for $n = 0, 1, 2, \ldots, 42$. The current record holder in this regard

$$k(n) = 36n^2 - 810n + 2753$$

does slightly better by giving a string of 45 prime values.

The failure of the previous functions to be prime-producing is no accident, for it is easy to prove that there is no nonconstant polynomial $f(n)$ with integral

coefficients that takes on just prime values for integral n. We assume that such a polynomial $f(n)$ actually does exist and argue until a contradiction is reached. Let

$$f(n) = a_k n^k + a_{k-1} n^{k-1} + \cdots + a_2 n^2 + a_1 n + a_0$$

where all the coefficients $a_0, a_1, \ldots, a_k$ are integers, and $a_k \neq 0$. For a fixed value of (n_0), $p = f(n_0)$ is a prime number. Now, for any integer t, we consider the following expression:

$$\begin{aligned} f(n_0 + tp) &= a_k(n_0 + tp)^k + \cdots + a_1(n_0 + tp) + a_0 \\ &= (a_k n_0^k + \cdots + a_1 n_0 + a_0) + pQ(t) \\ &= f(n_0) + pQ(t) \\ &= p + pQ(t) = p(1 + Q(t)) \end{aligned}$$

where $Q(t)$ is a polynomial in t having integral coefficients. Our reasoning shows that $p \mid f(n_0 + tp)$; hence, from our own assumption that $f(n)$ takes on only prime values, $f(n_0 + tp) = p$ for any integer t. Because a polynomial of degree k cannot assume the same value more than k times, we have obtained the required contradiction.

Recent years have seen a measure of success in the search for prime-producing functions. W. H. Mills proved (1947) that there exists a positive real number r such that the expression $f(n) = [r^{3^n}]$ is prime for $n = 1, 2, 3, \ldots$ (the brackets indicate the greatest integer function). Needless to say, this is strictly an existence theorem and nothing is known about the actual value of r. Mills's function does not produce all the primes.

PROBLEMS 3.3

1. Verify that the integers 1949 and 1951 are twin primes.
2. (a) If 1 is added to a product of twin primes, prove that a perfect square is always obtained.
   (b) Show that the sum of twin primes p and $p + 2$ is divisible by 12, provided that $p > 3$.
3. Find all pairs of primes p and q satisfying $p - q = 3$.
4. Sylvester (1896) rephrased the Goldbach conjecture: Every even integer $2n$ greater than 4 is the sum of two primes, one larger than $n/2$ and the other less than $3n/2$. Verify this version of the conjecture for all even integers between 6 and 76.
5. In 1752, Goldbach submitted the following conjecture to Euler: Every odd integer can be written in the form $p + 2a^2$, where p is either a prime or 1 and $a \geq 0$. Show that the integer 5777 refutes this conjecture.
6. Prove that the Goldbach conjecture that every even integer greater than 2 is the sum of two primes is equivalent to the statement that every integer greater than 5 is the sum of three primes.
   [*Hint:* If $2n - 2 = p_1 + p_2$, then $2n = p_1 + p_2 + 2$ and $2n + 1 = p_1 + p_2 + 3$.]
7. A conjecture of Lagrange (1775) asserts that every odd integer greater than 5 can be written as a sum $p_1 + 2p_2$, where p_1, p_2 are both primes. Confirm this for all odd integers through 75.
8. Given a positive integer n, it can be shown that there exists an even integer a that is representable as the sum of two odd primes in n different ways. Confirm that the integers

60, 78, and 84 can be written as the sum of two primes in six, seven, and eight ways, respectively.

9. (a) For $n > 3$, show that the integers $n, n+2, n+4$ cannot all be prime.
(b) Three integers $p, p+2, p+6$, which are all prime, are called a *prime-triplet*. Find five sets of prime-triplets.

10. Establish that the sequence

$$(n+1)!-2, (n+1)!-3, \ldots, (n+1)!-(n+1)$$

produces n consecutive composite integers for $n > 2$.

11. Find the smallest positive integer n for which the function $f(n) = n^2 + n + 17$ is composite. Do the same for the functions $g(n) = n^2 + 21n + 1$ and $h(n) = 3n^2 + 3n + 23$.

12. Let p_n denote the nth prime number. For $n \geq 3$, prove that $p_{n+3}^2 < p_n p_{n+1} p_{n+2}$.
[*Hint:* Note that $p_{n+3}^2 < 4p_{n+2}^2 < 8p_{n+1}p_{n+2}$.]

13. Apply the same method of proof as in Theorem 3.6 to show that there are infinitely many primes of the form $6n + 5$.

14. Find a prime divisor of the integer $N = 4(3 \cdot 7 \cdot 11) - 1$ of the form $4n + 3$. Do the same for $N = 4(3 \cdot 7 \cdot 11 \cdot 15) - 1$.

15. Another unanswered question is whether there exist an infinite number of sets of five consecutive odd integers of which four are primes. Find five such sets of integers.

16. Let the sequence of primes, with 1 adjoined, be denoted by $p_0 = 1$, $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$. For each $n \geq 1$, it is known that there exists a suitable choice of coefficients $\epsilon_k = \pm 1$ such that

$$p_{2n} = p_{2n-1} + \sum_{k=0}^{2n-2} \epsilon_k p_k \qquad p_{2n+1} = 2p_{2n} + \sum_{k=0}^{2n-1} \epsilon_k p_k$$

To illustrate:

$$13 = 1 + 2 - 3 - 5 + 7 + 11$$

and

$$17 = 1 + 2 - 3 - 5 + 7 - 11 + 2 \cdot 13$$

Determine similar representations for the primes 23, 29, 31, and 37.

17. In 1848, de Polignac claimed that every odd integer is the sum of a prime and a power of 2. For example, $55 = 47 + 2^3 = 23 + 2^5$. Show that the integers 509 and 877 discredit this claim.

18. (a) If p is a prime and $p \nmid b$, prove that in the arithmetic progression

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

every pth term is divisible by p.
[*Hint:* Because $\gcd(p, b) = 1$, there exist integers r and s satisfying $pr + bs = 1$. Put $n_k = kp - as$ for $k = 1, 2, \ldots$ and show that $p \mid (a + n_k b)$.]
(b) From part (a), conclude that if b is an odd integer, then every other term in the indicated progression is even.

19. In 1950, it was proved that any integer $n > 9$ can be written as a sum of distinct odd primes. Express the integers 25, 69, 81, and 125 in this fashion.

20. If p and $p^2 + 8$ are both prime numbers, prove that $p^3 + 4$ is also prime.

21. (a) For any integer $k > 0$, establish that the arithmetic progression

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

where $\gcd(a, b) = 1$, contains k consecutive terms that are composite.
[*Hint:* Put $n = (a + b)(a + 2b) \cdots (a + kb)$ and consider the k terms $a + (n + 1)b$, $a + (n + 2)b, \ldots, a + (n + k)b$.]

(b) Find five consecutive composite terms in the arithmetic progression

$$6, 11, 16, 21, 26, 31, 36, \ldots$$

22. Show that 13 is the largest prime that can divide two successive integers of the form $n^2 + 3$.

23. (a) The arithmetic mean of the twin primes 5 and 7 is the triangular number 6. Are there any other twin primes with a triangular mean?

(b) The arithmetic mean of the twin primes 3 and 5 is the perfect square 4. Are there any other twin primes with a square mean?

24. Determine all twin primes p and $q = p + 2$ for which $pq - 2$ is also prime.

25. Let p_n denote the nth prime. For $n > 3$, show that

$$p_n < p_1 + p_2 + \cdots + p_{n-1}$$

[*Hint:* Use induction and the Bertrand conjecture.]

26. Verify the following:

(a) There exist infinitely many primes ending in 33, such as 233, 433, 733, 1033, ...
[*Hint:* Apply Dirichlet's theorem.]

(b) There exist infinitely many primes that do not belong to any pair of twin primes.
[*Hint:* Consider the arithmetic progression $21k + 5$ for $k = 1, 2, \ldots$.]

(c) There exists a prime ending in as many consecutive 1's as desired.
[*Hint:* To obtain a prime ending in n consecutive 1's, consider the arithmetic progression $10^n k + R_n$ for $k = 1, 2, \ldots$.]

(d) There exist infinitely many primes that contain but do not end in the block of digits 123456789.
[*Hint:* Consider the arithmetic progression $10^{11}k + 1234567891$ for $k = 1, 2, \ldots$.]

27. Prove that for every $n \geq 2$ there exists a prime p with $p \leq n < 2p$.
[*Hint:* In the case where $n = 2k + 1$, then by the Bertrand conjecture there exists a prime p such that $k < p < 2k$.]

28. (a) If $n > 1$, show that $n!$ is never a perfect square.

(b) Find the values of $n \geq 1$ for which

$$n! + (n + 1)! + (n + 2)!$$

is a perfect square.
[*Hint:* Note that $n! + (n + 1)! + (n + 2)! = n!(n + 2)^2$.]

CHAPTER 3

PRIMES AND THEIR DISTRIBUTION

Mighty are numbers, joined with art resistless.
EURIPIDES

3.1 THE FUNDAMENTAL THEOREM OF ARITHMETIC

Essential to everything discussed herein—in fact, essential to every aspect of number theory—is the notion of a prime number. We have previously observed that any integer $a > 1$ is divisible by ± 1 and $\pm a$; if these exhaust the divisors of a, then it is said to be a prime number. In Definition 3.1 we state this somewhat differently.

> **Definition 3.1.** An integer $p > 1$ is called a *prime number*, or simply a *prime*, if its only positive divisors are 1 and p. An integer greater than 1 that is not a prime is termed *composite*.

Among the first ten positive integers, 2, 3, 5, 7 are primes and 4, 6, 8, 9, 10 are composite numbers. Note that the integer 2 is the only even prime, and according to our definition the integer 1 plays a special role, being neither prime nor composite.

In the rest of this book, the letters p and q will be reserved, so far as is possible, for primes.

Proposition 14 of Book IX of Euclid's *Elements* embodies the result that later became known as the Fundamental Theorem of Arithmetic, namely, that every integer greater than 1 can, except for the order of the factors, be represented as a product of primes in one and only one way. To quote the proposition itself: "If a number be the least that is measured by prime numbers, it will not be measured by any other

prime except those originally measuring it." Because every number $a > 1$ is either a prime or, by the Fundamental Theorem, can be broken down into unique prime factors and no further, the primes serve as the building blocks from which all other integers can be made. Accordingly, the prime numbers have intrigued mathematicians through the ages, and although a number of remarkable theorems relating to their distribution in the sequence of positive integers have been proved, even more remarkable is what remains unproved. The open questions can be counted among the outstanding unsolved problems in all of mathematics.

To begin on a simpler note, we observe that the prime 3 divides the integer 36, where 36 may be written as any one of the products

$$6 \cdot 6 = 9 \cdot 4 = 12 \cdot 3 = 18 \cdot 2$$

In each instance, 3 divides at least one of the factors involved in the product. This is typical of the general situation, the precise result being Theorem 3.1.

Theorem 3.1. If p is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof. If $p \mid a$, then we need go no further, so let us assume that $p \nmid a$. Because the only positive divisors of p are 1 and p itself, this implies that $\gcd(p, a) = 1$. (In general, $\gcd(p, a) = p$ or $\gcd(p, a) = 1$ according as $p \mid a$ or $p \nmid a$.) Hence, citing Euclid's lemma, we get $p \mid b$.

This theorem easily extends to products of more than two terms.

Corollary 1. If p is a prime and $p \mid a_1 a_2 \cdots a_n$, then $p \mid a_k$ for some k, where $1 \leq k \leq n$.

Proof. We proceed by induction on n, the number of factors. When $n = 1$, the stated conclusion obviously holds; whereas when $n = 2$, the result is the content of Theorem 3.1. Suppose, as the induction hypothesis, that $n > 2$ and that whenever p divides a product of less than n factors, it divides at least one of the factors. Now let $p \mid a_1 a_2 \cdots a_n$. From Theorem 3.1, either $p \mid a_n$ or $p \mid a_1 a_2 \cdots a_{n-1}$. If $p \mid a_n$, then we are through. As regards the case where $p \mid a_1 a_2 \cdots a_{n-1}$, the induction hypothesis ensures that $p \mid a_k$ for some choice of k, with $1 \leq k \leq n - 1$. In any event, p divides one of the integers $a_1, a_2, \ldots, a_n$.

Corollary 2. If $p, q_1, q_2, \ldots, q_n$ are all primes and $p \mid q_1 q_2 \cdots q_n$, then $p = q_k$ for some k, where $1 \leq k \leq n$.

Proof. By virtue of Corollary 1, we know that $p \mid q_k$ for some k, with $1 \leq k \leq n$. Being a prime, q_k is not divisible by any positive integer other than 1 or q_k itself. Because $p > 1$, we are forced to conclude that $p = q_k$.

With this preparation out of the way, we arrive at one of the cornerstones of our development, the Fundamental Theorem of Arithmetic. As indicated earlier, this theorem asserts that every integer greater than 1 can be factored into primes in essentially one way; the linguistic ambiguity *essentially* means that $2 \cdot 3 \cdot 2$ is not considered as being a different factorization of 12 from $2 \cdot 2 \cdot 3$. We state this precisely in Theorem 3.2.

Theorem 3.2 Fundamental Theorem of Arithmetic. Every positive integer $n > 1$ can be expressed as a product of primes; this representation is unique, apart from the order in which the factors occur.

Proof. Either n is a prime or it is composite; in the former case, there is nothing more to prove. If n is composite, then there exists an integer d satisfying $d \mid n$ and $1 < d < n$. Among all such integers d, choose p_1 to be the smallest (this is possible by the Well-Ordering Principle). Then p_1 must be a prime number. Otherwise it too would have a divisor q with $1 < q < p_1$; but then $q \mid p_1$ and $p_1 \mid n$ imply that $q \mid n$, which contradicts the choice of p_1 as the smallest positive divisor, not equal to 1, of n.

We therefore may write $n = p_1 n_1$, where p_1 is prime and $1 < n_1 < n$. If n_1 happens to be a prime, then we have our representation. In the contrary case, the argument is repeated to produce a second prime number p_2 such that $n_1 = p_2 n_2$; that is,

$$n = p_1 p_2 n_2 \qquad 1 < n_2 < n_1$$

If n_2 is a prime, then it is not necessary to go further. Otherwise, write $n_2 = p_3 n_3$, with p_3 a prime:

$$n = p_1 p_2 p_3 n_3 \qquad 1 < n_3 < n_2$$

The decreasing sequence

$$n > n_1 > n_2 > \cdots > 1$$

cannot continue indefinitely, so that after a finite number of steps n_{k-1} is a prime, call it, p_k. This leads to the prime factorization

$$n = p_1 p_2 \cdots p_k$$

To establish the second part of the proof—the uniqueness of the prime factorization—let us suppose that the integer n can be represented as a product of primes in two ways; say,

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s \qquad r \leq s$$

where the p_i and q_j are all primes, written in increasing magnitude so that

$$p_1 \leq p_2 \leq \cdots \leq p_r \qquad q_1 \leq q_2 \leq \cdots \leq q_s$$

Because $p_1 \mid q_1 q_2 \cdots q_s$, Corollary 2 of Theorem 3.1 tells us that $p_1 = q_k$ for some k; but then $p_1 \geq q_1$. Similar reasoning gives $q_1 \geq p_1$, whence $p_1 = q_1$. We may cancel this common factor and obtain

$$p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$$

Now repeat the process to get $p_2 = q_2$ and, in turn,

$$p_3 p_4 \cdots p_r = q_3 q_4 \cdots q_s$$

Continue in this fashion. If the inequality $r < s$ were to hold, we would eventually arrive at

$$1 = q_{r+1} q_{r+2} \cdots q_s$$

which is absurd, because each $q_j > 1$. Hence, $r = s$ and

$$p_1 = q_1 \qquad p_2 = q_2, \ldots, p_r = q_r$$

making the two factorizations of n identical. The proof is now complete.

Of course, several of the primes that appear in the factorization of a given positive integer may be repeated, as is the case with $360 = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 5$. By collecting like primes and replacing them by a single factor, we can rephrase Theorem 3.2 as a corollary.

Corollary. Any positive integer $n > 1$ can be written uniquely in a *canonical form*

$$n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

where, for $i = 1, 2, \ldots, r$, each k_i is a positive integer and each p_i is a prime, with $p_1 < p_2 < \cdots < p_r$.

To illustrate, the canonical form of the integer 360 is $360 = 2^3 \cdot 3^2 \cdot 5$. As further examples we cite

$$4725 = 3^3 \cdot 5^2 \cdot 7 \qquad \text{and} \qquad 17460 = 2^3 \cdot 3^2 \cdot 5 \cdot 7^2$$

Theorem 3.2 should not be taken lightly because number systems do exist in which the factorization into "primes" is not unique. Perhaps the most elemental example is the set E of all positive even integers. Let us agree to call an even integer an e-prime if it is not the product of two other even integers. Thus, 2, 6, 10, 14, ... all are e-primes, whereas 4, 8, 12, 16, ... are not. It is not difficult to see that the integer 60 can be factored into e-primes in two distinct ways; namely,

$$60 = 2 \cdot 30 = 6 \cdot 10$$

Part of the difficulty arises from the fact that Theorem 3.1 is lacking in the set E; that is, $6 \mid 2 \cdot 30$, but $6 \nmid 2$ and $6 \nmid 30$.

This is an opportune moment to insert a famous result of Pythagoras. Mathematics as a science began with Pythagoras (569–500 B.C.), and much of the content of Euclid's *Elements* is due to Pythagoras and his School. The Pythagoreans deserve the credit for being the first to classify numbers into odd and even, prime and composite.

Theorem 3.3 Pythagoras. The number $\sqrt{2}$ is irrational.

Proof. Suppose, to the contrary, that $\sqrt{2}$ is a rational number, say, $\sqrt{2} = a/b$, where a and b are both integers with $\gcd(a, b) = 1$. Squaring, we get $a^2 = 2b^2$, so that $b \mid a^2$. If $b > 1$, then the Fundamental Theorem of Arithmetic guarantees the existence of a prime p such that $p \mid b$. It follows that $p \mid a^2$ and, by Theorem 3.1, that $p \mid a$; hence, $\gcd(a, b) \geq p$. We therefore arrive at a contradiction, unless $b = 1$. But if this happens, then $a^2 = 2$, which is impossible (we assume that the reader is willing to grant that no integer can be multiplied by itself to give 2). Our supposition that $\sqrt{2}$ is a rational number is untenable, and so $\sqrt{2}$ must be irrational.

There is an interesting variation on the proof of Theorem 3.3. If $\sqrt{2} = a/b$ with $\gcd(a, b) = 1$, there must exist integers r and s satisfying $ar + bs = 1$. As a result,

$$\sqrt{2} = \sqrt{2}(ar + bs) = (\sqrt{2}a)r + (\sqrt{2}b)s = 2br + as$$

This representation of $\sqrt{2}$ leads us to conclude that $\sqrt{2}$ is an integer, an obvious impossibility.

PROBLEMS 3.1

1. It has been conjectured that there are infinitely many primes of the form $n^2 - 2$. Exhibit five such primes.
2. Give an example to show that the following conjecture is not true: Every positive integer can be written in the form $p + a^2$, where p is either a prime or 1, and $a \geq 0$.
3. Prove each of the assertions below:
   (a) Any prime of the form $3n + 1$ is also of the form $6m + 1$.
   (b) Each integer of the form $3n + 2$ has a prime factor of this form.
   (c) The only prime of the form $n^3 - 1$ is 7.
   [*Hint:* Write $n^3 - 1$ as $(n - 1)(n^2 + n + 1)$.]
   (d) The only prime p for which $3p + 1$ is a perfect square is $p = 5$.
   (e) The only prime of the form $n^2 - 4$ is 5.
4. If $p \geq 5$ is a prime number, show that $p^2 + 2$ is composite.
   [*Hint:* p takes one of the forms $6k + 1$ or $6k + 5$.]
5. (a) Given that p is a prime and $p \mid a^n$, prove that $p^n \mid a^n$.
   (b) If $\gcd(a, b) = p$, a prime, what are the possible values of $\gcd(a^2, b^2)$, $\gcd(a^2, b)$ and $\gcd(a^3, b^2)$?
6. Establish each of the following statements:
   (a) Every integer of the form $n^4 + 4$, with $n > 1$, is composite.
   [*Hint:* Write $n^4 + 4$ as a product of two quadratic factors.]
   (b) If $n > 4$ is composite, then n divides $(n - 1)!$.
   (c) Any integer of the form $8^n + 1$, where $n \geq 1$, is composite.
   [*Hint:* $2^n + 1 \mid 2^{3n} + 1$.]
   (d) Each integer $n > 11$ can be written as the sum of two composite numbers.
   [*Hint:* If n is even, say $n = 2k$, then $n - 6 = 2(k - 3)$; for n odd, consider the integer $n - 9$.]
7. Find all prime numbers that divide 50!.
8. If $p \geq q \geq 5$ and p and q are both primes, prove that $24 \mid p^2 - q^2$.
9. (a) An unanswered question is whether there are infinitely many primes that are 1 more than a power of 2, such as $5 = 2^2 + 1$. Find two more of these primes.
   (b) A more general conjecture is that there exist infinitely many primes of the form $n^2 + 1$; for example, $257 = 16^2 + 1$. Exhibit five more primes of this type.
10. If $p \neq 5$ is an odd prime, prove that either $p^2 - 1$ or $p^2 + 1$ is divisible by 10.
11. Another unproven conjecture is that there are an infinitude of primes that are 1 less than a power of 2, such as $3 = 2^2 - 1$.
   (a) Find four more of these primes.
   (b) If $p = 2^k - 1$ is prime, show that k is an odd integer, except when $k = 2$.
   [*Hint:* $3 \mid 4^n - 1$ for all $n \geq 1$.]
12. Find the prime factorization of the integers 1234, 10140, and 36000.
13. If $n > 1$ is an integer not of the form $6k + 3$, prove that $n^2 + 2^n$ is composite.
   [*Hint:* Show that either 2 or 3 divides $n^2 + 2^n$.]
14. It has been conjectured that every even integer can be written as the difference of two consecutive primes in infinitely many ways. For example,

$$6 = 29 - 23 = 137 - 131 = 599 - 593 = 1019 - 1013 = \cdots$$

   Express the integer 10 as the difference of two consecutive primes in 15 ways.
15. Prove that a positive integer $a > 1$ is a square if and only if in the canonical form of a all the exponents of the primes are even integers.

16. An integer is said to be *square-free* if it is not divisible by the square of any integer greater than 1. Prove the following:
   (a) An integer $n > 1$ is square-free if and only if n can be factored into a product of distinct primes.
   (b) Every integer $n > 1$ is the product of a square-free integer and a perfect square. [*Hint:* If $n = p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$ is the canonical factorization of n, then write $k_i = 2q_i + r_i$ where $r_i = 0$ or 1 according as k_i is even or odd.]

17. Verify that any integer n can be expressed as $n = 2^k m$, where $k \geq 0$ and m is an odd integer.

18. Numerical evidence makes it plausible that there are infinitely many primes p such that $p + 50$ is also prime. List 15 of these primes.

19. A positive integer n is called *square-full*, or *powerful*, if $p^2 \mid n$ for every prime factor p of n (there are 992 square-full numbers less than 250,000). If n is square-full, show that it can be written in the form $n = a^2 b^3$, with a and b positive integers.

3.2 THE SIEVE OF ERATOSTHENES

Given a particular integer, how can we determine whether it is prime or composite and, in the latter case, how can we actually find a nontrivial divisor? The most obvious approach consists of successively dividing the integer in question by each of the numbers preceding it; if none of them (except 1) serves as a divisor, then the integer must be prime. Although this method is very simple to describe, it cannot be regarded as useful in practice. For even if one is undaunted by large calculations, the amount of time and work involved may be prohibitive.

There is a property of composite numbers that allows us to reduce materially the necessary computations—but still the process remains cumbersome. If an integer $a > 1$ is composite, then it may be written as $a = bc$, where $1 < b < a$ and $1 < c < a$. Assuming that $b \leq c$, we get $b^2 \leq bc = a$, and so $b \leq \sqrt{a}$. Because $b > 1$, Theorem 3.2 ensures that b has at least one prime factor p. Then $p \leq b \leq \sqrt{a}$; furthermore, because $p \mid b$ and $b \mid a$, it follows that $p \mid a$. The point is simply this: A composite number a will always possess a prime divisor p satisfying $p \leq \sqrt{a}$.

In testing the primality of a specific integer $a > 1$, it therefore suffices to divide a by those primes not exceeding $\sqrt{a}$ (presuming, of course, the availability of a list of primes up to $\sqrt{a}$). This may be clarified by considering the integer $a = 509$. Inasmuch as $22 < \sqrt{509} < 23$, we need only try out the primes that are not larger than 22 as possible divisors, namely, the primes 2, 3, 5, 7, 11, 13, 17, 19. Dividing 509 by each of these, in turn, we find that none serves as a divisor of 509. The conclusion is that 509 must be a prime number.

Example 3.1. The foregoing technique provides a practical means for determining the canonical form of an integer, say $a = 2093$. Because $45 < \sqrt{2093} < 46$, it is enough to examine the primes 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43. By trial, the first of these to divide 2093 is 7, and $2093 = 7 \cdot 299$. As regards the integer 299, the seven primes that are less than 18 (note that $17 < \sqrt{299} < 18$) are 2, 3, 5, 7, 11, 13, 17. The first prime divisor of 299 is 13 and, carrying out the required division, we obtain $299 = 13 \cdot 23$. But 23 is itself a prime, whence 2093 has exactly three prime factors, 7, 13, and 23:

$$2093 = 7 \cdot 13 \cdot 23$$

Another Greek mathematician whose work in number theory remains significant is Eratosthenes of Cyrene (276–194 B.C.). Although posterity remembers him mainly as the director of the world-famous library at Alexandria, Eratosthenes was gifted in all branches of learning, if not of first rank in any; in his own day, he was nicknamed "Beta" because, it was said, he stood at least second in every field. Perhaps the most impressive feat of Eratosthenes was the accurate measurement of the earth's circumference by a simple application of Euclidean geometry.

We have seen that if an integer $a > 1$ is not divisible by any prime $p \leq \sqrt{a}$, then a is of necessity a prime. Eratosthenes used this fact as the basis of a clever technique, called the *Sieve of Eratosthenes*, for finding all primes below a given integer n. The scheme calls for writing down the integers from 2 to n in their natural order and then systematically eliminating all the composite numbers by striking out all multiples $2p, 3p, 4p, 5p, \ldots$ of the primes $p \leq \sqrt{n}$. The integers that are left on the list—those that do not fall through the "sieve"—are primes.

To see an example of how this works, suppose that we wish to find all primes not exceeding 100. Consider the sequence of consecutive integers $2, 3, 4, \ldots, 100$. Recognizing that 2 is a prime, we begin by crossing out all even integers from our listing, except 2 itself. The first of the remaining integers is 3, which must be a prime. We keep 3, but strike out all higher multiples of 3, so that 9, 15, 21, ... are now removed (the even multiples of 3 having been removed in the previous step). The smallest integer after 3 that has not yet been deleted is 5. It is not divisible by either 2 or 3—otherwise it would have been crossed out—hence, it is also a prime. All proper multiples of 5 being composite numbers, we next remove 10, 15, 20, ... (some of these are, of course, already missing), while retaining 5 itself. The first surviving integer 7 is a prime, for it is not divisible by 2, 3, or 5, the only primes that precede it. After eliminating the proper multiples of 7, the largest prime less than $\sqrt{100} = 10$, all composite integers in the sequence $2, 3, 4, \ldots, 100$ have fallen through the sieve. The positive integers that remain, to wit, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, are all of the primes less than 100.

The following table represents the result of the completed sieve. The multiples of 2 are crossed out by \; the multiples of 3 are crossed out by /; the multiples of 5 are crossed out by —; the multiples of 7 are crossed out by ~.

	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	83	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~

By this point, an obvious question must have occurred to the reader. Is there a largest prime number, or do the primes go on forever? The answer is to be found in a remarkably simple proof given by Euclid in Book IX of his *Elements*. Euclid's argument is universally regarded as a model of mathematical elegance. Loosely

speaking, it goes like this: Given any finite list of prime numbers, one can always find a prime not on the list; hence, the number of primes is infinite. The actual details appear below.

Theorem 3.4 Euclid. There is an infinite number of primes.

Proof. Euclid's proof is by contradiction. Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7, \ldots$ be the primes in ascending order, and suppose that there is a last prime, called p_n. Now consider the positive integer

$$P = p_1 p_2 \cdots p_n + 1$$

Because $P > 1$, we may put Theorem 3.2 to work once again and conclude that P is divisible by some prime p. But $p_1, p_2, \ldots, p_n$ are the only prime numbers, so that p must be equal to one of $p_1, p_2, \ldots, p_n$. Combining the divisibility relation $p \mid p_1 p_2 \cdots p_n$ with $p \mid P$, we arrive at $p \mid P - p_1 p_2 \cdots p_n$ or, equivalently, $p \mid 1$. The only positive divisor of the integer 1 is 1 itself and, because $p > 1$, a contradiction arises. Thus, no finite list of primes is complete, whence the number of primes is infinite.

For a prime p, define $p^{\#}$ to be the product of all primes that are less than or equal to p. Numbers of the form $p^{\#} + 1$ might be termed *Euclidean numbers*, because they appear in Euclid's scheme for proving the infinitude of primes. It is interesting to note that in forming these integers, the first five, namely,

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

are all prime numbers. However,

$$\begin{aligned} 13^{\#} + 1 &= 59 \cdot 509 \\ 17^{\#} + 1 &= 19 \cdot 97 \cdot 277 \\ 19^{\#} + 1 &= 347 \cdot 27953 \end{aligned}$$

are not prime. A question whose answer is not known is whether there are infinitely many primes p for which $p^{\#} + 1$ is also prime. For that matter, are there infinitely many composite $p^{\#} + 1$?

At present, 19 primes of the form $p^{\#} + 1$ have been identified. These correspond to the values $p = 2, 3, 5, 7, 11, 31, 379, 1019, 1021, 2657, 3229, 4547, 4787, 11549, 13649, 18523, 23801, 24029$, and 42209; the largest of these, a number consisting of 18241 digits, was discovered in 2000. The integer $p^{\#} + 1$ is composite for all other $p \leq 120000$.

Euclid's theorem is too important for us to be content with a single proof. Here is a variation in the reasoning: Form the infinite sequence of positive integers

$$\begin{aligned} n_1 &= 2 \\ n_2 &= n_1 + 1 \\ n_3 &= n_1 n_2 + 1 \\ n_4 &= n_1 n_2 n_3 + 1 \\ &\vdots \\ n_k &= n_1 n_2 \cdots n_{k-1} + 1 \\ &\vdots \end{aligned}$$

Because each $n_k > 1$, each of these integers is divisible by a prime. But no two n_k can have the same prime divisor. To see this, let $d = \gcd(n_i, n_k)$ and suppose that $i < k$. Then d divides n_i and, hence, must divide $n_1 n_2 \cdots n_{k-1}$. Because $d \mid n_k$, Theorem 2.2 (g) tells us that $d \mid n_k - n_1 n_2 \cdots n_{k-1}$ or $d \mid 1$. The implication is that $d = 1$, and so the integers $n_k (k = 1, 2, \ldots)$ are pairwise relatively prime. The point we wish to make is that there are as many distinct primes as there are integers n_k, namely, infinitely many of them.

Let p_n denote the nth of the prime numbers in their natural order. Euclid's proof shows that the expression $p_1 p_2 \cdots p_n + 1$ is divisible by at least one prime. If there are several such prime divisors, then p_{n+1} cannot exceed the smallest of these so that $p_{n+1} \leq p_1 p_2 \cdots p_n + 1$ for $n \geq 1$. Another way of saying the same thing is that

$$p_n \leq p_1 p_2 \cdots p_{n-1} + 1 \qquad n \geq 2$$

With a slight modification of Euclid's reasoning, this inequality can be improved to give

$$p_n \leq p_1 p_2 \cdots p_{n-1} - 1 \qquad n \geq 3$$

For instance, when $n = 5$, this tells us that

$$11 = p_5 \leq 2 \cdot 3 \cdot 5 \cdot 7 - 1 = 209$$

We can see that the estimate is rather extravagant. A sharper limitation on the size of p_n is given by *Bonse's inequality*, which states that

$$p_n^2 < p_1 p_2 \cdots p_{n-1} \qquad n \geq 5$$

This inequality yields $p_5^2 < 210$, or $p_5 \leq 14$. A somewhat better size-estimate for p_5 comes from the inequality

$$p_{2n} \leq p_2 p_3 \cdots p_n - 2 \qquad n \geq 3$$

Here, we obtain

$$p_5 < p_6 \leq p_2 p_3 - 2 = 3 \cdot 5 - 2 = 13$$

To approximate the size of p_n from these formulas, it is necessary to know the values of $p_1, p_2, \ldots, p_{n-1}$. For a bound in which the preceding primes do not enter the picture, we have the following theorem.

Theorem 3.5. If p_n is the nth prime number, then $p_n \leq 2^{2^{n-1}}$.

Proof. Let us proceed by induction on n, the asserted inequality being clearly true when $n = 1$. As the hypothesis of the induction, we assume that $n > 1$ and that the result holds for all integers up to n. Then

$$\begin{aligned} p_{n+1} &\leq p_1 p_2 \cdots p_n + 1 \\ &\leq 2 \cdot 2^2 \cdots 2^{2^{n-1}} + 1 = 2^{1+2+2^2+\cdots+2^{n-1}} + 1 \end{aligned}$$

Recalling the identity $1 + 2 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$, we obtain

$$p_{n+1} \leq 2^{2^n-1} + 1$$

However, $1 \leq 2^{2^n-1}$ for all n; whence

$$\begin{aligned} p_{n+1} &\leq 2^{2^n-1} + 2^{2^n-1} \\ &= 2 \cdot 2^{2^n-1} = 2^{2^n} \end{aligned}$$

completing the induction step, and the argument.

There is a corollary to Theorem 3.5 that is of interest.

Corollary. For $n \geq 1$, there are at least $n + 1$ primes less than 2^{2^n}.

Proof. From the theorem, we know that $p_1, p_2, \ldots, p_{n+1}$ are all less than 2^{2^n}.

We can do considerably better than is indicated by Theorem 3.5. In 1845, Joseph Bertrand conjectured that the prime numbers are well-distributed in the sense that between $n \geq 2$ and $2n$ there is at least one prime. He was unable to establish his conjecture, but verified it for all $n \leq 3{,}000{,}000$. (One way of achieving this is to consider a sequence of primes 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259, 2503, 5003, 9973, 19937, 39869, 79699, 159389, ... each of which is less than twice the preceding.) Because it takes some real effort to substantiate this famous conjecture, let us content ourselves with saying that the first proof was carried out by the Russian mathematician P. L. Tchebycheff in 1852. Granting the result, it is not difficult to show that

$$p_n < 2^n \qquad n \geq 2$$

and as a direct consequence, $p_{n+1} < 2p_n$ for $n \geq 2$. In particular,

$$11 = p_5 < 2 \cdot p_4 = 14$$

To see that $p_n < 2^n$, we argue by induction on n. Clearly, $p_2 = 3 < 2^2$, so that the inequality is true here. Now assume that the inequality holds for an integer n, whence $p_n < 2^n$. Invoking Bertrand's conjecture, there exists a prime number p satisfying $2^n < p < 2^{n+1}$; that is, $p_n < p$. This immediately leads to the conclusion that $p_{n+1} \leq p < 2^{n+1}$, which completes the induction and the proof.

