A Word to Students

This course is likely to be different from your previous mathematics courses in several ways. There are very few equations to solve, even fewer formulas, and just a handful of procedures. Although there will be definitions and theorems to learn, rote memorization alone will not carry you through the course. Understanding concepts well enough to apply them in a variety of settings is essential for success.

The good news is that there is a wealth of interesting and useful material in this text. We have chosen topics that form a basis for applications in everyday life, mathematics, computer science, and other fields. We have also chosen the topics so that they fit together and build on each other; this will help you to master the concepts covered.

Two distinctive features of this course are a higher level of abstraction and more emphasis on proofs than you have perhaps encountered in earlier mathematics courses. Here is an example of what we mean by abstraction. When you studied algebra, you learned the distributive property of multiplication over addition. In this course, you will abstract the concept of a distributive property and investigate this idea for many pairs of operations, not just multiplication and addition.

The other feature is proofs. Before you close the book right here, let us tell you something about how proofs are handled in this book. The goals are for you to be able to read proofs intelligently and to produce proofs on your own. The way we help you to these goals may remind you of your composition classes. Learning to write a persuasive essay or a meaningful sonnet or other composition style is a complicated process. First, you read, analyze, and study many examples. Next you try your hand at the specific style. Typically this involves draft versions, revisions, critiques, polishing, and rewriting to produce a strong essay or a good sonnet or whatever form is required. There are no formulas or rote procedures for writing.

Proofs, like the products of a composition course, have structures and styles. We give you lots of proofs to read and analyze. Some exercises ask that you outline, analyze, or critique a proof. Other exercises require the completion of partial proofs. And finally, there are many opportunities for you to construct a proof on your own. Believe us, reading and writing proofs are learnable skills.

On a larger scale, we hope this text helps you to become an effective communicator, a critical thinker, a reflective learner, and an innovative problem solver.

Best wishes for a successful and interesting experience.

Bernard Kolman

Robert C. Busby

Sharon Cutler Ross

CHAPTER

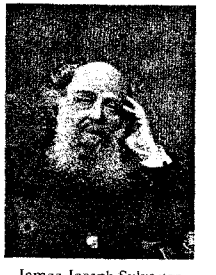
1 Fundamentals

Prerequisites: There are no formal prerequisites for this chapter; the reader is encouraged to read carefully and work through all examples.

In this chapter we introduce some of the basic tools of discrete mathematics. We begin with sets, subsets, and their operations, notions with which you may already be familiar. Next we deal with sequences, using both explicit and recursive patterns. Then we review some of the basic properties of the integers. Finally we introduce matrices and matrix operations. This gives us the background needed to begin our exploration of mathematical structures.

LOOKING BACK

Matrices The origin of matrices goes back to approximately 200 B.C.E., when they were used by the Chinese to solve linear systems of equations. After being in the shadows for nearly two thousand years, matrices came back into mathematics toward the end of the seventeenth century and from then research in this area proceeded at a rapid pace. The term "matrix" (the singular of "matrices") was coined in 1850 by James Joseph Sylvester (1814–1897), a British mathematician and lawyer. In 1851, Sylvester met Arthur Cayley (1821–1895), also a British lawyer with a strong interest in mathematics. Cayley quickly realized the importance of the notion of a matrix and in 1858 published a book showing the basic operations on matrices. He also discovered a number of important results in matrix theory.



James Joseph Sylvester

Arthur Cayley

1.1 Sets and Subsets

Sets

A **set** is any well-defined collection of objects called the **elements** or **members of the set**. For example, the collection of all wooden chairs, the collection of all one-legged black birds, or the collection of real numbers between zero and one are all sets. *Well-defined* just means that it is possible to decide if a given object belongs to the collection or not. Almost all mathematical objects are first of all sets, regardless of any additional properties they may possess. Thus set theory is, in a sense, the foundation on which virtually all of mathematics is constructed. In spite of this, set theory (at least the informal brand we need) is quite easy to learn and use.

One way of describing a set that has a finite number of elements is by listing the elements of the set between braces. Thus the set of all positive integers that are less than 4 can be written as

$$\{1, 2, 3\}. \tag{1}$$

The order in which the elements of a set are listed is not important. Thus $\{1, 3, 2\}$, $\{3, 2, 1\}$, $\{3, 1, 2\}$, $\{2, 1, 3\}$, and $\{2, 3, 1\}$ are all representations of the set given in (1). Moreover, repeated elements in the listing of the elements of a set can be ignored. Thus, $\{1, 3, 2, 3, 1\}$ is another representation of the set given in (1).

We use uppercase letters such as A, B, C to denote sets, and lowercase letters such as a, b, c, x, y, z, t to denote the members (or elements) of sets.

We indicate the fact that x is an element of the set A by writing $x \in A$, and we indicate the fact that x is not an element of A by writing $x \notin A$.

EXAMPLE 1 Let $A = \{1, 3, 5, 7\}$. Then $1 \in A$, $3 \in A$, but $2 \notin A$. ■

Sometimes it is inconvenient or impossible to describe a set by listing all of its elements. Another useful way to define a set is by specifying a property that the elements of the set have in common. We use the notation $P(x)$ to denote a sentence or statement P concerning the variable object x. The set defined by $P(x)$, written $\{x \mid P(x)\}$, is just the collection of all objects for which P is sensible and true. For example, $\{x \mid x$ is a positive integer less than $4\}$ is the set $\{1, 2, 3\}$ described in (1) by listing its elements.

EXAMPLE 2 The set consisting of all the letters in the word "byte" can be denoted by {b, y, t, e} or by $\{x \mid x$ is a letter in the word "byte"$\}$. ■

EXAMPLE 3 We introduce here several sets and their notations that will be used throughout this book.

(a) $Z^+ = \{x \mid x$ is a positive integer$\}$.
Thus Z^+ consists of the numbers used for counting: $1, 2, 3 \ldots$.

(b) $N = \{x \mid x$ is a positive integer or zero$\}$.
Thus N consists of the positive integers and zero: $0, 1, 2, \ldots$.

(c) $Z = \{x \mid x$ is an integer$\}$.
Thus Z consists of all the integers: $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$.

(d) $\mathbb{Q} = \{x \mid x$ is a rational number$\}$.
Thus $\mathbb{Q}$ consists of numbers that can be written as $\frac{a}{b}$, where a and b are integers and b is not 0.

(e) $\mathbb{R} = \{x \mid x$ is a real number$\}$.

(f) The set that has no elements in it is denoted either by $\{\ \}$ or the symbol $\varnothing$ and is called the **empty set**. ■

EXAMPLE 4 Since the square of a real number is always nonnegative,

$$\{x \mid x \text{ is a real number and } x^2 = -1\} = \varnothing. \quad ■$$

Sets are completely known when their members are all known. Thus we say two sets A and B are **equal** if they have the same elements, and we write $A = B$.

EXAMPLE 5 If $A = \{1, 2, 3\}$ and $B = \{x \mid x$ is a positive integer and $x^2 < 12\}$, then $A = B$. ■

EXAMPLE 6 If A = {BASIC, PASCAL, ADA} and B = {ADA, BASIC, PASCAL}, then $A = B$. ■

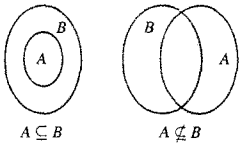


Figure 1.1

Subsets

If every element of A is also an element of B, that is, if whenever $x \in A$ then $x \in B$, we say that A is a **subset** of B or that A is **contained in** B, and we write $A \subseteq B$. If A is not a subset of B, we write $A \not\subseteq B$. (See Figure 1.1.)

Diagrams, such as those in Figure 1.1, which are used to show relationships between sets, are called **Venn diagrams** after the British logician John Venn. Venn diagrams will be used extensively in Section 1.2.

EXAMPLE 7 We have $Z^+ \subseteq Z$. Moreover, if $\mathbb{Q}$ denotes the set of rational numbers, then $Z \subseteq \mathbb{Q}$. ■

EXAMPLE 8 Let $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{2, 4, 5\}$, and $C = \{1, 2, 3, 4, 5\}$. Then $B \subseteq A$, $B \subseteq C$, and $C \subseteq A$. However, $A \not\subseteq B$, $A \not\subseteq C$, and $C \not\subseteq B$. ■

EXAMPLE 9 If A is any set, then $A \subseteq A$. That is, every set is a subset of itself. ■

EXAMPLE 10 Let A be a set and let $B = \{A, \{A\}\}$. Then, since A and $\{A\}$ are elements of B, we have $A \in B$ and $\{A\} \in B$. It follows that $\{A\} \subseteq B$ and $\{\{A\}\} \subseteq B$. However, it is not true that $A \subseteq B$. ■

For any set A, since there are no elements of $\varnothing$ that are not in A, we have $\varnothing \subseteq A$. (We will look at this again in Section 2.1.)

It is easy to see that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

The collection of everything, it turns out, cannot be considered a set without presenting serious logical difficulties. To avoid this and other problems, which need not concern us here, we will assume that for each discussion there is a "universal set" U (which will vary with the discussion) containing all objects for which the discussion is meaningful. Any other set mentioned in the discussion will automatically be assumed to be a subset of U. Thus, if we are discussing real numbers and we mention sets A and B, then A and B must (we assume) be sets of real numbers, not matrices, electronic circuits, or rhesus monkeys. In most problems, a universal set will be apparent from the setting of the problem. In Venn diagrams, the universal set U will be denoted by a rectangle, while sets within U will be denoted by circles as shown in Figure 1.2.

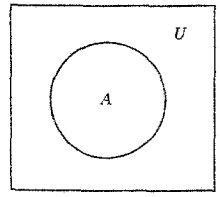


Figure 1.2

A set A is called **finite** if it has n distinct elements, where $n \in N$. In this case, n is called the **cardinality** of A and is denoted by $|A|$. Thus, the sets of Examples 1, 2, 4, 5, and 6 are finite. A set that is not finite is called **infinite**. The sets introduced in Example 3 (except $\varnothing$) are infinite sets.

If A is a set, then the set of all subsets of A is called the **power set** of A and is denoted by $P(A)$.

EXAMPLE 11 Let $A = \{1, 2, 3\}$. Then $P(A)$ consists of the following subsets of A: $\{\ \}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$ (or A). In a later section, we will count the number of subsets that a set can have. ■

1.1 Exercises

1. Let $A = \{1, 2, 4, a, b, c\}$. Identify each of the following as true or false.
(a) $2 \in A$ (b) $3 \in A$ (c) $c \notin A$
(d) $\varnothing \in A$ (e) $\{\ \} \notin A$ (f) $A \in A$

2. Let $A = \{x \mid x \text{ is a real number and } x < 6\}$. Identify each of the following as true or false.
(a) $3 \in A$ (b) $6 \in A$ (c) $5 \notin A$
(d) $8 \notin A$ (e) $-8 \in A$ (f) $3.4 \notin A$

3. In each part, give the set of letters in each word by listing the elements of the set.
(a) AARDVARK (b) BOOK
(c) MISSISSIPPI

4. Give the set by listing its elements.
(a) The set of all positive integers that are less than ten.
(b) $\{x \mid x \in Z \text{ and } x^2 < 12\}$

5. Let $A = \{1, \{2, 3\}, 4\}$. Identify each of the following as true or false.
(a) $3 \in A$ (b) $\{1, 4\} \subseteq A$ (c) $\{2, 3\} \subseteq A$
(d) $\{2, 3\} \in A$ (e) $\{4\} \in A$ (f) $\{1, 2, 3\} \subseteq A$

In Exercises 6 through 9, write the set in the form $\{x \mid P(x)\}$, where $P(x)$ is a property that describes the elements of the set.

6. $\{2, 4, 6, 8, 10\}$ **7.** $\{a, e, i, o, u\}$
8. $\{1, 8, 27, 64, 125\}$ **9.** $\{-2, -1, 0, 1, 2\}$

10. Let $A = \{1, 2, 3, 4, 5\}$. Which of the following sets are equal to A?
(a) $\{4, 1, 2, 3, 5\}$ (b) $\{2, 3, 4\}$ (c) $\{1, 2, 3, 4, 5, 6\}$
(d) $\{x \mid x \text{ is an integer and } x^2 \leq 25\}$
(e) $\{x \mid x \text{ is a positive integer and } x \leq 5\}$
(f) $\{x \mid x \text{ is a positive rational number and } x \leq 5\}$

11. Which of the following sets are the empty set?
(a) $\{x \mid x \text{ is a real number and } x^2 - 1 = 0\}$
(b) $\{x \mid x \text{ is a real number and } x^2 + 1 = 0\}$
(c) $\{x \mid x \text{ is a real number and } x^2 = -9\}$
(d) $\{x \mid x \text{ is a real number and } x = 2x + 1\}$
(e) $\{x \mid x \text{ is a real number and } x = x + 1\}$

12. List all the subsets of $\{a, b\}$.

13. List all the subsets of {BASIC, PASCAL, ADA}.

14. List all the subsets of $\{\ \}$.

15. Let $A = \{1, 2, 5, 8, 11\}$. Identify each of the following as true or false.
(a) $\{5, 1\} \subseteq A$ (b) $\{8, 1\} \in A$
(c) $\{1, 8, 2, 11, 5\} \nsubseteq A$ (d) $\varnothing \subseteq A$
(e) $\{1, 6\} \nsubseteq A$ (f) $\{2\} \subseteq A$
(g) $\{3\} \notin A$ (h) $A \subseteq \{11, 2, 5, 1, 8, 4\}$

16. Let $A = \{x \mid x \text{ is an integer and } x^2 < 16\}$. Identify each of the following as true or false.
(a) $\{0, 1, 2, 3\} \subseteq A$ (b) $\{-3, -2, -1\} \subseteq A$
(c) $\{\ \} \subseteq A$
(d) $\{x \mid x \text{ is an integer and } |x| < 4\} \subseteq A$
(e) $A \subseteq \{-3, -2, -1, 0, 1, 2, 3\}$

17. Let $A = \{1\}$, $B = \{1, a, 2, b, c\}$, $C = \{b, c\}$, $D = \{a, b\}$, and $E = \{1, a, 2, b, c, d\}$. For each part, replace the symbol $\square$ with either $\subseteq$ or $\nsubseteq$ to give a true statement.
(a) $A \square B$ (b) $\varnothing \square A$ (c) $B \square C$
(d) $C \square E$ (e) $D \square C$ (f) $B \square E$

In Exercises 18 through 20, find the set of smallest cardinality that contains the given sets as subsets.