Primes of special form have been of perennial interest. Among these, the repunit primes are outstanding in their simplicity. A *repunit* is an integer written (in decimal notation) as a string of 1's, such as 11, 111, or 1111. Each such integer must have the form $(10^n - 1)/9$. We use the symbol R_n to denote the repunit consisting of n consecutive 1's. A peculiar feature of these numbers is the apparent scarcity of primes among them. So far, only R_2, R_{19}, R_{23}, R_{317}, R_{1031}, R_{49081}, and R_{86453}

have been identified as primes (the last one in 2001). It is known that the only possible repunit primes R_n for all $n \leq 45000$ are the seven numbers just indicated. No conjecture has been made as to the existence of any others. For a repunit R_n to be prime, the subscript n must be a prime; that this is not a sufficient condition is shown by

$$R_5 = 11111 = 41 \cdot 271 \qquad R_7 = 1111111 = 239 \cdot 4649$$

PROBLEMS 3.2

1. Determine whether the integer 701 is prime by testing all primes $p \leq \sqrt{701}$ as possible divisors. Do the same for the integer 1009.
2. Employing the Sieve of Eratosthenes, obtain all the primes between 100 and 200.
3. Given that $p \nmid n$ for all primes $p \leq \sqrt[3]{n}$, show that $n > 1$ is either a prime or the product of two primes.
   [*Hint:* Assume to the contrary that n contains at least three prime factors.]
4. Establish the following facts:
   (a) $\sqrt{p}$ is irrational for any prime p.
   (b) If $a > 0$ and $\sqrt[n]{a}$ is rational, then $\sqrt[n]{a}$ must be an integer.
   (c) For $n \geq 2$, $\sqrt[n]{n}$ is irrational.
   [*Hint:* Use the fact that $2^n > n$.]
5. Show that any composite three-digit number must have a prime factor less than or equal to 31.
6. Fill in any missing details in this sketch of a proof of the infinitude of primes: Assume that there are only finitely many primes, say $p_1, p_2, \ldots, p_n$. Let A be the product of any r of these primes and put $B = p_1 p_2 \cdots p_n / A$. Then each p_k divides either A or B, but not both. Because $A + B > 1$, $A + B$ has a prime divisor different from any of the p_k, which is a contradiction.
7. Modify Euclid's proof that there are infinitely many primes by assuming the existence of a largest prime p and using the integer $N = p! + 1$ to arrive at a contradiction.
8. Give another proof of the infinitude of primes by assuming that there are only finitely many primes, say $p_1, p_2, \ldots, p_n$, and using the following integer to arrive at a contradiction:

$$N = p_2 p_3 \cdots p_n + p_1 p_3 \cdots p_n + \cdots + p_1 p_2 \cdots p_{n-1}$$

9. (a) Prove that if $n > 2$, then there exists a prime p satisfying $n < p < n!$.
   [*Hint:* If $n! - 1$ is not prime, then it has a prime divisor p; and $p \leq n$ implies $p \mid n!$, leading to a contradiction.]
   (b) For $n > 1$, show that every prime divisor of $n! + 1$ is an odd integer that is greater than n.
10. Let q_n be the smallest prime that is strictly greater than $P_n = p_1 p_2 \cdots p_n + 1$. It has been conjectured that the difference $q_n - (p_1 p_2 \cdots p_n)$ is always a prime. Confirm this for the first five values of n.
11. If p_n denotes the nth prime number, put $d_n = p_{n+1} - p_n$. An open question is whether the equation $d_n = d_{n+1}$ has infinitely many solutions. Give five solutions.
12. Assuming that p_n is the nth prime number, establish each of the following statements:
   (a) $p_n > 2n - 1$ for $n \geq 5$.
   (b) None of the integers $P_n = p_1 p_2 \cdots p_n + 1$ is a perfect square.
   [*Hint:* Each P_n is of the form $4k + 3$ for $n > 1$.]

(c) The sum

$$\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n}$$

is never an integer.

13. For the repunits R_n, verify the assertions below:

(a) If $n \mid m$, then $R_n \mid R_m$.

[*Hint:* If $m = kn$, consider the identity

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

(b) If $d \mid R_n$ and $d \mid R_m$, then $d \mid R_{n+m}$.

[*Hint:* Show that $R_{m+n} = R_n 10^m + R_m$.]

(c) If $\gcd(n, m) = 1$, then $\gcd(R_n, R_m) = 1$.

14. Use the previous problem to obtain the prime factors of the repunit R_{10}.

3.3 THE GOLDBACH CONJECTURE

Although there is an infinitude of primes, their distribution within the positive integers is most mystifying. Repeatedly in their distribution we find hints or, as it were, shadows of a pattern; yet an actual pattern amenable to precise description remains elusive. The difference between consecutive primes can be small, as with the pairs 11 and 13, 17 and 19, or for that matter 1000000000061 and 1000000000063. At the same time there exist arbitrarily long intervals in the sequence of integers that are totally devoid of any primes.

It is an unanswered question whether there are infinitely many pairs of *twin primes*; that is, pairs of successive odd integers p and $p + 2$ that are both primes. Numerical evidence leads us to suspect an affirmative conclusion. Electronic computers have discovered 152892 pairs of twin primes less than 30000000 and 20 pairs between 10^{12} and $10^{12}+$ 10000, which hints at their growing scarcity as the positive integers increase in magnitude. Many examples of immense twins are known. The largest twins to date, each 51090 digits long,

$$33218925 \cdot 2^{169690} \pm 1$$

were discovered in 2002.

Consecutive primes cannot only be close together, but also can be far apart; that is, arbitrarily large gaps can occur between consecutive primes. Stated precisely: Given any positive integer n, there exist n consecutive integers, all of which are composite. To prove this, we simply need to consider the integers

$$(n + 1)! + 2, (n + 1)! + 3, \ldots, (n + 1)! + (n + 1)$$

where $(n + 1)! = (n + 1) \cdot n \cdots 3 \cdot 2 \cdot 1$. Clearly, there are n integers listed and they are consecutive. What is important is that each integer is composite. Indeed, $(n + 1)! + 2$ is divisible by 2, $(n + 1)! + 3$ is divisible by 3, and so on.

For instance, if a sequence of four consecutive composite integers is desired, then the previous argument produces 122, 123, 124, and 125:

$$5! + 2 = 122 = 2 \cdot 61$$
$$5! + 3 = 123 = 3 \cdot 41$$
$$5! + 4 = 124 = 4 \cdot 31$$
$$5! + 5 = 125 = 5 \cdot 25$$

Of course, we can find other sets of four consecutive composites, such as 24, 25, 26, 27 or 32, 33, 34, 35.

As this example suggests, our procedure for constructing gaps between two consecutive primes gives a gross overestimate of where they occur among the integers. The first occurrences of prime gaps of specific lengths, where all the intervening integers are composite, have been the subject of computer searches. For instance, there is a gap of length 778 (that is, $p_{n+1} - p_n = 778$) following the prime 42842283925351. No gap of this size exists between two smaller primes. The largest effectively calculated gap between consecutive prime numbers has length 1132, with a string of 1131 composites immediately after the prime

$$1693182318746371$$

Interestingly, computer researchers have not identified gaps of every possible width up to 1132. The smallest missing gap size is 796. The conjecture is that there is a prime gap (a string of $2k - 1$ consecutive composites between two primes) for every even integer $2k$.

This brings us to another unsolved problem concerning the primes, the Goldbach conjecture. In a letter to Leonhard Euler in the year 1742, Christian Goldbach hazarded the guess that every even integer is the sum of two numbers that are either primes or 1. A somewhat more general formulation is that every even integer greater than 4 can be written as a sum of two odd prime numbers. This is easy to confirm for the first few even integers:

$$\begin{aligned}
2 &= 1 + 1 \\
4 &= 2 + 2 = 1 + 3 \\
6 &= 3 + 3 = 1 + 5 \\
8 &= 3 + 5 = 1 + 7 \\
10 &= 3 + 7 = 5 + 5 \\
12 &= 5 + 7 = 1 + 11 \\
14 &= 3 + 11 = 7 + 7 = 1 + 13 \\
16 &= 3 + 13 = 5 + 11 \\
18 &= 5 + 13 = 7 + 11 = 1 + 17 \\
20 &= 3 + 17 = 7 + 13 = 1 + 19 \\
22 &= 3 + 19 = 5 + 17 = 11 + 11 \\
24 &= 5 + 19 = 7 + 17 = 11 + 13 = 1 + 23 \\
26 &= 3 + 23 = 7 + 19 = 13 + 13 \\
28 &= 5 + 23 = 11 + 17 \\
30 &= 7 + 23 = 11 + 19 = 13 + 17 = 1 + 29
\end{aligned}$$

Although it seems that Euler never tried to prove the result, upon writing to Goldbach at a later date, Euler countered with a conjecture of his own: Any even integer (≥ 6) of the form $4n + 2$ is a sum of two numbers each being either a prime of the form $4n + 1$ or 1.

The numerical data suggesting the truth of Goldbach's conjecture are overwhelming. It has been verified by computers for all even integers less than $4 \cdot 10^{14}$.

As the integers become larger, the number of different ways in which $2n$ can be expressed as the sum of two primes increases. For example, there are 219400 such representations for the even integer 100000000. Although this supports the feeling that Goldbach was correct in his conjecture, it is far from a mathematical proof, and all attempts to obtain a proof have been completely unsuccessful. One of the most famous number theorists of the last century, G. H. Hardy, in his address to the Mathematical Society of Copenhagen in 1921, stated that the Goldbach conjecture appeared "... probably as difficult as any of the unsolved problems in mathematics." It is currently known that every even integer is the sum of six or fewer primes.

We remark that if the conjecture of Goldbach is true, then each odd number larger than 7 must be the sum of three odd primes. To see this, take n to be an odd integer greater than 7, so that $n - 3$ is even and greater than 4; if $n - 3$ could be expressed as the sum of two odd primes, then n would be the sum of three.

The first real progress on the conjecture in nearly 200 years was made by Hardy and Littlewood in 1922. On the basis of a certain unproved hypothesis, the so-called generalized Riemann hypothesis, they showed that every sufficiently large odd number is the sum of three odd primes. In 1937, the Russian mathematician I. M. Vinogradov was able to remove the dependence on the generalized Riemann hypothesis, thereby giving an unconditional proof of this result; that is to say, he established that all odd integers greater than some effectively computable n_0 can be written as the sum of three odd primes.

$$n = p_1 + p_2 + p_3 \qquad (n \text{ odd}, n \text{ sufficiently large})$$

Vinogradov was unable to decide how large n_0 should be, but Borozdkin (1956) proved that $n_0 < 3^{3^{15}}$. In 2002, the bound on n_0 was reduced to 10^{1346}. It follows immediately that every even integer from some point on is the sum of either two or four primes. Thus, it is enough to answer the question for every odd integer n in the range $9 \leq n \leq n_0$, which, for a given integer, becomes a matter of tedious computation (unfortunately, n_0 is so large that this exceeds the capabilities of the most modern electronic computers).

Because of the strong evidence in favor of the famous Goldbach conjecture, we readily become convinced that it is true. Nevertheless, it might be false. Vinogradov showed that if $A(x)$ is the number of even integers $n \leq x$ that are not the sum of two primes, then

$$\lim_{x \to \infty} A(x)/x = 0$$

This allows us to say that "almost all" even integers satisfy the conjecture. As Edmund Landau so aptly put it, "The Goldbach conjecture is false for at most 0% of all even integers; this *at most* 0% does not exclude, of course, the possibility that there are infinitely many exceptions."

Having digressed somewhat, let us observe that according to the Division Algorithm, every positive integer can be written uniquely in one of the forms

$$4n \qquad 4n + 1 \qquad 4n + 2 \qquad 4n + 3$$

for some suitable $n \geq 0$. Clearly, the integers $4n$ and $4n + 2 = 2(2n + 1)$ are both even. Thus, all odd integers fall into two progressions: one containing integers of the form $4n + 1$, and the other containing integers of the form $4n + 3$.

The question arises as to how these two types of primes are distributed within the set of positive integers. Let us display the first few odd prime numbers in consecutive order, putting the $4n + 3$ primes in the top row and the $4n + 1$ primes under them:

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

At this point, one might have the general impression that primes of the form $4n + 3$ are more abundant than are those of the form $4n + 1$. To obtain more precise information, we require the help of the function $\pi_{a,b}(x)$, which counts the number of primes of the form $p = an + b$ not exceeding x. Our small table, for instance, indicates that $\pi_{4,1}(89) = 10$ and $\pi_{4,3}(89) = 13$.

In a famous letter written in 1853, Tchebycheff remarked that $\pi_{4,1}(x) \leq \pi_{4,3}(x)$ for small values of x. He also implied that he had a proof that the inequality always held. In 1914, J. E. Littlewood showed that the inequality fails infinitely often, but his method gave no indication of the value of x for which this first happens. It turned out to be quite difficult to find. Not until 1957 did a computer search reveal that $x = 26861$ is the smallest prime for which $\pi_{4,1}(x) > \pi_{4,3}(x)$; here, $\pi_{4,1}(x) = 1473$ and $\pi_{4,3}(x) = 1472$. This is an isolated situation, because the next prime at which a reversal occurs is $x = 616{,}841$. Remarkably, $\pi_{4,1}(x) > \pi_{4,3}(x)$ for the 410 million successive integers x lying between 18540000000 and 18950000000.

The behavior of primes of the form $3n \pm 1$ provided more of a computational challenge: the inequality $\pi_{3,1}(x) \leq \pi_{3,2}(x)$ holds for all x until one reaches $x = 608981813029$.

This furnishes a pleasant opportunity for a repeat performance of Euclid's method for proving the existence of an infinitude of primes. A slight modification of his argument reveals that there is an infinite number of primes of the form $4n + 3$. We approach the proof through a simple lemma.

Lemma. The product of two or more integers of the form $4n + 1$ is of the same form.

Proof. It is sufficient to consider the product of just two integers. Let us take $k = 4n + 1$ and $k' = 4m + 1$. Multiplying these together, we obtain

$$\begin{aligned} kk' &= (4n + 1)(4m + 1) \\ &= 16nm + 4n + 4m + 1 = 4(4nm + n + m) + 1 \end{aligned}$$

which is of the desired form.

This paves the way for Theorem 3.6.

Theorem 3.6. There are an infinite number of primes of the form $4n + 3$.

Proof. In anticipation of a contradiction, let us assume that there exist only finitely many primes of the form $4n + 3$; call them $q_1, q_2, \ldots, q_s$. Consider the positive integer

$$N = 4q_1q_2 \cdots q_s - 1 = 4(q_1q_2 \cdots q_s - 1) + 3$$

and let $N = r_1r_2 \cdots r_t$ be its prime factorization. Because N is an odd integer, we have $r_k \neq 2$ for all k, so that each r_k is either of the form $4n + 1$ or $4n + 3$. By the lemma, the product of any number of primes of the form $4n + 1$ is again an integer of this type. For N to take the form $4n + 3$, as it clearly does, N must contain at least one prime factor r_i of the form $4n + 3$. But r_i cannot be found among the listing $q_1, q_2, \ldots, q_s$, for this would lead to the contradiction that $r_i \mid 1$. The only possible conclusion is that there are infinitely many primes of the form $4n + 3$.

Having just seen that there are infinitely many primes of the form $4n + 3$, we might reasonably ask: Is the number of primes of the form $4n + 1$ also infinite? This answer is likewise in the affirmative, but a demonstration must await the development of the necessary mathematical machinery. Both these results are special cases of a remarkable theorem by P. G. L. Dirichlet on primes in arithmetic progressions, established in 1837. The proof is much too difficult for inclusion here, so that we must content ourselves with the mere statement.

Theorem 3.7 Dirichlet. If a and b are relatively prime positive integers, then the arithmetic progression

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

contains infinitely many primes.

Dirichlet's theorem tells us, for instance, that there are infinitely many prime numbers ending in 999, such as 1999, 100999, 1000999, ... for these appear in the arithmetic progression determined by $1000n + 999$, where $\gcd(1000, 999) = 1$.

There is no arithmetic progression $a, a + b, a + 2b, \ldots$ that consists solely of prime numbers. To see this, suppose that $a + nb = p$, where p is a prime. If we put $n_k = n + kp$ for $k = 1, 2, 3, \ldots$ then the n_kth term in the progression is

$$a + n_kb = a + (n + kp)b = (a + nb) + kpb = p + kpb$$

Because each term on the right-hand side is divisible by p, so is $a + n_kb$. In other words, the progression must contain infinitely many composite numbers.

It is an old, but still unsolved question of whether there exist arbitrarily long but finite arithmetic progressions consisting only of prime numbers (not necessarily consecutive primes). The longest progression found to date is composed of the 22 primes:

$$11410337850553 + 4609098694200n \qquad 0 \leq n \leq 21$$

The prime factorization of the common difference between the terms is

$$2^3 \cdot 3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 1033$$

which is divisible by 9699690, the product of the primes less than 22. This takes place according to Theorem 3.8.

Theorem 3.8. If all the $n > 2$ terms of the arithmetic progression

$$p, p + d, p + 2d, \ldots, p + (n - 1)d$$

are prime numbers, then the common difference d is divisible by every prime $q < n$.

Proof. Consider a prime number $q < n$ and assume to the contrary that $q \nmid d$. We claim that the first q terms of the progression

$$p, p + d, p + 2d, \ldots, p + (q - 1)d \tag{1}$$

will leave different remainders when divided by q. Otherwise there exist integers j and k, with $0 \leq j < k \leq q - 1$, such that the numbers $p + jd$ and $p + kd$ yield the same remainder upon division by q. Then q divides their difference $(k - j)d$. But $\gcd(q , d) = 1$, and so Euclid's lemma leads to $q \mid k - j$, which is nonsense in light of the inequality $k - j \leq q - 1$.

Because the q different remainders produced from Eq. (1) are drawn from the q integers $0, 1, \ldots, q - 1$, one of these remainders must be zero. This means that $q \mid p + td$ for some t satisfying $0 \leq t \leq q - 1$. Because of the inequality $q < n \leq p \leq p + td$, we are forced to conclude that $p + td$ is composite. (If p were less than n, one of the terms of the progression would be $p + pd = p(1 + d)$.) With this contradiction, the proof that $q \mid d$ is complete.

It has been conjectured that there exist arithmetic progressions of finite (but otherwise arbitrary) length, composed of consecutive prime numbers. Examples of such progressions consisting of three and four primes, respectively, are 47, 53, 59, and 251, 257, 263, 269.

Most recently a sequence of 10 consecutive primes was discovered in which each term exceeds its predecessor by just 210; the smallest of these primes has 93 digits. Finding an arithmetic progression consisting of 11 consecutive primes is likely to be out of reach for some time. Absent the restriction that the primes involved be consecutive, strings of 11-term arithmetic progressions are easily located. One such is

$$110437 + 13860n \qquad 0 \leq n \leq 10$$

In the interest of completeness, we might mention another famous problem that, so far, has resisted the most determined attack. For centuries, mathematicians have sought a simple formula that would yield every prime number or, failing this, a formula that would produce nothing but primes. At first glance, the request seems modest enough: Find a function $f(n)$ whose domain is, say, the nonnegative integers and whose range is some infinite subset of the set of all primes. It was widely believed years ago that the quadratic polynomial

$$f(n) = n^2 + n + 41$$

assumed only prime values. This was shown to be false by Euler, in 1772. As evidenced by the following table, the claim is a correct one for $n = 0, 1, 2, \ldots, 39$.

n	$f(n)$	n	$f(n)$	n	$f(n)$
0	41	14	251	28	853
1	43	15	281	29	911
2	47	16	313	30	971
3	53	17	347	31	1033
4	61	18	383	32	1097
5	71	19	421	33	1163
6	83	20	461	34	1231
7	97	21	503	35	1301
8	113	22	547	36	1373
9	131	23	593	37	1447
10	151	24	641	38	1523
11	173	25	691	39	1601
12	197	26	743		
13	223	27	797		

However, this provocative conjecture is shattered in the cases $n = 40$ and $n = 41$, where there is a factor of 41:

$$f(40) = 40 \cdot 41 + 41 = 41^2$$

and

$$f(41) = 41 \cdot 42 + 41 = 41 \cdot 43$$

The next value $f(42) = 1847$ turns out to be prime once again. In fact, for the first 100 integer values of n, the so-called Euler polynomial represents 86 primes. Although it starts off very well in the production of primes, there are other quadratics such as

$$g(n) = n^2 + n + 27941$$

that begin to best $f(n)$ as the values of n become larger. For example, $g(n)$ is prime for 286129 values of $0 \leq n \leq 10^6$, whereas its famous rival yields 261081 primes in this range.

It has been shown that no polynomial of the form $n^2 + n + q$, with q a prime, can do better than the Euler polynomial in giving primes for successive values of n. Indeed, until fairly recently no other quadratic polynomial of any kind was known to produce more than 40 successive prime values. The polynomial

$$h(n) = 103n^2 - 3945n + 34381$$

found in 1988, produces 43 distinct prime values for $n = 0, 1, 2, \ldots, 42$. The current record holder in this regard

$$k(n) = 36n^2 - 810n + 2753$$

does slightly better by giving a string of 45 prime values.

The failure of the previous functions to be prime-producing is no accident, for it is easy to prove that there is no nonconstant polynomial $f(n)$ with integral

coefficients that takes on just prime values for integral n. We assume that such a polynomial $f(n)$ actually does exist and argue until a contradiction is reached. Let

$$f(n) = a_k n^k + a_{k-1} n^{k-1} + \cdots + a_2 n^2 + a_1 n + a_0$$

where all the coefficients $a_0, a_1, \ldots, a_k$ are integers, and $a_k \neq 0$. For a fixed value of (n_0), $p = f(n_0)$ is a prime number. Now, for any integer t, we consider the following expression:

$$\begin{aligned} f(n_0 + tp) &= a_k(n_0 + tp)^k + \cdots + a_1(n_0 + tp) + a_0 \\ &= (a_k n_0^k + \cdots + a_1 n_0 + a_0) + pQ(t) \\ &= f(n_0) + pQ(t) \\ &= p + pQ(t) = p(1 + Q(t)) \end{aligned}$$

where $Q(t)$ is a polynomial in t having integral coefficients. Our reasoning shows that $p \mid f(n_0 + tp)$; hence, from our own assumption that $f(n)$ takes on only prime values, $f(n_0 + tp) = p$ for any integer t. Because a polynomial of degree k cannot assume the same value more than k times, we have obtained the required contradiction.

Recent years have seen a measure of success in the search for prime-producing functions. W. H. Mills proved (1947) that there exists a positive real number r such that the expression $f(n) = [r^{3^n}]$ is prime for $n = 1, 2, 3, \ldots$ (the brackets indicate the greatest integer function). Needless to say, this is strictly an existence theorem and nothing is known about the actual value of r. Mills's function does not produce all the primes.

PROBLEMS 3.3

1. Verify that the integers 1949 and 1951 are twin primes.
2. (a) If 1 is added to a product of twin primes, prove that a perfect square is always obtained.
   (b) Show that the sum of twin primes p and $p + 2$ is divisible by 12, provided that $p > 3$.
3. Find all pairs of primes p and q satisfying $p - q = 3$.
4. Sylvester (1896) rephrased the Goldbach conjecture: Every even integer $2n$ greater than 4 is the sum of two primes, one larger than $n/2$ and the other less than $3n/2$. Verify this version of the conjecture for all even integers between 6 and 76.
5. In 1752, Goldbach submitted the following conjecture to Euler: Every odd integer can be written in the form $p + 2a^2$, where p is either a prime or 1 and $a \geq 0$. Show that the integer 5777 refutes this conjecture.
6. Prove that the Goldbach conjecture that every even integer greater than 2 is the sum of two primes is equivalent to the statement that every integer greater than 5 is the sum of three primes.
   [*Hint:* If $2n - 2 = p_1 + p_2$, then $2n = p_1 + p_2 + 2$ and $2n + 1 = p_1 + p_2 + 3$.]
7. A conjecture of Lagrange (1775) asserts that every odd integer greater than 5 can be written as a sum $p_1 + 2p_2$, where p_1, p_2 are both primes. Confirm this for all odd integers through 75.
8. Given a positive integer n, it can be shown that there exists an even integer a that is representable as the sum of two odd primes in n different ways. Confirm that the integers

60, 78, and 84 can be written as the sum of two primes in six, seven, and eight ways, respectively.

9. (a) For $n > 3$, show that the integers $n, n+2, n+4$ cannot all be prime.
(b) Three integers $p, p+2, p+6$, which are all prime, are called a *prime-triplet*. Find five sets of prime-triplets.

10. Establish that the sequence

$$(n+1)!-2, (n+1)!-3, \ldots, (n+1)!-(n+1)$$

produces n consecutive composite integers for $n > 2$.

11. Find the smallest positive integer n for which the function $f(n) = n^2 + n + 17$ is composite. Do the same for the functions $g(n) = n^2 + 21n + 1$ and $h(n) = 3n^2 + 3n + 23$.

12. Let p_n denote the nth prime number. For $n \geq 3$, prove that $p_{n+3}^2 < p_n p_{n+1} p_{n+2}$.
[*Hint:* Note that $p_{n+3}^2 < 4p_{n+2}^2 < 8p_{n+1}p_{n+2}$.]

13. Apply the same method of proof as in Theorem 3.6 to show that there are infinitely many primes of the form $6n + 5$.

14. Find a prime divisor of the integer $N = 4(3 \cdot 7 \cdot 11) - 1$ of the form $4n + 3$. Do the same for $N = 4(3 \cdot 7 \cdot 11 \cdot 15) - 1$.

15. Another unanswered question is whether there exist an infinite number of sets of five consecutive odd integers of which four are primes. Find five such sets of integers.

16. Let the sequence of primes, with 1 adjoined, be denoted by $p_0 = 1$, $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$. For each $n \geq 1$, it is known that there exists a suitable choice of coefficients $\epsilon_k = \pm 1$ such that

$$p_{2n} = p_{2n-1} + \sum_{k=0}^{2n-2} \epsilon_k p_k \qquad p_{2n+1} = 2p_{2n} + \sum_{k=0}^{2n-1} \epsilon_k p_k$$

To illustrate:

$$13 = 1 + 2 - 3 - 5 + 7 + 11$$

and

$$17 = 1 + 2 - 3 - 5 + 7 - 11 + 2 \cdot 13$$

Determine similar representations for the primes 23, 29, 31, and 37.

17. In 1848, de Polignac claimed that every odd integer is the sum of a prime and a power of 2. For example, $55 = 47 + 2^3 = 23 + 2^5$. Show that the integers 509 and 877 discredit this claim.

18. (a) If p is a prime and $p \nmid b$, prove that in the arithmetic progression

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

every pth term is divisible by p.
[*Hint:* Because $\gcd(p, b) = 1$, there exist integers r and s satisfying $pr + bs = 1$. Put $n_k = kp - as$ for $k = 1, 2, \ldots$ and show that $p \mid (a + n_k b)$.]
(b) From part (a), conclude that if b is an odd integer, then every other term in the indicated progression is even.

19. In 1950, it was proved that any integer $n > 9$ can be written as a sum of distinct odd primes. Express the integers 25, 69, 81, and 125 in this fashion.

20. If p and $p^2 + 8$ are both prime numbers, prove that $p^3 + 4$ is also prime.

21. (a) For any integer $k > 0$, establish that the arithmetic progression

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

where $\gcd(a, b) = 1$, contains k consecutive terms that are composite.
[*Hint:* Put $n = (a + b)(a + 2b) \cdots (a + kb)$ and consider the k terms $a + (n + 1)b$, $a + (n + 2)b, \ldots, a + (n + k)b$.]

(b) Find five consecutive composite terms in the arithmetic progression

$$6, 11, 16, 21, 26, 31, 36, \ldots$$

22. Show that 13 is the largest prime that can divide two successive integers of the form $n^2 + 3$.

23. (a) The arithmetic mean of the twin primes 5 and 7 is the triangular number 6. Are there any other twin primes with a triangular mean?

(b) The arithmetic mean of the twin primes 3 and 5 is the perfect square 4. Are there any other twin primes with a square mean?

24. Determine all twin primes p and $q = p + 2$ for which $pq - 2$ is also prime.

25. Let p_n denote the nth prime. For $n > 3$, show that

$$p_n < p_1 + p_2 + \cdots + p_{n-1}$$

[*Hint:* Use induction and the Bertrand conjecture.]

26. Verify the following:

(a) There exist infinitely many primes ending in 33, such as 233, 433, 733, 1033, ...
[*Hint:* Apply Dirichlet's theorem.]

(b) There exist infinitely many primes that do not belong to any pair of twin primes.
[*Hint:* Consider the arithmetic progression $21k + 5$ for $k = 1, 2, \ldots$.]

(c) There exists a prime ending in as many consecutive 1's as desired.
[*Hint:* To obtain a prime ending in n consecutive 1's, consider the arithmetic progression $10^n k + R_n$ for $k = 1, 2, \ldots$.]

(d) There exist infinitely many primes that contain but do not end in the block of digits 123456789.
[*Hint:* Consider the arithmetic progression $10^{11}k + 1234567891$ for $k = 1, 2, \ldots$.]

27. Prove that for every $n \geq 2$ there exists a prime p with $p \leq n < 2p$.
[*Hint:* In the case where $n = 2k + 1$, then by the Bertrand conjecture there exists a prime p such that $k < p < 2k$.]

28. (a) If $n > 1$, show that $n!$ is never a perfect square.

(b) Find the values of $n \geq 1$ for which

$$n! + (n + 1)! + (n + 2)!$$

is a perfect square.
[*Hint:* Note that $n! + (n + 1)! + (n + 2)! = n!(n + 2)^2$.]

CHAPTER 4

THE THEORY OF CONGRUENCES

Gauss once said "Mathematics is the queen of the sciences and number-theory the queen of mathematics." If this be true we may add that the Disquisitiones *is the Magna Charta of number-theory.*
M. CANTOR

4.1 CARL FRIEDRICH GAUSS

Another approach to divisibility questions is through the arithmetic of remainders, or the *theory of congruences* as it is now commonly known. The concept, and the notation that makes it such a powerful tool, was first introduced by the German mathematician Carl Friedrich Gauss (1777–1855) in his *Disquisitiones Arithmeticae*; this monumental work, which appeared in 1801 when Gauss was 24 years old, laid the foundations of modern number theory. Legend has it that a large part of the *Disquisitiones Arithmeticae* had been submitted as a memoir to the French Academy the previous year and had been rejected in a manner that, even if the work had been as worthless as the referees believed, would have been inexcusable. (In an attempt to lay this defamatory tale to rest, the officers of the Academy made an exhaustive search of their permanent records in 1935 and concluded that the *Disquisitiones* was never submitted, much less rejected.) "It is really astonishing," said Kronecker, "to think that a single man of such young years was able to bring to light such a wealth of results, and above all to present such a profound and well-organized treatment of an entirely new discipline."

Carl Friedrich Gauss
(1777–1855)

(Dover Publications, Inc.)

Gauss was one of those remarkable infant prodigies whose natural aptitude for mathematics soon becomes apparent. As a child of age three, according to a well-authenticated story, he corrected an error in his father's payroll calculations. His arithmetical powers so overwhelmed his schoolmasters that, by the time Gauss was 7 years old, they admitted that there was nothing more they could teach the boy. It is said that in his first arithmetic class Gauss astonished his teacher by instantly solving what was intended to be a "busy work" problem: Find the sum of all the numbers from 1 to 100. The young Gauss later confessed to having recognized the pattern

$$1 + 100 = 101, 2 + 99 = 101, 3 + 98 = 101, \ldots, 50 + 51 = 101$$

Because there are 50 pairs of numbers, each of which adds up to 101, the sum of all the numbers must be $50 \cdot 101 = 5050$. This technique provides another way of deriving the formula

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

for the sum of the first n positive integers. One need only display the consecutive integers 1 through n in two rows as follows:

$$\begin{array}{ccccccc} 1 & 2 & 3 & \cdots & n-1 & n \\ n & n-1 & n-2 & \cdots & 2 & 1 \end{array}$$

Addition of the vertical columns produces n terms, each of which is equal to $n + 1$; when these terms are added, we get the value $n(n + 1)$. Because the same sum is obtained on adding the two rows horizontally, what occurs is the formula $n(n + 1) = 2(1 + 2 + 3 + \cdots + n)$.

Gauss went on to a succession of triumphs, each new discovery following on the heels of a previous one. The problem of constructing regular polygons with only "Euclidean tools," that is to say, with ruler and compass alone, had long been laid aside in the belief that the ancients had exhausted all the possible constructions. In 1796, Gauss showed that the 17-sided regular polygon is so constructible, the first

advance in this area since Euclid's time. Gauss' doctoral thesis of 1799 provided a rigorous proof of the Fundamental Theorem of Algebra, which had been stated first by Girard in 1629 and then proved imperfectly by d'Alembert (1746), and later by Euler (1749). The theorem (it asserts that a polynomial equation of degree n has exactly n complex roots) was always a favorite of Gauss', and he gave, in all, four distinct demonstrations of it. The publication of *Disquisitiones Arithmeticae* in 1801 at once placed Gauss in the front rank of mathematicians.

The most extraordinary achievement of Gauss was more in the realm of theoretical astronomy than of mathematics. On the opening night of the 19th century, January 1, 1801, the Italian astronomer Piazzi discovered the first of the so-called minor planets (planetoids or asteroids), later called Ceres. But after the course of this newly found body—visible only by telescope—passed the sun, neither Piazzi nor any other astronomer could locate it again. Piazzi's observations extended over a period of 41 days, during which the orbit swept out an angle of only nine degrees. From the scanty data available, Gauss was able to calculate the orbit of Ceres with amazing accuracy, and the elusive planet was rediscovered at the end of the year in almost exactly the position he had forecasted. This success brought Gauss worldwide fame, and led to his appointment as director of Göttingen Observatory.

By the middle of the 19th century, mathematics had grown into an enormous and unwieldy structure, divided into a large number of fields in which only the specialist knew his way. Gauss was the last complete mathematician, and it is no exaggeration to say that he was in some degree connected with nearly every aspect of the subject. His contemporaries regarded him as Princeps Mathematicorum (Prince of Mathematicians), on a par with Archimedes and Isaac Newton. This is revealed in a small incident: On being asked who was the greatest mathematician in Germany, Laplace answered, "Why, Pfaff." When the questioner indicated that he would have thought Gauss was, Laplace replied, "Pfaff is by far the greatest in Germany, but Gauss is the greatest in all Europe."

Although Gauss adorned every branch of mathematics, he always held number theory in high esteem and affection. He insisted that, "Mathematics is the Queen of the Sciences, and the theory of numbers is the Queen of Mathematics."

4.2 BASIC PROPERTIES OF CONGRUENCE

In the first chapter of *Disquisitiones Arithmeticae*, Gauss introduces the concept of congruence and the notation that makes it such a powerful technique (he explains that he was induced to adopt the symbol $\equiv$ because of the close analogy with algebraic equality). According to Gauss, "If a number n measures the difference between two numbers a and b, then a and b are said to be congruent with respect to n; if not, incongruent." Putting this into the form of a definition, we have Definition 4.1.

Definition 4.1. Let n be a fixed positive integer. Two integers a and b are said to be *congruent modulo* n, symbolized by

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

if n divides the difference $a - b$; that is, provided that $a - b = kn$ for some integer k.

To fix the idea, consider $n = 7$. It is routine to check that

$$3 \equiv 24 \pmod 7 \qquad -31 \equiv 11 \pmod 7 \qquad -15 \equiv -64 \pmod 7$$

because $3 - 24 = (-3)7$, $-31 - 11 = (-6)7$, and $-15 - (-64) = 7 \cdot 7$. When $n \nmid (a - b)$, we say that a is *incongruent to b modulo n*, and in this case we write $a \not\equiv b \pmod n$. For a simple example: $25 \not\equiv 12 \pmod 7$, because 7 fails to divide $25 - 12 = 13$.

It is to be noted that any two integers are congruent modulo 1, whereas two integers are congruent modulo 2 when they are both even or both odd. Inasmuch as congruence modulo 1 is not particularly interesting, the usual practice is to assume that $n > 1$.

Given an integer a, let q and r be its quotient and remainder upon division by n, so that

$$a = qn + r \qquad 0 \leq r < n$$

Then, by definition of congruence, $a \equiv r \pmod n$. Because there are n choices for r, we see that every integer is congruent modulo n to exactly one of the values $0, 1, 2, \ldots, n - 1$; in particular, $a \equiv 0 \pmod n$ if and only if $n \mid a$. The set of n integers $0, 1, 2, \ldots, n - 1$ is called the set of *least nonnegative residues modulo n*.

In general, a collection of n integers $a_1, a_2, \ldots, a_n$ is said to form a *complete set of residues* (or a *complete system of residues*) *modulo n* if every integer is congruent modulo n to one and only one of the a_k. To put it another way, $a_1, a_2, \ldots, a_n$ are congruent modulo n to $0, 1, 2, \ldots, n - 1$, taken in some order. For instance,

$$-12, -4, 11, 13, 22, 82, 91$$

constitute a complete set of residues modulo 7; here, we have

$$-12 \equiv 2 \quad -4 \equiv 3 \quad 11 \equiv 4 \quad 13 \equiv 6 \quad 22 \equiv 1 \quad 82 \equiv 5 \quad 91 \equiv 0$$

all modulo 7. An observation of some importance is that any n integers form a complete set of residues modulo n if and only if no two of the integers are congruent modulo n. We shall need this fact later.

Our first theorem provides a useful characterization of congruence modulo n in terms of remainders upon division by n.

Theorem 4.1. For arbitrary integers a and b, $a \equiv b \pmod n$ if and only if a and b leave the same nonnegative remainder when divided by n.

Proof. First take $a \equiv b \pmod n$, so that $a = b + kn$ for some integer k. Upon division by n, b leaves a certain remainder r; that is, $b = qn + r$, where $0 \leq r < n$. Therefore,

$$a = b + kn = (qn + r) + kn = (q + k)n + r$$

which indicates that a has the same remainder as b.

On the other hand, suppose we can write $a = q_1 n + r$ and $b = q_2 n + r$, with the same remainder r $(0 \leq r < n)$. Then

$$a - b = (q_1 n + r) - (q_2 n + r) = (q_1 - q_2)n$$

whence $n \mid a - b$. In the language of congruences, we have $a \equiv b \pmod n$.

Example 4.1. Because the integers -56 and -11 can be expressed in the form

$$-56 = (-7)9 + 7 \qquad -11 = (-2)9 + 7$$

with the same remainder 7, Theorem 4.1 tells us that $-56 \equiv -11 \pmod 9$. Going in the other direction, the congruence $-31 \equiv 11 \pmod 7$ implies that -31 and 11 have the same remainder when divided by 7; this is clear from the relations

$$-31 = (-5)7 + 4 \qquad 11 = 1 \cdot 7 + 4$$

Congruence may be viewed as a generalized form of equality, in the sense that its behavior with respect to addition and multiplication is reminiscent of ordinary equality. Some of the elementary properties of equality that carry over to congruences appear in the next theorem.

Theorem 4.2. Let $n > 1$ be fixed and a, b, c, d be arbitrary integers. Then the following properties hold:

(a) $a \equiv a \pmod n$.
(b) If $a \equiv b \pmod n$, then $b \equiv a \pmod n$.
(c) If $a \equiv b \pmod n$ and $b \equiv c \pmod n$, then $a \equiv c \pmod n$.
(d) If $a \equiv b \pmod n$ and $c \equiv d \pmod n$, then $a + c \equiv b + d \pmod n$ and $ac \equiv bd \pmod n$.
(e) If $a \equiv b \pmod n$, then $a + c \equiv b + c \pmod n$ and $ac \equiv bc \pmod n$.
(f) If $a \equiv b \pmod n$, then $a^k \equiv b^k \pmod n$ for any positive integer k.

Proof. For any integer a, we have $a - a = 0 \cdot n$, so that $a \equiv a \pmod n$. Now if $a \equiv b \pmod n$, then $a - b = kn$ for some integer k. Hence, $b - a = -(kn) = (-k)n$ and because $-k$ is an integer, this yields property (b).

Property (c) is slightly less obvious: Suppose that $a \equiv b \pmod n$ and also $b \equiv c \pmod n$. Then there exist integers h and k satisfying $a - b = hn$ and $b - c = kn$. It follows that

$$a - c = (a - b) + (b - c) = hn + kn = (h + k)n$$

which is $a \equiv c \pmod n$ in congruence notation.

In the same vein, if $a \equiv b \pmod n$ and $c \equiv d \pmod n$, then we are assured that $a - b = k_1 n$ and $c - d = k_2 n$ for some choice of k_1 and k_2. Adding these equations, we obtain

$$\begin{aligned}(a + c) - (b + d) &= (a - b) + (c - d) \\ &= k_1 n + k_2 n = (k_1 + k_2)n\end{aligned}$$

or, as a congruence statement, $a + c \equiv b + d \pmod n$. As regards the second assertion of property (d), note that

$$ac = (b + k_1 n)(d + k_2 n) = bd + (bk_2 + dk_1 + k_1 k_2 n)n$$

Because $bk_2 + dk_1 + k_1 k_2 n$ is an integer, this says that $ac - bd$ is divisible by n, whence $ac \equiv bd \pmod n$.

The proof of property (e) is covered by (d) and the fact that $c \equiv c \pmod n$. Finally, we obtain property (f) by making an induction argument. The statement certainly holds for $k = 1$, and we will assume it is true for some fixed k. From (d), we know

that $a \equiv b \pmod{n}$ and $a^k \equiv b^k \pmod{n}$ together imply that $aa^k \equiv bb^k \pmod{n}$, or equivalently $a^{k+1} \equiv b^{k+1} \pmod{n}$. This is the form the statement should take for $k+1$, and so the induction step is complete.

Before going further, we should illustrate that congruences can be a great help in carrying out certain types of computations.

Example 4.2. Let us endeavor to show that 41 divides $2^{20} - 1$. We begin by noting that $2^5 \equiv -9 \pmod{41}$, whence $(2^5)^4 \equiv (-9)^4 \pmod{41}$ by Theorem 4.2(f); in other words, $2^{20} \equiv 81 \cdot 81 \pmod{41}$. But $81 \equiv -1 \pmod{41}$, and so $81 \cdot 81 \equiv 1 \pmod{41}$. Using parts (b) and (e) of Theorem 4.2, we finally arrive at

$$2^{20} - 1 \equiv 81 \cdot 81 - 1 \equiv 1 - 1 \equiv 0 \pmod{41}$$

Thus, $41 \mid 2^{20} - 1$, as desired.

Example 4.3. For another example in the same spirit, suppose that we are asked to find the remainder obtained upon dividing the sum

$$1! + 2! + 3! + 4! + \cdots + 99! + 100!$$

by 12. Without the aid of congruences this would be an awesome calculation. The observation that starts us off is that $4! \equiv 24 \equiv 0 \pmod{12}$; thus, for $k \geq 4$,

$$k! \equiv 4! \cdot 5 \cdot 6 \cdots k \equiv 0 \cdot 5 \cdot 6 \cdots k \equiv 0 \pmod{12}$$

In this way, we find that

$$\begin{aligned} 1! + 2! + 3! + 4! + \cdots + 100! \\ \equiv 1! + 2! + 3! + 0 + \cdots + 0 \equiv 9 \pmod{12} \end{aligned}$$

Accordingly, the sum in question leaves a remainder of 9 when divided by 12.

In Theorem 4.1 we saw that if $a \equiv b \pmod{n}$, then $ca \equiv cb \pmod{n}$ for any integer c. The converse, however, fails to hold. As an example, perhaps as simple as any, note that $2 \cdot 4 \equiv 2 \cdot 1 \pmod{6}$, whereas $4 \not\equiv 1 \pmod{6}$. In brief: One cannot unrestrictedly cancel a common factor in the arithmetic of congruences.

With suitable precautions, cancellation can be allowed; one step in this direction, and an important one, is provided by the following theorem.

Theorem 4.3. If $ca \equiv cb \pmod{n}$, then $a \equiv b \pmod{n/d}$, where $d = \gcd(c, n)$.

Proof. By hypothesis, we can write

$$c(a - b) = ca - cb = kn$$

for some integer k. Knowing that $\gcd(c, n) = d$, there exist relatively prime integers r and s satisfying $c = dr$, $n = ds$. When these values are substituted in the displayed equation and the common factor d canceled, the net result is

$$r(a - b) = ks$$

Hence, $s \mid r(a - b)$ and $\gcd(r, s) = 1$. Euclid's lemma yields $s \mid a - b$, which may be recast as $a \equiv b \pmod{s}$; in other words, $a \equiv b \pmod{n/d}$.

Theorem 4.3 gets its maximum force when the requirement that $\gcd(c, n) = 1$ is added, for then the cancellation may be accomplished without a change in modulus.

Corollary 1. If $ca \equiv cb \pmod{n}$ and $\gcd(c, n) = 1$, then $a \equiv b \pmod{n}$.

We take a moment to record a special case of Corollary 1 that we shall have frequent occasion to use, namely, Corollary 2.

Corollary 2. If $ca \equiv cb \pmod{p}$ and $p \nmid c$, where p is a prime number, then $a \equiv b \pmod{p}$.

Proof. The conditions $p \nmid c$ and p a prime imply that $\gcd(c, p) = 1$.

Example 4.4. Consider the congruence $33 \equiv 15 \pmod{9}$ or, if one prefers, $3 \cdot 11 \equiv 3 \cdot 5 \pmod{9}$. Because $\gcd(3, 9) = 3$, Theorem 4.3 leads to the conclusion that $11 \equiv 5 \pmod{3}$. A further illustration is given by the congruence $-35 \equiv 45 \pmod{8}$, which is the same as $5 \cdot (-7) \equiv 5 \cdot 9 \pmod{8}$. The integers 5 and 8 being relatively prime, we may cancel the factor 5 to obtain a correct congruence $-7 \equiv 9 \pmod{8}$.

Let us call attention to the fact that, in Theorem 4.3, it is unnecessary to stipulate that $c \not\equiv 0 \pmod{n}$. Indeed, if $c \equiv 0 \pmod{n}$, then $\gcd(c, n) = n$ and the conclusion of the theorem would state that $a \equiv b \pmod{1}$; but, as we remarked earlier, this holds trivially for all integers a and b.

There is another curious situation that can arise with congruences: The product of two integers, neither of which is congruent to zero, may turn out to be congruent to zero. For instance, $4 \cdot 3 \equiv 0 \pmod{12}$, but $4 \not\equiv 0 \pmod{12}$ and $3 \not\equiv 0 \pmod{12}$. It is a simple matter to show that if $ab \equiv 0 \pmod{n}$ and $\gcd(a, n) = 1$, then $b \equiv 0 \pmod{n}$: Corollary 1 permits us legitimately to cancel the factor a from both sides of the congruence $ab \equiv a \cdot 0 \pmod{n}$. A variation on this is that when $ab \equiv 0 \pmod{p}$, with p a prime, then either $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$.

PROBLEMS 4.2

1. Prove each of the following assertions:
   (a) If $a \equiv b \pmod{n}$ and $m \mid n$, then $a \equiv b \pmod{m}$.
   (b) If $a \equiv b \pmod{n}$ and $c > 0$, then $ca \equiv cb \pmod{cn}$.
   (c) If $a \equiv b \pmod{n}$ and the integers a, b, n are all divisible by $d > 0$, then $a/d \equiv b/d \pmod{n/d}$.
2. Give an example to show that $a^2 \equiv b^2 \pmod{n}$ need not imply that $a \equiv b \pmod{n}$.
3. If $a \equiv b \pmod{n}$, prove that $\gcd(a, n) = \gcd(b, n)$.
4. (a) Find the remainders when 2^{50} and 41^{65} are divided by 7.
   (b) What is the remainder when the following sum is divided by 4?

$$1^5 + 2^5 + 3^5 + \cdots + 99^5 + 100^5$$

5. Prove that the integer $53^{103} + 103^{53}$ is divisible by 39, and that $111^{333} + 333^{111}$ is divisible by 7.

6. For $n \geq 1$, use congruence theory to establish each of the following divisibility statements:
(a) $7 \mid 5^{2n} + 3 \cdot 2^{5n-2}$.
(b) $13 \mid 3^{n+2} + 4^{2n+1}$.
(c) $27 \mid 2^{5n+1} + 5^{n+2}$.
(d) $43 \mid 6^{n+2} + 7^{2n+1}$.

7. For $n \geq 1$, show that

$$(-13)^{n+1} \equiv (-13)^n + (-13)^{n-1} \pmod{181}$$

[*Hint:* Notice that $(-13)^2 \equiv -13 + 1 \pmod{181}$; use induction on n.]

8. Prove the assertions below:
(a) If a is an odd integer, then $a^2 \equiv 1 \pmod 8$.
(b) For any integer a, $a^3 \equiv 0, 1$, or $6 \pmod 7$.
(c) For any integer a, $a^4 \equiv 0$ or $1 \pmod 5$.
(d) If the integer a is not divisible by 2 or 3, then $a^2 \equiv 1 \pmod{24}$.

9. If p is a prime satisfying $n < p < 2n$, show that

$$\binom{2n}{n} \equiv 0 \pmod p$$

10. If $a_1, a_2, \ldots, a_n$ is a complete set of residues modulo n and $\gcd(a, n) = 1$, prove that $aa_1, aa_2, \ldots, aa_n$ is also a complete set of residues modulo n.
[*Hint:* It suffices to show that the numbers in question are incongruent modulo n.]

11. Verify that $0, 1, 2, 2^2, 2^3, \ldots, 2^9$ form a complete set of residues modulo 11, but that $0, 1^2, 2^2, 3^2, \ldots, 10^2$ do not.

12. Prove the following statements:
(a) If $\gcd(a, n) = 1$, then the integers

$$c, c + a, c + 2a, c + 3a, \ldots, c + (n - 1)a$$

form a complete set of residues modulo n for any c.
(b) Any n consecutive integers form a complete set of residues modulo n.
[*Hint:* Use part (a).]
(c) The product of any set of n consecutive integers is divisible by n.

13. Verify that if $a \equiv b \pmod{n_1}$ and $a \equiv b \pmod{n_2}$, then $a \equiv b \pmod n$, where the integer $n = \operatorname{lcm}(n_1, n_2)$. Hence, whenever n_1 and n_2 are relatively prime, $a \equiv b \pmod{n_1 n_2}$.

14. Give an example to show that $a^k \equiv b^k \pmod n$ and $k \equiv j \pmod n$ need not imply that $a^j \equiv b^j \pmod n$.

15. Establish that if a is an odd integer, then for any $n \geq 1$

$$a^{2^n} \equiv 1 \pmod{2^{n+2}}$$

[*Hint:* Proceed by induction on n.]

16. Use the theory of congruences to verify that

$$89 \mid 2^{44} - 1 \qquad \text{and} \qquad 97 \mid 2^{48} - 1$$

17. Prove that whenever $ab \equiv cd \pmod n$ and $b \equiv d \pmod n$, with $\gcd(b, n) = 1$, then $a \equiv c \pmod n$.

18. If $a \equiv b \pmod{n_1}$ and $a \equiv c \pmod{n_2}$, prove that $b \equiv c \pmod n$, where the integer $n = \gcd(n_1, n_2)$.

4.3 BINARY AND DECIMAL REPRESENTATIONS OF INTEGERS

One of the more interesting applications of congruence theory involves finding special criteria under which a given integer is divisible by another integer. At their heart, these divisibility tests depend on the notational system used to assign "names" to integers and, more particularly, to the fact that 10 is taken as the base for our number system. Let us, therefore, start by showing that, given an integer $b > 1$, any positive integer N can be written uniquely in terms of powers of b as

$$N = a_m b^m + a_{m-1} b^{m-1} + \cdots + a_2 b^2 + a_1 b + a_0$$

where the coefficients a_k can take on the b different values $0, 1, 2, \ldots, b-1$. For the Division Algorithm yields integers q_1 and a_0 satisfying

$$N = q_1 b + a_0 \qquad 0 \leq a_0 < b$$

If $q_1 \geq b$, we can divide once more, obtaining

$$q_1 = q_2 b + a_1 \qquad 0 \leq a_1 < b$$

Now substitute for q_1 in the earlier equation to get

$$N = (q_2 b + a_1)b + a_0 = q_2 b^2 + a_1 b + a_0$$

As long as $q_2 \geq b$, we can continue in the same fashion. Going one more step: $q_2 = q_3 b + a_2$, where $0 \leq a_2 < b$; hence

$$N = q_3 b^3 + a_2 b^2 + a_1 b + a_0$$

Because $N > q_1 > q_2 > \cdots \geq 0$ is a strictly decreasing sequence of integers, this process must eventually terminate, say, at the $(m-1)$th stage, where

$$q_{m-1} = q_m b + a_{m-1} \qquad 0 \leq a_{m-1} < b$$

and $0 \leq q_m < b$. Setting $a_m = q_m$, we reach the representation

$$N = a_m b^m + a_{m-1} b^{m-1} + \cdots + a_1 b + a_0$$

which was our aim.

To show uniqueness, let us suppose that N has two distinct representations, say,

$$N = a_m b^m + \cdots + a_1 b + a_0 = c_m b^m + \cdots + c_1 b + c_0$$

with $0 \leq a_i < b$ for each i and $0 \leq c_j < b$ for each j (we can use the same m by simply adding terms with coefficients $a_i = 0$ or $c_j = 0$, if necessary). Subtracting the second representation from the first gives the equation

$$0 = d_m b^m + \cdots + d_1 b + d_0$$

where $d_i = a_i - c_i$ for $i = 0, 1, \ldots, m$. Because the two representations for N are assumed to be different, we must have $d_i \neq 0$ for some value of i. Take k to be the smallest subscript for which $d_k \neq 0$. Then

$$0 = d_m b^m + \cdots + d_{k+1} b^{k+1} + d_k b^k$$

and so, after dividing by b^k,

$$d_k = -b(d_m b^{m-k-1} + \cdots + d_{k+1})$$

This tells us that $b \mid d_k$. Now the inequalities $0 \leq a_k < b$ and $0 \leq c_k < b$ lead us to $-b < a_k - c_k < b$, or $|d_k| < b$. The only way of reconciling the conditions $b \mid d_k$ and $|d_k| < b$ is to have $d_k = 0$, which is impossible. From this contradiction, we conclude that the representation of N is unique.

The essential feature in all of this is that the integer N is completely determined by the ordered array $a_m, a_{m-1}, \ldots, a_1, a_0$ of coefficients, with the plus signs and the powers of b being superfluous. Thus, the number

$$N = a_m b^m + a_{m-1} b^{m-1} + \cdots + a_2 b^2 + a_1 b + a_0$$

may be replaced by the simpler symbol

$$N = (a_m a_{m-1} \cdots a_2 a_1 a_0)_b$$

(the right-hand side is not to be interpreted as a product, but only as an abbreviation for N). We call this the *base b place-value notation for N*.

Small values of b give rise to lengthy representation of numbers, but have the advantage of requiring fewer choices for coefficients. The simplest case occurs when the base $b = 2$, and the resulting system of enumeration is called the *binary number system* (from the Latin *binarius*, two). The fact that when a number is written in the binary system only the integers 0 and 1 can appear as coefficients means that every positive integer is expressible in exactly one way as a sum of distinct powers of 2. For example, the integer 105 can be written as

$$\begin{aligned} 105 &= 1 \cdot 2^6 + 1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2 + 1 \\ &= 2^6 + 2^5 + 2^3 + 1 \end{aligned}$$

or, in abbreviated form,

$$105 = (1101001)_2$$

In the other direction, $(1001111)_2$ translates into

$$1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2 + 1 = 79$$

The binary system is most convenient for use in modern electronic computing machines, because binary numbers are represented by strings of zeros and ones; 0 and 1 can be expressed in the machine by a switch (or a similar electronic device) being either on or off.

We shall frequently wish to calculate the value of $a^k \pmod{n}$ when k is large. Is there a more efficient way of obtaining the least positive residue than multiplying a by itself k times before reducing modulo n? One such procedure, called the *binary exponential algorithm*, relies on successive squarings, with a reduction modulo n after each squaring. More specifically, the exponent k is written in binary form, as $k = (a_m a_{m-1} \ldots a_2 a_1 a_0)_2$, and the values $a^{2^j} \pmod{n}$ are calculated for the powers of 2, which correspond to the 1's in the binary representation. These partial results are then multiplied together to give the final answer.

An illustration should make this process clear.

Example 4.5. To calculate $5^{110} \pmod{131}$, first note that the exponent 110 can be expressed in binary form as

$$110 = 64 + 32 + 8 + 4 + 2 = (110110)_2$$

Thus, we obtain the powers $5^{2^j} \pmod{131}$ for $0 \leq j \leq 6$ by repeatedly squaring while at each stage reducing each result modulo 131:

$$\begin{aligned} 5^2 &\equiv 25 \pmod{131} & 5^{16} &\equiv 27 \pmod{131} \\ 5^4 &\equiv 101 \pmod{131} & 5^{32} &\equiv 74 \pmod{131} \\ 5^8 &\equiv 114 \pmod{131} & 5^{64} &\equiv 105 \pmod{131} \end{aligned}$$

When the appropriate partial results—those corresponding to the 1's in the binary expansion of 110—are multiplied, we see that

$$\begin{aligned} 5^{110} &= 5^{64+32+8+4+2} \\ &= 5^{64} \cdot 5^{32} \cdot 5^8 \cdot 5^4 \cdot 5^2 \\ &\equiv 105 \cdot 74 \cdot 114 \cdot 101 \cdot 25 \equiv 60 \pmod{131} \end{aligned}$$

As a minor variation of the procedure, one might calculate, modulo 131, the powers $5, 5^2, 5^3, 5^6, 5^{12}, 5^{24}, 5^{48}, 5^{96}$ to arrive at

$$5^{110} = 5^{96} \cdot 5^{12} \cdot 5^2 \equiv 41 \cdot 117 \cdot 25 \equiv 60 \pmod{131}$$

which would require two fewer multiplications.

We ordinarily record numbers in the *decimal system* of notation, where $b = 10$, omitting the 10-subscript that specifies the base. For instance, the symbol 1492 stands for the more awkward expression

$$1 \cdot 10^3 + 4 \cdot 10^2 + 9 \cdot 10 + 2$$

The integers 1, 4, 9, and 2 are called the *digits* of the given number, 1 being the thousands digit, 4 the hundreds digit, 9 the tens digit, and 2 the units digit. In technical language we refer to the representation of the positive integers as sums of powers of 10, with coefficients at most 9, as their *decimal representation* (from the Latin *decem*, ten).

We are about ready to derive criteria for determining whether an integer is divisible by 9 or 11, without performing the actual division. For this, we need a result having to do with congruences involving polynomials with integral coefficients.

Theorem 4.4. Let $P(x) = \sum_{k=0}^{m} c_k x^k$ be a polynomial function of x with integral coefficients c_k. If $a \equiv b \pmod{n}$, then $P(a) \equiv P(b) \pmod{n}$.

Proof. Because $a \equiv b \pmod{n}$, part (f) of Theorem 4.2 can be applied to give $a^k \equiv b^k \pmod{n}$ for $k = 0, 1, \ldots, m$. Therefore,

$$c_k a^k \equiv c_k b^k \pmod{n}$$

for all such k. Adding these $m + 1$ congruences, we conclude that

$$\sum_{k=0}^{m} c_k a^k \equiv \sum_{k=0}^{m} c_k b^k \pmod{n}$$

or, in different notation, $P(a) \equiv P(b) \pmod{n}$.

If $P(x)$ is a polynomial with integral coefficients, we say that a is a solution of the congruence $P(x) \equiv 0 \pmod{n}$ if $P(a) \equiv 0 \pmod{n}$.

Corollary. If a is a solution of $P(x) \equiv 0 \pmod{n}$ and $a \equiv b \pmod{n}$, then b also is a solution.

Proof. From the last theorem, it is known that $P(a) \equiv P(b) \pmod{n}$. Hence, if a is a solution of $P(x) \equiv 0 \pmod{n}$, then $P(b) \equiv P(a) \equiv 0 \pmod{n}$, making b a solution.

One divisibility test that we have in mind is this. A positive integer is divisible by 9 if and only if the sum of the digits in its decimal representation is divisible by 9.

Theorem 4.5. Let $N = a_m 10^m + a_{m-1}10^{m-1} + \cdots + a_1 10 + a_0$ be the decimal expansion of the positive integer N, $0 \leq a_k < 10$, and let $S = a_0 + a_1 + \cdots + a_m$. Then $9 \mid N$ if and only if $9 \mid S$.

Proof. Consider $P(x) = \sum_{k=0}^{m} a_k x^k$, a polynomial with integral coefficients. The key observation is that $10 \equiv 1 \pmod{9}$, whence by Theorem 4.4, $P(10) \equiv P(1) \pmod{9}$. But $P(10) = N$ and $P(1) = a_0 + a_1 + \cdots + a_m = S$, so that $N \equiv S \pmod{9}$. It follows that $N \equiv 0 \pmod{9}$ if and only if $S \equiv 0 \pmod{9}$, which is what we wanted to prove.

Theorem 4.4 also serves as the basis for a well-known test for divisibility by 11: an integer is divisible by 11 if and only if the alternating sum of its digits is divisible by 11. We state this more precisely by Theorem 4.6.

Theorem 4.6. Let $N = a_m 10^m + a_{m-1}10^{m-1} + \cdots + a_1 10 + a_0$ be the decimal expansion of the positive integer N, $0 \leq a_k < 10$, and let $T = a_0 - a_1 + a_2 - \cdots + (-1)^m a_m$. Then $11 \mid N$ if and only if $11 \mid T$.

Proof. As in the proof of Theorem 4.5, put $P(x) = \sum_{k=0}^{m} a_k x^k$. Because $10 \equiv -1 \pmod{11}$, we get $P(10) \equiv P(-1) \pmod{11}$. But $P(10) = N$, whereas $P(-1) = a_0 - a_1 + a_2 - \cdots + (-1)^m a_m = T$, so that $N \equiv T \pmod{11}$. The implication is that either both N and T are divisible by 11 or neither is divisible by 11.

Example 4.6. To see an illustration of the last two results, take the integer $N = 1{,}571{,}724$. Because the sum

$$1 + 5 + 7 + 1 + 7 + 2 + 4 = 27$$

is divisible by 9, Theorem 4.5 guarantees that 9 divides N. It also can be divided by 11; for, the alternating sum

$$4 - 2 + 7 - 1 + 7 - 5 + 1 = 11$$

is divisible by 11.

Congruence theory is frequently used to append an extra check digit to identification numbers, in order to recognize transmission errors or forgeries. Personal

identification numbers of some kind appear on passports, credit cards, bank accounts, and a variety of other settings.

Some banks use an eight-digit identification number $a_1a_2 \ldots a_8$ together with a final check digit a_9. The check digit is usually obtained by multiplying the digits $a_i (1 \leq i \leq 8)$ by certain "weights" and calculating the sum of the weighted products modulo 10. For instance, the check digit might be chosen to satisfy

$$a_9 \equiv 7a_1 + 3a_2 + 9a_3 + 7a_4 + 3a_5 + 9a_6 + 7a_7 + 3a_8 \pmod{10}$$

The identification number 81504216 would then have check digit

$$a_9 \equiv 7 \cdot 8 + 3 \cdot 1 + 9 \cdot 5 + 7 \cdot 0 + 3 \cdot 4 + 9 \cdot 2 + 7 \cdot 1 + 3 \cdot 6 \equiv 9 \pmod{10}$$

so that 815042169 would be printed on the check.

This weighting scheme for assigning check digits detects any single-digit error in the identification number. For suppose that the digit a_i is replaced by a different a_i'. By the manner in which the check digit is calculated, the difference between the correct a_9 and the new a_9' is

$$a_9 - a_9' \equiv k(a_i - a_i') \pmod{10}$$

where k is 7, 3, or 9 depending on the position of a_i'. Because $k(a_i - a_i') \not\equiv 0 \pmod{10}$, it follows that $a_9 \neq a_9'$ and the error is apparent. Thus, if the valid number 81504216 were incorrectly entered as 81504316 into a computer programmed to calculate check digits, an 8 would come up rather than the expected 9.

The modulo 10 approach is not entirely effective, for it does not always detect the common error of transposing distinct adjacent entries a and b within the string of digits. To illustrate: the identification numbers 81504216 and 81504261 have the same check digit 9 when our example weights are used. (The problem occurs when $|a - b| = 5$.) More sophisticated methods are available, with larger moduli and different weights, that would prevent this possible error.

PROBLEMS 4.3

1. Use the binary exponentiation algorithm to compute both $19^{53} \pmod{503}$ and $141^{47} \pmod{1537}$.
2. Prove the following statements:
   (a) For any integer a, the units digit of a^2 is 0, 1, 4, 5, 6, or 9.
   (b) Any one of the integers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 can occur as the units digit of a^3.
   (c) For any integer a, the units digit of a^4 is 0, 1, 5, or 6.
   (d) The units digit of a triangular number is 0, 1, 3, 5, 6, or 8.
3. Find the last two digits of the number 9^{9^9}.
   [*Hint:* $9^9 \equiv 9 \pmod{10}$; hence, $9^{9^9} = 9^{9+10k}$; now use the fact that $9^9 \equiv 89 \pmod{100}$.]
4. Without performing the divisions, determine whether the integers 176,521,221 and 149,235,678 are divisible by 9 or 11.
5. (a) Obtain the following generalization of Theorem 4.6: If the integer N is represented in the base b by

$$N = a_m b^m + \cdots + a_2 b^2 + a_1 b + a_0 \qquad 0 \leq a_k \leq b - 1$$

   then $b - 1 \mid N$ if and only if $b - 1 \mid (a_m + \cdots + a_2 + a_1 + a_0)$.

(b) Give criteria for the divisibility of N by 3 and 8 that depend on the digits of N when written in the base 9.
(c) Is the integer $(447836)_9$ divisible by 3 and 8?

6. Working modulo 9 or 11, find the missing digits in the calculations below:
(a) $51840 \cdot 273581 = 1418243x040$.
(b) $2x99561 = [3(523 + x)]^2$.
(c) $2784x = x \cdot 5569$.
(d) $512 \cdot 1x53125 = 1000000000$.

7. Establish the following divisibility criteria:
(a) An integer is divisible by 2 if and only if its units digit is 0, 2, 4, 6, or 8.
(b) An integer is divisible by 3 if and only if the sum of its digits is divisible by 3.
(c) An integer is divisible by 4 if and only if the number formed by its tens and units digits is divisible by 4.
[*Hint:* $10^k \equiv 0 \pmod 4$ for $k \geq 2$.]
(d) An integer is divisible by 5 if and only if its units digit is 0 or 5.

8. For any integer a, show that $a^2 - a + 7$ ends in one of the digits 3, 7, or 9.

9. Find the remainder when 4444^{4444} is divided by 9.
[*Hint:* Observe that $2^3 \equiv -1 \pmod 9$.]

10. Prove that no integer whose digits add up to 15 can be a square or a cube.
[*Hint:* For any a, $a^3 \equiv 0$, 1, or 8 (mod 9).]

11. Assuming that 495 divides $273x49y5$, obtain the digits x and y.

12. Determine the last three digits of the number 7^{999}.
[*Hint:* $7^{4n} \equiv (1 + 400)^n \equiv 1 + 400n \pmod{1000}$.]

13. If t_n denotes the nth triangular number, show that $t_{n+2k} \equiv t_n \pmod k$; hence, t_n and t_{n+20} must have the same last digit.

14. For any $n \geq 1$, prove that there exists a prime with at least n of its digits equal to 0.
[*Hint:* Consider the arithmetic progression $10^{n+1}k + 1$ for $k = 1, 2, \ldots$.]

15. Find the values of $n \geq 1$ for which $1! + 2! + 3! + \cdots + n!$ is a perfect square.
[*Hint:* Problem 2(a).]

16. Show that 2^n divides an integer N if and only if 2^n divides the number made up of the last n digits of N.
[*Hint:* $10^k = 2^k 5^k \equiv 0 \pmod{2^n}$ for $k \geq n$.]

17. Let $N = a_m 10^m + \cdots + a_2 10^2 + a_1 10 + a_0$, where $0 \leq a_k \leq 9$, be the decimal expansion of a positive integer N.
(a) Prove that 7, 11, and 13 all divide N if and only if 7, 11, and 13 divide the integer

$$M = (100a_2 + 10a_1 + a_0) - (100a_5 + 10a_4 + a_3) + (100a_8 + 10a_7 + a_6) - \cdots$$

[*Hint:* If n is even, then $10^{3n} \equiv 1$, $10^{3n+1} \equiv 10$, $10^{3n+2} \equiv 100 \pmod{1001}$; if n is odd, then $10^{3n} \equiv -1$, $10^{3n+1} \equiv -10$, $10^{3n+2} \equiv -100 \pmod{1001}$.]
(b) Prove that 6 divides N if and only if 6 divides the integer

$$M = a_0 + 4a_1 + 4a_2 + \cdots + 4a_m$$

18. Without performing the divisions, determine whether the integer 1010908899 is divisible by 7, 11, and 13.

19. (a) Given an integer N, let M be the integer formed by reversing the order of the digits of N (for example, if $N = 6923$, then $M = 3296$). Verify that $N - M$ is divisible by 9.