18. $\{a, b, c\}, \{a, d, e, f\}, \{b, c, e, g\}$

19. $\{1, 2\}, \{1, 3\}, \varnothing$

20. $\{2, 4, 6, \ldots, 20\}, \{3, 6, 9, \ldots, 21\}$

21. Is it possible to have two different (appropriate) universal sets for a collection of sets? Would having different universal sets create any problems? Explain.

22. Use the Venn diagram in Figure 1.3 to identify each of the following as true or false.
(a) $A \subseteq B$ (b) $B \subseteq A$ (c) $C \subseteq B$
(d) $x \in B$ (e) $x \in A$ (f) $y \in B$

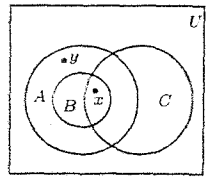


Figure 1.3

23. Use the Venn diagram in Figure 1.4 to identify each of the following as true or false.
(a) $B \subseteq A$ (b) $A \subseteq C$ (c) $C \subseteq B$
(d) $w \in A$ (e) $t \in A$ (f) $w \in B$

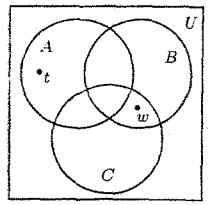


Figure 1.4

24. (a) Complete the following statement. A generic Venn diagram for a single set has _____ regions. Describe them in words.
(b) Complete the following statement. A generic Venn diagram for two sets has _____ regions. Describe them in words.

25. Complete the following statement. A generic Venn diagram for three sets has _____ regions. Describe them in words.

26. (a) If $A = \{3, 7\}$, find $P(A)$.
(b) What is $|A|$? (c) What is $|P(A)|$?

27. If $P(B) = \{\{\ \}, \{m\}, \{n\}, \{m, n\}\}$, then find B.

28. (a) If $A = \{3, 7, 2\}$, find $P(A)$.
(b) What is $|A|$? (c) What is $|P(A)|$?

29. If $P(B) = \{\{a\}, \{\ \}, \{c\}, \{b, c\}, \{a, b\}, \ldots\}$ and $|P(B)| = 8$, then $B =$ _____.

In Exercises 30 through 32, draw a Venn diagram that represents these relationships.

30. $A \subseteq B$, $A \subseteq C$, $B \nsubseteq C$, and $C \nsubseteq B$

31. $x \in A$, $x \in B$, $x \notin C$, $y \in B$, $y \in C$, and $y \notin A$

32. $A \subseteq B$, $x \notin A$, $x \in B$, $A \nsubseteq C$, $y \in B$, $y \in C$

33. Describe all the subset relationships that hold for the sets given in Example 3.

34. Show that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

35. The statement about sets in Exercise 34 can be restated as "Any subset of _____ is also a subset of any set that contains _____."

36. Suppose we know that set A has n subsets, $S_1, S_2, \ldots, S_n$. If set B consists of the elements of A and one more element so $|B| = |A| + 1$, show that B must have $2n$ subsets.

37. Compare the results of Exercises 12, 13, 26, and 28 and complete the following: Any set with two elements has _____ subsets. Any set with three elements has _____ subsets.

1.2 Operations on Sets

In this section we will discuss several operations that will combine given sets to yield new sets. These operations, which are analogous to the familiar operations on the real numbers, will play a key role in the many applications and ideas that follow.

If A and B are sets, we define their **union** as the set consisting of all elements that belong to A or B and denote it by $A \cup B$. Thus

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

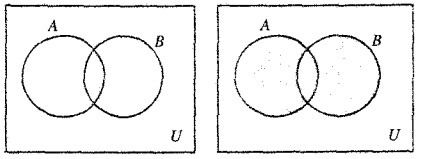
Observe that $x \in A \cup B$ if $x \in A$ or $x \in B$ or x belongs to both A and B.

EXAMPLE 1 Let $A = \{a, b, c, e, f\}$ and $B = \{b, d, r, s\}$. Find $A \cup B$.

Solution

Since $A \cup B$ consists of all the elements that belong to either A or B, $A \cup B = \{a, b, c, d, e, f, r, s\}$. ■

We can illustrate the union of two sets with a Venn diagram as follows. If A and B are the sets in Figure 1.5(a), then $A \cup B$ is the set represented by the shaded region in Figure 1.5(b).



(a) (b) $A \cup B$

Figure 1.5

If A and B are sets, we define their **intersection** as the set consisting of all elements that belong to both A and B and denote it by $A \cap B$. Thus

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

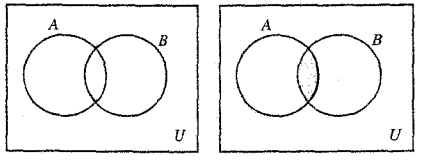
EXAMPLE 2 Let $A = \{a, b, c, e, f\}$, $B = \{b, e, f, r, s\}$, and $C = \{a, t, u, v\}$. Find $A \cap B$, $A \cap C$, and $B \cap C$.

Solution

The elements b, e, and f are the only ones that belong to both A and B, so $A \cap B = \{b, e, f\}$. Similarly, $A \cap C = \{a\}$. There are no elements that belong to both B and C, so $B \cap C = \{\ \}$. ■

Two sets that have no common elements, such as B and C in Example 2, are called **disjoint sets.**

We can illustrate the intersection of two sets by a Venn diagram as follows. If A and B are the sets given in Figure 1.6(a), then $A \cap B$ is the set represented by the shaded region in Figure 1.6(b). Figure 1.7 illustrates a Venn diagram for two disjoint sets.



(a) (b) $A \cap B$

Figure 1.6

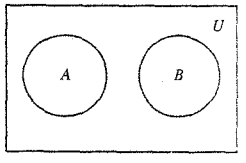


Figure 1.7

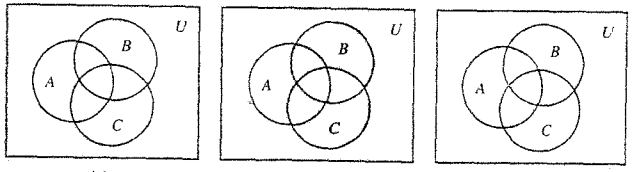
The operations of union and intersection can be defined for three or more sets in an obvious manner:

$$A \cup B \cup C = \{x \mid x \in A \text{ or } x \in B \text{ or } x \in C\}$$

and

$$A \cap B \cap C = \{x \mid x \in A \text{ and } x \in B \text{ and } x \in C\}.$$

The shaded region in Figure 1.8(b) is the union of the sets A, B, and C shown in Figure 1.8(a), and the shaded region in Figure 1.8(c) is the intersection of the sets A, B, and C. Note that Figure 1.8(a) says nothing about possible relationships between the sets, but allows for all possible relationships. In general, if $A_1, A_2, \ldots, A_n$ are subsets of U, then $A_1 \cup A_2 \cup \cdots \cup A_n$ will be denoted by $\bigcup_{k=1}^{n} A_k$ and $A_1 \cap A_2 \cap \cdots \cap A_n$ will be denoted by $\bigcap_{k=1}^{n} A_k$.



(a) (b) $A \cup B \cup C$ (c) $A \cap B \cap C$

Figure 1.8

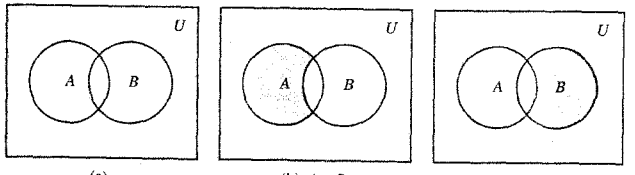
EXAMPLE 3 Let $A = \{1, 2, 3, 4, 5, 7\}$, $B = \{1, 3, 8, 9\}$, and $C = \{1, 3, 6, 8\}$. Then $A \cap B \cap C$ is the set of elements that belong to A, B, and C. Thus $A \cap B \cap C = \{1, 3\}$. ■

If A and B are two sets, we define the **complement of B with respect to A** as the set of all elements that belong to A but not to B, and we denote it by $A - B$. Thus

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}.$$

EXAMPLE 4 Let $A = \{a, b, c\}$ and $B = \{b, c, d, e\}$. Then $A - B = \{a\}$ and $B - A = \{d, e\}$. ■

If A and B are the sets in Figure 1.9(a), then $A - B$ and $B - A$ are represented by the shaded regions in Figures 1.9(b) and 1.9(c), respectively.



(a) (b) $A - B$ (c) $B - A$

Figure 1.9

If U is a universal set containing A, then $U - A$ is called the **complement** of A and is denoted by $\overline{A}$. Thus $\overline{A} = \{x \mid x \notin A\}$.

EXAMPLE 5 Let $A = \{x \mid x \text{ is an integer and } x \leq 4\}$ and $U = Z$. Then $\overline{A} = \{x \mid x \text{ is an integer and } x > 4\}$. ■

If A is the set in Figure 1.10, its complement is the shaded region in that figure.

If A and B are two sets, we define their **symmetric difference** as the set of all elements that belong to A or to B, but not to both A and B, and we denote it by $A \oplus B$. Thus

$$A \oplus B = \{x \mid (x \in A \text{ and } x \notin B) \text{ or } (x \in B \text{ and } x \notin A)\}.$$

EXAMPLE 6 Let $A = \{a, b, c, d\}$ and $B = \{a, c, e, f, g\}$. Then $A \oplus B = \{b, d, e, f, g\}$. ■

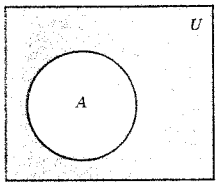
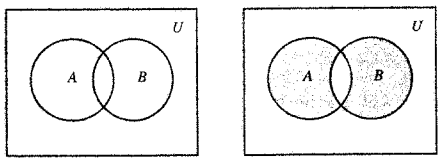


Figure 1.10

If A and B are as indicated in Figure 1.11(a), their symmetric difference is the shaded region shown in Figure 1.11(b). It is easy to see that

$$A \oplus B = (A - B) \cup (B - A).$$



(a) (b) $A \oplus B$

Figure 1.11

■ Algebraic Properties of Set Operations

The operations on sets that we have just defined satisfy many algebraic properties, some of which resemble the algebraic properties satisfied by the real numbers and their operations. All the principal properties listed here can be proved using the definitions given and the rules of logic. We shall prove only one of the properties and leave proofs of the remaining ones as exercises for the reader. Proofs are fundamental to mathematics. We discuss proof techniques in Chapter 2, but in this chapter some proofs are given as examples for later work. Some simple proofs are required in the exercises. Venn diagrams are often useful to suggest or justify the method of proof.

Theorem 1 The operations defined on sets satisfy the following properties:

Commutative Properties

1. $A \cup B = B \cup A$
2. $A \cap B = B \cap A$

Associative Properties

3. $A \cup (B \cup C) = (A \cup B) \cup C$
4. $A \cap (B \cap C) = (A \cap B) \cap C$

Distributive Properties

5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Idempotent Properties

7. $A \cup A = A$
8. $A \cap A = A$

Properties of the Complement

9. $(\overline{A}) = A$
10. $A \cup \overline{A} = U$
11. $A \cap \overline{A} = \varnothing$
12. $\overline{\varnothing} = U$
13. $\overline{U} = \varnothing$
14. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
15. $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Properties 14 and 15 are known as De Morgan's laws.

Properties of a Universal Set

16. $A \cup U = U$
17. $A \cap U = A$

Properties of the Empty Set

18. $A \cup \varnothing = A$ or $A \cup \{\ \} = A$
19. $A \cap \varnothing = \varnothing$ or $A \cap \{\ \} = \{\ \}$

Proof

We prove Property 14 here and leave proofs of the remaining properties as exercises for the reader. A common style of proof for statements about sets is to choose an element in one of the sets and see what we know about it. Suppose that $x \in \overline{A \cup B}$. Then we know that $x \notin A \cup B$, so $x \notin A$ and $x \notin B$. (Why?) This means $x \in \overline{A} \cap \overline{B}$ (why?), so each element of $\overline{A \cup B}$ belongs to $\overline{A} \cap \overline{B}$. Thus $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$. Conversely, suppose that $x \in \overline{A} \cap \overline{B}$. Then $x \notin A$ and $x \notin B$ (why?), so $x \notin A \cup B$, which means that $x \in \overline{A \cup B}$. Thus each element of $\overline{A} \cap \overline{B}$ also belongs to $\overline{A \cup B}$, and $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$. Now we see that $\overline{A \cup B} = \overline{A} \cap \overline{B}$. ■

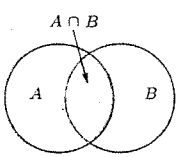


Figure 1.12

■ The Addition Principle

Suppose now that A and B are finite subsets of a universal set U. It is frequently useful to have a formula for $|A \cup B|$, the cardinality of the union. If A and B are disjoint sets, that is, if $A \cap B = \varnothing$, then each element of $A \cup B$ appears in either A or B, but not in both; therefore, $|A \cup B| = |A| + |B|$. If A and B overlap, as shown in Figure 1.12, then elements in $A \cap B$ belong to both sets, and the sum $|A| + |B|$ counts these elements twice. To correct for this double counting, we subtract $|A \cap B|$. Thus we have the following theorem, sometimes called the **addition principle**. Because of Figure 1.12, this is also called the **inclusion-exclusion principle.**

Theorem 2 If A and B are finite sets, then $|A \cup B| = |A| + |B| - |A \cap B|$. ■

EXAMPLE 7 Let $A = \{a, b, c, d, e\}$ and $B = \{c, e, f, h, k, m\}$. Verify Theorem 2.

Solution

We have $A \cup B = \{a, b, c, d, e, f, h, k, m\}$ and $A \cap B = \{c, e\}$. Also, $|A| = 5$, $|B| = 6$, $|A \cup B| = 9$, and $|A \cap B| = 2$. Then $|A| + |B| - |A \cap B| = 5 + 6 - 2$ or 9 and Theorem 2 is verified. ■

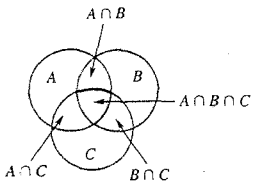


Figure 1.13

If A and B are disjoint sets, $A \cap B = \varnothing$ and $|A \cap B| = 0$, so the formula in Theorem 2 now becomes $|A \cup B| = |A| + |B|$. This special case can be stated in a way that is useful in a variety of counting situations.

If a task T_1 can be performed in exactly n ways, and a different task T_2 can be performed in exactly m ways, then the number of ways of performing task T_1 or task T_2 is $n + m$.

The situation for three sets is shown in Figure 1.13. We state the three-set addition principle without discussion.

Theorem 3 Let A, B, and C be finite sets. Then $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|$. ■

Theorem 3 can be generalized for more than three sets. This is done in Exercises 42 and 43.

EXAMPLE 8 Let $A = \{a, b, c, d, e\}$, $B = \{a, b, e, g, h\}$, and $C = \{b, d, e, g, h, k, m, n\}$. Verify Theorem 3.

Solution

We have $A \cup B \cup C = \{a, b, c, d, e, g, h, k, m, n\}$, $A \cap B = \{a, b, e\}$, $A \cap C = \{b, d, e\}$, $B \cap C = \{b, e, g, h\}$, and $A \cap B \cap C = \{b, e\}$, so $|A| = 5$, $|B| = 5$, $|C| = 8$, $|A \cup B \cup C| = 10$, $|A \cap B| = 3$, $|A \cap C| = 3$, $|B \cap C| = 4$, and $|A \cap B \cap C| = 2$. Thus $|A|+|B|+|C|-|A \cap B|-|B \cap C|-|A \cap C|+|A \cap B \cap C| = 5+5+8-3-3-4+2$ or 10, and Theorem 3 is verified. ■

EXAMPLE 9 A computer company wants to hire 25 programmers to handle systems programming jobs and 40 programmers for applications programming. Of those hired, ten will be expected to perform jobs of both types. How many programmers must be hired?

Solution

Let A be the set of systems programmers hired and B be the set of applications programmers hired. The company must have $|A| = 25$ and $|B| = 40$, and $|A \cap B| = 10$. The number of programmers that must be hired is $|A \cup B|$, but $|A \cup B| = |A| + |B| - |A \cap B|$. So the company must hire $25 + 40 - 10$ or 55 programmers. ■

EXAMPLE 10 A survey has been taken on methods of commuter travel. Each respondent was asked to check BUS, TRAIN, or AUTOMOBILE as a major method of traveling to work. More than one answer was permitted. The results reported were as follows: BUS, 30 people; TRAIN, 35 people; AUTOMOBILE, 100 people; BUS and TRAIN, 15 people; BUS and AUTOMOBILE, 15 people; TRAIN and AUTOMOBILE, 20 people; and all three methods, 5 people. How many people completed a survey form?

Solution

Let B, T, and A be the sets of people who checked BUS, TRAIN, and AUTOMOBILE, respectively. We know $|B| = 30$, $|T| = 35$, $|A| = 100$, $|B \cap T| = 15$, $|B \cap A| = 15$, $|T \cap A| = 20$, and $|B \cap T \cap A| = 5$. So $|B| + |T| + |A| - |B \cap T| - |B \cap A| - |T \cap A| + |B \cap T \cap A| = 30 + 35 + 100 - 15 - 15 - 20 + 5$ or 120 is $|A \cup B \cup C|$, the number of people who responded. ■

1.2 Exercises

In Exercises 1 through 4, let $U = \{a, b, c, d, e, f, g, h, k\}$, $A = \{a, b, c, g\}$, $B = \{d, e, f, g\}$, $C = \{a, c, f\}$, and $D = \{f, h, k\}$.

1. Compute
   (a) $A \cup B$ (b) $B \cup C$ (c) $A \cap C$
   (d) $B \cap D$ (e) $(A \cup B) - C$ (f) $A - B$
   (g) $\overline{A}$ (h) $A \oplus B$ (i) $A \oplus C$
   (j) $(A \cap B) - C$
2. Compute
   (a) $A \cup D$ (b) $B \cup D$ (c) $C \cap D$
   (d) $A \cap D$ (e) $(A \cup B) - (C \cup B)$
   (f) $B - C$ (g) $\overline{B}$ (h) $C - B$
   (i) $C \oplus D$ (j) $(A \cap B) - (B \cap D)$
3. Compute
   (a) $A \cup B \cup C$ (b) $A \cap B \cap C$
   (c) $A \cap (B \cup C)$ (d) $(A \cup B) \cap C$
   (e) $\overline{A \cup B}$ (f) $\overline{A \cap B}$
4. Compute
   (a) $A \cup \varnothing$ (b) $A \cup U$ (c) $B \cup B$
   (d) $C \cap \{\ \}$ (e) $\overline{C \cup D}$ (f) $\overline{C \cap D}$

In Exercises 5 through 8, let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{1, 2, 4, 6, 8\}$, $B = \{2, 4, 5, 9\}$, $C = \{x \mid x$ is a positive integer and $x^2 \leq 16\}$, and $D = \{7, 8\}$.

5. Compute
   (a) $A \cup B$ (b) $A \cup C$ (c) $A \cup D$
   (d) $B \cup C$ (e) $A \cap C$ (f) $A \cap D$
   (g) $B \cap C$ (h) $C \cap D$
6. Compute
   (a) $A - B$ (b) $B - A$ (c) $C - D$
   (d) $\overline{C}$ (e) $\overline{A}$ (f) $A \oplus B$
   (g) $C \oplus D$ (h) $B \oplus C$
7. Compute
   (a) $A \cup B \cup C$ (b) $A \cap B \cap C$
   (c) $A \cap (B \cup C)$ (d) $(A \cup B) \cap D$
   (e) $\overline{A \cup B}$ (f) $\overline{A \cap B}$
8. Compute
   (a) $B \cup C \cup D$ (b) $B \cap C \cap D$
   (c) $A \cup A$ (d) $A \cap \overline{A}$
   (e) $A \cup \overline{A}$ (f) $A \cap (\overline{C} \cup D)$

In Exercises 9 and 10, let $U = \{a, b, c, d, e, f, g, h\}$, $A = \{a, c, f, g\}$, $B = \{a, e\}$, and $C = \{b, h\}$.

9. Compute
   (a) $\overline{A}$ (b) $\overline{B}$ (c) $\overline{A \cup B}$
   (d) $\overline{A \cap B}$ (e) $\overline{U}$ (f) $A - B$
10. Compute
    (a) $\overline{A} \cap \overline{B}$ (b) $\overline{B} \cup \overline{C}$ (c) $\overline{A \cup A}$
    (d) $\overline{C \cap C}$ (e) $A \oplus B$ (f) $B \oplus C$
11. Let U be the set of real numbers, $A = \{x \mid x$ is a solution of $x^2 - 1 = 0\}$, and $B = \{-1, 4\}$. Compute
    (a) $\overline{A}$ (b) $\overline{B}$ (c) $\overline{A \cup B}$ (d) $\overline{A \cap B}$

In Exercises 12 and 13, refer to Figure 1.14.

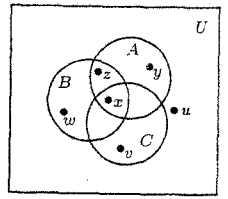


Figure 1.14

12. Identify the following as true or false.
    (a) $y \in A \cap B$ (b) $x \in B \cup C$
    (c) $w \in B \cap C$ (d) $u \notin C$
13. Identify the following as true or false.
    (a) $x \in A \cap B \cap C$ (b) $y \in A \cup B \cup C$
    (c) $z \in A \cap C$ (d) $v \in B \cap C$
14. Describe the shaded region shown in Figure 1.15 using unions and intersections of the sets A, B, and C. (Several descriptions are possible.)

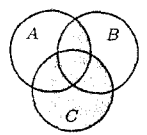


Figure 1.15

15. Let A, B, and C be finite sets with $|A| = 6$, $|B| = 8$, $|C| = 6$, $|A \cup B \cup C| = 11$, $|A \cap B| = 3$, $|A \cap C| = 2$, and $|B \cap C| = 5$. Find $|A \cap B \cap C|$.

In Exercises 16 through 18, verify Theorem 2 for the given sets.

16. (a) $A = \{1, 2, 3, 4\}$, $B = \{2, 3, 5, 6, 8\}$
    (b) $A = \{1, 2, 3, 4\}$, $B = \{5, 6, 7, 8, 9\}$
17. (a) $A = \{a, b, c, d, e, f\}$, $B = \{a, c, f, g, h, i, r\}$
    (b) $A = \{a, b, c, d, e\}$, $B = \{f, g, r, s, t, u\}$
18. (a) $A = \{x \mid x$ is a positive integer $< 8\}$,
    $B = \{x \mid x$ is an integer such that $2 \leq x \leq 5\}$
    (b) $A = \{x \mid x$ is a positive integer and $x^2 \leq 16\}$,
    $B = \{x \mid x$ is a negative integer and $x^2 \leq 25\}$

19. If A and B are disjoint sets such that $|A \cup B| = |A|$, what must be true about B?

In Exercises 20 through 22, verify Theorem 3 for the given sets.

20. $A = \{a, b, c, d, e\}$, $B = \{d, e, f, g, h, i, k\}$, $C = \{a, c, d, e, k, r, s, t\}$

21. $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{2, 4, 7, 8, 9\}$, $C = \{1, 2, 4, 7, 10, 12\}$

22. $A = \{x \mid x \text{ is a positive integer} < 8\}$, $B = \{x \mid x \text{ is an integer such that } 2 \le x \le 4\}$, $C = \{x \mid x \text{ is an integer such that } x^2 < 16\}$

23. In a survey of 260 college students, the following data were obtained:

64 had taken a mathematics course,

94 had taken a computer science course,

58 had taken a business course,

28 had taken both a mathematics and a business course,

26 had taken both a mathematics and a computer science course,

22 had taken both a computer science and a business course, and

14 had taken all three types of courses.

(a) How many students were surveyed who had taken none of the three types of courses?

(b) Of the students surveyed, how many had taken only a computer science course?

24. A survey of 500 television watchers produced the following information: 285 watch football games, 195 watch hockey games, 115 watch basketball games, 45 watch football and basketball games, 70 watch football and hockey games, 50 watch hockey and basketball games, and 50 do not watch any of the three kinds of games.

(a) How many people in the survey watch all three kinds of games?

(b) How many people watch exactly one of the sports?

25. The Journalism 101 class recently took a survey to determine where the city's people obtained their news. Unfortunately, some of the reports were damaged. What we know is that 88 people said they obtained their news from television, 73 from the local paper, and 46 from a news magazine. Thirty-four people reported that they obtained news from television and the local paper, 16 said they obtained their news from television and a news magazine, and 12 obtained theirs from the local paper and a news magazine. A total of five people reported that they used all three media. If 166 people were surveyed, how many use none of the three media to obtain their news? How many obtain their news from a news magazine exclusively?

26. The college catering service must decide if the mix of food that is supplied for receptions is appropriate. Of 100 people questioned, 37 say they eat fruits, 33 say they eat vegetables, 9 say they eat cheese and fruits, 12 eat cheese and vegetables, 10 eat fruits and vegetables, 12 eat only cheese, and 3 report they eat all three offerings. How many people surveyed eat cheese? How many do not eat any of the offerings?

27. In a psychology experiment, the subjects under study were classified according to body type and gender as follows:

	ENDO-MORPH	ECTO-MORPH	MESO-MORPH
Male	72	54	36
Female	62	64	38

(a) How many male subjects were there?

(b) How many subjects were ectomorphs?

(c) How many subjects were either female or endomorphs?

(d) How many subjects were not male mesomorphs?

(e) How many subjects were either male, ectomorph, or mesomorph?

28. The following table displays information about the sophomore, junior, and senior classes at Old U.

Class	Major Declared (D)	Major Undeclared (U)
Sophomore (S)	143	289
Junior (J)	245	158
Senior (R)	392	36

For each of the following tell how many students are in the set and describe those students in words.

(a) $D \cap J$ **(b)** $\overline{U \cup R}$ **(c)** $(D \cup S) \cap \overline{R}$

29. Create a Venn diagram that displays the information in the table in Exercise 28.

30. Complete the following proof that $A \subseteq A \cup B$. Suppose $x \in A$. Then $x \in A \cup B$, because _____. Thus by the definition of subset $A \subseteq A \cup B$.

31. Complete the following proof that $A \cap B \subseteq A$. Suppose $x \in A \cap B$. Then x belongs to _____. Thus $A \cap B \subseteq A$.

32. **(a)** Draw a Venn diagram to represent the situation $C \subseteq A$ and $C \subseteq B$.

(b) To prove $C \subseteq A \cup B$, we should choose an element from which set?

(c) Prove that if $C \subseteq A$ and $C \subseteq B$, then $C \subseteq A \cup B$.

33. **(a)** Draw a Venn diagram to represent the situation $A \subseteq C$ and $B \subseteq C$.

(b) To prove $A \cup B \subseteq C$, we should choose an element from which set?

(c) Prove that if $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.