(b) A *palindrome* is a number that reads the same backwards as forwards (for instance, 373 and 521125 are palindromes). Prove that any palindrome with an even number of digits is divisible by 11.

20. Given a repunit R_n, show that
(a) $9 \mid R_n$ if and only if $9 \mid n$.
(b) $11 \mid R_n$ if and only if n is even.

21. Factor the repunit $R_6 = 111111$ into a product of primes.
[*Hint:* Problem 17(a).]

22. Explain why the following curious calculations hold:

$$\begin{aligned}
1 \cdot 9 + 2 &= 11 \\
12 \cdot 9 + 3 &= 111 \\
123 \cdot 9 + 4 &= 1111 \\
1234 \cdot 9 + 5 &= 11111 \\
12345 \cdot 9 + 6 &= 111111 \\
123456 \cdot 9 + 7 &= 1111111 \\
1234567 \cdot 9 + 8 &= 11111111 \\
12345678 \cdot 9 + 9 &= 111111111 \\
123456789 \cdot 9 + 10 &= 1111111111
\end{aligned}$$

[*Hint:* Show that

$$(10^{n-1} + 2 \cdot 10^{n-2} + 3 \cdot 10^{n-3} + \cdots + n)(10 - 1) + (n + 1) = \frac{10^{n+1} - 1}{9}.]$$

23. An old and somewhat illegible invoice shows that 72 canned hams were purchased for $\$x\,67.9y$. Find the missing digits.

24. If 792 divides the integer $13xy\,45z$, find the digits x, y, and z.
[*Hint:* By Problem 17, $8 \mid 45z$.]

25. For any prime $p > 3$ prove that 13 divides $10^{2p} - 10^p + 1$.

26. Consider the eight-digit bank identification number $a_1a_2 \ldots a_8$, which is followed by a ninth check digit a_9 chosen to satisfy the congruence

$$a_9 \equiv 7a_1 + 3a_2 + 9a_3 + 7a_4 + 3a_5 + 9a_6 + 7a_7 + 3a_8 \pmod{10}$$

(a) Obtain the check digits that should be appended to the two numbers 55382006 and 81372439.
(b) The bank identification number $237a_418538$ has an illegible fourth digit. Determine the value of the obscured digit.

27. The International Standard Book Number (ISBN) used in many libraries consists of nine digits $a_1a_2 \ldots a_9$ followed by a tenth check digit a_{10}, which satisfies

$$a_{10} \equiv \sum_{k=1}^{9} ka_k \pmod{11}$$

Determine whether each of the ISBNs below is correct:
(a) 0-07-232569-0 (United States).
(b) 91-7643-497-5 (Sweden).
(c) 1-56947-303-10 (England).

28. When printing the ISBN $a_1a_2 \ldots a_9$, two unequal digits were transposed. Show that the check digits detected this error.

4.4 LINEAR CONGRUENCES AND THE CHINESE REMAINDER THEOREM

This is a convenient place in our development of number theory at which to investigate the theory of linear congruences: An equation of the form $ax \equiv b \pmod{n}$ is called a *linear congruence*, and by a solution of such an equation we mean an integer x_0 for which $ax_0 \equiv b \pmod{n}$. By definition, $ax_0 \equiv b \pmod{n}$ if and only if $n \mid ax_0 - b$ or, what amounts to the same thing, if and only if $ax_0 - b = ny_0$ for some integer y_0. Thus, the problem of finding all integers that will satisfy the linear congruence $ax \equiv b \pmod{n}$ is identical with that of obtaining all solutions of the linear Diophantine equation $ax - ny = b$. This allows us to bring the results of Chapter 2 into play.

It is convenient to treat two solutions of $ax \equiv b \pmod{n}$ that are congruent modulo n as being "equal" even though they are not equal in the usual sense. For instance, $x = 3$ and $x = -9$ both satisfy the congruence $3x \equiv 9 \pmod{12}$; because $3 \equiv -9 \pmod{12}$, they are not counted as different solutions. In short: When we refer to the number of solutions of $ax \equiv b \pmod{n}$, we mean the number of incongruent integers satisfying this congruence.

With these remarks in mind, the principal result is easy to state.

Theorem 4.7. The linear congruence $ax \equiv b \pmod{n}$ has a solution if and only if $d \mid b$, where $d = \gcd(a, n)$. If $d \mid b$, then it has d mutually incongruent solutions modulo n.

Proof. We already have observed that the given congruence is equivalent to the linear Diophantine equation $ax - ny = b$. From Theorem 2.9, it is known that the latter equation can be solved if and only if $d \mid b$; moreover, if it is solvable and x_0, y_0 is one specific solution, then any other solution has the form

$$x = x_0 + \frac{n}{d}t \qquad y = y_0 + \frac{a}{d}t$$

for some choice of t.

Among the various integers satisfying the first of these formulas, consider those that occur when t takes on the successive values $t = 0, 1, 2, \ldots, d - 1$:

$$x_0, x_0 + \frac{n}{d}, x_0 + \frac{2n}{d}, \ldots, x_0 + \frac{(d-1)n}{d}$$

We claim that these integers are incongruent modulo n, and all other such integers x are congruent to some one of them. If it happened that

$$x_0 + \frac{n}{d}t_1 \equiv x_0 + \frac{n}{d}t_2 \pmod{n}$$

where $0 \leq t_1 < t_2 \leq d - 1$, then we would have

$$\frac{n}{d}t_1 \equiv \frac{n}{d}t_2 \pmod{n}$$

Now $\gcd(n/d, n) = n/d$, and therefore by Theorem 4.3 the factor n/d could be canceled to arrive at the congruence

$$t_1 \equiv t_2 \pmod{d}$$

which is to say that $d \mid t_2 - t_1$. But this is impossible in view of the inequality $0 < t_2 - t_1 < d$.

It remains to argue that any other solution $x_0 + (n/d)t$ is congruent modulo n to one of the d integers listed above. The Division Algorithm permits us to write t as $t = qd + r$, where $0 \leq r \leq d - 1$. Hence

$$\begin{aligned} x_0 + \frac{n}{d}t &= x_0 + \frac{n}{d}(qd + r) \\ &= x_0 + nq + \frac{n}{d}r \\ &\equiv x_0 + \frac{n}{d}r \pmod{n} \end{aligned}$$

with $x_0 + (n/d)r$ being one of our d selected solutions. This ends the proof.

The argument that we gave in Theorem 4.7 brings out a point worth stating explicitly: If x_0 is any solution of $ax \equiv b \pmod{n}$, then the $d = \gcd(a, n)$ incongruent solutions are given by

$$x_0, x_0 + \frac{n}{d}, x_0 + 2\left(\frac{n}{d}\right), \ldots, x_0 + (d-1)\left(\frac{n}{d}\right)$$

For the reader's convenience, let us also record the form Theorem 4.7 takes in the special case in which a and n are assumed to be relatively prime.

Corollary. If $\gcd(a, n) = 1$, then the linear congruence $ax \equiv b \pmod{n}$ has a unique solution modulo n.

Given relatively prime integers a and n, the congruence $ax \equiv 1 \pmod{n}$ has a unique solution. This solution is sometimes called the (multiplicative) inverse of a modulo n.

We now pause to look at two concrete examples.

Example 4.7. First consider the linear congruence $18x \equiv 30 \pmod{42}$. Because $\gcd(18, 42) = 6$ and 6 surely divides 30, Theorem 4.7 guarantees the existence of exactly six solutions, which are incongruent modulo 42. By inspection, one solution is found to be $x = 4$. Our analysis tells us that the six solutions are as follows:

$$x \equiv 4 + (42/6)t \equiv 4 + 7t \pmod{42} \qquad t = 0, 1, \ldots, 5$$

or, plainly enumerated,

$$x \equiv 4, 11, 18, 25, 32, 39 \pmod{42}$$

Example 4.8. Let us solve the linear congruence $9x \equiv 21 \pmod{30}$. At the outset, because $\gcd(9, 30) = 3$ and $3 \mid 21$, we know that there must be three incongruent solutions.

One way to find these solutions is to divide the given congruence through by 3, thereby replacing it by the equivalent congruence $3x \equiv 7 \pmod{10}$. The relative primeness of 3 and 10 implies that the latter congruence admits a unique solution modulo 10. Although it is not the most efficient method, we could test the integers

0, 1, 2, . . . ,9 in turn until the solution is obtained. A better way is this: Multiply both sides of the congruence $3x \equiv 7 \pmod{10}$ by 7 to get

$$21x \equiv 49 \pmod{10}$$

which reduces to $x \equiv 9 \pmod{10}$. (This simplification is no accident, for the multiples $0 \cdot 3,\ 1 \cdot 3,\ 2 \cdot 3, \ldots, 9 \cdot 3$ form a complete set of residues modulo 10; hence, one of them is necessarily congruent to 1 modulo 10.) But the original congruence was given modulo 30, so that its incongruent solutions are sought among the integers 0, 1, 2, . . . , 29. Taking $t = 0, 1, 2$, in the formula

$$x = 9 + 10t$$

we obtain 9, 19, 29, whence

$$x \equiv 9 \pmod{30} \qquad x \equiv 19 \pmod{30} \qquad x \equiv 29 \pmod{30}$$

are the required three solutions of $9x \equiv 21 \pmod{30}$.

A different approach to the problem is to use the method that is suggested in the proof of Theorem 4.7. Because the congruence $9x \equiv 21 \pmod{30}$ is equivalent to the linear Diophantine equation

$$9x - 30y = 21$$

we begin by expressing $3 = \gcd(9, 30)$ as a linear combination of 9 and 30. It is found, either by inspection or by using the Euclidean Algorithm, that $3 = 9(-3) + 30 \cdot 1$, so that

$$21 = 7 \cdot 3 = 9(-21) - 30(-7)$$

Thus, $x = -21$, $y = -7$ satisfy the Diophantine equation and, in consequence, all solutions of the congruence in question are to be found from the formula

$$x = -21 + (30/3)t = -21 + 10t$$

The integers $x = -21 + 10t$, where $t = 0, 1, 2$, are incongruent modulo 30 (but all are congruent modulo 10); thus, we end up with the incongruent solutions

$$x \equiv -21 \pmod{30} \qquad x \equiv -11 \pmod{30} \qquad x \equiv -1 \pmod{30}$$

or, if one prefers positive numbers, $x \equiv 9, 19, 29 \pmod{30}$.

Having considered a single linear congruence, it is natural to turn to the problem of solving a system of simultaneous linear congruences:

$$a_1x \equiv b_1 \pmod{m_1},\ a_2x \equiv b_2 \pmod{m_2}, \ldots, a_rx \equiv b_r \pmod{m_r}$$

We shall assume that the moduli m_k are relatively prime in pairs. Evidently, the system will admit no solution unless each individual congruence is solvable; that is, unless $d_k \mid b_k$ for each k, where $d_k = \gcd(a_k, m_k)$. When these conditions are satisfied, the factor d_k can be canceled in the kth congruence to produce a new system having the same set of solutions as the original one:

$$a_1'x \equiv b_1' \pmod{n_1},\ a_2'x \equiv b_2' \pmod{n_2}, \ldots, a_r'x \equiv b_r' \pmod{n_r}$$

where $n_k = m_k/d_k$ and $\gcd(n_i, n_j) = 1$ for $i \neq j$; in addition, $\gcd(a_i', n_i) = 1$. The solutions of the individual congruences assume the form

$$x \equiv c_1 \pmod{n_1}, x \equiv c_2 \pmod{n_2}, \ldots, x \equiv c_r \pmod{n_r}$$

Thus, the problem is reduced to one of finding a simultaneous solution of a system of congruences of this simpler type.

The kind of problem that can be solved by simultaneous congruences has a long history, appearing in the Chinese literature as early as the 1st century A.D. Sun-Tsu asked: Find a number that leaves the remainders 2, 3, 2 when divided by 3, 5, 7, respectively. (Such mathematical puzzles are by no means confined to a single cultural sphere; indeed, the same problem occurs in the *Introductio Arithmeticae* of the Greek mathematician Nicomachus, circa 100 A.D.) In honor of their early contributions, the rule for obtaining a solution usually goes by the name of the Chinese Remainder Theorem.

Theorem 4.8 Chinese Remainder Theorem. Let $n_1, n_2, \ldots, n_r$ be positive integers such that $\gcd(n_i, n_j) = 1$ for $i \neq j$. Then the system of linear congruences

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_r \pmod{n_r} \end{aligned}$$

has a simultaneous solution, which is unique modulo the integer $n_1 n_2 \cdots n_r$.

Proof. We start by forming the product $n = n_1 n_2 \cdots n_r$. For each $k = 1, 2, \ldots, r$, let

$$N_k = \frac{n}{n_k} = n_1 \cdots n_{k-1} n_{k+1} \cdots n_r$$

In words, N_k is the product of all the integers n_i with the factor n_k omitted. By hypothesis, the n_i are relatively prime in pairs, so that $\gcd(N_k, n_k) = 1$. According to the theory of a single linear congruence, it is therefore possible to solve the congruence $N_k x \equiv 1 \pmod{n_k}$; call the unique solution x_k. Our aim is to prove that the integer

$$\bar{x} = a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_r N_r x_r$$

is a simultaneous solution of the given system.

First, observe that $N_i \equiv 0 \pmod{n_k}$ for $i \neq k$, because $n_k \mid N_i$ in this case. The result is

$$\bar{x} = a_1 N_1 x_1 + \cdots + a_r N_r x_r \equiv a_k N_k x_k \pmod{n_k}$$

But the integer x_k was chosen to satisfy the congruence $N_k x \equiv 1 \pmod{n_k}$, which forces

$$\bar{x} = a_k \cdot 1 = a_k \pmod{n_k}$$

This shows that a solution to the given system of congruences exists.

As for the uniqueness assertion, suppose that x' is any other integer that satisfies these congruences. Then

$$\bar{x} \equiv a_k \equiv x' \pmod{n_k} \qquad k = 1, 2, \ldots, r$$

and so $n_k \mid \bar{x} - x'$ for each value of k. Because $\gcd(n_i, n_j) = 1$, Corollary 2 to Theorem 2.4 supplies us with the crucial point that $n_1 n_2 \cdots n_r \mid \bar{x} - x'$; hence $\bar{x} \equiv x' \pmod{n}$. With this, the Chinese Remainder Theorem is proven.

Example 4.9. The problem posed by Sun-Tsu corresponds to the system of three congruences

$$x \equiv 2 \pmod 3$$
$$x \equiv 3 \pmod 5$$
$$x \equiv 2 \pmod 7$$

In the notation of Theorem 4.8, we have $n = 3 \cdot 5 \cdot 7 = 105$ and

$$N_1 = \frac{n}{3} = 35 \qquad N_2 = \frac{n}{5} = 21 \qquad N_3 = \frac{n}{7} = 15$$

Now the linear congruences

$$35x \equiv 1 \pmod 3 \qquad 21x \equiv 1 \pmod 5 \qquad 15x \equiv 1 \pmod 7$$

are satisfied by $x_1 = 2$, $x_2 = 1$, $x_3 = 1$, respectively. Thus, a solution of the system is given by

$$x = 2 \cdot 35 \cdot 2 + 3 \cdot 21 \cdot 1 + 2 \cdot 15 \cdot 1 = 233$$

Modulo 105, we get the unique solution $x = 233 \equiv 23 \pmod{105}$.

Example 4.10. For a second illustration, let us solve the linear congruence

$$17x \equiv 9 \pmod{276}$$

Because $276 = 3 \cdot 4 \cdot 23$, this is equivalent to finding a solution for the system of congruences

$$\begin{array}{lll} 17x \equiv 9 \pmod 3 & \text{or} & x \equiv 0 \pmod 3 \\ 17x \equiv 9 \pmod 4 & & x \equiv 1 \pmod 4 \\ 17x \equiv 9 \pmod{23} & & 17x \equiv 9 \pmod{23} \end{array}$$

Note that if $x \equiv 0 \pmod 3$, then $x = 3k$ for any integer k. We substitute into the second congruence of the system and obtain

$$3k \equiv 1 \pmod 4$$

Multiplication of both sides of this congruence by 3 gives us

$$k \equiv 9k \equiv 3 \pmod 4$$

so that $k = 3 + 4j$, where j is an integer. Then

$$x = 3(3 + 4j) = 9 + 12j$$

For x to satisfy the last congruence, we must have

$$17(9 + 12j) \equiv 9 \pmod{23}$$

or $204j \equiv -144 \pmod{23}$, which reduces to $3j \equiv 6 \pmod{23}$; in consequence, $j \equiv 2 \pmod{23}$. This yields $j = 2 + 23t$, with t an integer, whence

$$x = 9 + 12(2 + 23t) = 33 + 276t$$

All in all, $x \equiv 33 \pmod{276}$ provides a solution to the system of congruences and, in turn, a solution to $17x \equiv 9 \pmod{276}$.

We should say a few words about linear congruences in two variables; that is, congruences of the form

$$ax + by \equiv c \pmod{n}$$

In analogy with Theorem 4.7, such a congruence has a solution if and only if $\gcd(a, b, n)$ divides c. The condition for solvability holds if either $\gcd(a, n) = 1$ or $\gcd(b, n) = 1$. Say $\gcd(a, n) = 1$. When the congruence is expressed as

$$ax \equiv c - by \pmod{n}$$

the corollary to Theorem 4.7 guarantees a unique solution x for each of the n incongruent values of y. Take as a simple illustration $7x + 4y \equiv 5 \pmod{12}$, that would be treated as $7x \equiv 5 - 4y \pmod{12}$. Substitution of $y \equiv 5 \pmod{12}$ gives $7x \equiv -15 \pmod{12}$; but this is equivalent to $-5x \equiv -15 \pmod{12}$ so that $x \equiv 3 \pmod{12}$. It follows that $x \equiv 3 \pmod{12}, y \equiv 5 \pmod{12}$ is one of the 12 incongruent solutions of $7x + 4y \equiv 5 \pmod{12}$. Another solution having the same value of x is $x \equiv 3 \pmod{12}, y \equiv 8 \pmod{12}$.

The focus of our concern here is how to solve a system of two linear congruences in two variables with the same modulus. The proof of the coming theorem adopts the familiar procedure of eliminating one of the unknowns.

Theorem 4.9. The system of linear congruences

$$ax + by \equiv r \pmod{n}$$
$$cx + dy \equiv s \pmod{n}$$

has a unique solution modulo n whenever $\gcd(ad - bc, n) = 1$.

Proof. Let us multiply the first congruence of the system by d, the second congruence by b, and subtract the lower result from the upper. These calculations yield

$$(ad - bc)x \equiv dr - bs \pmod{n} \tag{1}$$

The assumption $\gcd(ad - bc, n) = 1$ ensures that the congruence

$$(ad - bc)z \equiv 1 \pmod{n}$$

posseses a unique solution; denote the solution by t. When congruence (1) is multiplied by t, we obtain

$$x \equiv t(dr - bs) \pmod{n}$$

A value for y is found by a similar elimination process. That is, multiply the first congruence of the system by c, the second one by a, and subtract to end up with

$$(ad - bc)y \equiv as - cr \pmod{n} \tag{2}$$

Multiplication of this congruence by t leads to

$$y \equiv t(as - cr) \pmod{n}$$

A solution of the system is now established.

We close this section with an example illustrating Theorem 4.9.

Example 4.11. Consider the system

$$7x + 3y \equiv 10 \pmod{16}$$
$$2x + 5y \equiv 9 \pmod{16}$$

Because $\gcd(7 \cdot 5 - 2 \cdot 3, 16) = \gcd(29, 16) = 1$, a solution exists. It is obtained by the method developed in the proof of Theorem 4.9. Multiplying the first congruence by 5, the second one by 3, and subtracting, we arrive at

$$29x \equiv 5 \cdot 10 - 3 \cdot 9 \equiv 23 \pmod{16}$$

or, what is the same thing, $13x \equiv 7 \pmod{16}$. Multiplication of this congruence by 5 (noting that $5 \cdot 13 \equiv 1 \pmod{16}$) produces $x \equiv 35 \equiv 3 \pmod{16}$. When the variable x is eliminated from the system of congruences in a like manner, it is found that

$$29y \equiv 7 \cdot 9 - 2 \cdot 10 \equiv 43 \pmod{16}$$

But then $13y \equiv 11 \pmod{16}$, which upon multiplication by 5, results in $y \equiv 55 \equiv 7 \pmod{16}$. The unique solution of our system turns out to be

$$x \equiv 3 \pmod{16} \qquad y \equiv 7 \pmod{16}$$

PROBLEMS 4.4

1. Solve the following linear congruences:
   (a) $25x \equiv 15 \pmod{29}$.
   (b) $5x \equiv 2 \pmod{26}$.
   (c) $6x \equiv 15 \pmod{21}$.
   (d) $36x \equiv 8 \pmod{102}$.
   (e) $34x \equiv 60 \pmod{98}$.
   (f) $140x \equiv 133 \pmod{301}$.
   [*Hint:* $\gcd(140, 301) = 7$.]
2. Using congruences, solve the Diophantine equations below:
   (a) $4x + 51y = 9$.
   [*Hint:* $4x \equiv 9 \pmod{51}$ gives $x = 15 + 51t$, whereas $51y \equiv 9 \pmod{4}$ gives $y = 3 + 4s$. Find the relation between s and t.]
   (b) $12x + 25y = 331$.
   (c) $5x - 53y = 17$.
3. Find all solutions of the linear congruence $3x - 7y \equiv 11 \pmod{13}$.
4. Solve each of the following sets of simultaneous congruences:
   (a) $x \equiv 1 \pmod{3}$, $x \equiv 2 \pmod{5}$, $x \equiv 3 \pmod{7}$.
   (b) $x \equiv 5 \pmod{11}$, $x \equiv 14 \pmod{29}$, $x \equiv 15 \pmod{31}$.
   (c) $x \equiv 5 \pmod{6}$, $x \equiv 4 \pmod{11}$, $x \equiv 3 \pmod{17}$.
   (d) $2x \equiv 1 \pmod{5}$, $3x \equiv 9 \pmod{6}$, $4x \equiv 1 \pmod{7}$, $5x \equiv 9 \pmod{11}$.
5. Solve the linear congruence $17x \equiv 3 \pmod{2 \cdot 3 \cdot 5 \cdot 7}$ by solving the system

$$17x \equiv 3 \pmod{2} \qquad 17x \equiv 3 \pmod{3}$$
$$17x \equiv 3 \pmod{5} \qquad 17x \equiv 3 \pmod{7}$$

6. Find the smallest integer $a > 2$ such that

$$2 \mid a,\ 3 \mid a + 1,\ 4 \mid a + 2, 5 \mid a + 3, 6 \mid a + 4$$

7. (a) Obtain three consecutive integers, each having a square factor.
[*Hint:* Find an integer a such that $2^2 \mid a, 3^2 \mid a+1, 5^2 \mid a+2$.]
(b) Obtain three consecutive integers, the first of which is divisible by a square, the second by a cube, and the third by a fourth power.

8. (Brahmagupta, 7th century A.D.) When eggs in a basket are removed 2, 3, 4, 5, 6 at a time there remain, respectively, 1, 2, 3, 4, 5 eggs. When they are taken out 7 at a time, none are left over. Find the smallest number of eggs that could have been contained in the basket.

9. The basket-of-eggs problem is often phrased in the following form: One egg remains when the eggs are removed from the basket 2, 3, 4, 5, or 6 at a time; but, no eggs remain if they are removed 7 at a time. Find the smallest number of eggs that could have been in the basket.

10. (Ancient Chinese Problem.) A band of 17 pirates stole a sack of gold coins. When they tried to divide the fortune into equal portions, 3 coins remained. In the ensuing brawl over who should get the extra coins, one pirate was killed. The wealth was redistributed, but this time an equal division left 10 coins. Again an argument developed in which another pirate was killed. But now the total fortune was evenly distributed among the survivors. What was the least number of coins that could have been stolen?

11. Prove that the congruences

$$x \equiv a \pmod{n} \qquad \text{and} \qquad x \equiv b \pmod{m}$$

admit a simultaneous solution if and only if $\gcd(n, m) \mid a - b$; if a solution exists, confirm that it is unique modulo $\operatorname{lcm}(n, m)$.

12. Use Problem 11 to show that the following system does not possess a solution:

$$x \equiv 5 \pmod{6} \qquad \text{and} \qquad x \equiv 7 \pmod{15}$$

13. If $x \equiv a \pmod{n}$, prove that either $x \equiv a \pmod{2n}$ or $x \equiv a + n \pmod{2n}$.

14. A certain integer between 1 and 1200 leaves the remainders 1, 2, 6 when divided by 9, 11, 13, respectively. What is the integer?

15. (a) Find an integer having the remainders 1, 2, 5, 5 when divided by 2, 3, 6, 12, respectively. (Yih-hing, died 717).
(b) Find an integer having the remainders 2, 3, 4, 5 when divided by 3, 4, 5, 6, respectively. (Bhaskara, born 1114).
(c) Find an integer having the remainders 3, 11, 15 when divided by 10, 13, 17, respectively. (Regiomontanus, 1436–1476).

16. Let t_n denote the nth triangular number. For which values of n does t_n divide

$$t_1^2 + t_2^2 + \cdots + t_n^2$$

[*Hint:* Because $t_1^2 + t_2^2 + \cdots + t_n^2 = t_n(3n^3 + 12n^2 + 13n + 2)/30$, it suffices to determine those n satisfying $3n^3 + 12n^2 + 13n + 2 \equiv 0 \pmod{2 \cdot 3 \cdot 5}$.]

17. Find the solutions of the system of congruences:

$$3x + 4y \equiv 5 \pmod{13}$$
$$2x + 5y \equiv 7 \pmod{13}$$

18. Obtain the two incongruent solutions modulo 210 of the system

$$2x \equiv 3 \pmod{5}$$
$$4x \equiv 2 \pmod{6}$$
$$3x \equiv 2 \pmod{7}$$

19. Obtain the eight incongruent solutions of the linear congruence $3x + 4y \equiv 5 \pmod 8$

20. Find the solutions of each of the following systems of congruences:

(a) $5x + 3y \equiv 1 \pmod 7$

$3x + 2y \equiv 4 \pmod 7$.

(b) $7x + 3y \equiv 6 \pmod{11}$

$4x + 2y \equiv 9 \pmod{11}$.

(c) $11x + 5y \equiv 7 \pmod{20}$

$6x + 3y \equiv 8 \pmod{20}$.

CHAPTER 5

FERMAT'S THEOREM

And perhaps posterity will thank me for having shown it that the ancients did not know everything.
P. DE FERMAT

5.1 PIERRE DE FERMAT

What the ancient world had known was largely forgotten during the intellectual torpor of the Dark Ages, and it was only after the 12th century that Western Europe again became conscious of mathematics. The revival of classical scholarship was stimulated by Latin translations from the Greek and, more especially, from the Arabic. The Latinization of Arabic versions of Euclid's great treatise, the *Elements*, first appeared in 1120. The translation was not a faithful rendering of the *Elements*, having suffered successive, inaccurate translations from the Greek—first into Arabic, then into Castilian, and finally into Latin—done by copyists not versed in the content of the work. Nevertheless, this much-used copy, with its accumulation of errors, served as the foundation of all editions known in Europe until 1505, when the Greek text was recovered.

With the fall of Constantinople to the Turks in 1453, the Byzantine scholars who had served as the major custodians of mathematics brought the ancient masterpieces of Greek learning to the West. It is reported that a copy of what survived of Diophantus' *Arithmetica* was found in the Vatican library around 1462 by Johannes Müller (better known as Regiomontanus from the Latin name of his native town, Königsberg). Presumably, it had been brought to Rome by the refugees from Byzantium. Regiomontanus observed that "In these books the very flower of the

Pierre de Fermat
(1601–1665)

(David Eugene Smith Collection, Rare Book and Manuscript Library, Columbia University)

whole of arithmetic lies hid," and tried to interest others in translating it. Notwithstanding the attention that was called to the work, it remained practically a closed book until 1572 when the first translation and printed edition was brought out by the German professor Wilhelm Holzmann, who wrote under the Grecian form of his name, Xylander. The *Arithmetica* became fully accessible to European mathematicians when Claude Bachet—borrowing liberally from Xylander—published (1621) the original Greek text, along with a Latin translation containing notes and comments. The Bachet edition probably has the distinction of being the work that first directed the attention of Fermat to the problems of number theory.

Few if any periods were so fruitful for mathematics as was the 17th century; Northern Europe alone produced as many men of outstanding ability as had appeared during the preceding millennium. At a time when such names as Desargues, Descartes, Pascal, Wallis, Bernoulli, Leibniz, and Newton were becoming famous, a certain French civil servant, Pierre de Fermat (1601–1665), stood as an equal among these brilliant scholars. Fermat, the "Prince of Amateurs," was the last great mathematician to pursue the subject as a sideline to a nonscientific career. By profession a lawyer and magistrate attached to the provincial parliament at Toulouse, he sought refuge from controversy in the abstraction of mathematics. Fermat evidently had no particular mathematical training and he evidenced no interest in its study until he was past 30; to him, it was merely a hobby to be cultivated in leisure time. Yet no practitioner of his day made greater discoveries or contributed more to the advancement of the discipline: one of the inventors of analytic geometry (the actual term was coined in the early 19th century), he laid the technical foundations of differential and integral calculus and, with Pascal, established the conceptual guidelines of the theory of probability. Fermat's real love in mathematics was undoubtedly number theory, which he rescued from the realm of superstition and occultism where it had long been imprisoned. His contributions here overshadow all else; it may well be said that the revival of interest in the abstract side of number theory began with Fermat.

Fermat preferred the pleasure he derived from mathematical research itself to any reputation that it might bring him; indeed, he published only one major manuscript during his lifetime and that just 5 years before his death, using the concealing initials M.P.E.A.S. Adamantly refusing to put his work in finished form, he thwarted several efforts by others to make the results available in print under his name. In partial compensation for his lack of interest in publication, Fermat carried on a voluminous correspondence with contemporary mathematicians. Most of what little we know about his investigations is found in the letters to friends with whom he exchanged problems and to whom he reported his successes. They did their best to publicize Fermat's talents by passing these letters from hand to hand or by making copies, which were dispatched over the Continent.

As his parliamentary duties demanded an ever greater portion of his time, Fermat was given to inserting notes in the margin of whatever book he happened to be using. Fermat's personal copy of the Bachet edition of Diophantus held in its margin many of his famous theorems in number theory. These were discovered by his son Samuel 5 years after Fermat's death. His son brought out a new edition of the *Arithmetica* incorporating Fermat's celebrated marginalia. Because there was little space available, Fermat's habit had been to jot down some result and omit all steps leading to the conclusion. Posterity has wished many times that the margins of the *Arithmetica* had been wider or that Fermat had been a little less secretive about his methods.

5.2 FERMAT'S LITTLE THEOREM AND PSEUDOPRIMES

The most significant of Fermat's correspondents in number theory was Bernhard Frénicle de Bessy (1605–1675), an official at the French mint who was renowned for his gift of manipulating large numbers. (Frénicle's facility in numerical calculation is revealed by the following incident: On hearing that Fermat had proposed the problem of finding cubes that when increased by their proper divisors become squares, as is the case with $7^3 + (1 + 7 + 7^2) = 20^2$, he immediately gave four different solutions, and supplied six more the next day.) Though in no way Fermat's equal as a mathematician, Frénicle alone among his contemporaries could challenge Fermat in number theory and Frénicle's challenges had the distinction of coaxing out of Fermat some of his carefully guarded secrets. One of the most striking is the theorem that states: If p is a prime and a is any integer not divisible by p, then p divides $a^{p-1} - 1$. Fermat communicated the result in a letter to Frénicle dated October 18, 1640, along with the comment, "I would send you the demonstration, if I did not fear its being too long." This theorem has since become known as "Fermat's Little Theorem," or just "Fermat's Theorem," to distinguish it from Fermat's "Great" or "Last Theorem," which is the subject of Chapter 12. Almost 100 years were to elapse before Euler published the first proof of the little theorem in 1736. Leibniz, however, seems not to have received his share of recognition, for he left an identical argument in an unpublished manuscript sometime before 1683.

We now proceed to a proof of Fermat's theorem.

Theorem 5.1 Fermat's theorem. Let p be a prime and suppose that $p \nmid a$. Then $a^{p-1} \equiv 1 \pmod{p}$.

Proof. We begin by considering the first $p-1$ positive multiples of a; that is, the integers

$$a, 2a, 3a, \ldots, (p-1)a$$

None of these numbers is congruent modulo p to any other, nor is any congruent to zero. Indeed, if it happened that

$$ra \equiv sa \pmod{p} \qquad 1 \leq r < s \leq p-1$$

then a could be canceled to give $r \equiv s \pmod{p}$, which is impossible. Therefore, the previous set of integers must be congruent modulo p to $1, 2, 3, \ldots, p-1$, taken in some order. Multiplying all these congruences together, we find that

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

whence

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

Once $(p-1)!$ is canceled from both sides of the preceding congruence (this is possible because since $p \nmid (p-1)!$), our line of reasoning culminates in the statement that $a^{p-1} \equiv 1 \pmod{p}$, which is Fermat's theorem.

This result can be stated in a slightly more general way in which the requirement that $p \nmid a$ is dropped.

Corollary. If p is a prime, then $a^p \equiv a \pmod{p}$ for any integer a.

Proof. When $p \mid a$, the statement obviously holds; for, in this setting, $a^p \equiv 0 \equiv a \pmod{p}$. If $p \nmid a$, then according to Fermat's theorem, we have $a^{p-1} \equiv 1 \pmod{p}$. When this congruence is multiplied by a, the conclusion $a^p \equiv a \pmod{p}$ follows.

There is a different proof of the fact that $a^p \equiv a \pmod{p}$, involving induction on a. If $a = 1$, the assertion is that $1^p \equiv 1 \pmod{p}$, which clearly is true, as is the case $a = 0$. Assuming that the result holds for a, we must confirm its validity for $a+1$. In light of the binomial theorem,

$$(a+1)^p = a^p + \binom{p}{1} a^{p-1} + \cdots + \binom{p}{k} a^{p-k} + \cdots + \binom{p}{p-1} a + 1$$

where the coefficient $\binom{p}{k}$ is given by

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} = \frac{p(p-1)\cdots(p-k+1)}{1 \cdot 2 \cdot 3 \cdots k}$$

Our argument hinges on the observation that $\binom{p}{k} \equiv 0 \pmod{p}$ for $1 \leq k \leq p-1$. To see this, note that

$$k! \binom{p}{k} = p(p-1)\cdots(p-k+1) \equiv 0 \pmod{p}$$

CHAPTER 12

CERTAIN NONLINEAR DIOPHANTINE EQUATIONS

He who seeks for methods without having a definite problem in mind seeks for the most part in vain.
D. HILBERT

12.1 THE EQUATION $x^2 + y^2 = z^2$

Fermat, whom many regard as a father of modern number theory, nevertheless, had a custom peculiarly ill-suited to this role. He published very little personally, preferring to communicate his discoveries in letters to friends (usually with no more than the terse statement that he possessed a proof) or to keep them to himself in notes. A number of such notes were jotted down in the margin of his copy of Bachet's translation of Diophantus's *Arithmetica*. By far the most famous of these marginal comments is the one—presumably written about 1637—which states:

> It is impossible to write a cube as a sum of two cubes, a fourth power as a sum of two fourth powers, and, in general, any power beyond the second as a sum of two similar powers. For this, I have discovered a truly wonderful proof, but the margin is too small to contain it.