34. Prove that $A - (A - B) \subseteq B$.

35. Suppose that $A \oplus B = A \oplus C$. Does this guarantee that $B = C$? Justify your conclusion.

36. Prove that $A - B = A \cap \overline{B}$.

37. If $A \cup B = A \cup C$, must $B = C$? Explain.

38. If $A \cap B = A \cap C$, must $B = C$? Explain.

39. Prove that if $A \subseteq B$ and $C \subseteq D$, then $A \cup C \subseteq B \cup D$ and $A \cap C \subseteq B \cap D$.

40. When is $A - B = B - A$? Explain.

41. Explain the last term in the sum in Theorem 3. Why is $|A \cap B \cap C|$ added and $|B \cap C|$ subtracted?

42. Write the four-set version of Theorem 3; that is, $|A \cup B \cup C \cup D| = \cdots$.

43. Describe in words the n-set version of Theorem 3.

1.3 Sequences

Some of the most important sets arise in connection with sequences. A **sequence** is simply a list of objects arranged in a definite order; a first element, second element, third element, and so on. The list may stop after n steps, $n \in N$, or it may go on forever. In the first case we say that the sequence is **finite**, and in the second case we say that it is **infinite**. The elements may all be different, or some may be repeated.

EXAMPLE 1 The sequence 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1 is a finite sequence with repeated items. The digit zero, for example, occurs as the second, third, fifth, seventh, and eighth elements of the sequence. ■

EXAMPLE 2 The list 3, 8, 13, 18, 23, ... is an infinite sequence. The three dots in the expression mean "and so on," that is, continue the pattern established by the first few elements. ■

EXAMPLE 3 Another infinite sequence is 1, 4, 9, 16, 25, ..., the list of the squares of all positive integers. ■

It may happen that how a sequence is to continue is not clear from the first few terms. Also, it may be useful to have a compact notation to describe a sequence. Two kinds of formulas are commonly used to describe sequences. In Example 2, a natural description of the sequence is that successive terms are produced by adding 5 to the previous term. If we use a subscript to indicate a term's position in the sequence, we can describe the sequence in Example 2 as $a_1 = 3$, $a_n = a_{n-1} + 5$, $2 \le n$. A formula, like this one, that refers to previous terms to define the next term is called **recursive**. Every recursive formula must include a starting place.

On the other hand, in Example 3 it is easy to describe a term using only its position number. In the nth position is the square of n; $b_n = n^2$, $1 \le n$. This type of formula is called **explicit**, because it tells us exactly what value any particular term has.

EXAMPLE 4 The recursive formula $c_1 = 5$, $c_n = 2c_{n-1}$, $2 \le n \le 6$, defines the finite sequence 5, 10, 20, 40, 80, 160. ■

EXAMPLE 5 The infinite sequence 3, 7, 11, 15, 19, 23, ... can be defined by the recursive formula $d_1 = 3$, $d_n = d_{n-1} + 4$. ■

EXAMPLE 6 The explicit formula $s_n = (-4)^n$, $1 \le n$, describes the infinite sequence $-4, 16, -64, 256, \ldots$. ■

EXAMPLE 7 The finite sequence 87, 82, 77, 72, 67 can be defined by the explicit formula $t_n = 92 - 5n$, $1 \le n \le 5$. ■

EXAMPLE 8 An ordinary English word such as "sturdy" can be viewed as the finite sequence

s, t, u, r, d, y

composed of letters from the ordinary English alphabet. ■

In examples such as Example 8, it is common to omit the commas and write the word in the usual way, if no confusion results. Similarly, even a meaningless word such as "abacabcd" may be regarded as a finite sequence of length 8. Sequences of letters or other symbols, written without the commas, are also referred to as **strings**.

EXAMPLE 9 An infinite string such as *abababab...* may be regarded as the infinite sequence $a, b, a, b, a, b, \ldots$. ■

EXAMPLE 10 The sentence "now is the time for the test" can be regarded as a finite sequence of English words: now, is, the, time, for, the, test. Here the elements of the sequence are themselves words of varying length, so we would not be able simply to omit the commas. The custom is to use spaces instead of commas in this case. ■

The **set corresponding to a sequence** is simply the set of all distinct elements in the sequence. Note that an essential feature of a sequence is the order in which the elements are listed. However, the order in which the elements of a set are listed is of no significance at all.

EXAMPLE 11

(a) The set corresponding to the sequence in Example 3 is $\{1, 4, 9, 16, 25, \ldots\}$.

(b) The set corresponding to the sequence in Example 9 is simply $\{a, b\}$. ■

The idea of a sequence is important in computer science, where a sequence is sometimes called a **linear array** or **list**. We will make a slight but useful distinction between a sequence and an array, and use a slightly different notation. If we have a sequence S: $s_1, s_2, s_3, \ldots$, we think of all the elements of S as completely determined. The element s_4, for example, is some fixed element of S, located in position four. Moreover, if we change any of the elements, we have a new sequence and will probably name it something other than S. Thus if we begin with the finite sequence S: 0, 1, 2, 3, 2, 1, 1 and we change the 3 to a 4, getting 0, 1, 2, 4, 2, 1, 1, we would think of this as a different sequence, say S'.

An array, on the other hand, may be viewed as a sequence of positions, which we represent in Figure 1.16 as boxes.

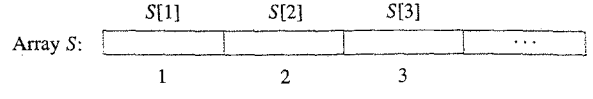


Figure 1.16

The positions form a finite or infinite list, depending on the desired size of the array. Elements from some set may be assigned to the positions of the array S. The element assigned to position n will be denoted by $S[n]$, and the sequence $S[1], S[2], S[3], \ldots$ will be called the **sequence of values** of the array S. The point is that S is considered to be a well-defined object, even if some of the positions have not been assigned values, or if some values are changed during the discussion. The following shows one use of arrays.

■ Characteristic Functions

A very useful concept for sets is the characteristic function. We discuss functions in Section 5.1, but for now we can proceed intuitively, and think of a function on a set as a rule that assigns some "value" to each element of the set. If A is a subset of a universal set U, the **characteristic function** f_A of A is defined for each $x \in U$ as follows:

$$f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

We may add and multiply characteristic functions, since their values are numbers, and these operations sometimes help us prove theorems about properties of subsets.

Theorem 1 Characteristic functions of subsets satisfy the following properties:

(a) $f_{A\cap B} = f_A f_B$; that is, $f_{A\cap B}(x) = f_A(x) f_B(x)$ for all x.

(b) $f_{A\cup B} = f_A + f_B - f_A f_B$; that is, $f_{A\cup B}(x) = f_A(x) + f_B(x) - f_A(x) f_B(x)$ for all x.

(c) $f_{A\oplus B} = f_A + f_B - 2 f_A f_B$; that is, $f_{A\oplus B}(x) = f_A(x) + f_B(x) - 2 f_A(x) f_B(x)$ for all x.

Proof

(a) $f_A(x) f_B(x)$ equals 1 if and only if both $f_A(x)$ and $f_B(x)$ are equal to 1, and this happens if and only if x is in A and x is in B, that is, x is in $A \cap B$. Since $f_A f_B$ is 1 on $A \cap B$ and 0 otherwise, it must be $f_{A\cap B}$.

(b) If $x \in A$, then $f_A(x) = 1$, so $f_A(x) + f_B(x) - f_A(x) f_B(x) = 1 + f_B(x) - f_B(x) = 1$. Similarly, when $x \in B$, $f_A(x) + f_B(x) - f_A(x) f_B(x) = 1$. If x is not in A or B, then $f_A(x)$ and $f_B(x)$ are 0, so $f_A(x) + f_B(x) - f_A(x) f_B(x) = 0$. Thus $f_A + f_B - f_A f_B$ is 1 on $A \cup B$ and 0 otherwise, so it must be $f_{A\cup B}$.

(c) We leave the proof of (c) as an exercise. ■

Note that the proof of Theorem 1 proceeds by direct application of the definition of the characteristic function.

■ Computer Representation of Sets and Subsets

Another use of characteristic functions is in representing sets in a computer. To represent a set in a computer, the elements of the set must be arranged in a sequence. The particular sequence selected is of no importance. When we list the set $A = \{a, b, c, \ldots, r\}$ we normally assume no particular ordering of the elements in A. Let us identify for now the set A with the sequence $a, b, c, \ldots, r$.

When a universal set U is finite, say $U = \{x_1, x_2, \ldots, x_n\}$, and A is a subset of U, then the characteristic function assigns 1 to an element that belongs to A and 0 to an element that does not belong to A. Thus f_A can be represented by a sequence of 0's and 1's of length n.

EXAMPLE 12 Let $U = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2\}$, $B = \{2, 4, 6\}$, and $C = \{4, 5, 6\}$. Then $f_A(x)$ has value 1 when x is 1 or 2, and otherwise is 0. Hence f_A corresponds to the sequence 1, 1, 0, 0, 0, 0. In a similar way, the finite sequence 0, 1, 0, 1, 0, 1 represents f_B and 0, 0, 0, 1, 1, 1 represents f_C. ■

Any set with n elements can be arranged in a sequence of length n, so each of its subsets corresponds to a sequence of zeros and ones of length n, representing the characteristic function of that subset. This fact allows us to represent a universal set in a computer as an array A of length n. Assignment of a zero or one to each location $A[k]$ of the array specifies a unique subset of U.

EXAMPLE 14

Let $U = \{a, b, e, g, h, r, s, w\}$. The array of length 8 shown in Figure 1.17 represents U, since $A[k] = 1$ for $1 \le k \le 8$.

If $S = \{a, e, r, w\}$, then

$$f_S(x) = \begin{cases} 1 & \text{if } x = a, e, r, w \\ 0 & \text{if } x = b, g, h, s. \end{cases}$$

1	1	1	1	1	1	1	1

Figure 1.17

Hence the array in Figure 1.18 represents the subset S.

1	0	1	0	0	1	0	1

Figure 1.18

A set is called **countable** if it is the set corresponding to some sequence. Informally, this means that the members of the set can be arranged in a list, with a first, second, third, . . . , element, and the set can therefore be "counted." We shall show in Section 2.4 that all finite sets are countable. However, not all infinite sets are countable. A set that is not countable is called **uncountable**.

The most accessible example of an uncountable set is the set of all real numbers that can be represented by an infinite decimal of the form $0.a_1a_2a_3\ldots$, where a_i is an integer and $0 \le a_i \le 9$. We shall now show that this set is uncountable. We will prove this result by contradiction; that is, we will show the countability of this set implies an impossible situation. (We will look more closely at proof by contradiction in Chapter 2.)

Assume that the set of all decimals $0.a_1a_2a_3\ldots$ is countable. Then we could form the following list (sequence), containing all such decimals:

$$\begin{aligned} d_1 &= 0.a_1a_2a_3\ldots \\ d_2 &= 0.b_1b_2b_3\ldots \\ d_3 &= 0.c_1c_2c_3\ldots \\ &\vdots \end{aligned}$$

Each of our infinite decimals must appear somewhere on this list. We shall establish a contradiction by constructing an infinite decimal of this type that is not on the list. Now construct a number x as follows: $x = 0.x_1x_2x_3\ldots$, where x_1 is 1 if $a_1 = 2$, otherwise x_1 is 2; $x_2 = 1$ if $b_2 = 2$, otherwise $x_2 = 2$; $x_3 = 1$ if $c_3 = 2$, otherwise $x_3 = 2$. This process can clearly be continued indefinitely. The resulting number is an infinite decimal consisting of 1's and 2's, but by its construction x differs from each number in the list at some position. Thus x is not on the list, a contradiction to our assumption. Hence no matter how the list is constructed, there is some real number of the form $0.x_1x_2x_3\ldots$ that is not in the list. On the other hand, it can be shown that the set of rational numbers is countable.

■ Strings and Regular Expressions

Given a set A, we can construct the set A^* consisting of all finite sequences of elements of A. Often, the set A is not a set of numbers, but some set of symbols. In this case, A is called an **alphabet**, and the finite sequences in A^* are called **words** from A, or sometimes strings from A. For this case in particular, the sequences in A^* are *not* written with commas. We assume that A^* contains the **empty sequence** or **empty string**, containing no symbols, and we denote this string by Λ. This string will be useful in Chapters 9 and 10.

EXAMPLE 14

Let $A = \{a, b, c, \ldots, z\}$, the usual English alphabet. Then A^* consists of all ordinary words, such as ape, sequence, antidisestablishmentarianism, and so on, as well as "words" such as yxaloble, zigadongdong, esy, and pqrst. All finite sequences from A are in A^*, whether they have meaning or not. ■

If $w_1 = s_1s_2s_3\ldots s_n$ and $w_2 = t_1t_2t_3\ldots t_k$ are elements of A^* for some set A, we define the **catenation** of w_1 and w_2 as the sequence $s_1s_2s_3\ldots s_nt_1t_2t_3\ldots t_k$. The catenation of w_1 with w_2 is written as $w_1 \cdot w_2$ or w_1w_2, and is another element of A^*. Note that if w belongs to A^*, then $w \cdot \Lambda = w$ and $\Lambda \cdot w = w$. This property is convenient and is one of the main reasons for defining the empty string Λ.

EXAMPLE 15

Let $A = \{\text{John, Sam, Jane, swims, runs, well, quickly, slowly}\}$. Then A^* contains real sentences such as "Jane swims quickly" and "Sam runs well," as well as nonsense sentences such as "Well swims Jane slowly John." Here we separate the elements in each sequence with spaces. This is often done when the elements of A are words. ■

The idea of a recursive formula for a sequence is useful in more general settings as well. In the formal languages and the finite state machines we discuss in Chapter 10, the concept of regular expression plays an important role, and regular expressions are defined recursively. A **regular expression over A** is a string constructed from the elements of A and the symbols $(,), \vee, *, \Lambda$, according to the following definition.

RE1. The symbol Λ is a regular expression.

RE2. If $x \in A$, the symbol x is a regular expression.

RE3. If α and β are regular expressions, then the expression $\alpha\beta$ is regular.

RE4. If α and β are regular expressions, then the expression $(\alpha \vee \beta)$ is regular.

RE5. If α is a regular expression, then the expression $(\alpha)^*$ is regular.

Note here that RE1 and RE2 provide initial regular expressions. The other parts of the definition are used repetitively to define successively larger sets of regular expressions from those already defined. Thus the definition is recursive.

By convention, if the regular expression α consists of a single symbol x, where $x \in A$, or if α begins and ends with parentheses, then we write $(\alpha)^*$ simply as α^*. When no confusion results, we will refer to a regular expression over A simply as a **regular expression** (omitting reference to A).

EXAMPLE 16

Let $A = \{0, 1\}$. Show that the following expressions are all regular expressions over A.

(a) $0^*(0 \vee 1)^*$ **(b)** $00^*(0 \vee 1)^*1$ **(c)** $(01)^*(01 \vee 1^*)$

Solution

(a) By RE2, 0 and 1 are regular expressions. Thus $(0 \vee 1)$ is regular by RE4, and so 0^* and $(0 \vee 1)^*$ are regular by RE5 (and the convention mentioned previously). Finally, we see that $0^*(0 \vee 1)^*$ is regular by RE3.

(b) We know that 0, 1, and $0^*(0 \vee 1)^*$ are all regular. Thus, using RE3 twice, $00^*(0 \vee 1)^*1$ must be regular.

(c) By RE3, 01 is a regular expression. Since 1^* is regular, $(01 \vee 1^*)$ is regular by RE4, and $(01)^*$ is regular by RE5. Then the regularity of $(01)^*(01 \vee 1^*)$ follows from RE3. ■

Associated with each regular expression over A, there is a corresponding subset of A^*. Such sets are called **regular subsets** of A^* or just **regular sets** if no reference to A is needed. To compute the regular set corresponding to a regular expression, we use the following correspondence rules.

1. The expression Λ corresponds to the set $\{\Lambda\}$, where Λ is the empty string in A^*.
2. If $x \in A$, then the regular expression x corresponds to the set $\{x\}$.
3. If α and β are regular expressions corresponding to the subsets M and N of A^*, then $\alpha\beta$ corresponds to $M \cdot N = \{s \cdot t \mid s \in M \text{ and } t \in N\}$. Thus $M \cdot N$ is the set of all catenations of strings in M with strings in N.
4. If the regular expressions α and β correspond to the subsets M and N of A^*, then $(\alpha \vee \beta)$ corresponds to $M \cup N$.
5. If the regular expression α corresponds to the subset M of A^*, then $(\alpha)^*$ corresponds to the set M^*. Note that M is a set of strings from A. Elements from M^* are finite sequences of such strings, and thus may themselves be interpreted as strings from A. Note also that we always have $\Lambda \in M^*$.

EXAMPLE 17 Let $A = \{a, b, c\}$. Then the regular expression a^* corresponds to the set of all finite sequences of a's, such as *aaa*, *aaaaaaa*, and so on. The regular expression $a(b \vee c)$ corresponds to the set $\{ab, ac\} \subseteq A^*$. Finally, the regular expression $ab(bc)^*$ corresponds to the set of all strings that begin with ab, and then repeat the symbols bc n times, where $n \geq 0$. This set includes the strings *ab*, *abbcbc*, *abbcbcbcbc*, and so on. ■

EXAMPLE 18 Let $A = \{0, 1\}$. Find regular sets corresponding to the three regular expressions in Example 16.

Solution

(a) The set corresponding to $0^*(0 \vee 1)^*$ consists of all sequences of 0's and 1's. Thus, the set is A^*.

(b) The expression $00^*(0 \vee 1)^*1$ corresponds to the set of all sequences of 0's and 1's that begin with at least one 0 and end with at least one 1.

(c) The expression $(01)^*(01 \vee 1^*)$ corresponds to the set of all sequences of 0's and 1's that either repeat the string 01 a total of $n \geq 1$ times, or begin with a total of $n \geq 0$ repetitions of 01 and end with some number $k \geq 0$ of 1's. This set includes, for example, the strings 1111, 01, 010101, 0101010111111, and 011. ■

1.3 Exercises

In Exercises 1 through 4, give the set corresponding to the sequence.

1. 1, 2, 1, 2, 1, 2, 1, 2, 1
2. 0, 2, 4, 6, 8, 10, ...
3. *aabbccddee...zz*
4. *abbcccdddd*
5. Give three different sequences that have $\{x, y, z\}$ as a corresponding set.
6. Give three different sequences that have $\{1, 2, 3, \ldots\}$ as a corresponding set.

In Exercises 7 through 14, write out the first four terms (begin with $n = 1$) of the sequence whose general term is given.

7. $a_n = 5^n$
8. $b_n = 3n^2 + 2n - 6$
9. $g_n = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$
10. $h_n = \dfrac{a^n - 1}{a - 1}, a \neq 1$
11. $c_1 = 2.5, c_n = c_{n-1} + 1.5$
12. $d_1 = -3, d_n = -2d_{n-1} + 1$
13. $e_1 = 0, e_n = e_{n-1} - 2$
14. $f_1 = 4, f_n = n \cdot f_{n-1}$

In Exercises 15 through 20, write a formula for the nth term of the sequence. Identify your formula as recursive or explicit.

15. 1, 3, 5, 7, ...
16. 0, 3, 8, 15, 24, 35, ...
17. 1, −1, 1, −1, 1, −1, ...
18. 0, 2, 0, 2, 0, 2, ...
19. 1, 4, 7, 10, 13, 16
20. $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots$
21. Write an explicit formula for the sequence 2, 5, 8, 11, 14, 17, ...
22. Write a recursive formula for the sequence 2, 5, 7, 12, 19, 31, ...
23. Let $A = \{x \mid x \text{ is a real number and } 0 < x < 1\}$, $B = \{x \mid x \text{ is a real number and } x^2 + 1 = 0\}$, $C = \{x \mid x = 4m, m \in Z\}$, $D = \{(x, 3) \mid x \text{ is an English word whose length is 3}\}$, and $E = \{x \mid x \in Z \text{ and } x^2 \leq 100\}$. Identify each set as finite, countable, or uncountable.
24. Let $A = W^*$ for $W = \{a, b\}$, $B = \{x \mid x \in \mathbb{R} \text{ and } x^2 + 41x + 41 = 0\}$, $C = \{x \mid x = \frac{m}{n}, m, n \in Z^+, n > 4\}$, $D = \{x \mid x \in \mathbb{R} \text{ and } x^2 + 3x + 2 \neq 0\}$, and $E = \{(x, y, z) \mid x \in Z, y \in \mathbb{R}^+, z \in Z^+\}$. Identify each set as finite, countable, or uncountable.
25. Let $A = \{ab, bc, ba\}$. In each part, tell whether the string belongs to A^*.
    (a) *ababab* (b) *abc* (c) *abba*
    (d) *abbcbaba* (e) *bcabbab* (f) *abbbcba*
26. Let U = {FORTRAN, PASCAL, ADA, COBOL, LISP, BASIC, C++, FORTH}, B = {C++, BASIC, ADA}, C = {PASCAL, ADA, LISP, C++}, D = {FORTRAN, PASCAL, ADA, BASIC, FORTH}, E = {PASCAL, ADA, COBOL, LISP, C++}. In each of the following, represent the given set by an array of zeros and ones.
    (a) $B \cup C$ (b) $C \cap D$
    (c) $B \cap (D \cap E)$ (d) $\overline{B} \cup E$
    (e) $\overline{C} \cap (B \cup E)$
27. Let $U = \{b, d, e, g, h, k, m, n\}$, $B = \{b\}$, $C = \{d, g, m, n\}$, and $D = \{d, k, n\}$.
    (a) What is $f_B(b)$? (b) What is $f_C(e)$?
    (c) Find the sequences of length 8 that correspond to f_B, f_C, and f_D.
    (d) Represent $B \cup C$, $C \cup D$, and $C \cap D$ by arrays of zeros and ones.
28. Complete the proof that $f_{A \oplus B} = f_A + f_B - 2f_A f_B$ [Theorem 1(c)]. Suppose $x \in A$ and $x \notin B$. Then $f_A(x) =$ ____, $f_B(x) =$ ____, and $f_A(x)f_B(x) =$ ____, so $f_A(x) + f_B(x) - 2f_A(x)f_B(x) =$ ____. Now suppose $x \notin A$ and $x \in B$. Then $f_A(x) =$ ____, $f_B(x) =$ ____, and $f_A(x)f_B(x) =$ ____, so $f_A(x) + f_B(x) - 2f_A(x)f_B(x) =$ ____. The remaining case to check is $x \notin A \oplus B$. If $x \notin A \oplus B$, then $x \in$ ____ and $f_A(x) + f_B(x) - 2f_A(x)f_B(x) =$ ____. Explain how these steps prove Theorem 1(c).
29. Using characteristic functions, prove that $(A \oplus B) \oplus C = A \oplus (B \oplus C)$.
30. Let $A = \{+, \times, a, b\}$. Show that the following expressions are regular over A.
    (a) $a + b(ab)^*(a \times b \vee a)$
    (b) $a + b \times (a^* \vee b)$
    (c) $(a^*b \vee +)^* \vee \times b^*$

In Exercises 31 and 32, let $A = \{a, b, c\}$. In each exercise a string in A^* is listed and a regular expression over A is given. In each case, tell whether or not the string on the left belongs to the regular set corresponding to the regular expression on the right.

31. (a) ac a^*b^*c (b) $abcc$ $(abc \vee c)^*$
    (c) $aaabc$ $((a \vee b) \vee c)^*$
32. (a) ac $(a^*b \vee c)$ (b) $abab$ $(ab)^*c$
    (c) $aaccc$ $(a^* \vee b)c^*$
33. Give three expressions that are not regular over the A given for Exercises 31 and 32.
34. Let $A = \{p, q, r\}$. Give the regular set corresponding to the regular expression given.
    (a) $(p \vee q)rq^*$ (b) $p(qq)^*r$
35. Let $S = \{0, 1\}$. Give the regular expression corresponding to the regular set given.

(a) {00, 010, 0110, 01110, ...}
(b) {0, 001, 000, 00001, 00000, 0000001, ...}

36. We define T-numbers recursively as follows:
1. 0 is a T-number.
2. If X is a T-number, $X + 3$ is a T-number.

Write a description of the set of T-numbers.

37. Define an S-number by
1. 8 is an S-number.
2. If X is an S-number and Y is a multiple of X, then Y is an S-number.
3. If X is an S-number and X is a multiple of Y, then Y is an S-number.

Describe the set of S-numbers.

38. Let F be a function defined for all nonnegative integers by the following recursive definition.

$$F(0) = 0. \qquad F(1) = 1$$
$$F(N+2) = 2F(N) + F(N+1), \quad N \geq 0$$

Compute the first six values of F; that is, write the values of $F(N)$ for $N = 0, 1, 2, 3, 4, 5$.

39. Let G be a function defined for all nonnegative integers by the following recursive definition.

$$G(0) = 1, \qquad G(1) = 2$$
$$G(N+2) = G(N)^2 + G(N+1), \quad N \geq 0$$

Compute the first five values of G.

1.4 Properties of Integers

We shall now discuss some results needed later about division and factoring in the integers. If m is an integer and n is a positive integer, we can plot the integer multiples of n on a line, and locate m as in Figure 1.19. If m is a multiple of n, say $m = qn$, then we can write $m = qn + r$, where r is 0. On the other hand (as shown in Figure 1.19), if m is not a multiple of n, we let qn be the first multiple of n lying to the left of m and let r be $m - qn$. Then r is the distance from qn to m, so clearly $0 < r < n$, and again we have $m = qn + r$. We state these observations as a theorem.

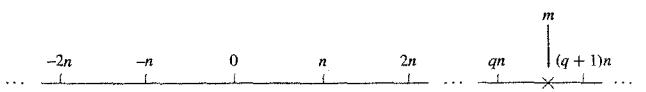


Figure 1.19

Theorem 1 If n and m are integers and $n > 0$, we can write $m = qn + r$ for integers q and r with $0 \leq r < n$. Moreover, there is just one way to do this. ■

EXAMPLE 1

(a) If n is 3 and m is 16, then $16 = 5(3) + 1$ so q is 5 and r is 1.
(b) If n is 10 and m is 3, then $3 = 0(10) + 3$ so q is 0 and r is 3.
(c) If n is 5 and m is -11, then $-11 = -3(5) + 4$ so q is -3 and r is 4. ■

If the r in Theorem 1 is zero, so that m is a multiple of n, we write $n \mid m$, which is read "n divides m." If $n \mid m$, then $m = qn$ and $n \leq |m|$. If m is not a multiple of n, we write $n \nmid m$, which is read "n does not divide m." We now prove some simple properties of divisibility.

Theorem 2 Let a, b, and c be integers.
(a) If $a \mid b$ and $a \mid c$, then $a \mid (b + c)$.
(b) If $a \mid b$ and $a \mid c$, where $b > c$, then $a \mid (b - c)$.
(c) If $a \mid b$ or $a \mid c$, then $a \mid bc$.
(d) If $a \mid b$ and $b \mid c$, then $a \mid c$.

Proof

(a) If $a \mid b$ and $a \mid c$, then $b = k_1 a$ and $c = k_2 a$ for integers k_1 and k_2. So $b + c = (k_1 + k_2)a$ and $a \mid (b + c)$.
(b) This can be proved in exactly the same way as (a).
(c) As in (a), we have $b = k_1 a$ or $c = k_2 a$. Then either $bc = k_1 ac$ or $bc = k_2 ab$, so in either case bc is a multiple of a and $a \mid bc$.
(d) If $a \mid b$ and $b \mid c$, we have $b = k_1 a$ and $c = k_2 b$, so $c = k_2 b = k_2(k_1 a) = (k_2 k_1)a$ and hence $a \mid c$. ■

Note that again we have a proof that proceeds directly from a definition by restating the original conditions. As a consequence of Theorem 2, we have that if $a \mid b$ and $a \mid c$, then $a \mid (mb + nc)$, for any integers m and n.