In this tantalizing aside, Fermat was simply asserting that, if $n > 2$, then the Diophantine equation

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

has no solution in the integers, other than the trivial solutions in which at least one of the variables is zero.

The quotation just cited has come to be known as Fermat's Last Theorem or, more accurately, Fermat's conjecture. By the 1800s, all the assertions appearing in the margin of his *Arithmetica* had either been proved or refuted—with the one exception of the Last Theorem (hence the name). The claim has fascinated many generations of mathematicians, professional and amateur alike, because it is so simple to understand yet so difficult to establish. If Fermat really did have a "truly wonderful proof," it has never come to light. Whatever demonstration he thought he possessed very likely contained a flaw. Indeed, Fermat himself may have subsequently discovered the error, for there is no reference to the proof in his correspondence with other mathematicians.

Fermat did, however, leave a proof of his Last Theorem for the case $n = 4$. To carry through the argument, we first undertake the task of identifying all solutions in the positive integers of the equation

$$x^2 + y^2 = z^2 \tag{1}$$

Because the length z of the hypotenuse of a right triangle is related to the lengths x and y of the sides by the famous Pythagorean equation $x^2 + y^2 = z^2$, the search for all positive integers that satisfy Eq. (1) is equivalent to the problem of finding all right triangles with sides of integral length. The latter problem was raised in the days of the Babylonians and was a favorite with the ancient Greek geometers. Pythagoras himself has been credited with a formula for infinitely many such triangles, namely,

$$x = 2n + 1 \qquad y = 2n^2 + 2n \qquad z = 2n^2 + 2n + 1$$

where n is an arbitrary positive integer. This formula does not account for all right triangles with integral sides, and it was not until Euclid wrote his *Elements* that a complete solution to the problem appeared.

The following definition gives us a concise way of referring to the solutions of Eq. (1).

Definition 12.1. A *Pythagorean triple* is a set of three integers x, y, z such that $x^2 + y^2 = z^2$; the triple is said to be *primitive* if $\gcd(x, y, z) = 1$.

Perhaps the best-known examples of primitive Pythagorean triples are 3, 4, 5 and 5, 12, 13, whereas a less obvious one is 12, 35, 37.

There are several points that need to be noted. Suppose that x, y, z is any Pythagorean triple and $d = \gcd(x, y, z)$. If we write $x = dx_1$, $y = dy_1$, $z = dz_1$, then it is easily seen that

$$x_1^2 + y_1^2 = \frac{x^2 + y^2}{d^2} = \frac{z^2}{d^2} = z_1^2$$

with $\gcd(x_1, y_1, z_1) = 1$. In short, x_1, y_1, z_1 form a primitive Pythagorean triple. Thus, it is enough to occupy ourselves with finding all primitive Pythagorean triples; any Pythagorean triple can be obtained from a primitive one upon multiplying by a suitable nonzero integer. The search may be confined to those primitive Pythagorean

triples x, y, z in which $x > 0$, $y > 0$, $z > 0$, inasmuch as all others arise from the positive ones through a simple change of sign.

Our development requires two preparatory lemmas, the first of which sets forth a basic fact regarding primitive Pythagorean triples.

Lemma 1. If x, y, z is a primitive Pythagorean triple, then one of the integers x or y is even, while the other is odd.

Proof. If x and y are both even, then $2 \mid (x^2 + y^2)$ or $2 \mid z^2$, so that $2 \mid z$. The inference is that $\gcd(x, y, z) \geq 2$, which we know to be false. If, on the other hand, x and y should both be odd, then $x^2 \equiv 1 \pmod 4$ and $y^2 \equiv 1 \pmod 4$, leading to

$$z^2 = x^2 + y^2 \equiv 2 \pmod 4$$

But this is equally impossible, because the square of any integer must be congruent either to 0 or to 1 modulo 4.

Given a primitive Pythagorean triple x, y, z, exactly one of these integers is even, the other two being odd (if x, y, z were all odd, then $x^2 + y^2$ would be even, whereas z^2 is odd). The foregoing lemma indicates that the even integer is either x or y; to be definite, we shall hereafter write our Pythagorean triples so that x is even and y is odd; then, of course, z is odd.

It is worth noticing (and we will use this fact) that each pair of the integers x, y, and z must be relatively prime. Were it the case that $\gcd(x, y) = d > 1$, then there would exist a prime p with $p \mid d$. Because $d \mid x$ and $d \mid y$, we would have $p \mid x$ and $p \mid y$, whence $p \mid x^2$ and $p \mid y^2$. But then $p \mid (x^2 + y^2)$, or $p \mid z^2$, giving $p \mid z$. This would conflict with the assumption that $\gcd(x, y, z) = 1$, and so $d = 1$. In like manner, one can verify that $\gcd(y, z) = \gcd(x, z) = 1$.

By virtue of Lemma 1, there exists no primitive Pythagorean triple x, y, z all of whose values are prime numbers. There are primitive Pythagorean triples in which z and one of x or y is a prime; for instance, 3, 4, 5; 11, 60, 61; and 19, 180, 181. It is unknown whether there exist infinitely many such triples.

The next hurdle that stands in our way is to establish that if a and b are relatively prime positive integers having a square as their product, then a and b are themselves squares. With an assist from the Fundamental Theorem of Arithmetic, we can prove considerably more, to wit, Lemma 2.

Lemma 2. If $ab = c^n$, where $\gcd(a, b) = 1$, then a and b are nth powers; that is, there exist positive integers a_1, b_1 for which $a = a_1^n$, $b = b_1^n$.

Proof. There is no harm in assuming that $a > 1$ and $b > 1$. If

$$a = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \qquad b = q_1^{j_1} q_2^{j_2} \cdots q_s^{j_s}$$

are the prime factorizations of a and b, then, bearing in mind that $\gcd(a, b) = 1$, no p_i can occur among the q_i. As a result, the prime factorization of ab is given by

$$ab = p_1^{k_1} \cdots p_r^{k_r} q_1^{j_1} \cdots q_s^{j_s}$$

Let us suppose that c can be factored into primes as $c = u_1^{l_1} u_2^{l_2} \cdots u_t^{l_t}$. Then the condition $ab = c^n$ becomes

$$p_1^{k_1} \cdots p_r^{k_r} q_1^{j_1} \cdots q_s^{j_s} = u_1^{nl_1} \cdots u_t^{nl_t}$$

From this we see that the primes $u_1, \ldots, u_t$ are $p_1, \ldots, p_r, q_1, \ldots, q_s$ (in some order) and $nl_1, \ldots, nl_t$ are the corresponding exponents $k_1, \ldots, k_r, j_1, \ldots, j_s$. The conclusion: Each of the integers k_i and j_i must be divisible by n. If we now put

$$a_1 = p_1^{k_1/n} p_2^{k_2/n} \cdots p_r^{k_r/n}$$
$$b_1 = q_1^{j_1/n} q_2^{j_2/n} \cdots q_s^{j_s/n}$$

then $a_1^n = a, b_1^n = b$, as desired.

With the routine work now out of the way, the characterization of all primitive Pythagorean triples is fairly straightforward.

Theorem 12.1. All the solutions of the Pythagorean equation

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

satisfying the conditions

$$\gcd(x, y, z) = 1 \qquad 2 \mid x \qquad x > 0, y > 0, z > 0$$

are given by the formulas

$$x = 2st \qquad y = s^2 - t^2 \qquad z = s^2 + t^2$$

for integers $s > t > 0$ such that $\gcd(s, t) = 1$ and $s \not\equiv t \pmod 2$.

Proof. To start, let x, y, z be a (positive) primitive Pythagorean triple. Because we have agreed to take x even, and y and z both odd, $z - y$ and $z + y$ are even integers; say, $z - y = 2u$ and $z + y = 2v$. Now the equation $x^2 + y^2 = z^2$ may be rewritten as

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

whence

$$\left(\frac{x}{2}\right)^2 = \left(\frac{z-y}{2}\right)\left(\frac{z+y}{2}\right) = uv$$

Notice that u and v are relatively prime; indeed, if $\gcd(u, v) = d > 1$, then $d \mid (u - v)$ and $d \mid (u + v)$, or equivalently, $d \mid y$ and $d \mid z$, which violates the fact that $\gcd(y, z) = 1$. Taking Lemma 2 into consideration, we may conclude that u and v are each perfect squares; to be specific, let

$$u = t^2 \qquad v = s^2$$

where s and t are positive integers. The result of substituting these values of u and v reads

$$z = v + u = s^2 + t^2$$
$$y = v - u = s^2 - t^2$$
$$x^2 = 4vu = 4s^2t^2$$

or, in the last case $x = 2st$. Because a common factor of s and t divides both y and z, the condition $\gcd(y, z) = 1$ forces $\gcd(s, t) = 1$. It remains for us to observe that if s and t were both even, or both odd, then this would make each of y and z even, which is an impossibility. Hence, exactly one of the pair s, t is even, and the other is odd; in symbols, $s \not\equiv t \pmod 2$.

Conversely, let s and t be two integers subject to the conditions described before. That $x = 2st$, $y = s^2 - t^2$, $z = s^2 + t^2$ form a Pythagorean triple follows from the easily verified identity

$$x^2 + y^2 = (2st)^2 + (s^2 - t^2)^2 = (s^2 + t^2)^2 = z^2$$

To see that this triple is primitive, we assume that $\gcd(x, y, z) = d > 1$ and take p to be any prime divisor of d. Observe that $p \neq 2$, because p divides the odd integer z (one of s and t is odd, and the other is even, hence, $s^2 + t^2 = z$ must be odd). From $p \mid y$ and $p \mid z$, we obtain $p \mid (z + y)$ and $p \mid (z - y)$, or put otherwise, $p \mid 2s^2$ and $p \mid 2t^2$. But then $p \mid s$ and $p \mid t$, which is incompatible with $\gcd(s, t) = 1$. The implication of all this is that $d = 1$ and so x, y, z constitutes a primitive Pythagorean triple. Theorem 12.1 is thus proven.

The table below lists some primitive Pythagorean triples arising from small values of s and t. For each value of $s = 2, 3, \ldots, 7$, we have taken those values of t that are relatively prime to s, less than s, and even whenever s is odd.

		x	y	z
s	t	$(2st)$	$(s^2 - t^2)$	$(s^2 + t^2)$
2	1	4	3	5
3	2	12	5	13
4	1	8	15	17
4	3	24	7	25
5	2	20	21	29
5	4	40	9	41
6	1	12	35	37
6	5	60	11	61
7	2	28	45	53
7	4	56	33	65
7	6	84	13	85

From this, or from a more extensive table, the reader might be led to suspect that if x, y, z is a primitive Pythagorean triple, then exactly one of the integers x or y is divisible by 3. This is, in fact, the case. For, by Theorem 12.1, we have

$$x = 2st \qquad y = s^2 - t^2 \qquad z = s^2 + t^2$$

where $\gcd(s, t) = 1$. If either $3 \mid s$ or $3 \mid t$, then evidently $3 \mid x$, and we need go no further. Suppose that $3 \nmid s$ and $3 \nmid t$. Fermat's theorem asserts that

$$s^2 \equiv 1 \pmod 3 \qquad t^2 \equiv 1 \pmod 3$$

and so

$$y = s^2 - t^2 \equiv 0 \pmod 3$$

In other words, y is divisible by 3, which is what we were required to show.

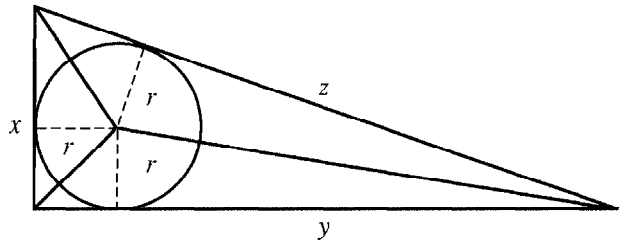
Let us define a *Pythagorean triangle* to be a right triangle whose sides are of integral length. Our findings lead to an interesting geometric fact concerning Pythagorean triangles, recorded as Theorem 12.2.

Theorem 12.2. The radius of the inscribed circle of a Pythagorean triangle is always an integer.

Proof. Let r denote the radius of the circle inscribed in a right triangle with hypotenuse of length z and sides of lengths x and y. The area of the triangle is equal to the sum of the areas of the three triangles having common vertex at the center of the circle; hence,

$$\frac{1}{2}xy = \frac{1}{2}rx + \frac{1}{2}ry + \frac{1}{2}rz = \frac{1}{2}r(x + y + z)$$

The situation is illustrated below:



Now $x^2 + y^2 = z^2$. But we know that the positive integral solutions of this equation are given by

$$x = 2kst \qquad y = k(s^2 - t^2) \qquad z = k(s^2 + t^2)$$

for an appropriate choice of positive integers k, s, t. Replacing x, y, z in the equation $xy = r(x + y + z)$ by these values and solving for r, it will be found that

$$\begin{aligned} r &= \frac{2k^2st(s^2 - t^2)}{k(2st + s^2 - t^2 + s^2 + t^2)} \\ &= \frac{kt(s^2 - t^2)}{s + t} \\ &= kt(s - t) \end{aligned}$$

which is an integer.

We take the opportunity to mention another result relating to Pythagorean triangles. Notice that it is possible for different Pythagorean triangles to have the same area; for instance, the right triangles associated with the primitive Pythagorean triples 20, 21, 29 and 12, 35, 37 each have an area equal to 210. Fermat proved: For any integer $n > 1$, there exist n Pythagorean triangles with different hypotenuses and the same area. The details of this are omitted.

PROBLEMS 12.1

1. (a) Find three different Pythagorean triples, not necessarily primitive, of the form $16, y, z$.

(b) Obtain all primitive Pythagorean triples x, y, z in which $x = 40$; do the same for $x = 60$.

2. If x, y, z is a primitive Pythagorean triple, prove that $x + y$ and $x - y$ are congruent modulo 8 to either 1 or 7.

3. (a) Prove that if $n \not\equiv 2 \pmod 4$, then there is a primitive Pythagorean triple x, y, z in which x or y equals n.

(b) If $n \geq 3$ is arbitrary, find a Pythagorean triple (not necessarily primitive) having n as one of its members.

[*Hint:* Assuming n is odd, consider the triple n, $\frac{1}{2}(n^2 - 1)$, $\frac{1}{2}(n^2 + 1)$; for n even, consider the triple n, $(n^2/4) - 1$, $(n^2/4) + 1$.]

4. Prove that in a primitive Pythagorean triple x, y, z, the product xy is divisible by 12, hence $60 \mid xyz$.

5. For a given positive integer n, show that there are at least n Pythagorean triples having the same first member.

[*Hint:* Let $y_k = 2^k(2^{2n-2k} - 1)$ and $z_k = 2^k(2^{2n-2k} + 1)$ for $k = 0, 1, 2, \ldots, n - 1$. Then 2^{n+1}, y_k, z_k are all Pythagorean triples.]

6. Verify that 3, 4, 5 is the only primitive Pythagorean triple involving consecutive positive integers.

7. Show that $3n, 4n, 5n$ where $n = 1, 2, \ldots$ are the only Pythagorean triples whose terms are in arithmetic progression.

[*Hint:* Call the triple in question $x - d, x, x + d$, and solve for x in terms of d.]

8. Find all Pythagorean triangles whose areas are equal to their perimeters.

[*Hint:* The equations $x^2 + y^2 = z^2$ and $x + y + z = \frac{1}{2}xy$ imply that $(x - 4)(y - 4) = 8$.]

9. (a) Prove that if x, y, z is a primitive Pythagorean triple in which x and z are consecutive positive integers, then

$$x = 2t(t + 1) \qquad y = 2t + 1 \qquad z = 2t(t + 1) + 1$$

for some $t > 0$.

[*Hint:* The equation $1 = z - x = s^2 + t^2 - 2st$ implies that $s - t = 1$.]

(b) Prove that if x, y, z is a primitive Pythagorean triple in which the difference $z - y = 2$, then

$$x = 2t \qquad y = t^2 - 1 \qquad z = t^2 + 1$$

for some $t > 1$.

10. Show that there exist infinitely many primitive Pythagorean triples x, y, z whose even member x is a perfect square.

[*Hint:* Consider the triple $4n^2$, $n^4 - 4$, $n^4 + 4$, where n is an arbitrary odd integer.]

11. For an arbitrary positive integer n, show that there exists a Pythagorean triangle the radius of whose inscribed circle is n.

[*Hint:* If r denotes the radius of the circle inscribed in the Pythagorean triangle having sides a and b and hypotenuse c, then $r = \frac{1}{2}(a + b - c)$. Now consider the triple $2n + 1$, $2n^2 + 2n$, $2n^2 + 2n + 1$.]

12. (a) Establish that there exist infinitely many primitive Pythagorean triples x, y, z in which x and y are consecutive positive integers. Exhibit five of these.

[*Hint:* If $x, x + 1, z$ forms a Pythagorean triple, then so does the triple $3x + 2z + 1$, $3x + 2z + 2$, $4x + 3z + 2$.]

(b) Show that there exist infinitely many Pythagorean triples x, y, z in which x and y are consecutive triangular numbers. Exhibit three of these.
[*Hint:* If x, $x+1$, z forms a Pythagorean triple, then so does t_{2x}, t_{2x+1}, $(2x+1)z$.]

13. Use Problem 12 to prove that there exist infinitely many triangular numbers that are perfect squares. Exhibit five such triangular numbers.
[*Hint:* If x, $x+1$, z forms a Pythagorean triple, then upon setting $u = z - x - 1$, $v = x + \frac{1}{2}(1-z)$, one obtains $u(u+1)/2 = v^2$.]

12.2 FERMAT'S LAST THEOREM

With our knowledge of Pythagorean triples, we are now prepared to take up the one case in which Fermat himself had a proof of his conjecture, the case $n = 4$. The technique used in the proof is a form of induction sometimes called "Fermat's method of infinite descent." In brief, the method may be described as follows: It is assumed that a solution of the problem in question is possible in the positive integers. From this solution, one constructs a new solution in smaller positive integers, which then leads to a still smaller solution, and so on. Because the positive integers cannot be decreased in magnitude indefinitely, it follows that the initial assumption must be false and therefore no solution is possible.

Instead of giving a proof of the Fermat conjecture for $n = 4$, it turns out to be easier to establish a fact that is slightly stronger, namely, the impossibility of solving the equation $x^4 + y^4 = z^2$ in the positive integers.

Theorem 12.3 Fermat. The Diophantine equation $x^4 + y^4 = z^2$ has no solution in positive integers x, y, z.

Proof. With the idea of deriving a contradiction, let us assume that there exists a positive solution x_0, y_0, z_0 of $x^4 + y^4 = z^2$. Nothing is lost in supposing also that $\gcd(x_0, y_0) = 1$; otherwise, put $\gcd(x_0, y_0) = d$, $x_0 = dx_1$, $y_0 = dy_1$, $z_0 = d^2 z_1$ to get $x_1^4 + y_1^4 = z_1^2$ with $\gcd(x_1, y_1) = 1$.

Expressing the supposed equation $x_0^4 + y_0^4 = z_0^2$ in the form

$$(x_0^2)^2 + (y_0^2)^2 = z_0^2$$

we see that x_0^2, y_0^2, z_0 meet all the requirements of a primitive Pythagorean triple, and therefore Theorem 12.1 can be brought into play. In such triples, one of the integers x_0^2 or y_0^2 is necessarily even, whereas the other is odd. Taking x_0^2 (and hence x_0) to be even, there exist relatively prime integers $s > t > 0$ satisfying

$$x_0^2 = 2st$$
$$y_0^2 = s^2 - t^2$$
$$z_0 = s^2 + t^2$$

where exactly one of s and t is even. If it happens that s is even, then we have

$$1 \equiv y_0^2 = s^2 - t^2 \equiv 0 - 1 \equiv 3 \pmod 4$$

which is an impossibility. Therefore, s must be the odd integer and, in consequence, t is the even one. Let us put $t = 2r$. Then the equation $x_0^2 = 2st$ becomes $x_0^2 = 4sr$,

which says that

$$\left(\frac{x_0}{2}\right)^2 = sr$$

But Lemma 2 asserts that the product of two relatively prime integers [note that $\gcd(s, t) = 1$ implies that $\gcd(s, r) = 1$] is a square only if each of the integers itself is a square; hence, $s = z_1^2$, $r = w_1^2$ for positive integers z_1, w_1.

We wish to apply Theorem 12.1 again, this time to the equation

$$t^2 + y_0^2 = s^2$$

Because $\gcd(s, t) = 1$, it follows that $\gcd(t, y_0, s) = 1$, making t, y_0, s a primitive Pythagorean triple. With t even, we obtain

$$t = 2uv$$
$$y_0 = u^2 - v^2$$
$$s = u^2 + v^2$$

for relatively prime integers $u > v > 0$. Now the relation

$$uv = \frac{t}{2} = r = w_1^2$$

signifies that u and v are both squares (Lemma 2 serves its purpose once more); say, $u = x_1^2$ and $v = y_1^2$. When these values are substituted into the equation for s, the result is

$$z_1^2 = s = u^2 + v^2 = x_1^4 + y_1^4$$

A crucial point is that, z_1 and t being positive, we also have the inequality

$$0 < z_1 \leq z_1^2 = s \leq s^2 < s^2 + t^2 = z_0$$

What has happened is this. Starting with one solution x_0, y_0, z_0 of $x^4 + y^4 = z^2$, we have constructed another solution x_1, y_1, z_1 such that $0 < z_1 < z_0$. Repeating the whole argument, our second solution would lead to a third solution x_2, y_2, z_2 with $0 < z_2 < z_1$, which, in turn, gives rise to a fourth. This process can be carried out as many times as desired to produce an infinite decreasing sequence of positive integers

$$z_0 > z_1 > z_2 > \cdots$$

Because there is only a finite supply of positive integers less than z_0, a contradiction occurs. We are forced to conclude that $x^4 + y^4 = z^2$ is not solvable in the positive integers.

As an immediate result, one gets the following corollary.

Corollary. The equation $x^4 + y^4 = z^4$ has no solution in the positive integers.

Proof. If x_0, y_0, z_0 were a positive solution of $x^4 + y^4 = z^4$, then x_0, y_0, z_0^2 would satisfy the equation $x^4 + y^4 = z^2$, in conflict with Theorem 12.3.

If $n > 2$, then n is either a power of 2 or divisible by an odd prime p. In the first case, $n = 4k$ for some $k \geq 1$ and the Fermat equation $x^n + y^n = z^n$ can be written as

$$(x^k)^4 + (y^k)^4 = (z^k)^4$$

We have just seen that this equation is impossible in the positive integers. When $n = pk$, the Fermat equation is the same as

$$(x^k)^p + (y^k)^p = (z^k)^p$$

If it could be shown that the equation $u^p + v^p = w^p$ has no solution, then, in particular, there would be no solution of the form $u = x^k$, $v = y^k$, $w = z^k$; hence, $x^n + y^n = z^n$ would not be solvable. Therefore, Fermat's conjecture reduces to this: For no odd prime p does the equation

$$x^p + y^p = z^p$$

admit a solution in the positive integers.

Although the problem has challenged the foremost mathematicians of the last 300 years, their efforts tended to produce partial results and proofs of individual cases. Euler gave the first proof of the Fermat conjecture for the prime $p = 3$ in the year 1770; the reasoning was incomplete at one stage, but Legendre later supplied the missing steps. Using the method of infinite descent, Dirichlet and Legendre independently settled the case $p = 5$ around 1825. Not long thereafter, in 1839, Lamé proved the conjecture for seventh powers. With the increasing complexity of the arguments came the realization that a successful resolution of the general case called for different techniques. The best hope seemed to lie in extending the meaning of "integer" to include a wider class of numbers and, by attacking the problem within this enlarged system, obtaining more information than was possible by using ordinary integers only.

The German mathematician Kummer made the major breakthrough. In 1843, he submitted to Dirichlet a purported proof of Fermat's conjecture based upon an extension of the integers to include the so-called "algebraic numbers" (that is, complex numbers satisfying polynomials with rational coefficients). Having spent considerable time on the problem himself, Dirichlet was immediately able to detect the flaw in the reasoning: Kummer had taken for granted that algebraic numbers admit a unique factorization similar to that of the ordinary integers, which is not always true.

But Kummer was undeterred by this perplexing situation and returned to his investigations with redoubled effort. To restore unique factorization to the algebraic numbers, he was led to invent the concept of *ideal numbers*. By adjoining these new entities to the algebraic numbers, Kummer successfully proved Fermat's conjecture for a large class of primes that he termed *regular primes* (that this represented an enormous achievement is reflected in the fact that the only irregular primes less than 100 are 37, 59, and 67). Unfortunately, it is still not known whether there are an infinite number of regular primes, whereas in the other direction, Jensen (1915) established that there exist infinitely many irregular ones. Almost all the subsequent progress on the problem was within the framework suggested by Kummer.

In 1983, a 29-year-old West German mathematician, Gerd Faltings, proved that for each exponent $n > 2$, the Fermat equation $x^n + y^n = z^n$ can have at most a finite number (as opposed to an infinite number) of integral solutions. At first glance, this may not seem like much of an advance; but if it could be shown that the finite number of solutions was zero in each case, then the Fermat's conjecture would be laid to rest once and for all.

Another striking result, established in 1987, was that Fermat's assertion is true for "almost all" values of n; that is, as n increases the percentage of cases in which the conjecture could fail approaches zero.

With the advent of computers, various numerical tests were devised to verify Fermat's conjecture for specific values of n. In 1977, S. S. Wagstaff took over 2 years, using computing time on four machines on weekends and holidays, to show that the conjecture held for all $n \leq 125000$. Since that time, the range of exponents for which the result was determined to be true has been extended repeatedly. By 1992, Fermat's conjecture was known to be true for exponents up to 4000000.

For a moment in the summer of 1993, it appeared that the final breakthrough had been made. At the conclusion of 3 days of lectures in Cambridge, England, Andrew Wiles of Princeton University stunned his colleagues by announcing that he could favorably resolve Fermat's conjecture. His proposed proof, which had taken 7 years to prepare, was an artful blend of many sophisticated techniques developed by other mathematicians only within the preceding decade. The key insight was to link equations of the kind posed by Fermat with the much-studied theory of elliptic curves; that is, curves determined by cubic polynomials of the form $y^2 = x^3 + ax + b$, where a and b are integers.

The overall structure and strategy of Wiles's argument was so compelling that mathematicians hailed it as almost certainly correct. But when the immensely complicated 200-page manuscript was carefully scrutinized for hidden errors, it revealed a subtle snag. No one claimed that the flaw was fatal, and bridging the gap was felt to be feasible. Over a year later, Wiles provided a corrected, refined, and shorter (125-page) version of his original proof to the enthusiastic reviewers. The revised argument was seen to be sound, and Fermat's seemingly simple claim was finally settled.

The failure of Wiles's initial attempt is not really surprising or unusual in mathematical research. Normally, proposed proofs are privately circulated and examined for possible flaws months in advance of any formal announcement. In Wiles's case, the notoriety of one of number theory's most elusive conjectures brought premature publicity and temporary disappointment to the mathematical community.

To round out our historical digression, we might mention that in 1908 a prize of 100,000 marks was bequeathed to the Academy of Science at Göttingen to be paid for the first complete proof of Fermat's conjecture. The immediate result was a deluge of incorrect demonstrations by amateur mathematicians. Because only printed solutions were eligible, Fermat's conjecture is reputed to be the mathematical problem for which the greatest number of false proofs have been published; indeed, between 1908 and 1912 over 1000 alleged proofs appeared, mostly printed as private pamphlets. Suffice it to say, interest declined as the German inflation of the 1920s wiped out the monetary value of the prize. (With the introduction of the Reichsmark and Deutsche Mark [DM] and after various currency revaluations, the award was worth about DM 75,000 or $40,000 when it was presented to Wiles in 1997.)

From $x^4 + y^4 = z^2$, we move on to a closely related Diophantine equation, namely, $x^4 - y^4 = z^2$. The proof of its insolubility parallels that of Theorem 12.3, but we give a slight variation in the method of infinite descent.

Theorem 12.4 Fermat. The Diophantine equation $x^4 - y^4 = z^2$ has no solution in positive integers x, y, z.

Proof. The proof proceeds by contradiction. Let us assume that the equation admits a solution in the positive integers and among these solutions x_0, y_0, z_0 is one with a least value of x; in particular, this supposition forces x_0 to be odd (Why?). Were $\gcd(x_0, y_0) = d > 1$, then putting $x_0 = dx_1, y_0 = dy_1$, we would have $d^4(x_1^4 - y_1^4) = z_0^2$, whence $d^2 \mid z_0$ or $z_0 = d^2 z_1$ for some $z_1 > 0$. It follows that x_1, y_1, z_1 provides a solution to the equation under consideration with $0 < x_1 < x_0$, which is an impossible situation. Thus, we are free to assume a solution x_0, y_0, z_0 in which $\gcd(x_0, y_0) = 1$. The ensuing argument falls into two stages, depending on whether y_0 is odd or even.

First, consider the case of an odd integer y_0. If the equation $x_0^4 - y_0^4 = z_0^2$ is written in the form $z_0^2 + (y_0^2)^2 = (x_0^2)^2$, we see that z_0, y_0^2, x_0^2 constitute a primitive Pythagorean triple. Theorem 12.1 asserts the existence of relatively prime integers $s > t > 0$ for which

$$\begin{aligned} z_0 &= 2st \\ y_0^2 &= s^2 - t^2 \\ x_0^2 &= s^2 + t^2 \end{aligned}$$

Thus, it appears that

$$s^4 - t^4 = (s^2 + t^2)(s^2 - t^2) = x_0^2 y_0^2 = (x_0 y_0)^2$$

making $s, t, x_0 y_0$ a (positive) solution to the equation $x^4 - y^4 = z^2$. Because

$$0 < s < \sqrt{s^2 + t^2} = x_0$$

we arrive at a contradiction to the minimal nature of x_0.

For the second part of the proof, assume that y_0 is an even integer. Using the formulas for primitive Pythagorean triples, we now write

$$\begin{aligned} y_0^2 &= 2st \\ z_0 &= s^2 - t^2 \\ x_0^2 &= s^2 + t^2 \end{aligned}$$

where s may be taken to be even and t to be odd. Then, in the relation $y_0^2 = 2st$, we have $\gcd(2s, t) = 1$. The now-customary application of Lemma 2 tells us that $2s$ and t are each squares of positive integers; say, $2s = w^2, t = v^2$. Because w must of necessity be an even integer, set $w = 2u$ to get $s = 2u^2$. Therefore,

$$x_0^2 = s^2 + t^2 = 4u^4 + v^4$$

and so $2u^2, v^2, x_0$ forms a primitive Pythagorean triple. Falling back on Theorem 12.1 again, there exist integers $a > b > 0$ for which

$$\begin{aligned} 2u^2 &= 2ab \\ v^2 &= a^2 - b^2 \\ x_0 &= a^2 + b^2 \end{aligned}$$

where $\gcd(a, b) = 1$. The equality $u^2 = ab$ ensures that a and b are perfect squares, so that $a = c^2$ and $b = d^2$. Knowing this, the rest of the proof is easy; for, upon substituting,

$$v^2 = a^2 - b^2 = c^4 - d^4$$

The result is a new solution c, d, v of the given equation $x^4 - y^4 = z^2$ and what is more, a solution in which

$$0 < c = \sqrt{a} < a^2 + b^2 = x_0$$

contrary to our assumption regarding x_0.

The only resolution of these contradictions is that the equation $x^4 - y^4 = z^2$ cannot be satisfied in the positive integers.

In the margin of his copy of Diophantus's *Arithmetica*, Fermat states and proves the following: The area of a right triangle with rational sides cannot be the square of a rational number. Clearing of fractions, this reduces to a theorem about Pythagorean triangles, to wit, Theorem 12.5.

Theorem 12.5. The area of a Pythagorean triangle can never be equal to a perfect (integral) square.

Proof. Consider a Pythagorean triangle whose hypotenuse has length z and other two sides have lengths x and y, so that $x^2 + y^2 = z^2$. The area of the triangle in question is $\frac{1}{2}xy$, and if this were a square, say u^2, it would follow that $2xy = 4u^2$. By adding and subtracting the last-written equation from $x^2 + y^2 = z^2$, we are led to

$$(x + y)^2 = z^2 + 4u^2 \qquad \text{and} \qquad (x - y)^2 = z^2 - 4u^2$$

When these last two equations are multiplied together, the outcome is that two fourth powers have as their difference a square:

$$(x^2 - y^2)^2 = z^4 - 16u^4 = z^4 - (2u)^4$$

Because this amounts to an infringement on Theorem 12.4, there can be no Pythagorean triangle whose area is a square.

There are a number of simple problems pertaining to Pythagorean triangles that still await solution. The corollary to Theorem 12.3 may be expressed by saying that there exists no Pythagorean triangle all the sides of which are squares. However, it is not difficult to produce Pythagorean triangles whose sides, if increased by 1, are squares; for instance, the triangles associated with the triples $13^2 - 1$, $10^2 - 1$, $14^2 - 1$, and $287^2 - 1$, $265^2 - 1$, $329^2 - 1$. An obvious—and as yet unanswered—question is whether there are an infinite number of such triangles. We can find Pythagorean triangles each side of which is a triangular number. [By a triangular number, we mean an integer of the form $t_n = n(n + 1)/2$.] An example of such is the triangle corresponding to t_{132}, t_{143}, t_{164}. It is not known if infinitely many Pythagorean triangles of this type exist.

As a closing comment, we should observe that all the effort expended on attempting to prove Fermat's conjecture has been far from wasted. The new mathematics that was developed as a by-product laid the foundations for algebraic number theory and the ideal theory of modern abstract algebra. It seems fair to say that the value of these far exceeds that of the conjecture itself.

Another challenge to number theorists, somewhat akin to Fermat's conjecture, concerns the Catalan equation. Consider for the moment the squares and cubes of

positive integers in increasing order:

$$1, 4, 8, 9, 16, 25, 27, 36, 49, 64, 81, 100, \ldots$$

We notice that 8 and 9 are consecutive integers in this sequence. The medieval astronomer Levi ben Gerson (1288–1344) proved that there are no other consecutive powers of 2 and 3; to put it another way, he showed that if $3^m - 2^n = \pm 1$, with $m > 1$ and $n > 1$, then $m = 2$ and $n = 3$. In 1738, Euler, using Fermat's method of infinite descent, dealt with the equation $x^3 - y^2 = \pm 1$, proving that $x = 2$ and $y = 3$. Catalan himself contributed little more to the consecutive-power problem than the assertion (1844) that the only solution of the equation $x^m - y^n = 1$ in integers x, y, m, n, all greater than 1, is $m = y = 2, n = x = 3$. This statement, now known as Catalan's conjecture, was proved, in 2002.