A number $p > 1$ in Z^+ is called **prime** if the only positive integers that divide p are p and 1.

EXAMPLE 2 The numbers 2, 3, 5, 7, 11, and 13 are prime, while 4, 10, 16, and 21 are not prime. ■

It is easy to write a set of steps, or an **algorithm**,* to determine if a positive integer $n > 1$ is a prime number. First we check to see if n is 2. If $n > 2$, we could divide by every integer from 2 to $n - 1$, and if none of these is a divisor of n, then n is prime. To make the process more efficient, we note that if $mk = n$, then either m or k is less than or equal to $\sqrt{n}$. This means that if n is not prime, it has a divisor k satisfying the inequality $1 < k \leq \sqrt{n}$, so we need only test for divisors in this range. Also, if n has any even number as a divisor, it must have 2 as a divisor. Thus after checking for divisibility by 2, we may skip all even integers.

Algorithm

To test whether an integer $N > 1$ is prime:

Step 1 Check whether N is 2. If so, N is prime. If not, proceed to
Step 2 Check whether $2 \mid N$. If so, N is not prime; otherwise, proceed to
Step 3 Compute the largest integer $K \leq \sqrt{N}$. Then
Step 4 Check whether $D \mid N$, where D is any odd number such that $1 < D \leq K$. If $D \mid N$, then N is not prime; otherwise, N is prime. ●

Testing whether an integer is prime is a common task for computers. The algorithm given here is too inefficient for testing very large numbers, but there are many other algorithms for testing whether an integer is prime.

Theorem 3 Every positive integer $n > 1$ can be written uniquely as $p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$, where $p_1 < p_2 < \cdots < p_s$ are distinct primes that divide n and the k's are positive integers giving the number of times each prime occurs as a factor of n. ■

We leave the proof of Theorem 3 to Section 2.4, but we give several illustrations.

EXAMPLE 3

(a) $9 = 3 \cdot 3 = 3^2$
(b) $24 = 8 \cdot 3 = 2 \cdot 2 \cdot 2 \cdot 3 = 2^3 \cdot 3$
(c) $30 = 2 \cdot 3 \cdot 5$ ■

*Algorithms are discussed in Appendix A.

■ Greatest Common Divisor

If a, b, and k are in Z^+, and $k \mid a$ and $k \mid b$, we say that k is a **common divisor** of a and b. If d is the largest such k, d is called the **greatest common divisor**, or GCD, of a and b, and we write $d = \text{GCD}(a, b)$. This number has some interesting properties. It can be written as a combination of a and b, and it is not only larger than all the other common divisors, it is a multiple of each of them.

Theorem 4 If d is GCD(a, b), then

(a) $d = sa + tb$ for some integers s and t. (These are not necessarily positive.)

(b) If c is any other common divisor of a and b, then $c \mid d$.

Proof

Let x be the smallest positive integer that can be written as $sa + tb$ for some integers s and t, and let c be a common divisor of a and b. Since $c \mid a$ and $c \mid b$, it follows from Theorem 2 that $c \mid x$, so $c \leq x$. If we can show that x is a common divisor of a and b, it will then be the greatest common divisor of a and b and both parts of the theorem will have been proved. By Theorem 1, $a = qx + r$ with $0 \leq r < x$. Solving for r, we have

$$r = a - qx = a - q(sa + tb) = a - qsa - qtb = (1 - qs)a + (-qt)b.$$

If r is not zero, then since $r < x$ and r is the sum of a multiple of a and a multiple of b, we will have a contradiction to the fact that x is the smallest positive number that is a sum of multiples of a and b. Thus r must be 0 and $x \mid a$. In the same way we can show that $x \mid b$, and this completes the proof. ∎

This proof is more complex than the earlier ones. At this stage you should focus on understanding the details of each step. We will discuss the structure of this proof later.

From the definition of greatest common divisor and Theorem 4(b), we have the following result: Let a, b, and d be in Z^+. The integer d is the greatest common divisor of a and b if and only if

(a) $d \mid a$ and $d \mid b$.

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

EXAMPLE 4

(a) The common divisors of 12 and 30 are 1, 2, 3, and 6, so that

$$\text{GCD}(12, 30) = 6 \quad \text{and} \quad 6 = 1 \cdot 30 + (-2) \cdot 12.$$

(b) It is clear that GCD$(17, 95) = 1$ since 17 is prime and $17 \nmid 95$, and the reader may verify that $1 = 28 \cdot 17 + (-5) \cdot 95$. ■

If GCD$(a, b) = 1$, as in Example 4(b), we say a and b are **relatively prime**.

One remaining question is that of how to compute the GCD conveniently in general. Repeated application of Theorem 1 provides the key to doing this.

We now present a procedure, called the **Euclidean algorithm**, for finding GCD(a, b). Suppose that $a > b > 0$ (otherwise interchange a and b). Then by Theorem 1, we may write

$$a = k_1 b + r_1, \qquad \text{where } k_1 \text{ is in } Z^+ \text{ and } 0 \leq r_1 < b. \tag{1}$$

Now Theorem 2 tells us that if n divides a and b, then it must divide r_1, since $r_1 = a - k_1 b$. Similarly, if n divides b and r_1, then it must divide a. We see that the common divisors of a and b are the same as the common divisors of b and r_1, so GCD$(a, b) = \text{GCD}(b, r_1)$.

We now continue using Theorem 1 as follows:

$$\begin{array}{rlll}
\text{divide } b \text{ by } r_1: & b = k_2 r_1 + r_2 & & 0 \leq r_2 < r_1 \\
\text{divide } r_1 \text{ by } r_2: & r_1 = k_3 r_2 + r_3 & & 0 \leq r_3 < r_2 \\
\text{divide } r_2 \text{ by } r_3: & r_2 = k_4 r_3 + r_4 & & 0 \leq r_4 < r_3 \\
\vdots & \vdots & & \vdots \\
\text{divide } r_{n-2} \text{ by } r_{n-1}: & r_{n-2} = k_n r_{n-1} + r_n & & 0 \leq r_n < r_{n-1} \\
\text{divide } r_{n-1} \text{ by } r_n: & r_{n-1} = k_{n+1} r_n + r_{n+1} & & 0 \leq r_{n-1} < r_n.
\end{array} \tag{2}$$

Since $a > b > r_1 > r_2 > r_3 > r_4 > \cdots$, the remainder will eventually become zero, so at some point we have $r_{n+1} = 0$.

We now show that $r_n = \text{GCD}(a, b)$. We saw previously that

$$\text{GCD}(a, b) = \text{GCD}(b, r_1).$$

Repeating this argument with b and r_1, we see that

$$\text{GCD}(b, r_1) = \text{GCD}(r_1, r_2).$$

Upon continuing, we have

$$\text{GCD}(a, b) = \text{GCD}(b, r_1) = \text{GCD}(r_1, r_2) = \cdots = \text{GCD}(r_{n-1}, r_n).$$

Since $r_{n-1} = k_{n+1} r_n$, we see that GCD$(r_{n-1}, r_n) = r_n$. Hence $r_n = \text{GCD}(a, b)$.

EXAMPLE 5

Let a be 190 and b be 34. Then, using the Euclidean algorithm, we

$$\begin{array}{ll}
\text{divide 190 by 34:} & 190 = 5 \cdot 34 + 20 \\
\text{divide 34 by 20:} & 34 = 1 \cdot 20 + 14 \\
\text{divide 20 by 14:} & 20 = 1 \cdot 14 + 6 \\
\text{divide 14 by 6:} & 14 = 2 \cdot 6 + 2 \\
\text{divide 6 by 2:} & 6 = 3 \cdot 2 + 0
\end{array}$$

so GCD$(190, 34) = 2$, the last of the nonzero divisors. ■

In Theorem 4(a), we observed that if $d = \text{GCD}(a, b)$, we can find integers s and t such that $d = sa + tb$. The integers s and t can be found as follows. Solve the next-to-last equation in (2) for r_n:

$$r_n = r_{n-2} - k_n r_{n-1}. \tag{3}$$

Now solve the second-to-last equation in (2), $r_{n-3} = k_{n-1} r_{n-2} + r_{n-1}$ for r_{n-1}:

$$r_{n-1} = r_{n-3} - k_{n-1} r_{n-2}$$

and substitute this expression in (3):

$$r_n = r_{n-2} - k_n [r_{n-3} - k_{n-1} r_{n-2}].$$

Continue to work up through the equations in (2) and (1), replacing r_i by an expression involving r_{i-1} and r_{i-2}, and finally arriving at an expression involving only a and b.

EXAMPLE 6

(a) Let $a = 190$ and $b = 34$ as in Example 5. Then

$$\begin{aligned} \text{GCD}(190, 34) &= 2 = 14 - 2(6) \\ &= 14 - 2[20 - 1(14)] && 6 = 20 - 1 \cdot 14 \\ &= 3(14) - 2(20) \\ &= 3[34 - 1(20)] - 2(20) && 14 = 34 - 1 \cdot 20 \\ &= 3(34) - 5(190 - 5 \cdot 34) && 20 = 190 - 5 \cdot 34 \\ &= 28(34) - 5(190). \end{aligned}$$

Hence $s = -5$ and $t = 28$. Note that the key is to carry out the arithmetic only partially.

(b) Let $a = 108$ and $b = 60$. Then

$$\begin{aligned} \text{GCD}(108, 60) &= 12 = 60 - 1(48) \\ &= 60 - 1[108 - 1(60)] && 48 = 108 - 1 \cdot 60 \\ &= 2(60) - 108. \end{aligned}$$

Hence $s = -1$ and $t = 2$. ■

Theorem 5 If a and b are in Z^+, $b > a$, then $\text{GCD}(a, b) = \text{GCD}(b, b \pm a)$.

Proof

If c divides a and b, it divides $b \pm a$, by Theorem 2. Since $a = b - (b - a) = -b + (b + a)$, we see, also by Theorem 2, that a common divisor of b and $b \pm a$ also divides a and b. Since a and b have the same common divisors as b and $b \pm a$, they must have the same greatest common divisor. ■

This is another direct proof, but one that uses a previous theorem as well as definitions.

■ Least Common Multiple

If a, b, and k are in Z^+, and $a \mid k$, $b \mid k$, we say k is a **common multiple** of a and b. The smallest such k, call it c, is called the **least common multiple**, or LCM, of a and b, and we write $c = \text{LCM}(a, b)$. The following result shows that we can obtain the least common multiple from the greatest common divisor, so we do not need a separate procedure for finding the least common multiple.

Theorem 6 If a and b are two positive integers, then $\text{GCD}(a, b) \cdot \text{LCM}(a, b) = ab$.

Proof

Let $p_1, p_2, \ldots, p_k$ be all the prime factors of either a or b. Then we can write

$$a = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} \quad \text{and} \quad b = p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$$

where some of the a_i and b_i may be zero. It then follows that

$$\text{GCD}(a, b) = p_1^{\min(a_1, b_1)} p_2^{\min(a_2, b_2)} \cdots p_k^{\min(a_k, b_k)}$$

and

$$\text{LCM}(a, b) = p_1^{\max(a_1, b_1)} p_2^{\max(a_2, b_2)} \cdots p_k^{\max(a_k, b_k)}.$$

Hence

$$\begin{aligned} \text{GCD}(a, b) \cdot \text{LCM}(a, b) &= p_1^{a_1 + b_1} p_2^{a_2 + b_2} \cdots p_k^{a_k + b_k} \\ &= (p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}) \cdot (p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}) \\ &= ab. \end{aligned}$$

■

EXAMPLE 7

Let $a = 540$ and $b = 504$. Factoring a and b into primes, we obtain

$$a = 540 = 2^2 \cdot 3^3 \cdot 5 \quad \text{and} \quad b = 504 = 2^3 \cdot 3^2 \cdot 7.$$

Thus all the prime numbers that are factors of either a or b are $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and $p_4 = 7$. Then $a = 2^2 \cdot 3^3 \cdot 5^1 \cdot 7^0$ and $b = 2^3 \cdot 3^2 \cdot 5^0 \cdot 7^1$. We then have

$$\begin{aligned} \text{GCD}(540, 504) &= 2^{\min(2,3)} \cdot 3^{\min(3,2)} \cdot 5^{\min(1,0)} \cdot 7^{\min(0,1)} \\ &= 2^2 \cdot 3^2 \cdot 5^0 \cdot 7^0 \\ &= 2^2 \cdot 3^2 \text{ or } 36. \end{aligned}$$

Also,

$$\begin{aligned} \text{LCM}(540, 504) &= 2^{\max(2,3)} \cdot 3^{\max(3,2)} \cdot 5^{\max(1,0)} \cdot 7^{\max(0,1)} \\ &= 2^3 \cdot 3^3 \cdot 5^1 \cdot 7^1 \text{ or } 7560. \end{aligned}$$

Then

$$\text{GCD}(540, 504) \cdot \text{LCM}(540, 504) = 36 \cdot 7560 = 272{,}160 = 540 \cdot 504.$$

As a verification, we can also compute GCD(540, 504) by the Euclidean algorithm and obtain the same result. ■

If n and m are integers and $n > 1$, Theorem 1 tells us we can write $m = qn + r$, $0 \leq r < n$. Sometimes the remainder r is more important than the quotient q.

EXAMPLE 8

If the time is now 4 o'clock, what time will it be 101 hours from now?

Solution

Let $n = 12$ and $m = 4 + 101$ or 105. Then we have $105 = 8 \cdot 12 + 9$. The remainder 9 answers the question. In 101 hours it will be 9 o'clock. ■

For each $n \in Z^+$, we define a function f_n, the mod-n function, as follows: If z is a nonnegative integer, $f_n(z) = r$, the remainder when z is divided by n. (Again, functions are formally defined in Section 5.1, but as in Section 1.3, we need only think of a function as a rule that assigns some "value" to each member of a set.) The naming of these functions is made clear in Section 4.5.