Over the years, the Catalan equation $x^m - y^n = 1$ had been shown to be impossible of solution for special values of m and n. For example in 1850, V. A. Lebesgue proved that $x^m - y^2 = 1$ admits no solution in the positive integers for $m \neq 3$; but, it remained until 1964 to show that the more difficult equation $x^2 - y^n = 1$ is not solvable for $n \neq 3$. The cases $x^3 - y^n = 1$ and $x^m - y^3 = 1$, with $m \neq 2$, were successfully resolved in 1921. The most striking result, obtained by R. Tijdeman in 1976, is that $x^m - y^n = 1$ has only a finite number of solutions, all of which are smaller than some computable constant $C > 0$; that is, $x^m, y^n < C$.

Suppose that Catalan's equation did have a solution other than $3^2 - 2^3 = 1$. If p and q are primes dividing m and n respectively, then $x^{m/p}$ and $y^{n/q}$ would provide a solution to the equation $u^p - v^q = 1$. What needed to be shown was that this equation was not solvable in integers $u, v \geq 2$ and distinct primes $p, q \geq 5$. One approach called for obtaining explicit bounds on the possible size of the exponents. A series of investigations continually sharpened the restrictions until by the year 2000 it was known that $3 \cdot 10^8 < p < (7.15)10^{11}$ and $3.10^8 < q < (7.75)10^{16}$. Thus, the Catalan conjecture could in principle be settled by exhaustive computer calculations; but until the upper bound was lowered, this would take a long time.

In 2000, Preda Mihailescu proved that for a Catalan solution to exist, p and q must satisfy the simultaneous congruences

$$p^{q-1} \equiv 1 (\mathrm{mod}\, q^2) \quad \text{and} \quad q^{p-1} \equiv 1 (\mathrm{mod}\, p^2)$$

These are known as double Wieferich primes, after Arthur Wieferich, who investigated (1909) the congruence $2^{p-1} \equiv 1 \pmod{p^2}$. Such pairs of primes are rare, with only six pairs having been identified by the year 2001. Furthermore, as each of these 12 primes is less than $3 \cdot 10^8$, none satisfied the known restrictions. Taking advantage of his results on Wieferich primes, Mihailescu continued to work on the problem. He finally settled the famous question early in the following year: the only consecutive powers are 8 and 9.

One interesting consequence of these results is that no Fermat number $F_n = 2^{2^n} + 1$ can be a power of another integer, the exponent being greater than 1. For if $F_n = a^m$, with $m \geq 2$, then $a^m - (2^{2^{n-1}})^2 = 1$, which would imply that the equation $x^m - y^2 = 1$ has a solution.

PROBLEMS 12.2

1. Show that the equation $x^2 + y^2 = z^3$ has infinitely many solutions for x, y, z positive integers.
   [*Hint:* For any $n \geq 2$, let $x = n(n^2 - 3)$ and $y = 3n^2 - 1$.]
2. Prove the theorem: The only solutions in nonnegative integers of the equation $x^2 + 2y^2 = z^2$, with $\gcd(x, y, z) = 1$, are given by
   $$x = \pm(2s^2 - t^2) \qquad y = 2st \qquad z = 2s^2 + t^2$$
   where s, t are arbitrary nonnegative integers.
   [*Hint:* If u, v, w are such that $y = 2w$, $z + x = 2u$, $z - x = 2v$, then the equation becomes $2w^2 = uv$.]
3. In a Pythagorean triple x, y, z, prove that not more than one of x, y, or z can be a perfect square.
4. Prove each of the following assertions:
   (a) The system of simultaneous equations
   $$x^2 + y^2 = z^2 - 1 \qquad \text{and} \qquad x^2 - y^2 = w^2 - 1$$
   has infinitely many solutions in positive integers x, y, z, w.
   [*Hint:* For any integer $n \geq 1$, take $x = 2n^2$ and $y = 2n$.]
   (b) The system of simultaneous equations
   $$x^2 + y^2 = z^2 \qquad \text{and} \qquad x^2 - y^2 = w^2$$
   admits no solution in positive integers x, y, z, w.
   (c) The system of simultaneous equations
   $$x^2 + y^2 = z^2 + 1 \qquad \text{and} \qquad x^2 - y^2 = w^2 + 1$$
   has infinitely many solutions in positive integers x, y, z, w.
   [*Hint:* For any integer $n \geq 1$, take $x = 8n^4 + 1$ and $y = 8n^3$.]
5. Use Problem 4 to establish that there is no solution in positive integers of the simultaneous equations
   $$x^2 + y^2 = z^2 \qquad \text{and} \qquad x^2 + 2y^2 = w^2$$
   [*Hint:* Any solution of the given system also satisfies $z^2 + y^2 = w^2$ and $z^2 - y^2 = x^2$.]
6. Show that there is no solution in positive integers of the simultaneous equations
   $$x^2 + y^2 = z^2 \qquad \text{and} \qquad x^2 + z^2 = w^2$$
   hence, there exists no Pythagorean triangle whose hypotenuse and one of whose sides form the sides of another Pythagorean triangle.
   [*Hint:* Any solution of the given system also satisfies $x^4 + (wy)^2 = z^4$.]
7. Prove that the equation $x^4 - y^4 = 2z^2$ has no solutions in positive integers x, y, z.
   [*Hint:* Because x, y must be both odd or both even, $x^2 + y^2 = 2a^2$, $x + y = 2b^2$, $x - y = 2c^2$ for some a, b, c; hence, $a^2 = b^4 + c^4$.]
8. Verify that the only solution in relatively prime positive integers of the equation $x^4 + y^4 = 2z^2$ is $x = y = z = 1$.
   [*Hint:* Any solution of the given equation also satisfies the equation
   $$z^4 - (xy)^4 = \left(\frac{x^4 - y^4}{2}\right)^2 .]$$

9. Prove that the Diophantine equation $x^4 - 4y^4 = z^2$ has no solution in positive integers x, y, z.
[*Hint:* Rewrite the given equation as $(2y^2)^2 + z^2 = (x^2)^2$ and appeal to Theorem 12.1.]

10. Use Problem 9 to prove that there exists no Pythagorean triangle whose area is twice a perfect square.
[*Hint:* Assume to the contrary that $x^2 + y^2 = z^2$ and $\frac{1}{2}xy = 2w^2$. Then $(x + y)^2 = z^2 + 8w^2$, and $(x - y)^2 = z^2 - 8w^2$. This leads to $z^4 - 4(2w)^4 = (x^2 - y^2)^2$.]

11. Prove the theorem: The only solutions in positive integers of the equation

$$\frac{1}{x^2} + \frac{1}{y^2} = \frac{1}{z^2} \qquad \gcd(x\,,\ y\,,\ z) = 1$$

are given by

$$x = 2st(s^2 + t^2) \qquad y = s^4 - t^4 \qquad z = 2st(s^2 - t^2)$$

where s, t are relatively prime positive integers, one of which is even, with $s > t$.

12. Show that the equation $1/x^4 + 1/y^4 = 1/z^2$ has no solution in positive integers.

CHAPTER 13

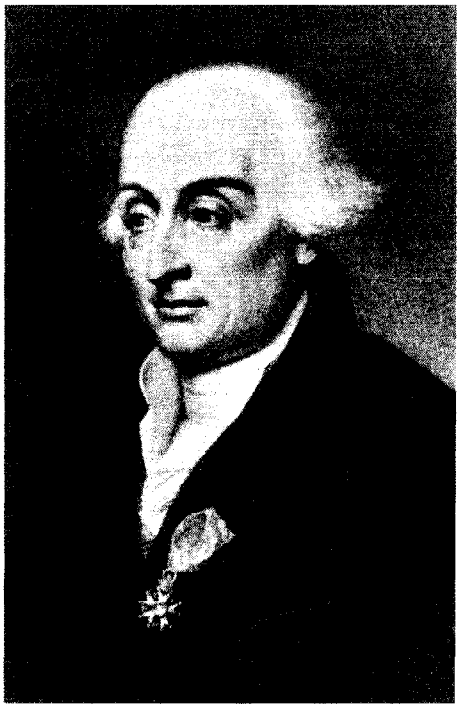
REPRESENTATION OF INTEGERS AS SUMS OF SQUARES

The object of pure Physic is the unfolding of the laws of the intelligible world; the object of pure Mathematic that of unfolding the laws of human intelligence.
J. J. SYLVESTER

13.1 JOSEPH LOUIS LAGRANGE

After the deaths of Descartes, Pascal, and Fermat, no French mathematician of comparable stature appeared for over a century. In England, meanwhile, mathematics was being pursued with restless zeal, first by Newton, then by Taylor, Stirling, and Maclaurin, while Leibniz came upon the scene in Germany. Mathematical activity in Switzerland was marked by the work of the Bernoullis and Euler. Toward the end of the 18th century, Paris did again become the center of mathematical studies, as Lagrange, Laplace, and Legendre brought fresh glory to France.

An Italian by birth, German by adoption, and Frenchman by choice, Joseph Louis Lagrange (1736–1813) was, next to Euler, the foremost mathematician of the 18th century. When he entered the University of Turin, his great interest was in physics, but, after chancing to read a tract by Halley on the merits of Newtonian calculus, he became excited about the new mathematics that was transforming celestial mechanics. He applied himself with such energy to mathematical studies that he was appointed, at the age of 18, Professor of Geometry at the Royal Artillery School in Turin. The French Academy of Sciences soon became accustomed to including Lagrange among the competitors for its biennial prizes: between 1764 and 1788, he



Joseph Louis Lagrange
(1736–1813)

(*Dover Publications, Inc.*)

won five of the coveted prizes for his applications of mathematics to problems in astronomy.

In 1766, when Euler left Berlin for St. Petersburg, Frederick the Great arranged for Lagrange to fill the vacated post, accompanying his invitation with a modest message that said, "It is necessary that the greatest geometer of Europe should live near the greatest of Kings." (To D'Alembert, who had suggested Lagrange's name, the King wrote, "To your care and recommendation am I indebted for having replaced a half-blind mathematician with a mathematician with both eyes, which will especially please the anatomical members of my academy.") For the next 20 years, Lagrange served as director of the mathematics section of the Berlin Academy, producing work of high distinction that culminated in his monumental treatise, the *Mécanique Analytique* (published in 1788 in four volumes). In this work he unified general mechanics and made of it, as the mathematician Hamilton was later to say, "a kind of scientific poem." Holding that mechanics was really a branch of pure mathematics, Lagrange so completely banished geometric ideas from the *Mécanique Analytique* that he could boast in the preface that not a single diagram appeared in its pages.

Frederick the Great died in 1786, and Lagrange, no longer finding a sympathetic atmosphere at the Prussian court, decided to accept the invitation of Louis XVI to settle in Paris, where he took French citizenship. But the years of constant activity had taken their toll: Lagrange fell into a deep mental depression that destroyed his interest in mathematics. So profound was his loathing for the subject that the first printed copy of the *Mécanique Analytique*—the work of a quarter century—lay unexamined on his desk for more than 2 years. Strange to say, it was the turmoil of the French Revolution that helped to awaken him from his lethargy. Following

the abolition of all the old French universities (the Academy of Sciences was also suppressed) in 1793, the revolutionists created two new schools, with the humble titles of École Normale and École Polytechnique, and Lagrange was invited to lecture on analysis. Although he had not lectured since his early days in Turin, having been under royal patronage in the interim, he seemed to welcome the appointment. Subject to constant surveillance, the instructors were pledged "neither to read nor repeat from memory" and transcripts of their lectures as delivered were inspected by the authorities. Despite the petty harassments, Lagrange gained a reputation as an inspiring teacher. His lecture notes on differential calculus formed the basis of another classic in mathematics, the *Théorie des Fonctions Analytique* (1797).

Although Lagrange's research covered an extraordinarily wide spectrum, he possessed, much like Diophantus and Fermat before him, a special talent for the theory of numbers. His work here included: the first proof of Wilson's theorem that if n is a prime, then $(n-1)! \equiv -1 \pmod{n}$; the investigation of the conditions under which ± 2 and ± 5 are quadratic residues or nonresidues of an odd prime (-1 and ± 3 having been discussed by Euler); finding all integral solutions of the equation $x^2 - ay^2 = 1$; and the solution of a number of problems posed by Fermat to the effect that certain primes can be represented in particular ways (typical of these is the result that asserts that every prime $p \equiv 3 \pmod 8$ is of the form $p = a^2 + 2b^2$). This chapter focuses on the discovery for which Lagrange has acquired his greatest renown in number theory, the proof that every positive integer can be expressed as the sum of four squares.

13.2 SUMS OF TWO SQUARES

Historically, a problem that has received a good deal of attention has been that of representing numbers as sums of squares. In the present chapter, we develop enough material to settle completely the following question: What is the smallest value n such that every positive integer can be written as the sum of not more than n squares? Upon examining the first few positive integers, we find that

$$\begin{aligned}
1 &= 1^2 \\
2 &= 1^2 + 1^2 \\
3 &= 1^2 + 1^2 + 1^2 \\
4 &= 2^2 \\
5 &= 2^2 + 1^2 \\
6 &= 2^2 + 1^2 + 1^2 \\
7 &= 2^2 + 1^2 + 1^2 + 1^2
\end{aligned}$$

Because four squares are needed in the representation of 7, a partial answer to our question is that $n \geq 4$. Needless to say, there remains the possibility that some integers might require more than four squares. A justly famous theorem of Lagrange, proved in 1770, asserts that four squares are sufficient; that is, every positive integer is realizable as the sum of four squared integers, some of which may be $0 = 0^2$. This is our Theorem 13.7.

To begin with simpler things, we first find necessary and sufficient conditions that a positive integer be representable as the sum of two squares. The problem may be reduced to the consideration of primes by the following lemma.

Lemma. If m and n are each the sum of two squares, then so is their product mn.

Proof. If $m = a^2 + b^2$ and $n = c^2 + d^2$ for integers a, b, c, d, then

$$mn = (a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

It is clear that not every prime can be written as the sum of two squares; for instance, $3 = a^2 + b^2$ has no solution for integral a and b. More generally, one can prove Theorem 13.1.

Theorem 13.1. No prime p of the form $4k + 3$ is a sum of two squares.

Proof. Modulo 4, we have $a \equiv 0, 1, 2,$ or 3 for any integer a; consequently, $a^2 \equiv 0$ or 1 (mod 4). It follows that, for arbitrary integers a and b,

$$a^2 + b^2 \equiv 0, 1, \text{or } 2 \pmod 4$$

Because $p \equiv 3 \pmod 4$, the equation $p = a^2 + b^2$ is impossible.

On the other hand, any prime that is congruent to 1 modulo 4 is expressible as the sum of two squared integers. The proof, in the form we shall give it, employs a theorem on congruences due to the Norwegian mathematician Axel Thue. This, in its turn, relies on Dirichlet's famous pigeonhole principle.

Pigeonhole principle. If n objects are placed in m boxes (or pigeonholes) and if $n > m$, then some box will contain at least two objects.

Phrased in more mathematical terms, this simple principle asserts that if a set with n elements is the union of m of its subsets and if $n > m$, then some subset has more than one element.

Lemma Thue's lemma. Let p be a prime and let $\gcd(a, p) = 1$. Then the congruence

$$ax \equiv y \pmod p$$

admits a solution x_0, y_0, where

$$0 < |x_0| < \sqrt{p} \qquad \text{and} \qquad 0 < |y_0| < \sqrt{p}$$

Proof. Let $k = [\sqrt{p}] + 1$, and consider the set of integers

$$S = \{ax - y | 0 \le x \le k - 1, 0 \le y \le k - 1\}$$

Because $ax - y$ takes on $k^2 > p$ possible values, the pigeonhole principle guarantees that at least two members of S must be congruent modulo p; call them $ax_1 - y_1$ and

$ax_2 - y_2$, where $x_1 \neq x_2$ or $y_1 \neq y_2$. Then we can write

$$a(x_1 - x_2) \equiv y_1 - y_2 \pmod{p}$$

Setting $x_0 = x_1 - x_2$ and $y_0 = y_1 - y_2$, it follows that x_0 and y_0 provide a solution to the congruence $ax \equiv y \pmod{p}$. If either x_0 or y_0 is equal to zero, then the fact that $\gcd(a, p) = 1$ can be used to show that the other must also be zero, contrary to assumption. Hence, $0 < |x_0| \leq k - 1 < \sqrt{p}$ and $0 < |y_0| \leq k - 1 < \sqrt{p}$.

We are now ready to derive the theorem of Fermat that every prime of the form $4k + 1$ can be expressed as the sum of squares of two integers. (In terms of priority, Albert Girard recognized this fact several years earlier and the result is sometimes referred to as Girard's theorem.) Fermat communicated his theorem in a letter to Mersenne, dated December 25, 1640, stating that he possessed an irrefutable proof. However, the first published proof was given by Euler in 1754, who in addition succeeded in showing that the representation is unique.

Theorem 13.2 Fermat. An odd prime p is expressible as a sum of two squares if and only if $p \equiv 1 \pmod{4}$.

Proof. Although the "only if" part is covered by Theorem 13.1, let us give a different proof here. Suppose that p can be written as the sum of two squares, let us say $p = a^2 + b^2$. Because p is a prime, we have $p \nmid a$ and $p \nmid b$. (If $p \mid a$, then $p \mid b^2$, and so $p \mid b$, leading to the contradiction that $p^2 \mid p$.) Thus, by the theory of linear congruences, there exists an integer c for which $bc \equiv 1 \pmod{p}$. Modulo p, the relation $(ac)^2 + (bc)^2 = pc^2$ becomes

$$(ac)^2 \equiv -1 \pmod{p}$$

making -1 a quadratic residue of p. At this point, the corollary to Theorem 9.2 comes to our aid, for $(-1/p) = 1$ only when $p \equiv 1 \pmod{4}$.

For the converse, assume that $p \equiv 1 \pmod{4}$. Because -1 is a quadratic residue of p, we can find an integer a satisfying $a^2 \equiv -1 \pmod{p}$; in fact, by Theorem 5.4, $a = [(p - 1)/2]!$ is one such integer. Now $\gcd(a, p) = 1$, so that the congruence

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

admits a solution x_0, y_0 for which the conclusion of Thue's lemma holds. As a result,

$$-x_0^2 \equiv a^2 x_0^2 \equiv (ax_0)^2 \equiv y_0^2 \pmod{p}$$

or $x_0^2 + y_0^2 \equiv 0 \pmod{p}$. This says that

$$x_0^2 + y_0^2 = kp$$

for some integer $k \geq 1$. Inasmuch as $0 < |x_0| < \sqrt{p}$ and $0 < |y_0| < \sqrt{p}$, we obtain $0 < x_0^2 + y_0^2 < 2p$, the implication of which is that $k = 1$. Consequently, $x_0^2 + y_0^2 = p$, and we are finished.

Counting a^2 and $(-a)^2$ as the same, we have the following corollary.

Corollary. Any prime p of the form $4k + 1$ can be represented uniquely (aside from the order of the summands) as a sum of two squares.

Proof. To establish the uniqueness assertion, suppose that

$$p = a^2 + b^2 = c^2 + d^2$$

where a, b, c, d are all positive integers. Then

$$a^2d^2 - b^2c^2 = p(d^2 - b^2) \equiv 0 \pmod{p}$$

whence $ad \equiv bc \pmod{p}$ or $ad \equiv -bc \pmod{p}$. Because a, b, c, d are all less than $\sqrt{p}$, these relations imply that

$$ad - bc = 0 \qquad \text{or} \qquad ad + bc = p$$

If the second equality holds, then we would have $ac = bd$; for,

$$\begin{aligned} p^2 = (a^2 + b^2)(c^2 + d^2) &= (ad + bc)^2 + (ac - bd)^2 \\ &= p^2 + (ac - bd)^2 \end{aligned}$$

and so $ac - bd = 0$. It follows that either

$$ad = bc \qquad \text{or} \qquad ac = bd$$

Suppose, for instance, that $ad = bc$. Then $a \mid bc$, with $\gcd(a, b) = 1$, which forces $a \mid c$; say, $c = ka$. The condition $ad = bc = b(ka)$ then reduces to $d = bk$. But

$$p = c^2 + d^2 = k^2(a^2 + b^2)$$

implies that $k = 1$. In this case, we get $a = c$ and $b = d$. By a similar argument, the condition $ac = bd$ leads to $a = d$ and $b = c$. What is important is that, in either event, our two representations of the prime p turn out to be identical.

Let us follow the steps in Theorem 13.2, using the prime $p = 13$. One choice for the integer a is $6! = 720$. A solution of the congruence $720x \equiv y \pmod{13}$, or rather,

$$5x \equiv y \pmod{13}$$

is obtained by considering the set

$$S = \{5x - y \mid 0 \leq x, y < 4\}$$

The elements of S are just the integers

$$\begin{array}{rrrr} 0 & 5 & 10 & 15 \\ -1 & 4 & 9 & 14 \\ -2 & 3 & 8 & 13 \\ -3 & 2 & 7 & 12 \end{array}$$

which, modulo 13, become

$$\begin{array}{rrrr} 0 & 5 & 10 & 2 \\ 12 & 4 & 9 & 1 \\ 11 & 3 & 8 & 0 \\ 10 & 2 & 7 & 12 \end{array}$$

Among the various possibilities, we have

$$5 \cdot 1 - 3 \equiv 2 \equiv 5 \cdot 3 - 0 \pmod{13}$$

or

$$5(1 - 3) \equiv 3 \pmod{13}$$

Thus, we may take $x_0 = -2$ and $y_0 = 3$ to obtain

$$13 = x_0^2 + y_0^2 = 2^2 + 3^2$$

Remark. Some authors would claim that any prime $p \equiv 1 \pmod 4$ can be written as a sum of squares in eight ways. For with $p = 13$, we have

$$\begin{aligned} 13 &= 2^2 + 3^3 = 2^2 + (-3)^2 = (-2)^2 + 3^2 = (-2)^2 + (-3)^2 \\ &= 3^2 + 2^2 = 3^2 + (-2)^2 = (-3)^2 + 2^2 = (-3)^2 + (-2)^2 \end{aligned}$$

Because all eight representations can be obtained from any one of them by interchanging the signs of 2 and 3 or by interchanging the summands, there is "essentially" only one way of doing this. Thus, from our point of view, 13 is uniquely representable as the sum of two squares.

We have shown that every prime p such that $p \equiv 1 \pmod 4$ is expressible as the sum of two squares. But other integers also enjoy this property; for instance,

$$10 = 1^2 + 3^2$$

The next step in our program is to characterize explicitly those positive integers that can be realized as the sum of two squares.

Theorem 13.3. Let the positive integer n be written as $n = N^2 m$, where m is square-free. Then n can be represented as the sum of two squares if and only if m contains no prime factor of the form $4k + 3$.

Proof. To start, suppose that m has no prime factor of the form $4k + 3$. If $m = 1$, then $n = N^2 + 0^2$ and we are through. In the case in which $m > 1$, let $m = p_1 p_2 \cdots p_r$ be the factorization of m into a product of distinct primes. Each of these primes p_i, being equal to 2 or of the form $4k + 1$, can be written as the sum of two squares. Now, the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

shows that the product of two (and, by induction, any finite number) integers, each of which is representable as a sum of two squares, is likewise so representable. Thus, there exist integers x and y satisfying $m = x^2 + y^2$. We end up with

$$n = N^2 m = N^2(x^2 + y^2) = (Nx)^2 + (Ny)^2$$

a sum of two squares.

Now for the opposite direction. Assume that n can be represented as the sum of two squares

$$n = a^2 + b^2 = N^2 m$$

and let p be any odd prime divisor of m (without loss of generality, it may be assumed that $m > 1$). If $d = \gcd(a, b)$, then $a = rd$, $b = sd$, where $\gcd(r, s) = 1$. We get

$$d^2(r^2 + s^2) = N^2 m$$

and so, m being square-free, $d^2 \mid N^2$. But then

$$r^2 + s^2 = \left(\frac{N^2}{d^2}\right) m = tp$$

for some integer t, which leads to

$$r^2 + s^2 \equiv 0 \pmod{p}$$

Now the condition $\gcd(r, s) = 1$ implies that one of r or s, say r, is relatively prime to p. Let r' satisfy the congruence

$$rr' \equiv 1 \pmod{p}$$

When the equation $r^2 + s^2 \equiv 0 \pmod{p}$ is multiplied by $(r')^2$, we obtain

$$(sr')^2 + 1 \equiv 0 \pmod{p}$$

or, to put it differently, $(-1/p) = 1$. Because -1 is a quadratic residue of p, Theorem 9.2 ensures that $p \equiv 1 \pmod{4}$. The implication of our reasoning is that there is no prime of the form $4k + 3$ that divides m.

The following is a corollary to the preceding analysis.

Corollary. A positive integer n is representable as the sum of two squares if and only if each of its prime factors of the form $4k + 3$ occurs to an even power.

Example 13.1. The integer 459 cannot be written as the sum of two squares, because $459 = 3^3 \cdot 17$, with the prime 3 occurring to an odd exponent. On the other hand, $153 = 3^2 \cdot 17$ admits the representation

$$153 = 3^2(4^2 + 1^2) = 12^2 + 3^2$$

Somewhat more complicated is the example $n = 5 \cdot 7^2 \cdot 13 \cdot 17$. In this case, we have

$$n = 7^2 \cdot 5 \cdot 13 \cdot 17 = 7^2(2^2 + 1^2)(3^2 + 2^2)(4^2 + 1^2)$$

Two applications of the identity appearing in Theorem 13.3 give

$$(3^2 + 2^2)(4^2 + 1^2) = (12 + 2)^2 + (3 - 8)^2 = 14^2 + 5^2$$

and

$$(2^2 + 1^2)(14^2 + 5^2) = (28 + 5)^2 + (10 - 14)^2 = 33^2 + 4^2$$

When these are combined, we end up with

$$n = 7^2(33^2 + 4^2) = 231^2 + 28^2$$

There exist certain positive integers (obviously, not primes of the form $4k + 1$) that can be represented in more than one way as the sum of two squares. The smallest is

$$25 = 4^2 + 3^2 = 5^2 + 0^2$$

If $a \equiv b \pmod 2$, then the relation

$$ab = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$$

allows us to manufacture a variety of such examples. Take $n = 153$ as an illustration; here,

$$153 = 17 \cdot 9 = \left(\frac{17+9}{2}\right)^2 - \left(\frac{17-9}{2}\right)^2 = 13^2 - 4^2$$

and

$$153 = 51 \cdot 3 = \left(\frac{51+3}{2}\right)^2 - \left(\frac{51-3}{2}\right)^2 = 27^2 - 24^2$$

so that

$$13^2 - 4^2 = 27^2 - 24^2$$

This yields the two distinct representations

$$27^2 + 4^2 = 24^2 + 13^2 = 745$$

At this stage, a natural question should suggest itself: What positive integers admit a representation as the difference of two squares? We answer this below.

Theorem 13.4. A positive integer n can be represented as the difference of two squares if and only if n is not of the form $4k + 2$.

Proof. Because $a^2 \equiv 0$ or $1 \pmod 4$ for all integers a, it follows that

$$a^2 - b^2 \equiv 0, 1, \text{or } 3 \pmod 4$$

Thus, if $n \equiv 2 \pmod 4$, we cannot have $n = a^2 - b^2$ for any choice of a and b.

Turning affairs around, suppose that the integer n is not of the form $4k + 2$; that is to say, $n \equiv 0, 1, \text{or } 3 \pmod 4$. If $n \equiv 1 \text{ or } 3 \pmod 4$, then $n + 1$ and $n - 1$ are both even integers; hence, n can be written as

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

a difference of squares. If $n \equiv 0 \pmod 4$, then we have

$$n = \left(\frac{n}{4} + 1\right)^2 - \left(\frac{n}{4} - 1\right)^2$$

Corollary. An odd prime is the difference of two successive squares.

Examples of this last corollary are afforded by

$$11 = 6^2 - 5^2 \qquad 17 = 9^2 - 8^2 \qquad 29 = 15^2 - 14^2$$

Another point worth mentioning is that the representation of a given prime p as the difference of two squares is unique. To see this, suppose that

$$p = a^2 - b^2 = (a - b)(a + b)$$

where $a > b > 0$. Because 1 and p are the only factors of p, necessarily we have

$$a - b = 1 \qquad \text{and} \qquad a + b = p$$

from which it may be inferred that

$$a = \frac{p+1}{2} \qquad \text{and} \qquad b = \frac{p-1}{2}$$

Thus, any odd prime p can be written as the difference of the squares of two integers in precisely one way; namely, as

$$p = \left(\frac{p+1}{2}\right)^2 - \left(\frac{p-1}{2}\right)^2$$

A different situation occurs when we pass from primes to arbitrary integers. Suppose that n is a positive integer that is neither prime nor of the form $4k + 2$. Starting with a divisor d of n, put $d' = n/d$ (it is harmless to assume that $d \geq d'$). Now if d and d' are both even, or both odd, then $(d + d')/2$ and $(d - d')2$ are integers. Furthermore, we may write

$$n = dd' = \left(\frac{d+d'}{2}\right)^2 - \left(\frac{d-d'}{2}\right)^2$$

By way of illustration, consider the integer $n = 24$. Here,

$$24 = 12 \cdot 2 = \left(\frac{12+2}{2}\right)^2 - \left(\frac{12-2}{2}\right)^2 = 7^2 - 5^2$$

and

$$24 = 6 \cdot 4 = \left(\frac{6+4}{2}\right)^2 - \left(\frac{6-4}{2}\right)^2 = 5^2 - 1^2$$

giving us two representations for 24 as the difference of squares.

PROBLEMS 13.2

1. Represent each of the primes 113, 229, and 373 as a sum of two squares.
2. (a) It has been conjectured that there exist infinitely many prime numbers p such that $p = n^2 + (n + 1)^2$ for some positive integer n; for example, $5 = 1^2 + 2^2$ and $13 = 2^2 + 3^2$. Find five more of these primes.
   (b) Another conjecture is that there are infinitely many prime numbers p of the form $p = 2^2 + p_1^2$, where p_1 is a prime. Find five such primes.
3. Establish each of the following assertions:
   (a) Each of the integers 2^n, where $n = 1, 2, 3, \ldots$, is a sum of two squares.
   (b) If $n \equiv 3$ or 6 (mod 9), then n cannot be represented as a sum of two squares.
   (c) If n is the sum of two triangular numbers, then $4n + 1$ is the sum of two squares.

(d) Every Fermat number $F_n = 2^{2^n} + 1$, where $n \geq 1$, can be expressed as the sum of two squares.

(e) Every odd perfect number (if one exists) is the sum of two squares.
[*Hint:* See the Corollary to Theorem 11.7.]

4. Prove that a prime p can be written as a sum of two squares if and only if the congruence $x^2 + 1 \equiv 0 \pmod{p}$ admits a solution.

5. (a) Show that a positive integer n is a sum of two squares if and only if $n = 2^m a^2 b$, where $m \geq 0$, a is an odd integer, and every prime divisor of b is of the form $4k + 1$.

(b) Write the integers $3185 = 5 \cdot 7^2 \cdot 13$; $39690 = 2 \cdot 3^4 \cdot 5 \cdot 7^2$; and $62920 = 2^3 \cdot 5 \cdot 11^2 \cdot 13$ as a sum of two squares.

6. Find a positive integer having at least three different representations as the sum of two squares, disregarding signs and the order of the summands.
[*Hint:* Choose an integer that has three distinct prime factors, each of the form $4k + 1$.]

7. If the positive integer n is not the sum of squares of two integers, show that n cannot be represented as the sum of two squares of rational numbers.
[*Hint:* By Theorem 13.3, there is a prime $p \equiv 3 \pmod 4$ and an odd integer k such that $p^k \mid n$, whereas $p^{k+1} \nmid n$. If $n = (a/b)^2 + (c/d)^2$, then p will occur to an odd power on the left-hand side of the equation $n(bd)^2 = (ad)^2 + (bc)^2$, but not on the right-hand side.]

8. Prove that the positive integer n has as many representations as the sum of two squares as does the integer $2n$.
[*Hint:* Starting with a representation of n as a sum of two squares, obtain a similar representation for $2n$, and conversely.]

9. (a) If n is a triangular number, show that each of the three successive integers $8n^2$, $8n^2 + 1$, $8n^2 + 2$ can be written as a sum of two squares.

(b) Prove that of any four consecutive integers, at least one is not representable as a sum of two squares.

10. Prove the following:

(a) If a prime number is the sum of two or four squares of different primes, then one of these primes must be equal to 2.

(b) If a prime number is the sum of three squares of different primes, then one of these primes must be equal to 3.

11. (a) Let p be an odd prime. If $p \mid a^2 + b^2$, where $\gcd(a, b) = 1$, prove that the prime $p \equiv 1 \pmod 4$.
[*Hint:* Raise the congruence $a^2 \equiv -b^2 \pmod p$ to the power $(p - 1)/2$ and apply Fermat's theorem to conclude that $(-1)^{(p-1)/2} = 1$.]

(b) Use part (a) to show that any positive divisor of a sum of two relatively prime squares is itself a sum of two squares.

12. Establish that every prime number p of the form $8k + 1$ or $8k + 3$ can be written as $p = a^2 + 2b^2$ for some choice of integers a and b.
[*Hint:* Mimic the proof of Theorem 13.2.]

13. Prove the following:

(a) A positive integer is representable as the difference of two squares if and only if it is the product of two factors that are both even or both odd.

(b) A positive even integer can be written as the difference of two squares if and only if it is divisible by 4.

14. Verify that 45 is the smallest positive integer admitting three distinct representations as the difference of two squares.
[*Hint:* See part (a) of the previous problem.]

15. For any $n > 0$, show that there exists a positive integer that can be expressed in n distinct ways as the difference of two squares.
[*Hint:* Note that, for $k = 1, 2, \ldots, n$,

$$2^{2n+1} = (2^{2n-k} + 2^{k-1})^2 - (2^{2n-k} - 2^{k-1})^2.]$$

16. Prove that every prime $p \equiv 1 \pmod 4$ divides the sum of two relatively prime squares, where each square exceeds 3.
[*Hint:* Given an odd primitive root r of p, we have $r^k \equiv 2 \pmod p$ for some k; hence $r^{2[k+(p-1)/4]} \equiv -4 \pmod p$.]

17. For a prime $p \equiv 1$ or $3 \pmod 8$, show that the equation $x^2 + 2y^2 = p$ has a solution.

18. The English number theorist G. H. Hardy relates the following story about his young protégé Ramanujan: "I remember going to see him once when he was lying ill in Putney. I had ridden in taxi-cab No. 1729, and remarked that the number seemed to me rather a dull one, and that I hoped it was not an unfavorable omen. 'No,' he reflected, 'it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways.'" Verify Ramanujan's assertion.

13.3 SUMS OF MORE THAN TWO SQUARES

Although not every positive integer can be written as the sum of two squares, what about their representation in terms of three squares (0^2 still permitted)? With an extra square to add, it seems reasonable that there should be fewer exceptions. For instance, when only two squares are allowed, we have no representation for such integers as 14, 33, and 67, but

$$14 = 3^2 + 2^2 + 1^2 \qquad 33 = 5^2 + 2^2 + 2^2 \qquad 67 = 7^2 + 3^2 + 3^2$$

It is still possible to find integers that are not expressible as the sum of three squares. Theorem 13.5 speaks to this point.

Theorem 13.5. No positive integer of the form $4^n(8m + 7)$ can be represented as the sum of three squares.

Proof. To start, let us show that the integer $8m + 7$ is not expressible as the sum of three squares. For any integer a, we have $a^2 \equiv 0, 1$, or $4 \pmod 8$. It follows that

$$a^2 + b^2 + c^2 \equiv 0, 1, 2, 3, 4, 5, \text{or } 6 \pmod 8$$

for any choice of integers a, b, c. Because we have $8m + 7 \equiv 7 \pmod 8$, the equation $a^2 + b^2 + c^2 = 8m + 7$ is impossible.

Next, let us suppose that $4^n(8m + 7)$, where $n \geq 1$, can be written as

$$4^n(8m + 7) = a^2 + b^2 + c^2$$

Then each of the integers a, b, c must be even. Putting $a = 2a_1, b = 2b_1, c = 2c_1$, we get

$$4^{n-1}(8m + 7) = a_1^2 + b_1^2 + c_1^2$$

If $n - 1 \geq 1$, the argument may be repeated until $8m + 7$ is eventually represented as the sum of three squared integers; this, of course, contradicts the result of the first paragraph.

We can prove that the condition of Theorem 13.5 is also sufficient in order that a positive integer be realizable as the sum of three squares; however, the argument

is much too difficult for inclusion here. Part of the trouble is that, unlike the case of two (or even four) squares, there is no algebraic identity that expresses the product of sums of three squares as a sum of three squares.

With this trace of ignorance left showing, let us make a few historical remarks. Diophantus conjectured, in effect, that no number of the form $8m + 7$ is the sum of three squares, a fact easily verified by Descartes in 1638. It seems fair to credit Fermat with being the first to state in full the criterion that a number can be written as a sum of three squared integers if and only if it is not of the form $4^n(8m + 7)$, where m and n are nonnegative integers. This was proved in a complicated manner by Legendre in 1798 and more clearly (but by no means easily) by Gauss in 1801.

As just indicated, there exist positive integers that are not representable as the sum of either two or three squares (take 7 and 15, for simple examples). Things change dramatically when we turn to four squares: There are no exceptions at all!

The first explicit reference to the fact that every positive integer can be written as the sum of four squares, counting 0^2, was made by Bachet (in 1621) and he checked this conjecture for all integers up to 325. Fifteen years later Fermat claimed that he had a proof using his favorite method of infinite descent; however, as usual, he gave no details. Both Bachet and Fermat felt that Diophantus must have known the result; the evidence is entirely conjectural: Diophantus gave necessary conditions in order that a number be the sum of two or three squares, while making no mention of a condition for a representation as a sum of four squares.

One measure of the difficulty of the problem is the fact that Euler, despite his brilliant achievements, wrestled with it for more than 40 years without success. Nonetheless, his contribution toward the eventual solution was substantial; Euler discovered the fundamental identity that allows one to express the product of two sums of four squares as such a sum, and the crucial result that the congruence $x^2 + y^2 + 1 \equiv 0 \pmod{p}$ is solvable for any prime p. A complete proof of the four-square conjecture was published by Lagrange in 1772, who acknowledged his indebtedness to the ideas of Euler. The next year, Euler offered a much simpler demonstration, which is essentially the version to be presented here.

It is convenient to establish two preparatory lemmas, so as not to interrupt the main argument at an awkward stage. The proof of the first contains the algebraic identity (Euler's identity) that allows us to reduce the four-square problem to the consideration of prime numbers only.

Lemma 1 Euler. If the integers m and n are each the sum of four squares, then mn is likewise so representable.

Proof. If $m = a_1^2 + a_2^2 + a_3^2 + a_4^2$ and $n = b_1^2 + b_2^2 + b_3^2 + b_4^2$ for integers a_i, b_i, then

$$\begin{aligned} mn &= (a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) \\ &= (a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4)^2 \\ &\quad + (a_1b_2 - a_2b_1 + a_3b_4 - a_4b_3)^2 \\ &\quad + (a_1b_3 - a_2b_4 - a_3b_1 + a_4b_2)^2 \\ &\quad + (a_1b_4 + a_2b_3 - a_3b_2 - a_4b_1)^2 \end{aligned}$$

We confirm this cumbersome identity by brute force: Just multiply everything out and compare terms. The details are not suitable for the printed page.

Another basic ingredient in our development is Lemma 2.

Lemma 2. If p is an odd prime, then the congruence

$$x^2 + y^2 + 1 \equiv 0 \pmod{p}$$

has a solution x_0, y_0 where $0 \leq x_0 \leq (p-1)/2$ and $0 \leq y_0 \leq (p-1)/2$.

Proof. The idea of the proof is to consider the following two sets:

$$S_1 = \left\{1 + 0^2, 1 + 1^2, 1 + 2^2, \ldots, 1 + \left(\frac{p-1}{2}\right)^2\right\}$$

$$S_2 = \left\{-0^2, -1^2, -2^2, \ldots, -\left(\frac{p-1}{2}\right)^2\right\}$$

No two elements of the set S_1 are congruent modulo p. For if $1 + x_1^2 \equiv 1 + x_2^2 \pmod{p}$, then either $x_1 \equiv x_2 \pmod{p}$ or $x_1 \equiv -x_2 \pmod{p}$. But the latter consequence is impossible, because $0 < x_1 + x_2 < p$ (unless $x_1 = x_2 = 0$), whence $x_1 \equiv x_2 \pmod{p}$, which implies that $x_1 = x_2$. In the same vein, no two elements of S_2 are congruent modulo p.

Together S_1 and S_2 contain $2[1 + \frac{1}{2}(p-1)] = p + 1$ integers. By the pigeonhole principle, some integer in S_1 must be congruent modulo p to some integer in S_2; that is, there exist x_0, y_0 such that

$$1 + x_0^2 \equiv -y_0^2 \pmod{p}$$

where $0 \leq x_0 \leq (p-1)/2$ and $0 \leq y_0 \leq (p-1)/2$.

Corollary. Given an odd prime p, there exists an integer $k < p$ such that kp is the sum of four squares.

Proof. According to the theorem, we can find integers x_0 and y_0,

$$0 \leq x_0 < \frac{p}{2} \qquad 0 \leq y_0 < \frac{p}{2}$$

such that

$$x_0^2 + y_0^2 + 1^2 + 0^2 = kp$$

for a suitable choice of k. The restrictions on the size of x_0 and y_0 imply that

$$kp = x_0^2 + y_0^2 + 1 < \frac{p^2}{4} + \frac{p^2}{4} + 1 < p^2$$

and so $k < p$, as asserted in the corollary.

Example 13.2. We digress for a moment to look at an example. If we take $p = 17$, then the sets S_1 and S_2 become

$$S_1 = \{1, 2, 5, 10, 17, 26, 37, 50, 65\}$$

and

$$S_2 = \{0, -1, -4, -9, -16, -25, -36, -49, -64\}$$

Modulo 17, the set S_1 consists of the integers 1, 2, 5, 10, 0, 9, 3, 16, 14, and those in S_2 are 0, 16, 13, 8, 1, 9, 15, 2, 4. Lemma 2 tells us that some member $1 + x^2$ of the first set is congruent to some member $-y^2$ of the second set. We have, among the various possibilities,

$$1 + 5^2 \equiv 9 \equiv -5^2 \pmod{17}$$

or $1 + 5^2 + 5^2 \equiv 0 \pmod{17}$. It follows that

$$3 \cdot 17 = 1^2 + 5^2 + 5^2 + 0^2$$

is a multiple of 17 written as a sum of four squares.

The last lemma is so essential to our work that it is worth pointing out another approach, this one involving the theory of quadratic residues. If $p \equiv 1 \pmod 4$, we may choose x_0 to be a solution of $x^2 \equiv -1 \pmod p$ (this is permissible by the corollary to Theorem 9.2) and $y_0 = 0$ to get

$$x_0^2 + y_0^2 + 1 \equiv 0 \pmod p$$

Thus, it suffices to concentrate on the case $p \equiv 3 \pmod 4$. We first pick the integer a to be the smallest positive quadratic nonresidue of p (keep in mind that $a \geq 2$, because 1 is a quadratic residue). Then

$$(-a/p) = (-1/p)(a/p) = (-1)(-1) = 1$$

so that $-a$ is a quadratic residue of p. Hence, the congruence

$$x^2 \equiv -a \pmod p$$

admits a solution x_0, with $0 < x_0 \leq (p-1)/2$. Now $a - 1$, being positive and smaller than a, must itself be a quadratic residue of p. Thus, there exists an integer y_0, where $0 < y_0 \leq (p-1)/2$, satisfying

$$y^2 \equiv a - 1 \pmod p$$

The conclusion is

$$x_0^2 + y_0^2 + 1 \equiv -a + (a-1) + 1 \equiv 0 \pmod p$$

With these two lemmas among our tools, we now have the necessary information to carry out a proof of the fact that any prime can be realized as the sum of four squared integers.

Theorem 13.6. Any prime p can be written as the sum of four squares.

Proof. The theorem is certainly true for $p = 2$, because $2 = 1^2 + 1^2 + 0^2 + 0^2$. Thus, we may hereafter restrict our attention to odd primes. Let k be the smallest positive integer such that kp is the sum of four squares; say,

$$kp = x^2 + y^2 + z^2 + w^2$$

By virtue of the foregoing corollary, $k < p$. The crux of our argument is that $k = 1$.

We make a start by showing that k is an odd integer. For a proof by contradiction, assume that k is even. Then x, y, z, w are all even; or all are odd; or two are even and two are odd. In any event, we may rearrange them, so that

$$x \equiv y \pmod 2 \qquad \text{and} \qquad z \equiv w \pmod 2$$

It follows that

$$\frac{1}{2}(x-y) \qquad \frac{1}{2}(x+y) \qquad \frac{1}{2}(z-w) \qquad \frac{1}{2}(z+w)$$

are all integers and

$$\frac{1}{2}(kp) = \left(\frac{x-y}{2}\right)^2 + \left(\frac{x+y}{2}\right)^2 + \left(\frac{z-w}{2}\right)^2 + \left(\frac{z+w}{2}\right)^2$$

is a representation of $(k/2)p$ as a sum of four squares. This violates the minimal nature of k, giving us our contradiction.

There still remains the problem of showing that $k = 1$. Assume that $k \neq 1$; then k, being an odd integer, is at least 3. It is therefore possible to choose integers a, b, c, d such that

$$a \equiv x \pmod k \qquad b \equiv y \pmod k \qquad c \equiv z \pmod k \qquad d \equiv w \pmod k$$

and

$$|a| < \frac{k}{2} \qquad |b| < \frac{k}{2} \qquad |c| < \frac{k}{2} \qquad |d| < \frac{k}{2}$$

(To obtain the integer a, for instance, find the remainder r when x is divided by k; put $a = r$ or $a = r - k$ according as $r < k/2$ or $r > k/2$.) Then

$$a^2 + b^2 + c^2 + d^2 \equiv x^2 + y^2 + z^2 + w^2 \equiv 0 \pmod k$$

and therefore

$$a^2 + b^2 + c^2 + d^2 = nk$$

for some nonnegative integer n. Because of the restrictions on the size of a, b, c, d,

$$0 \leq nk = a^2 + b^2 + c^2 + d^2 < 4\left(\frac{k}{2}\right)^2 = k^2$$

We cannot have $n = 0$, because this would signify that $a = b = c = d = 0$ and, in consequence, that k divides each of the integers x, y, z, w. Then $k^2 \mid kp$, or $k \mid p$, which is impossible in light of the inequality $1 < k < p$. The relation $nk < k^2$ also allows us to conclude that $n < k$. In summary: $0 < n < k$. Combining the various pieces, we get

$$\begin{aligned} k^2np = (kp)(kn) &= (x^2 + y^2 + z^2 + w^2)(a^2 + b^2 + c^2 + d^2) \\ &= r^2 + s^2 + t^2 + u^2 \end{aligned}$$

where

$$\begin{aligned} r &= xa + yb + zc + wd \\ s &= xb - ya + zd - wc \\ t &= xc - yd - za + wb \\ u &= xd + yc - zb - wa \end{aligned}$$

It is important to observe that all four of r, s, t, u are divisible by k. In the case of the integer r, for example, we have

$$r = xa + yb + zc + wd \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{k}$$

Similarly, $s \equiv t \equiv u \equiv 0 \pmod{k}$. This leads to the representation

$$np = \left(\frac{r}{k}\right)^2 + \left(\frac{s}{k}\right)^2 + \left(\frac{t}{k}\right)^2 + \left(\frac{u}{k}\right)^2$$

where $r/k, s/k, t/k, u/k$ are all integers. Because $0 < n < k$, we therefore arrive at a contradiction to the choice of k as the smallest positive integer for which kp is the sum of four squares. With this contradiction, $k = 1$, and the proof is finally complete.

This brings us to our ultimate objective, the classical result of Lagrange.

Theorem 13.7 Lagrange. Any positive integer n can be written as the sum of four squares, some of which may be zero.

Proof. Clearly, the integer 1 is expressible as $1 = 1^2 + 0^2 + 0^2 + 0^2$, a sum of four squares. Assume that $n > 1$ and let $n = p_1 p_2 \cdots p_r$ be the factorization of n into (not necessarily distinct) primes. Because each p_i is realizable as a sum of four squares, Euler's identity permits us to express the product of any two primes as a sum of four squares. This, by induction, extends to any finite number of prime factors, so that applying the identity $r - 1$ times, we obtain the desired representation for n.

Example 13.3. To write the integer $459 = 3^3 \cdot 17$ as the sum of four squares, we use Euler's identity as follows:

$$\begin{aligned}
459 &= 3^2 \cdot 3 \cdot 17 \\
&= 3^2(1^2 + 1^2 + 1^2 + 0^2)(4^2 + 1^2 + 0^2 + 0^2) \\
&= 3^2[(4 + 1 + 0 + 0)^2 + (1 - 4 + 0 - 0)^2 \\
&\quad + (0 - 0 - 4 + 0)^2 + (0 + 0 - 1 - 0)^2] \\
&= 3^2[5^2 + 3^2 + 4^2 + 1^2] \\
&= 15^2 + 9^2 + 12^2 + 3^2
\end{aligned}$$

Although squares have received all our attention so far, many of the ideas involved generalize to higher powers.

In his book, *Meditationes Algebraicae* (1770), Edward Waring stated that each positive integer is expressible as a sum of at most 9 cubes, also a sum of at most 19 fourth powers, and so on. This assertion has been interpreted to mean the following: Can each positive integer be written as the sum of no more than a fixed number $g(k)$ of kth powers, where $g(k)$ depends only on k, not the integer being represented? In other words, for a given k, a number $g(k)$ is sought such that every $n > 0$ can be represented in at least one way as

$$n = a_1^k + a_2^k + \cdots + a_{g(k)}^k$$

where the a_i are nonnegative integers, not necessarily distinct. The resulting problem was the starting point of a large body of research in number theory on what has

become known as "Waring's problem." There seems little doubt that Waring had limited numerical grounds in favor of his assertion and no shadow of a proof.

As we have reported in Theorem 13.7, $g(2) = 4$. Except for squares, the first case of a Waring-type theorem actually proved is attributed to Liouville (1859): Every positive integer is a sum of at most 53 fourth powers. This bound for $g(4)$ is somewhat inflated, and through the years it was progressively reduced. The existence of $g(k)$ for each value of k was resolved in the affirmative by Hilbert in 1909; unfortunately, his proof relies on heavy machinery (including a 25-fold integral at one stage) and is in no way constructive.

Once it is known that Waring's problem admits a solution, a natural question to pose is "How big is $g(k)$?" There is an extensive literature on this aspect of the problem, but the question itself is still open. A sample result, due to Leonard Dickson, is that $g(3) = 9$, whereas

$$23 = 2^3 + 2^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3$$

and

$$239 = 4^3 + 4^3 + 3^3 + 3^3 + 3^3 + 3^3 + 1^3 + 1^3 + 1^3$$

are the only integers that actually require as many as 9 cubes in their representation; each integer greater than 239 can be realized as the sum of at most 8 cubes. In 1942, Linnik proved that only a finite number of integers need 8 cubes; from some point onward 7 will suffice. Whether 6 cubes are also sufficient to obtain all but finitely many positive integers is still unsettled.

The cases $k = 4$ and $k = 5$ have turned out to be the most subtle. For many years, the best-known result was that $g(4)$ lay somewhere in the range $19 \leq g(4) \leq 35$, whereas $g(5)$ satisfied $37 \leq g(5) \leq 54$. Subsequent work (1964) has shown that $g(5) = 37$. The upper bound on $g(4)$ was decreased dramatically during the 1970s, the sharpest estimate being $g(4) \leq 22$. It was also proved that every integer less than 10^{140} or greater than 10^{367} can be written as a sum of at most 19 fourth powers; thus, in principle, $g(4)$ could be calculated. The relatively recent (1986) announcement that, in fact, 19 fourth powers suffice to represent all integers settled this case completely. As far as $k \geq 6$ is concerned, it has been established that the formula

$$g(k) = [(3/2)^k] + 2^k - 2$$

holds, except possibly for a finite number of values of k. There is considerable evidence to suggest that this expression is correct for all k.

For $k \geq 3$, all sufficiently large integers require fewer than $g(k)$ kth powers in their representations. This suggests a general definition: Let $G(k)$ denote the smallest integer r with the property that every sufficiently large integer is the sum of at most r kth powers. Clearly, $G(k) \leq g(k)$. Exact values of $G(k)$ are known only in two cases; namely, $G(2) = 4$ and $G(4) = 16$. Linnik's result on cubes indicates that $G(3) \leq 7$, while as far back as 1851 Jacobi conjectured that $G(3) \leq 5$. Although more than half a century has passed without an improvement in the size of $G(3)$, nevertheless, it is felt that $G(3) = 4$. In recent years, the bounds $G(5) \leq 17$ and $G(6) \leq 24$ have been established.

Below are listed known values and estimates for the first few $g(k)$ and $G(k)$:

$$\begin{aligned} g(2) &= 4 & G(2) &= 4 \\ g(3) &= 9 & 4 \le G(3) &\le 7 \\ g(4) &= 19 & G(4) &= 16 \\ g(5) &= 37 & 6 \le G(5) &\le 17 \\ g(6) &= 73 & 9 \le G(6) &\le 24 \\ g(7) &= 143 & 8 \le G(7) &\le 33 \\ g(8) &= 279 & 32 \le G(8) &\le 42 \end{aligned}$$

Another problem that has attracted considerable attention is whether an nth power can be written as a sum of n nth powers, with $n > 3$. Progress was first made in 1911 with the discovery of the smallest solution in fourth powers,

$$353^4 = 30^4 + 120^4 + 272^4 + 315^4$$

In fifth powers, the smallest solution is

$$72^5 = 19^5 + 43^5 + 46^5 + 47^5 + 67^5$$

However, for sixth or higher powers no solution is yet known.

There is a related question; it may be asked, "Can an nth power ever be the sum of fewer than n nth powers?" Euler conjectured that this is impossible; however, in 1968, Lander and Parkin came across the representation

$$144^5 = 27^5 + 84^5 + 110^5 + 133^5$$

With the subsequent increase in computer power and sophistication, N. Elkies was able to show (1987) that for fourth powers there are infinitely many counterexamples to Euler's conjecture. The one with the smallest value is

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

PROBLEMS 13.3

1. Without actually adding the squares, confirm that the following relations hold:
   (a) $1^2 + 2^2 + 3^2 + \cdots + 23^2 + 24^2 = 70^2$.
   (b) $18^2 + 19^2 + 20^2 + \cdots + 27^2 + 28^2 = 77^2$.
   (c) $2^2 + 5^2 + 8^2 + \cdots + 23^2 + 26^2 = 48^2$.
   (d) $6^2 + 12^2 + 18^2 + \cdots + 42^2 + 48^2 = 95^2 - 41^2$.
2. Regiomontanus proposed the problem of finding 20 squares whose sum is a square greater than 300,000. Furnish two solutions.
   [*Hint:* Consider the identity
   $$\begin{aligned}(a_1^2 + a_2^2 + \cdots + a_n^2)^2 = (a_1^2 + a_2^2 + \cdots + a_{n-1}^2 - a_n^2)^2 \\ +(2a_1a_n)^2 + (2a_2a_n)^2 + \cdots + (2a_{n-1}a_n)^2.]\end{aligned}$$
3. If $p = q_1^2 + q_2^2 + q_3^2$, where p, q_1, q_2, and q_3 are all primes, show that some $q_i = 3$.
4. Establish that the equation $a^2 + b^2 + c^2 + a + b + c = 1$ has no solution in the integers.
   [*Hint:* The equation in question is equivalent to the equation $(2a + 1)^2 + (2b + 1)^2 + (2c + 1)^2 = 7$.]

5. For a given positive integer n, show that n or $2n$ is a sum of three squares.
6. An unanswered question is whether there exist infinitely many prime numbers p such that $p = n^2 + (n+1)^2 + (n+2)^2$, for some $n > 0$. Find three of these primes.
7. In our examination of $n = 459$, no representation as a sum of two squares was found. Express 459 as a sum of three squares.
8. Verify each of the statements below:
   (a) Every positive odd integer is of the form $a^2 + b^2 + 2c^2$, where a, b, c are integers.
   [*Hint:* Given $n > 0$, $4n + 2$ can be written as $4n + 2 = x^2 + y^2 + z^2$, with x and y odd and z even. Then
   $$2n + 1 = \left(\frac{x+y}{2}\right)^2 + \left(\frac{x-y}{2}\right)^2 + 2\left(\frac{z}{2}\right)^2.]$$
   (b) Every positive integer is either of the form $a^2 + b^2 + c^2$ or $a^2 + b^2 + 2c^2$, where a, b, c are integers.
   [*Hint:* If $n > 0$ cannot be written as a sum $a^2 + b^2 + c^2$, then it is of the form $4^m(8k + 7)$. Apply part (a) to the odd integer $8k + 7$.]
   (c) Every positive integer is of the form $a^2 + b^2 - c^2$, where a, b, c are integers.
   [*Hint:* Given $n > 0$, choose a such that $n - a^2$ is a positive odd integer and use Theorem 13.4.]
9. Establish the following:
   (a) No integer of the form $9k + 4$ or $9k + 5$ can be the sum of three or fewer cubes.
   [*Hint:* Notice that $a^3 \equiv 0, 1$, or $8 \pmod 9$ for any integer a.]
   (b) The only prime p that is representable as the sum of two positive cubes is $p = 2$.
   [*Hint:* Use the identity
   $$a^3 + b^3 = (a+b)((a-b)^2 + ab).]$$
   (c) A prime p can be represented as the difference of two cubes if and only if it is of the form $p = 3k(k+1) + 1$, for some k.
10. Express each of the primes 7, 19, 37, 61, and 127 as the difference of two cubes.
11. Prove that every positive integer can be represented as a sum of three or fewer triangular numbers.
    [*Hint:* Given $n > 0$, express $8n + 3$ as a sum of three odd squares and then solve for n.]
12. Show that there are infinitely many primes p of the form $p = a^2 + b^2 + c^2 + 1$, where a, b, c are integers.
    [*Hint:* By Theorem 9.8, there are infinitely many primes of the form $p = 8k + 7$. Write $p - 1 = 8k + 6 = a^2 + b^2 + c^2$ for some a, b, c.]
13. Express the integers $231 = 3 \cdot 7 \cdot 11$, $391 = 17 \cdot 23$, and $2109 = 37 \cdot 57$ as sums of four squares.
14. (a) Prove that every integer $n \geq 170$ is a sum of five squares, none of which are equal to zero.
    [*Hint:* Write $n - 169 = a^2 + b^2 + c^2 + d^2$ for some integers a, b, c, d and consider the cases in which one or more of a, b, c is zero.]
    (b) Prove that any positive multiple of 8 is a sum of eight odd squares.
    [*Hint:* Assuming $n = a^2 + b^2 + c^2 + d^2$, then $8n + 8$ is the sum of the squares of $2a \pm 1$, $2b \pm 1$, $2c \pm 1$, and $2d \pm 1$.]
15. From the fact that $n^3 \equiv n \pmod 6$ conclude that every integer n can be represented as the sum of the cubes of five integers, allowing negative cubes.
    [*Hint:* Utilize the identity
    $$n^3 - 6k = n^3 - (k+1)^3 - (k-1)^3 + k^3 + k^3.]$$

16. Prove that every odd integer is the sum of four squares, two of which are consecutive.
[*Hint:* For $n > 0$, $4n + 1$ is a sum of three squares, only one being odd; notice that $4n + 1 = (2a)^2 + (2b)^2 + (2c + 1)^2$ gives $2n + 1 = (a + b)^2 + (a - b)^2 + c^2 + (c + 1)^2$.]

17. Prove that there are infinitely many triangular numbers that are simultaneously expressible as the sum of two cubes and the difference of two cubes. Exhibit the representations for one such triangular number.
[*Hint:* In the identity

$$\begin{aligned}(27k^6)^2 - 1 &= (9k^4 - 3k)^3 + (9k^3 - 1)^3 \\ &= (9k^4 + 3k)^3 - (9k^3 + 1)^3\end{aligned}$$

take k to be an odd integer to get

$$(2n + 1)^2 - 1 = (2a)^3 + (2b)^3 = (2c)^3 - (2d)^3$$

or equivalently, $t_n = a^3 + b^3 = c^3 - d^3$.]

18. (a) If $n - 1$ and $n + 1$ are both primes, establish that the integer $2n^2 + 2$ can be represented as the sum of 2, 3, 4, and 5 squares.
(b) Illustrate the result of part (a) in the cases in which $n = 4$, 6, and 12.

1

The Propositional Calculus

1.1 Propositional Connectives: Truth Tables

Sentences may be combined in various ways to form more complicated sentences. We shall consider only *truth-functional* combinations, in which the truth or falsity of the new sentence is determined by the truth or falsity of its component sentences.

Negation is one of the simplest operations on sentences. Although a sentence in a natural language may be negated in many ways, we shall adopt a uniform procedure: placing a sign for negation, the symbol ¬, in front of the entire sentence. Thus, if A is a sentence, then $\neg A$ denotes the negation of A.

The truth-functional character of negation is made apparent in the following *truth table*:

A	$\neg A$
T	F
F	T

When A is true, $\neg A$ is false; when A is false, $\neg A$ is true. We use T and F to denote the *truth values* true and false.

Another common truth-functional operation is the *conjunction*: "and." The conjunction of sentences A and B will be designated by $A \wedge B$ and has the following truth table:

A	B	$A \wedge B$
T	T	T
F	T	F
T	F	F
F	F	F

$A \wedge B$ is true when and only when both A and B are true. A and B are called the *conjuncts* of $A \wedge B$. Note that there are four rows in the table, corresponding to the number of possible assignments of truth values to A and B.

In natural languages, there are two distinct uses of "or": the inclusive and the exclusive. According to the inclusive usage, "A or B" means "A or B or both," whereas according to the exclusive usage, the meaning is "A or B, but

not both," We shall introduce a special sign, $\vee$, for the inclusive connective. Its truth table is as follows:

A	B	$A \vee B$
T	T	T
F	T	T
T	F	T
F	F	F

Thus, $A \vee B$ is false when and only when both A and B are false. "$A \vee B$" is called a *disjunction*, with the *disjuncts* A and B.

Another important truth-functional operation is the *conditional*: "if A, then B." Ordinary usage is unclear here. Surely, "if A, then B" is false when the *antecedent* A is true and the *consequent* B is false. However, in other cases, there is no well-defined truth value. For example, the following sentences would be considered neither true nor false:

1. If $1 + 1 = 2$, then Paris is the capital of France.
2. If $1 + 1 \neq 2$, then Paris is the capital of France.
3. If $1 + 1 \neq 2$, then Rome is the capital of France.

Their meaning is unclear, since we are accustomed to the assertion of some sort of relationship (usually causal) between the antecedent and the consequent. We shall make the convention that "if A, then B" is false when and only when A is true and B is false. Thus, sentences 1–3 are assumed to be true. Let us denote "if A, then B" by "$A \Rightarrow B$." An expression "$A \Rightarrow B$" is called a *conditional*. Then $\Rightarrow$ has the following truth table:

A	B	$A \Rightarrow B$
T	T	T
F	T	T
T	F	F
F	F	T

This sharpening of the meaning of "if A, then B" involves no conflict with ordinary usage, but rather only an extension of that usage.*

* There is a common non-truth-functional interpretation of "if A, then B" connected with causal laws. The sentence "if this piece of iron is placed in water at time t, then the iron will dissolve" is regarded as false even in the case that the piece of iron is not placed in water at time t—that is, even when the antecedent is false. Another non-truth-functional usage occurs in so-called counterfactual conditionals, such as "if Sir Walter Scott had not written any novels, then there would have been no War Between the States." (This was Mark Twain's contention in *Life on the Mississippi*: "Sir Walter had so large a hand in making Southern character, as it existed before the war, that he is in great measure responsible for the war.") This sentence might be asserted to be false even though the antecedent is admittedly false. However, causal laws and counterfactual conditions seem not to be needed in mathematics and logic. For a clear treatment of conditionals and other connectives, see Quine (1951). (The quotation from *Life on the Mississippi* was brought to my attention by Professor J.C. Owings, Jr.)

A justification of the truth table for $\Rightarrow$ is the fact that we wish "if A and B, then B" to be true in all cases. Thus, the case in which A and B are true justifies the first line of our truth table for $\Rightarrow$, since (A and B) and B are both true. If A is false and B true, then (A and B) is false while B is true. This corresponds to the second line of the truth table. Finally, if A is false and B is false, (A and B) is false and B is false. This gives the fourth line of the table. Still more support for our definition comes from the meaning of statements such as "for every x, if x is an odd positive integer, then x^2 is an odd positive integer." This asserts that, for every x, the statement "if x is an odd positive integer, then x^2 is an odd positive integer" is true. Now we certainly do not want to consider cases in which x is not an odd positive integer as counterexamples to our general assertion. This supports the second and fourth lines of our truth table. In addition, any case in which x is an odd positive integer and x^2 is an odd positive integer confirms our general assertion. This corresponds to the first line of the table.

Let us denote "A if and only if B" by "$A \Leftrightarrow B$." Such an expression is called a *biconditional*. Clearly, $A \Leftrightarrow B$ is true when and only when A and B have the same truth value. Its truth table, therefore is:

A	B	$A \Leftrightarrow B$
T	T	T
F	T	F
T	F	F
F	F	T

The symbols $\neg$, $\wedge$, $\vee$, $\Rightarrow$, and $\Leftrightarrow$ will be called *propositional connectives*.* Any sentence built up by application of these connectives has a truth value that depends on the truth values of the constituent sentences. In order to make this dependence apparent, let us apply the name *statement form* to an expression built up from the *statement letters* A, B, C, and so on by appropriate applications of the propositional connectives.

1. All statement letters (capital italic letters) and such letters with numerical subscripts† are statement forms.
2. If $\mathscr{B}$ and $\mathscr{C}$ are statement forms, then so are $(\neg\mathscr{B})$, $(\mathscr{B} \wedge \mathscr{C})$, $(\mathscr{B} \vee \mathscr{C})$, $(\mathscr{B} \Rightarrow \mathscr{C})$, and $(\mathscr{B} \Leftrightarrow \mathscr{C})$.

* We have been avoiding and shall in the future avoid the use of quotation marks to form names whenever this is not likely to cause confusion. The given sentence should have quotation marks around each of the connectives. See Quine (1951, pp. 23–27).

† For example, A_1, A_2, A_{17}, B_{31}, C_2, ….

3. Only those expressions are statement forms that are determined to be so by means of conditions 1 and 2.* Some examples of statement forms are B, $(\neg C_2)$, $(D_3 \wedge (\neg B))$, $(((\neg B_1) \vee B_2) \Rightarrow (A_1 \wedge C_2))$, and $(((\neg A) \Leftrightarrow A) \Leftrightarrow (C \Rightarrow (B \vee C)))$.

For every assignment of truth values T or F to the statement letters that occur in a statement form, there corresponds, by virtue of the truth tables for the propositional connectives, a truth value for the statement form. Thus, each statement form determines a *truth function*, which can be graphically represented by a truth table for the statement form. For example, the statement form $(((\neg A) \vee B) \Rightarrow C)$ has the following truth table:

A	B	C	$(\neg A)$	$((\neg A) \vee B)$	$(((\neg A) \vee B) \Rightarrow C)$
T	T	T	F	T	T
F	T	T	T	T	T
T	F	T	F	F	T
F	F	T	T	T	T
T	T	F	F	T	F
F	T	F	T	T	F
T	F	F	F	F	T
F	F	F	T	T	F

Each row represents an assignment of truth values to the statement letters A, B, and C and the corresponding truth values assumed by the statement forms that appear in the construction of $(((\neg A) \vee B) \Rightarrow C)$.

The truth table for $((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$ is as follows:

A	B	$(A \Leftrightarrow B)$	$(\neg A)$	$((\neg A) \wedge B)$	$((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$
T	T	T	F	F	F
F	T	F	T	T	T
T	F	F	F	F	T
F	F	T	T	F	F

If there are n distinct letters in a statement form, then there are 2^n possible assignments of truth values to the statement letters and, hence, 2^n rows in the truth table.

---

* This can be rephrased as follows: $\mathscr{C}$ is a statement form if and only if there is a finite sequence $\mathscr{B}_1, \ldots, \mathscr{B}_n$ $(n \geq 1)$ such that $\mathscr{B}_n = \mathscr{C}$ and, if $1 \leq i \leq n$, $\mathscr{B}_i$ is either a statement letter or a negation, conjunction, disjunction, conditional, or biconditional constructed from previous expressions in the sequence. Notice that we use script letters $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, … to stand for arbitrary expressions, whereas italic letters are used as statement letters.

A truth table can be abbreviated by writing only the full statement form, putting the truth values of the statement letters underneath all occurrences of these letters, and writing, step by step, the truth values of each component statement form under the principal connective of the form.* As an example, for $((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$, we obtain

$((A$	$\Leftrightarrow$	$B)$	$\Rightarrow$	$((\neg A)$	$\wedge$	$B))$
T	T	T	F	FT	F	T
F	F	T	T	TF	T	T
T	F	F	T	FT	F	F
F	T	F	F	TF	F	F

Exercises

1.1 Let $\oplus$ designate the exclusive use of "or." Thus, $A \oplus B$ stands for "A or B but not both." Write the truth table for $\oplus$.

1.2 Construct truth tables for the statement forms $((A \Rightarrow B) \vee (\neg A))$ and $((A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C)))$.