EXAMPLE 9

(a) $f_3(14) = 2$, because $14 = 4 \cdot 3 + 2$.

(b) $f_7(153) = 6$. ■

■ Pseudocode Versions

An alternative to expressing an algorithm in ordinary English as we did in this section is to express it in something like a computer language. Throughout the book we use a **pseudocode** language, which is described fully in Appendix A. Here we give pseudocode versions for an algorithm that determines if an integer is prime and for an algorithm that calculates the greatest common divisor of two integers.

In the pseudocode for the algorithm to determine if an integer is prime, we assume the existence of functions SQR and INT, where SQR(N) returns the greatest integer not exceeding $\sqrt{N}$, and INT(X) returns the greatest integer not exceeding X. For example, SQR(10) = 3, SQR(25) = 5, INT(7.124) = 7, and INT(8) = 8.

```
SUBROUTINE PRIME(N)
1. IF (N = 2) THEN
     a. PRINT ('PRIME')
     b. RETURN
2. ELSE
     a. IF (N/2 = INT(N/2)) THEN
          1. PRINT ('NOT PRIME')
          2. RETURN
     b. ELSE
          1. FOR D = 3 THRU SQR(N) BY 2
               a. IF (N/D = INT(N/D)) THEN
                    1. PRINT ('NOT PRIME')
                    2. RETURN
          2. PRINT ('PRIME')
          3. RETURN
END OF SUBROUTINE PRIME
```

The following gives a pseudocode program for finding the greatest common divisor of two positive integers. This procedure is different from the Euclidean algorithm, but in Chapter 2, we will see how to prove that this algorithm does indeed find the greatest common divisor.

```
FUNCTION GCD(X,Y)
1. WHILE (X ≠ Y)
     a. IF (X > Y) THEN
          1. X ← X − Y
     b. ELSE
          1. Y ← Y − X
2. RETURN (X)
END OF FUNCTION GCD
```

EXAMPLE 10 Use the pseudocode for GCD to calculate the greatest common divisor of 190 and 34 (Example 5).

Solution

The following table gives the values of X, Y, $X - Y$, or $Y - X$ as we go through the program.

X	Y	$X - Y$	$Y - X$
190	34	156	
156	34	122	
122	34	88	
88	34	54	
54	34	20	
20	34		14
20	14	6	
6	14		8
6	8		2
6	2	4	
4	2	2	
2	2		

Since the last value of X is 2, the greatest common divisor of 190 and 34 is 2. ■

■ Representations of Integers

The decimal representation of an integer is so familiar that we sometimes regard it as the symbol, or name for that integer. For example, when we write the integer 3264, we are saying that the number is the result of adding 3 times 10^3, 2 times 10^2, 6 times 10^1, and 4, or 4 times 10^0. We say 3264 is the **base 10 expansion of n** or the **decimal expansion of n**; 10 is called the **base** of this expansion.

There is nothing special about using the number 10 as a base, and it is likely the result of our having 10 fingers on which to count. Any positive integer $b > 1$ can be used in a similar way, and these expansions are often of much greater use than the standard decimal expansion. The bases 2, 8, and 16 are frequently used in computer science, and the base 26 is sometimes used in **cryptology** (the science of producing and deciphering secret codes).

Theorem 7 If $b > 1$ is an integer, then every positive integer n can be uniquely expressed in the form

$$n = d_k b^k + d_{k-1} b^{k-1} + \cdots + d_1 b + d_0, \tag{4}$$

where $0 \leq d_i < b$, $i = 0, 1, \ldots, k$, and $d_k \neq 0$. The sequence $d_k d_{k-1} \ldots d_1 d_0$ is called the **base b expansion of n**. If we need to explicitly indicate the base b, we will write the above sequence as $(d_k d_{k-1} \ldots d_1 d_0)_b$.

Proof

Suppose that k is the largest nonnegative integer so that $b^k \leq n$ (k could be 0). By Theorem 1 we can uniquely write $n = qb^k + r$, where $0 \leq r < b^k$. Let $d_k = q$. If $k = 0$, then $r = 0$ and we are done ($n = d_0$). Otherwise we have

$$n = d_k b^k + r.$$

Repeat this process, using r in place of n. Let s be the largest nonnegative integer so that $b^s \leq r$, write $r = qb^s + r_1$, where $0 \leq r_1 < b^s$, and define d_s to be q. If $s < k$, define $d_{s-1}, \ldots, d_{k-1}$ to be 0. Then

$$n = d_k b^k + d_{k-1} b^{k-1} + \cdots + d_s b^s + r_1.$$

Continuing this process using $r_1, r_2, \ldots$, we will eventually arrive at (4). ■

Now that we know that the base b expansion exists, we can find the digits $d_k, d_{k-1}, \ldots, d_1, d_0$ by a more direct method, easily implemented on a computer. Note that

$$n = (d_k b^{k-1} + d_{k-1} b^{k-2} + \cdots + d_1)b + d_0$$

so that d_0 is the remainder after dividing n by b, and the quotient is $d_k b^{k-1} + d_{k-1} b^{k-2} + \cdots + d_1$. Similarly, if this quotient is divided by b, the remainder is d_1. By repeatedly dividing the quotients by b and saving the remainders, we produce the digits of the base b representation of n from right to left.

EXAMPLE 11 Find the base 4 representation of 158.

Solution

We repeatedly divide by 4 and save the remainder:

$$\begin{array}{r|ll} 4 & 158 & 2 \\ 4 & 39 & 3 \\ 4 & 9 & 1 \\ 4 & 2 & 2 \\ & 0 & \end{array}$$

Thus $158 = (2132)_4$. ■

The following pseudocode algorithm returns the digits of the base b expansion of an integer. We use the expression m **mod** n to denote the mod-n function value for m, that is, the remainder after dividing m by n. The mod-n functions are commonly implemented in most programming languages.

```
SUBROUTINE EXPANSION(N)
1. Q ← N
2. K ← 0
3. WHILE (Q ≠ 0)
    a. D_K ← Q mod B
    b. Q ← INT(Q/B)
    c. K ← K + 1
4. RETURN
END OF SUBROUTINE EXPANSION
```

When this subroutine ends, the base B expansion of N consists of the integers D_i, which can then be further processed.

No matter what base is used to represent integers, the elementary rules of addition and multiplication are still valid. Only the appearance of the numbers changes.

EXAMPLE 12 Let $m = (313)_4$ and $n = (322)_4$. Find the base 4 expansion of $m + n$.

Solution

Adding the digits in the last column, we have $3 + 2 = (11)_4$, so we record a 1 and carry a 1 to the next column.

$$\begin{array}{r} 1 \\ 313 \\ +\ 322 \\ \hline 1 \end{array}$$

Adding the digits in the second column, we have $1 + 1 + 2 = (10)_4$, so we record a 0 and carry the 1.

$$\begin{array}{r} 11 \\ 313 \\ +\ 322 \\ \hline 01 \end{array}$$

Finally, adding the digits in the first column, we obtain $3 + 3 + 1 = (13)_4$, so the answer is $(1301)_4$. ■

The most common expansion used in computer work is the base 2 or binary expansion. Since the only remainders of division by 2 are 0 and 1, the binary expansion of every number needs only the digits 0 and 1 and so is easily implemented and manipulated by the on-off digital circuits of a computer.

EXAMPLE 13

(a) The binary expansion of 39 is $(100111)_2$.
(b) $(110101)_2 = 32 + 16 + 4 + 1 = 53$. ■

Binary addition and multiplication are usually implemented in computer circuitry rather than by software. Another common base in computer science is base 16, or hexadecimal (or hex) representation. This representation requires six additional symbols to use with 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 to represent the digits usually written 10, 11, 12, 13, 14, 15. It is customary to choose the letters A, B, C, D, E, F for this purpose.

EXAMPLE 14 To use 26 as a base, we can use the letters A, B, ..., Z of the English alphabet to represent the digits 0, 1, ..., 25. In this way we can interpret any text as the base 26 representation of an integer. Thus we can interpret "TWO" as the base 26 representation of $(T \times 26^2) + (W \times 26) + O = (19 \times 676) + (22 \times 26) + 14 = 13430$. The ability to "add" words can be made the basis of cryptographic encoding. We will explore this further in Chapter 11. ■

EXAMPLE 15 As an example of cryptology, we consider a remarkable code due to Sir Francis Bacon, the English writer and philosopher. Suppose that we wish to encode a message, say FLEE NOW. We first write the base 2 representation of the position of each letter of the message in the English alphabet, starting with A in position 0, and ending with Z in position 25. Thus we have the following table.

F	L	E	E	N	O	W
00101	01011	00100	00100	01101	01110	10110

Since F is in position 5, its binary representation is $(00101)_2$, and so on. Now choose an unrelated "dummy" message exactly five times as long (padded with a few extra letters if necessary). Place the dummy message in a third row. The letters of the dummy message correspond exactly to the string of 0's and 1's in the second row of the above table. We agree to write each letter in one font (say Times Roman) if it corresponds to a 0 in the table, and another font (say Times Roman Italic) if it corresponds to a 1 (Bacon used fonts that were even more similar). Thus if the dummy message is ONCE UPON A TIME IN THE WEST THERE WAS A TOWN, we would write that message as ONCE *UPON A TIME* IN *THE* WEST *THERE WAS A TOWN*. Note that when the letters in this message are arranged in a third row of the table, the patterns of nonitalic for 0 and italic for 1 allow us to

decode the message.

F	L	E	E	N	O	W
00101	01011	00100	00100	01101	01110	10110
ONCE*U*	*PONAT*	IME*IN*	THE*WE*	*STTHE*	*REWAS*	*ATOWN*

■

EXAMPLE 16 Suppose that we wish to decode the following dummy message, using the Bacon code

NOW IS THE TIME FOR ALL GOOD MEN TO AID THE COUNTRY

Since there are 40 letters, the true message must have 8 letters. Arrange the dummy message, 5 letters at a time in the following table, and then list the corresponding binary digits, using 1 for italic and 0 for plain text.

NOWIS	*THETI*	MEFOR	*ALLGO*	*ODMEN*	*TOAID*	*THECO*	UNTRY
10010	10011	00000	10001	10011	01001	01110	00001

The binary representations in the second row correspond respectively to the numbers 18, 19, 0, 17, 19, 9, 14, and 1, and therefore represent the letters STARTJOB. Thus the decoded message is START JOB. ■

If the fonts used are quite similar, it will not be obvious that the text is a coded message. Thus this example also illustrates **steganography**, the science of concealment of coded information. Modern versions include hiding information in stray bits of a digital photograph. One use of this is to watermark copyrighted artistic material.

1.4 Exercises

In Exercises 1 through 4, for the given integers m and n, write m as qn + r, with $0 \le r < n$.

1. $m = 20, n = 3$
2. $m = 64, n = 37$
3. $m = 3, n = 22$
4. $m = 48, n = 12$
5. Write each integer as a product of powers of primes (as in Theorem 3).
   (a) 828 (b) 1666 (c) 1781
   (d) 1125 (e) 107

In Exercises 6 through 9, find the greatest common divisor d of the integers a and b, and write d as sa + tb.

6. $a = 60, b = 100$
7. $a = 45, b = 33$
8. $a = 34, b = 58$
9. $a = 77, b = 128$

In Exercises 10 through 13, find the least common multiple of the integers.

10. 72, 108
11. 150, 70
12. 175, 245
13. 32, 27
14. If f is the mod-7 function, compute each of the following.
   (a) $f(17)$ (b) $f(48)$ (c) $f(1207)$
   (d) $f(130)$ (e) $f(93)$ (f) $f(169)$
15. If f is the mod-11 function, compute each of the following.
   (a) $f(39)$ (b) $f(386)$ (c) $f(1232)$
   (d) $f(573)$ (e) $2f(87)$ (f) $f(175) + 4$
16. If f is the mod-7 function, compute each of the following.
   (a) $f(752 + 793)$ (b) $f(752) + f(793)$
   (c) $f(3 \cdot 1759)$ (d) $3 \cdot f(1759)$
17. If f is the mod-12 function, compute each of the following.
   (a) $f(1259 + 743)$ (b) $f(1259) + f(743)$
   (c) $f(2 \cdot 319)$ (d) $2 \cdot f(319)$
18. Let f be the mod-n function for a fixed n. What do the results of Exercises 16 and 17 suggest about the relationship between $k \cdot f(a)$ and $f(k \cdot a)$?
19. Let f be the mod-n function for a fixed n. Based on the results of Exercises 16 and 17, explain why $f(a+b)$ does not always equal $f(a) + f(b)$.
20. Let f be the mod-n function for a fixed n. Explain when $f(a + b) = f(a) + f(b)$ is true.

21. If g is the mod-5 function, solve each of the following.
   (a) $g(n) = 2$ (b) $g(n) = 4$
22. If g is the mod-6 function, solve each of the following.
   (a) $g(n) = 3$ (b) $g(n) = 1$
23. Complete the following proof. Let a and b be integers. If p is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$. We need to show that if $p \nmid a$, then p must divide b. If $p \nmid a$, then GCD$(a, p) = 1$, because _____. By Theorem 4, we can write $1 = sa + tp$ for some integers s and t. Then $b = sab + tpb$. (Why?) Then p must divide $sab + tpb$, because _____. So $p \mid b$. (Why?)
24. Show that if GCD$(a, c) = 1$ and $c \mid ab$, then $c \mid b$. (*Hint*: Model the proof on the one in Exercise 23.)
25. Show that if GCD$(a, c) = 1$, $a \mid m$, and $c \mid m$, then $ac \mid m$. (*Hint*: Use Exercise 24.)
26. Show that if $d = \text{GCD}(a, b)$, $a \mid b$, and $c \mid b$, then $ac \mid bd$.
27. Show that GCD$(ca, cb) = c\,\text{GCD}(a, b)$.
28. Show that LCM$(a, ab) = ab$.
29. Show that if GCD$(a, b) = 1$, then LCM$(a, b) = ab$.
30. Let $c = \text{LCM}(a, b)$. Show that if $a \mid k$ and $b \mid k$, then $c \mid k$.
31. Prove that if a and b are positive integers such that $a \mid b$ and $b \mid a$, then $a = b$.
32. Let a be an integer and let p be a positive integer. Prove that if $p \mid a$, then $p = \text{GCD}(a, p)$.
33. Theorem 2(c) says that if $a \mid b$ or $a \mid c$, then $a \mid bc$. Is the converse true: that is, if $a \mid bc$, then $a \mid b$ or $a \mid c$? Justify your conclusion.
34. Prove that if m and n are relatively prime and mn is a perfect square, then m and n are each perfect squares.
35. Is the statement in Exercise 34 true for cubes? For any fixed power? Justify your conclusion.