1.3 Write abbreviated truth tables for $((A \Rightarrow B) \wedge A)$ and $((A \vee (\neg C)) \Leftrightarrow B)$.

1.4 Write the following sentences as statement forms, using statement letters to stand for the *atomic sentences*—that is, those sentences that are not built up out of other sentences.

a. If Mr Jones is happy, Mrs Jones is not happy, and if Mr Jones is not happy, Mrs Jones is not happy.

b. Either Sam will come to the party and Max will not, or Sam will not come to the party and Max will enjoy himself.

c. A sufficient condition for x to be odd is that x is prime.

d. A necessary condition for a sequence s to converge is that s be bounded.

e. A necessary and sufficient condition for the sheikh to be happy is that he has wine, women, and song.

f. Fiorello goes to the movies only if a comedy is playing.

g. The bribe will be paid if and only if the goods are delivered.

h. If x is positive, x^2 is positive.

i. Karpov will win the chess tournament unless Kasparov wins today.

* The *principal connective* of a statement form is the one that is applied last in constructing the form.

1.2 Tautologies

A *truth function of n arguments* is defined to be a function of n arguments, the arguments and values of which are the truth values T or F. As we have seen, any statement form containing n distinct statement letters determines a corresponding truth function of n arguments.*

A statement form that is always true, no matter what the truth values of its statement letters may be, is called a *tautology*. A statement form is a tautology if and only if its corresponding truth function takes only the value T, or equivalently, if, in its truth table, the column under the statement form contains only *T*s. An example of a tautology is $(A \vee (\neg A))$, the so-called *law of the excluded middle*. Other simple examples are $(\neg(A \wedge (\neg A)))$, $(A \Leftrightarrow (\neg(\neg A)))$, $((A \wedge B) \Rightarrow A)$, and $(A \Rightarrow (A \vee B))$.

$\mathscr{B}$ is said to *logically imply* $\mathscr{C}$ (or, synonymously, $\mathscr{C}$ is a *logical consequence* of $\mathscr{B}$) if and only if every truth assignment to the statement letters of $\mathscr{B}$ and $\mathscr{C}$ that makes $\mathscr{B}$ true also makes $\mathscr{C}$ true. For example, $(A \wedge B)$ logically implies A, A logically implies $(A \vee B)$, and $(A \wedge (A \Rightarrow B))$ logically implies B.

$\mathscr{B}$ and $\mathscr{C}$ are said to be *logically equivalent* if and only if $\mathscr{B}$ and $\mathscr{C}$ receive the same truth value under every assignment of truth values to the statement letters of $\mathscr{B}$ and $\mathscr{C}$. For example, A and $(\neg(\neg A))$ are logically equivalent, as are $(A \wedge B)$ and $(B \wedge A)$.

---

* To be precise, enumerate all statement letters as follows: $A, B, \ldots, Z; A_1, B_1, \ldots, Z_1; A_2, \ldots,$. If a statement form contains the $i_{1^{\text{th}}}, \ldots, i_{n^{\text{th}}}$ statement letters in this enumeration, where $i_1 < \cdots < i_n$, then the corresponding truth function is to have $x_{i_1}, \ldots, x_{i_n}$, in that order, as its arguments, where x_{i_j} corresponds to the $i_{j^{\text{th}}}$ statement letter. For example, $(A \Rightarrow B)$ generates the truth function:

x_1	x_2	$f(x_1, x_2)$
T	T	T
F	T	T
T	F	F
F	F	T

whereas $(B \Rightarrow A)$ generates the truth function:

x_1	x_2	$g(x_1, x_2)$
T	T	T
F	T	F
T	F	T
F	F	T

Proposition 1.1

a. $\mathscr{B}$ logically implies $\mathscr{C}$ if and only if $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology.
b. $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is a tautology.

Proof

a. (i) Assume $\mathscr{B}$ logically implies $\mathscr{C}$. Hence, every truth assignment that makes $\mathscr{B}$ true also makes $\mathscr{C}$ true. Thus, no truth assignment makes $\mathscr{B}$ true and $\mathscr{C}$ false. Therefore, no truth assignment makes $(\mathscr{B} \Rightarrow \mathscr{C})$ false, that is, every truth assignment makes $(\mathscr{B} \Rightarrow \mathscr{C})$ true. In other words, $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. (ii) Assume $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. Then, for every truth assignment, $(\mathscr{B} \Rightarrow \mathscr{C})$ is true, and, therefore, it is not the case that $\mathscr{B}$ is true and $\mathscr{C}$ false. Hence, every truth assignment that makes $\mathscr{B}$ true makes $\mathscr{C}$ true, that is, $\mathscr{B}$ logically implies $\mathscr{C}$.
b. $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is a tautology if and only if every truth assignment makes $(\mathscr{B} \Leftrightarrow \mathscr{C})$ true, which is equivalent to saying that every truth assignment gives $\mathscr{B}$ and $\mathscr{C}$ the same truth value, that is, $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent.

By means of a truth table, we have an effective procedure for determining whether a statement form is a tautology. Hence, by Proposition 1.1, we have effective procedures for determining whether a given statement form logically implies another given statement form and whether two given statement forms are logically equivalent.

To see whether a statement form is a tautology, there is another method that is often shorter than the construction of a truth table.

Examples

1. Determine whether $((A \Leftrightarrow ((\neg B) \vee C)) \Rightarrow ((\neg A) \Rightarrow B))$ is a tautology.

Assume that the statement form sometimes is F (line 1). Then $(A \Leftrightarrow ((\neg B) \vee C))$ is T and $((\neg A) \Rightarrow B)$ is F (line 2). Since $((\neg A) \Rightarrow B)$ is F, $(\neg A)$ is T and B is F (line 3). Since $(\neg A)$is T, A is F (line 4). Since A is F and $(A \Leftrightarrow ((\neg B) \vee C))$ is T, $((\neg B) \vee C)$ is F (line 5). Since $((\neg B) \vee C)$ is F, $(\neg B)$ and C are F (line 6). Since $(\neg B)$ is F, B is T (line 7). But B is both T and F (lines 7 and 3). Hence, it is impossible for the form to be false.

$((A \Leftrightarrow ((\neg B) \vee C)) \Rightarrow ((\neg A) \Rightarrow B))$

				F				1
T						F		2
					T		F	3
F					F			4
		F						5
	F		F					6
		T						7

2. Determine whether $((A \Rightarrow (B \vee C)) \vee (A \Rightarrow B))$ is a tautology.

Assume that the form is F (line 1). Then $(A \Rightarrow (B \vee C))$ and $(A \Rightarrow B)$ are F (line 2). Since $(A \Rightarrow B)$ is F, A is T and B is F (line 3). Since $(A \Rightarrow (B \vee C))$ is F, A is T and $(B \vee C)$ is F (line 4). Since $(B \vee C)$ is F, B and C are F (line 5). Thus, when A is T, B is F, and C is F, the form is F. Therefore, it is not a tautology.

$((A \Rightarrow (B \vee C)) \vee (A \Rightarrow B))$

				F				1
	F					F		2
					T		F	3
T			F					4
		F		F				5

Exercises

1.5 Determine whether the following are tautologies.

a. $(((A \Rightarrow B) \Rightarrow B) \Rightarrow B)$

b. $(((A \Rightarrow B) \Rightarrow B) \Rightarrow A)$

c. $(((A \Rightarrow B) \Rightarrow A) \Rightarrow A)$

d. $(((B \Rightarrow C) \Rightarrow (A \Rightarrow B)) \Rightarrow (A \Rightarrow B))$

e. $((A \vee (\neg(B \wedge C))) \Rightarrow ((A \Leftrightarrow C) \vee B))$

f. $(A \Rightarrow (B \Rightarrow (B \Rightarrow A)))$

g. $((A \wedge B) \Rightarrow (A \vee C))$

h. $((A \Leftrightarrow B) \Leftrightarrow (A \Leftrightarrow (B \Leftrightarrow A)))$

i. $((A \Rightarrow B) \vee (B \Rightarrow A))$

j. $((\neg(A \Rightarrow B)) \Rightarrow A)$

1.6 Determine whether the following pairs are logically equivalent.

a. $((A \Rightarrow B) \Rightarrow A)$ and A

b. $(A \Leftrightarrow B)$ and $((A \Rightarrow B) \wedge (B \Rightarrow A))$

c. $((\neg A) \vee B)$ and $((\neg B) \vee A)$

d. $(\neg(A \Leftrightarrow B))$ and $(A \Leftrightarrow (\neg B))$

e. $(A \vee (B \Leftrightarrow C))$ and $((A \vee B) \Leftrightarrow (A \vee C))$

f. $(A \Rightarrow (B \Leftrightarrow C))$ and $((A \Rightarrow B) \Leftrightarrow (A \Rightarrow C))$

g. $(A \wedge (B \Leftrightarrow C))$ and $((A \wedge B) \Leftrightarrow (A \wedge C))$

1.7 Prove:

a. $(A \Rightarrow B)$ is logically equivalent to $((\neg A) \vee B)$.

b. $(A \Rightarrow B)$ is logically equivalent to $(\neg(A \wedge (\neg B)))$.

1.8 Prove that $\mathscr{B}$ is logically equivalent to $\mathscr{C}$ if and only if $\mathscr{B}$ logically implies $\mathscr{C}$ and $\mathscr{C}$ logically implies $\mathscr{B}$.

1.9 Show that $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if, in their truth tables, the columns under $\mathscr{B}$ and $\mathscr{C}$ are the same.

1.10 Prove that $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if $(\neg\mathscr{B})$ and $(\neg\mathscr{C})$ are logically equivalent.

1.11 Which of the following statement forms are logically implied by $(A \wedge B)$?

a. A

b. B

c. $(A \vee B)$

d. $((\neg A) \vee B)$

e. $((\neg B) \Rightarrow A)$

f. $(A \Leftrightarrow B)$

g. $(A \Rightarrow B)$

h. $((\neg B) \Rightarrow (\neg A))$

i. $(A \wedge (\neg B))$

1.12 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \Rightarrow B)$ and by $(\neg(A \Rightarrow B))$, respectively.

1.13 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \vee B)$.

1.14 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \Leftrightarrow B)$ and by $(\neg(A \Leftrightarrow B))$, respectively.

A statement form that is false for all possible truth values of its statement letters is said to be *contradictory*. Its truth table has only Fs in the column under the statement form. One example is $(A \Leftrightarrow (\neg A))$:

A	$(\neg A)$	$(A \Leftrightarrow (\neg A))$
T	F	F
F	T	F

Another is $(A \wedge (\neg A))$.

Notice that a statement form $\mathscr{B}$ is a tautology if and only if $(\neg\mathscr{B})$ is contradictory, and vice versa.

A sentence (in some natural language like English or in a formal theory)* that arises from a tautology by the substitution of sentences for all the statement letters, with occurrences of the same statement letter being replaced by the same sentence, is said to be *logically true* (according to the propositional calculus). Such a sentence may be said to be true by virtue of its truth-functional structure alone. An example is the English sentence, "If it is raining or it is snowing, and it is not snowing, then it is raining," which arises by substitution from the tautology $(((A \vee B) \wedge (\neg B)) \Rightarrow A)$. A sentence that comes from

* By a formal theory we mean an artificial language in which the notions of *meaningful expressions, axioms*, and *rules of inference* are precisely described (see page 27).

a contradictory statement form by means of substitution is said to be *logically false* (according to the propositional calculus).

Now let us prove a few general facts about tautologies.

Proposition 1.2

If $\mathscr{B}$ and $(\mathscr{B} \Rightarrow \mathscr{C})$ are tautologies, then so is $\mathscr{C}$.

Proof

Assume that $\mathscr{B}$ and $(\mathscr{B} \Rightarrow \mathscr{C})$ are tautologies. If $\mathscr{C}$ took the value F for some assignment of truth values to the statement letters of $\mathscr{B}$ and $\mathscr{C}$, then, since $\mathscr{B}$ is a tautology, $\mathscr{B}$ would take the value T and, therefore, $(\mathscr{B} \Rightarrow \mathscr{C})$ would have the value F for that assignment. This contradicts the assumption that $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. Hence, $\mathscr{C}$ never takes the value F.

Proposition 1.3

If $\mathscr{T}$ is a tautology containing as statement letters $A_1, A_2, \ldots, A_n$, and $\mathscr{B}$ arises from $\mathscr{T}$ by substituting statement forms $\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_n$ for $A_1, A_2, \ldots, A_n$, respectively, then $\mathscr{B}$ is a tautology; that is, substitution in a tautology yields a tautology.

Example
Let $\mathscr{T}$ be $((A_1 \wedge A_2) \Rightarrow A_1)$, let $\mathscr{S}_1$ be $(B \vee C)$ and let $\mathscr{S}_2$ be $(C \wedge D)$. Then $\mathscr{B}$ is $(((B \vee C) \wedge (C \wedge D)) \Rightarrow (B \vee C))$.

Proof

Assume that $\mathscr{T}$ is a tautology. For any assignment of truth values to the statement letters in $\mathscr{B}$, the forms $\mathscr{S}_1, \ldots, \mathscr{S}_n$ have truth values $x_1, \ldots, x_n$ (where each x_i is T or F). If we assign the values $x_1, \ldots, x_n$ to $A_1, \ldots, A_n$, respectively, then the resulting truth value of $\mathscr{T}$ is the truth value of $\mathscr{B}$ for the given assignment of truth values. Since $\mathscr{T}$ is a tautology, this truth value must be T. Thus, $\mathscr{B}$ always takes the value T.

Proposition 1.4

If $\mathscr{C}_1$ arises from $\mathscr{B}_1$ by substitution of $\mathscr{C}$ for one or more occurrences of $\mathscr{B}$, then $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{B}_1 \Leftrightarrow \mathscr{C}_1))$ is a tautology. Hence, if $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent, then so are $\mathscr{B}_1$ and $\mathscr{C}_1$.

Example
Let $\mathscr{A}_1$ be $(\mathscr{C} \vee D)$, let $\mathscr{B}$ be $\mathscr{C}$, and let $\mathscr{C}$ be $(\neg(\neg\mathscr{C}))$. Then $\mathscr{C}_1$ is $((\neg(\neg\mathscr{C})) \vee D)$. Since $\mathscr{C}$ and $(\neg(\neg\mathscr{C}))$ are logically equivalent, $(\mathscr{C} \vee D)$ and $((\neg(\neg\mathscr{C})) \vee D)$ are also logically equivalent.

Proof

Consider any assignment of truth values to the statement letters. If $\mathscr{B}$ and $\mathscr{C}$ have opposite truth values under this assignment, then $(\mathscr{B} \Leftrightarrow \mathscr{C})$ takes the value F, and, hence, $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{A}_1 \Leftrightarrow \mathscr{C}_1))$ is T. If $\mathscr{B}$ and $\mathscr{C}$ take the same truth values, then so do $\mathscr{A}_1$ and $\mathscr{C}_1$, since $\mathscr{C}_1$ differs from $\mathscr{A}_1$ only in containing $\mathscr{C}$ in some places where $\mathscr{A}_1$ contains $\mathscr{B}$. Therefore, in this case, $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is T, $(\mathscr{A}_1 \Leftrightarrow \mathscr{C}_1)$ is T, and, thus, $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{A}_1 \Leftrightarrow \mathscr{C}_1))$ is T.

Parentheses

It is profitable at this point to agree on some conventions to avoid the use of so many parentheses in writing formulas. This will make the reading of complicated expressions easier.

First, we may omit the outer pair of parentheses of a statement form. (In the case of statement letters, there is no outer pair of parentheses.)

Second, we arbitrarily establish the following decreasing order of strength of the connectives: $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$. Now we shall explain a step-by-step process for restoring parentheses to an expression obtained by eliminating some or all parentheses from a statement form. (The basic idea is that, where possible, we first apply parentheses to negations, then to conjunctions, then to disjunctions, then to conditionals, and finally to biconditionals.) Find the leftmost occurrence of the strongest connective that has not yet been processed.

i. If the connective is $\neg$ and it precedes a statement form $\mathscr{B}$, restore left and right parentheses to obtain $(\neg\mathscr{B})$.
ii. If the connective is a binary connective C and it is preceded by a statement form $\mathscr{B}$ and followed by a statement form $\mathscr{D}$, restore left and right parentheses to obtain $(\mathscr{B} C \mathscr{D})$.
iii. If neither (i) nor (ii) holds, ignore the connective temporarily and find the leftmost occurrence of the strongest of the remaining unprocessed connectives and repeat (i–iii) for that connective.

Examples
Parentheses are restored to the expression in the first line of each of the following in the steps shown:

1. $A \Leftrightarrow (\neg B) \vee C \Rightarrow A$
   $A \Leftrightarrow ((\neg B) \vee C) \Rightarrow A$
   $A \Leftrightarrow (((\neg B) \vee C) \Rightarrow A)$
   $(A \Leftrightarrow (((\neg B) \vee C) \Rightarrow A))$

2. $A \Rightarrow \neg B \Rightarrow C$
   $A \Rightarrow (\neg B) \Rightarrow C$
   $(A \Rightarrow (\neg B)) \Rightarrow C$
   $((A \Rightarrow (\neg B)) \Rightarrow C)$
3. $B \Rightarrow \neg\neg A$
   $B \Rightarrow \neg(\neg A)$
   $B \Rightarrow (\neg(\neg A))$
   $(B \Rightarrow (\neg(\neg A)))$
4. $A \vee \neg(B \Rightarrow A \vee B)$
   $A \vee \neg(B \Rightarrow (A \vee B))$
   $A \vee (\neg(B \Rightarrow (A \vee B)))$
   $(A \vee (\neg(B \Rightarrow (A \vee B))))$

Not every form can be represented without the use of parentheses. For example, parentheses cannot be further eliminated from $A \Rightarrow (B \Rightarrow C)$, since $A \Rightarrow B \Rightarrow C$ stands for $((A \Rightarrow B) \Rightarrow C)$. Likewise, the remaining parentheses cannot be removed from $\neg(A \vee B)$ or from $A \wedge (B \Rightarrow C)$.

Exercises

1.15 Eliminate as many parentheses as possible from the following forms.

a. $((B \Rightarrow (\neg A)) \wedge C)$
b. $(A \vee (B \vee C))$
c. $(((A \wedge (\neg B)) \wedge C) \vee D)$
d. $((B \vee (\neg C)) \vee (A \wedge B))$
e. $((A \Leftrightarrow B) \Leftrightarrow (\neg(C \vee D)))$
f. $((\neg(\neg(\neg(B \vee C)))) \Leftrightarrow (B \Leftrightarrow C))$
g. $(\neg((\neg(\neg(B \vee C))) \Leftrightarrow (B \Leftrightarrow C)))$
h. $((((A \Rightarrow B) \Rightarrow (C \Rightarrow D)) \wedge (\neg A)) \vee C)$

1.16 Restore parentheses to the following forms.

a. $C \vee \neg A \wedge B$
b. $B \Rightarrow \neg\neg\neg A \wedge C$
c. $C \Rightarrow \neg(A \wedge B \Rightarrow C) \wedge A \Leftrightarrow B$
d. $C \Rightarrow A \Rightarrow A \Leftrightarrow \neg A \vee B$

1.17 Determine whether the following expressions are abbreviations of statement forms and, if so, restore all parentheses.

a. $\neg\neg A \Leftrightarrow A \Leftrightarrow B \vee C$
b. $\neg(\neg A \Leftrightarrow A) \Leftrightarrow B \vee C$
c. $\neg(A \Rightarrow B) \vee C \vee D \Rightarrow B$

d. $A \Leftrightarrow (\neg A \vee B) \Rightarrow (A \wedge (B \vee C)))$

e. $\neg A \vee B \vee C \wedge D \Leftrightarrow A \wedge \neg A$

f. $((A \Rightarrow B \wedge (C \vee D) \wedge (A \vee D))$

1.18 If we write $\neg \mathscr{B}$ instead of $(\neg \mathscr{B})$, $\Rightarrow \mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \Rightarrow \mathscr{C})$, $\wedge \mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \wedge \mathscr{C})$, $\vee \mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \vee \mathscr{C})$, and $\Leftrightarrow \mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \Leftrightarrow \mathscr{C})$, then there is no need for parentheses. For example, $((\neg A) \wedge (B \Rightarrow (\neg D)))$, which is ordinarily abbreviated as $\neg A \wedge (B \Rightarrow \neg D)$, becomes $\wedge \neg A \Rightarrow B \neg D$. This way of writing forms is called *Polish notation*.

a. Write $((C \Rightarrow (\neg A)) \vee B)$ and $(C \vee ((B \wedge (\neg D)) \Rightarrow C))$ in this notation.

b. If we count $\Rightarrow$, $\wedge$, $\vee$, and $\Leftrightarrow$ each as +1, each statement letter as −1 and $\neg$ as 0, prove that an expression $\mathscr{B}$ in this parenthesis-free notation is a statement form if and only if (i) the sum of the symbols of $\mathscr{B}$ is −1 and (ii) the sum of the symbols in any proper initial segment of $\mathscr{B}$ is nonnegative. (If an expression $\mathscr{B}$ can be written in the form $\mathscr{C}\mathscr{D}$, where $\mathscr{C} \neq \mathscr{B}$, then $\mathscr{C}$ is called a *proper initial segment* of $\mathscr{B}$.)

c. Write the statement forms of Exercise 1.15 in Polish notation.

d. Determine whether the following expressions are statement forms in Polish notation. If so, write the statement forms in the standard way.

i. $\neg\Rightarrow ABC \vee AB \neg C$

ii. $\Rightarrow\Rightarrow AB \Rightarrow\Rightarrow BC \Rightarrow\neg AC$

iii. $\vee \wedge \vee \neg A \neg BC \wedge \vee AC \vee \neg C \neg A$

iv. $\vee \wedge B \wedge BBB$

1.19 Determine whether each of the following is a tautology, is contradictory, or neither.

a. $B \Leftrightarrow (B \vee B)$

b. $((A \Rightarrow B) \wedge B) \Rightarrow A$

c. $(\neg A) \Rightarrow (A \wedge B)$

d. $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$

e. $(A \Leftrightarrow \neg B) \Rightarrow A \vee B$

f. $A \wedge (\neg(A \vee B))$

g. $(A \Rightarrow B) \Leftrightarrow ((\neg A) \vee B)$

h. $(A \Rightarrow B) \Leftrightarrow \neg(A \wedge (\neg B))$

i. $(B \Leftrightarrow (B \Leftrightarrow A)) \Rightarrow A$

j. $A \wedge \neg A \Rightarrow B$

1.20 If A and B are true and C is false, what are the truth values of the following statement forms?

a. $A \vee C$

b. $A \wedge C$

c. $\neg A \wedge \neg C$

d. $A \Leftrightarrow \neg B \vee C$

e. $B \vee \neg C \Rightarrow A$

f. $(B \vee A) \Rightarrow (B \Rightarrow \neg C)$

g. $(B \Rightarrow \neg A) \Leftrightarrow (A \Leftrightarrow C)$

h. $(B \Rightarrow A) \Rightarrow ((A \Rightarrow \neg C) \Rightarrow (\neg C \Rightarrow B))$

1.21 If $A \Rightarrow B$ is T, what can be deduced about the truth values of the following?

a. $A \vee C \Rightarrow B \vee C$

b. $A \wedge C \Rightarrow B \wedge C$

c. $\neg A \wedge B \Leftrightarrow A \vee B$

1.22 What further truth values can be deduced from those shown?

a. $\neg A \vee (A \Rightarrow B)$
F

b. $\neg(A \wedge B) \Leftrightarrow \neg A \Rightarrow \neg B$
T

c. $(\neg A \vee B) \Rightarrow (A \Rightarrow \neg C)$
F

d. $(A \Leftrightarrow B) \Leftrightarrow (C \Rightarrow \neg A)$
F T

1.23 If $A \Leftrightarrow B$ is F, what can be deduced about the truth values of the following?

a. $A \wedge B$

b. $A \vee B$

c. $A \Rightarrow B$

d. $A \wedge C \Leftrightarrow B \wedge C$

1.24 Repeat Exercise 1.23, but assume that $A \Leftrightarrow B$ is T.

1.25 What further truth values can be deduced from those given?

a. $(A \wedge B) \Leftrightarrow (A \vee B)$
F F

b. $(A \Rightarrow \neg B) \Rightarrow (C \Rightarrow B)$
F

1.26 a. Apply Proposition 1.3 when $\mathscr{T}$ is $A_1 \Rightarrow A_1 \vee A_2$, $\mathscr{S}_1$ is $B \wedge D$, and $\mathscr{S}_2$ is $\neg B$.

b. Apply Proposition 1.4 when $\mathscr{B}_1$ is $(B \Rightarrow C) \wedge D$, $\mathscr{B}$ is $B \Rightarrow C$, and $\mathscr{C}$ is $\neg B \vee C$.

1.27 Show that each statement form in column I is logically equivalent to the form next to it in column II.

	I	II	
a.	$A \Rightarrow (B \Rightarrow C)$	$(A \wedge B) \Rightarrow C$	
b.	$A \wedge (B \vee C)$	$(A \wedge B) \vee (A \wedge C)$	(Distributive law)
c.	$A \vee (B \wedge C)$	$(A \vee B) \wedge (A \vee C)$	(Distributive law)
d.	$(A \wedge B) \vee \neg B$	$A \vee \neg B$	
e.	$(A \vee B) \wedge \neg B$	$A \wedge \neg B$	
f.	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$	(Law of the contrapositive)
g.	$A \Leftrightarrow B$	$B \Leftrightarrow A$	(Biconditional commutativity)
h.	$(A \Leftrightarrow B) \Leftrightarrow C$	$A \Leftrightarrow (B \Leftrightarrow C)$	(Biconditional associativity)
i.	$A \Leftrightarrow B$	$(A \wedge B) \vee (\neg A \wedge \neg B)$	
j.	$\neg(A \Leftrightarrow B)$	$A \Leftrightarrow \neg B$	
k.	$\neg(A \vee B)$	$(\neg A) \wedge (\neg B)$	(De Morgan's law)
l.	$\neg(A \wedge B)$	$(\neg A) \vee (\neg B)$	(De Morgan's law)
m.	$A \vee (A \wedge B)$	A	
n.	$A \wedge (A \vee B)$	A	
o.	$A \wedge B$	$B \wedge A$	(Commutativity of conjunction)
p.	$A \vee B$	$B \vee A$	(Commutativity of disjunction)
q.	$(A \wedge B) \wedge C$	$A \wedge (B \wedge C)$	(Associativity of conjunction)
r.	$(A \vee B) \vee C$	$A \vee (B \vee C)$	(Associativity of disjunction)
s.	$A \oplus B$	$B \oplus A$	(Commutativity of exclusive "or")
t.	$A \oplus B) \oplus C$	$A \oplus (B \oplus C)$	(Associativity of exclusive "or")
u.	$A \wedge (B \oplus C)$	$(A \wedge B) \oplus (A \wedge C)$	(Distributive law)

1.28 Show the logical equivalence of the following pairs.

a. $\mathscr{T} \wedge \mathscr{B}$ and $\mathscr{B}$, where $\mathscr{T}$ is a tautology.

b. $\mathscr{T} \vee \mathscr{B}$ and $\mathscr{T}$, where $\mathscr{T}$ is a tautology.

c. $\mathscr{F} \wedge \mathscr{B}$ and $\mathscr{F}$, where $\mathscr{F}$ is contradictory.

d. $\mathscr{F} \vee \mathscr{B}$ and $\mathscr{B}$, where $\mathscr{F}$ is contradictory.

1.29 a. Show the logical equivalence of $\neg(A \Rightarrow B)$ and $A \wedge \neg B$.

b. Show the logical equivalence of $\neg(A \Leftrightarrow B)$ and $(A \wedge \neg B) \vee (\neg A \wedge B)$.

c. For each of the following statement forms, find a statement form that is logically equivalent to its negation and in which negation signs apply only to statement letters.

i. $A \Rightarrow (B \Leftrightarrow \neg C)$

ii. $\neg A \vee (B \Rightarrow C)$

iii. $A \wedge (B \vee \neg C)$

1.30 (*Duality*)

a. If $\mathscr{B}$ is a statement form involving only $\neg$, $\wedge$, and $\vee$, and $\mathscr{B}'$ results from $\mathscr{B}$ by replacing each $\wedge$ by $\vee$ and each $\vee$ by $\wedge$, show that $\mathscr{B}$ is a tautology if and only if $\neg\mathscr{B}'$ is a tautology. Then prove that, if $\mathscr{B} \Rightarrow \mathscr{C}$ is a tautology, then so is $\mathscr{C}' \Rightarrow \mathscr{B}'$, and if $\mathscr{B} \Leftrightarrow \mathscr{C}$ is a tautology, then so is $\mathscr{B}' \Leftrightarrow \mathscr{C}'$. (Here $\mathscr{C}$ is also assumed to involve only $\neg$, $\wedge$, and $\vee$.)

b. Among the logical equivalences in Exercise 1.27, derive (c) from (b), (e) from (d), (l) from (k), (p) from (o), and (r) from (q).

c. If $\mathscr{B}$ is a statement form involving only $\neg$, $\wedge$, and $\vee$, and $\mathscr{B}^*$ results from $\mathscr{B}$ by interchanging $\wedge$ and $\vee$ and replacing every statement letter by its negation, show that $\mathscr{B}^*$ is logically equivalent to $\neg\mathscr{B}$. Find a statement form that is logically equivalent to the negation of $(A \vee B \vee C) \wedge (\neg A \vee \neg B \vee D)$, in which $\neg$ applies only to statement letters.

1.31 a. Prove that a statement form that contains $\Leftrightarrow$ as its only connective is a tautology if and only if each statement letter occurs an even number of times.

b. Prove that a statement form that contains $\neg$ and $\Leftrightarrow$ as its only connectives is a tautology if and only if $\neg$ and each statement letter occur an even number of times.

1.32 (Shannon, 1938) An electric circuit containing only on–off switches (when a switch is on, it passes current; otherwise it does not) can be represented by a diagram in which, next to each switch, we put a letter representing a necessary and sufficient condition for the switch to be on (see Figure 1.1). The condition that a current flows through this network can be given by the statement form $(A \wedge B) \vee (C \wedge \neg A)$. A statement form representing the circuit shown in Figure 1.2 is $(A \wedge B) \vee ((C \vee A) \wedge \neg B)$, which is logically equivalent to each of the following forms by virtue of the indicated logical equivalence of Exercise 1.27.

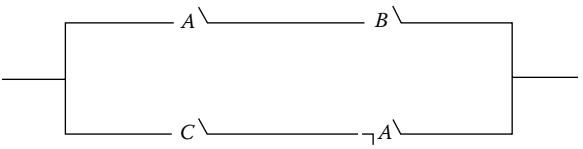


FIGURE 1.1

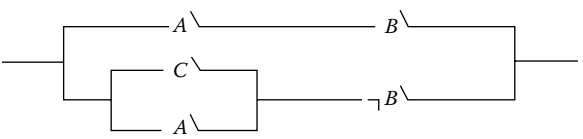


FIGURE 1.2

$$((A \wedge B) \vee (C \vee A)) \wedge ((A \wedge B) \vee \neg B) \quad \text{(c)}$$

$$((A \wedge B) \vee (C \vee A)) \wedge (A \vee \neg B) \quad \text{(d)}$$

$$((A \wedge B) \vee (A \vee C)) \wedge (A \vee \neg B) \quad \text{(p)}$$

$$(((A \wedge B) \vee A) \vee C) \wedge (A \vee \neg B) \quad \text{(r)}$$

$$(A \vee C) \wedge (A \vee \neg B) \quad \text{(p), (m)}$$

$$A \vee (C \wedge \neg B) \quad \text{(c)}$$

Hence, the given circuit is equivalent to the simpler circuit shown in Figure 1.3. (Two circuits are said to be *equivalent* if current flows through one if and only if it flows through the other, and one circuit is *simpler* if it contains fewer switches.)

a. Find simpler equivalent circuits for those shown in Figures 1.4 through 1.6.

b. Assume that each of the three members of a committee votes *yes* on a proposal by pressing a button. Devise as simple a circuit as you can that will allow current to pass when and only when at least two of the members vote in the affirmative.

c. We wish a light to be controlled by two different wall switches in a room in such a way that flicking either one of these switches will turn the light on if it is off and turn it off if it is on. Construct a simple circuit to do the required job.

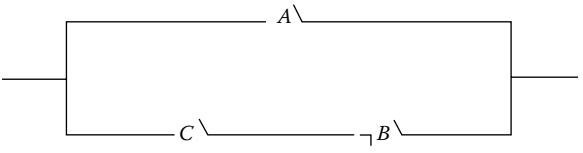


FIGURE 1.3

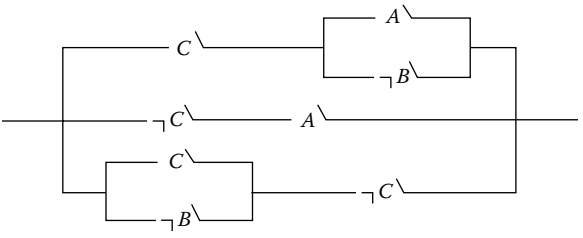


FIGURE 1.4

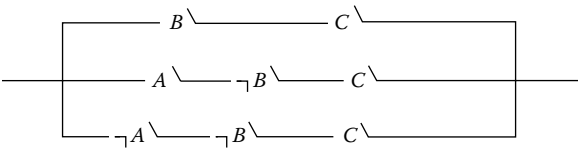


FIGURE 1.5

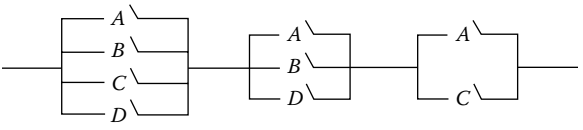


FIGURE 1.6

1.33 Determine whether the following arguments are logically correct by representing each sentence as a statement form and checking whether the conclusion is logically implied by the conjunction of the assumptions. (To do this, assign T to each assumption and F to the conclusion, and determine whether a contradiction results.)

a. If Jones is a communist, Jones is an atheist. Jones is an atheist. Therefore, Jones is a communist.

b. If the temperature and air pressure remained constant, there was no rain. The temperature did remain constant. Therefore, if there was rain, then the air pressure did not remain constant.

c. If Gorton wins the election, then taxes will increase if the deficit will remain high. If Gorton wins the election, the deficit will remain high. Therefore, if Gorton wins the election, taxes will increase.

d. If the number x ends in 0, it is divisible by 5. x does not end in 0. Hence, x is not divisible by 5.

e. If the number x ends in 0, it is divisible by 5. x is not divisible by 5. Hence, x does not end in 0.

f. If $a = 0$ or $b = 0$, then $ab = 0$. But $ab \neq 0$. Hence, $a \neq 0$ and $b \neq 0$.

g. A sufficient condition for f to be integrable is that g be bounded. A necessary condition for h to be continuous is that f is integrable. Hence, if g is bounded or h is continuous, then f is integrable.

h. Smith cannot both be a running star and smoke cigarettes. Smith is not a running star. Therefore, Smith smokes cigarettes.

i. If Jones drove the car, Smith is innocent. If Brown fired the gun, then Smith is not innocent. Hence, if Brown fired the gun, then Jones did not drive the car.