In Exercises 36 through 38, let $U = \{1, 2, 3, \ldots, 1689\}$, $A = \{x \mid x \in U \text{ and } 3 \mid x\}$, $B = \{y \mid y \in U \text{ and } 5 \mid y\}$, and $C = \{z \mid z \in U \text{ and } 11 \mid z\}$. Compute each of the following.

36. (a) $|A|$ (b) $|B|$ (c) $|C|$
37. (a) The number of elements in U that are divisible by 15
   (b) The number of elements of U that are divisible by 165
   (c) The number of elements of U that are divisible by 55
38. Use the results of Exercises 36 and 37 to compute each of the following.
   (a) $|A \cup B|$ (b) $|A \cup B \cup C|$
39. (a) Write the expansion in base 5 of each of the following numbers.
   (i) 29 (ii) 73 (iii) 215 (iv) 732
   (b) Write the expansion in base 10 of each of the following numbers.
   (i) $(144)_5$ (ii) $(320)_5$
   (iii) $(1242)_5$ (iv) $(11231)_5$
40. (a) Write the expansion in base 7 of each of the following numbers.
   (i) 29 (ii) 73 (iii) 215 (iv) 732
   (b) Write the expansion in base 10 of each of the following numbers.
   (i) $(102)_7$ (ii) $(161)_7$
   (iii) $(460)_7$ (iv) $(1613)_7$
41. For each of the following, write the expansion in the specified base.
   (i) 29 (ii) 73 (iii) 215 (iv) 732
   (a) 2 (b) 4 (c) 16
42. (a) How are the numbers 2, 4, and 16 related?
   (b) Because of the way 2, 4, and 16 are related, it is possible to change the expansion of a number relative to one of these numbers to the expansion relative to another directly without using the number's base 10 expansion. Examine the results of Exercise 41 (and other examples, if needed) and describe how to change from
   (i) a base 2 expansion to a base 4 expansion.
   (ii) a base 16 expansion to a base 2 expansion.
   (iii) a base 4 expansion to a base 16 expansion.
43. Use Bacon's code as given in Example 15
   (a) to create a dummy message for COME BACK
   (b) to decode *WHEN THE MOON COMES OVER THE MOUNTAIN, THE OWLS FLY HIGH.*
44. (a) For Bacon's code as given in Example 15, why should the dummy message be five times as long as the true message?
   (b) Modify Bacon's code so that it handles the spaces between words and the digits 0, 1, 2, . . . , 9.
45. If ONE and TWO are the base 26 representations of two integers, then what is the base 26 representation of the sum ONE + TWO?
46. Use Bacon's code, as given in Example 15, to decode *DO YOU KNOW THAT THE NUMBER PI IS NOW KNOWN TO MORE THAN FOUR HUNDRED MILLION DECIMAL PLACES.*

1.5 Matrices

A **matrix** is a rectangular array of numbers arranged in m horizontal **rows** and n vertical **columns**:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1}$$

The ***i*th row** of $\mathbf{A}$ is $\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$, $1 \le i \le m$, and the ***j*th column** of $\mathbf{A}$ is $\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$, $1 \le j \le n$. We say that $\mathbf{A}$ is ***m* by *n***, written $m \times n$. If $m = n$, we say $\mathbf{A}$ is a **square matrix** of order n and that the numbers $a_{11}, a_{22}, \ldots, a_{nn}$ form the **main diagonal** of $\mathbf{A}$. We refer to the number a_{ij}, which is in the ith row and jth column of $\mathbf{A}$ as the ***i*, *j*th element of A** or as the **(*i*, *j*) entry of A**, and we often write (1) as $\mathbf{A} = [\, a_{ij} \,]$.

EXAMPLE 1 Let

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 5 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & -1 & 3 & 4 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 2 & 3 \\ 2 & 4 & 5 \end{bmatrix}.$$

Then $\mathbf{A}$ is 2×3 with $a_{12} = 3$ and $a_{23} = 2$, $\mathbf{B}$ is 2×2 with $b_{21} = 4$, $\mathbf{C}$ is 1×4, $\mathbf{D}$ is 3×1, and $\mathbf{E}$ is 3×3. ■

A square matrix $\mathbf{A} = [\, a_{ij} \,]$ for which every entry off the main diagonal is zero, that is, $a_{ij} = 0$ for $i \neq j$, is called a **diagonal matrix**.

EXAMPLE 2 Each of the following is a diagonal matrix.

$$\mathbf{F} = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{and} \quad \mathbf{H} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$ ■

Matrices are used in many applications in computer science, and we shall see them in our study of relations and graphs. At this point we present the following simple application showing how matrices can be used to display data in a tabular form.

EXAMPLE 3 The following matrix gives the airline distances between the cities indicated.

	London	Madrid	New York	Tokyo
London	0	785	3469	5959
Madrid	785	0	3593	6706
New York	3469	3593	0	6757
Tokyo	5959	6706	6757	0

■

Two $m \times n$ matrices $\mathbf{A} = [\, a_{ij} \,]$ and $\mathbf{B} = [\, b_{ij} \,]$ are said to be **equal** if $a_{ij} = b_{ij}$, $1 \le i \le m$, $1 \le j \le n$; that is, if corresponding elements are the same. Notice how easy it is to state the definition using generic elements a_{ij}, b_{ij}.

EXAMPLE 4 If

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & -1 \\ 0 & 5 & 2 \\ 4 & -4 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & x & -1 \\ y & 5 & 2 \\ 4 & -4 & z \end{bmatrix},$$

then $\mathbf{A} = \mathbf{B}$ if and only if $x = -3$, $y = 0$, and $z = 6$. ■

If $\mathbf{A} = [\, a_{ij} \,]$ and $\mathbf{B} = [\, b_{ij} \,]$ are $m \times n$ matrices, then the **sum** of $\mathbf{A}$ and $\mathbf{B}$ is the matrix $\mathbf{C} = [\, c_{ij} \,]$ defined by $c_{ij} = a_{ij} + b_{ij}$, $1 \le i \le m$, $1 \le j \le n$. That is, $\mathbf{C}$ is obtained by adding the corresponding elements of $\mathbf{A}$ and $\mathbf{B}$. Once again the use of generic elements makes it easy to state the definition.

EXAMPLE 5 Let $\mathbf{A} = \begin{bmatrix} 3 & 4 & -1 \\ 5 & 0 & -2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 4 & 5 & 3 \\ 0 & -3 & 2 \end{bmatrix}$. Then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 3+4 & 4+5 & -1+3 \\ 5+0 & 0+(-3) & -2+2 \end{bmatrix} = \begin{bmatrix} 7 & 9 & 2 \\ 5 & -3 & 0 \end{bmatrix}.$$ ■

Observe that the sum of the matrices $\mathbf{A}$ and $\mathbf{B}$ is defined only when $\mathbf{A}$ and $\mathbf{B}$ have the same number of rows and the same number of columns. We agree to write $\mathbf{A} + \mathbf{B}$ only when the sum is defined.

A matrix all of whose entries are zero is called a **zero matrix** and is denoted by $\mathbf{0}$.

EXAMPLE 6 Each of the following is a zero matrix.

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$ ■

The following theorem gives some basic properties of matrix addition; the proofs are omitted.

Theorem 1
(a) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
(b) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
(c) $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$ ■

If $\mathbf{A} = [\, a_{ij} \,]$ is an $m \times p$ matrix and $\mathbf{B} = [\, b_{ij} \,]$ is a $p \times n$ matrix, then the **product** of $\mathbf{A}$ and $\mathbf{B}$, denoted $\mathbf{AB}$, is the $m \times n$ matrix $\mathbf{C} = [\, c_{ij} \,]$ defined by

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} \qquad 1 \le i \le m, 1 \le j \le n. \tag{2}$$

Let us explain (2) in more detail. The elements $a_{i1}, a_{i2}, \ldots, a_{ip}$ form the ith row of $\mathbf{A}$, and the elements $b_{1j}, b_{2j}, \ldots, b_{pj}$ form the jth column of $\mathbf{B}$. Then (2) states that for any i and j, the element c_{ij} of $\mathbf{C} = \mathbf{AB}$ can be computed in the following way, illustrated in Figure 1.20.

1. Select row i of $\mathbf{A}$ and column j of $\mathbf{B}$, and place them side by side.
2. Multiply corresponding entries and add all the products.

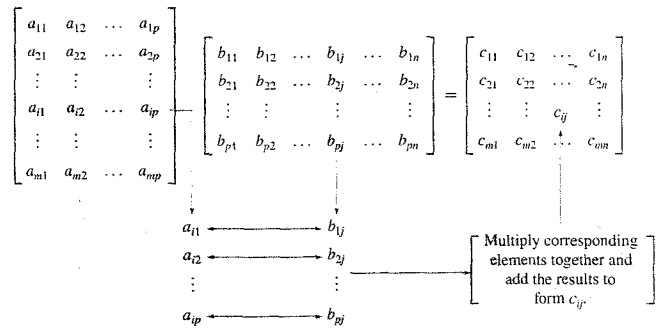


Figure 1.20

EXAMPLE 7 Let $\mathbf{A} = \begin{bmatrix} 2 & 3 & -4 \\ 1 & 2 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 1 \\ -2 & 2 \\ 5 & -3 \end{bmatrix}$. Then

$$\mathbf{AB} = \begin{bmatrix} (2)(3) + (3)(-2) + (-4)(5) & (2)(1) + (3)(2) + (-4)(-3) \\ (1)(3) + (2)(-2) + (3)(5) & (1)(1) + (2)(2) + (3)(-3) \end{bmatrix} = \begin{bmatrix} -20 & 20 \\ 14 & -4 \end{bmatrix}.$$

■

An **array of dimension two** is a modification of the idea of a matrix, in the same way that a linear array is a modification of the idea of a sequence. By an ***m* × *n* array A** we will mean an $m \times n$ matrix $\mathbf{A}$ of mn positions. We may assign numbers to these positions later, make further changes in these assignments, and still refer to the array as $\mathbf{A}$. This is a model for two-dimensional storage of information in a computer. The number assigned to row i and column j of an array $\mathbf{A}$ will be denoted $\mathbf{A}[i, j]$.

As we have seen, the properties of matrix addition resemble the familiar properties for the addition of real numbers. However, some of the properties of matrix multiplication do not resemble those of real number multiplication. First, observe that if $\mathbf{A}$ is an $m \times p$ matrix and $\mathbf{B}$ is a $p \times n$ matrix, then $\mathbf{AB}$ can be computed and is an $m \times n$ matrix. As for $\mathbf{BA}$, we have the following four possibilities:

1. $\mathbf{BA}$ may not be defined; we may have $n \neq m$.
2. $\mathbf{BA}$ may be defined and then $\mathbf{BA}$ is $p \times p$, while $\mathbf{AB}$ is $m \times m$ and $p \neq m$. Thus $\mathbf{AB}$ and $\mathbf{BA}$ are not equal.
3. $\mathbf{AB}$ and $\mathbf{BA}$ may both be the same size, but not be equal as matrices.
4. $\mathbf{AB} = \mathbf{BA}$.

We agree as before to write $\mathbf{AB}$ only when the product is defined.

EXAMPLE 8 Let $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & -2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & -1 \\ 2 & -3 \end{bmatrix}$. Then $\mathbf{AB} = \begin{bmatrix} 4 & -5 \\ -1 & 3 \end{bmatrix}$ and $\mathbf{BA} = \begin{bmatrix} -1 & 3 \\ -5 & 8 \end{bmatrix}$. ■

The basic properties of matrix multiplication are given by the following theorem.

Theorem 2
(a) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
(b) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
(c) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ ■

The $n \times n$ diagonal matrix

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

all of whose diagonal elements are 1, is called the **identity matrix** of order n. If $\mathbf{A}$ is an $m \times n$ matrix, it is easy to verify that $\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$. If $\mathbf{A}$ is an $n \times n$ matrix and p is a positive integer, we define

$$\mathbf{A}^p = \underbrace{\mathbf{A} \cdot \mathbf{A} \cdots \mathbf{A}}_{p \text{ factors}} \quad \text{and} \quad \mathbf{A}^0 = \mathbf{I}_n.$$

If p and q are nonnegative integers, we can prove the following laws of exponents for matrices:

$$\mathbf{A}^p\mathbf{A}^q = \mathbf{A}^{p+q} \quad \text{and} \quad (\mathbf{A}^p)^q = \mathbf{A}^{pq}.$$

Observe that the rule $(\mathbf{AB})^p = \mathbf{A}^p\mathbf{B}^p$ does *not* hold for all square matrices. However, if $\mathbf{AB} = \mathbf{BA}$, then $(\mathbf{AB})^p = \mathbf{A}^p\mathbf{B}^p$.

If $\mathbf{A} = [a_{ij}]$ is an $m \times n$ matrix, then the $n \times m$ matrix $\mathbf{A}^T = [a_{ij}^T]$, where $a_{ij}^T = a_{ji}$, $1 \leq i \leq m$, $1 \leq j \leq n$, is called the **transpose of A**. Thus the transpose of $\mathbf{A}$ is obtained by interchanging the rows and columns of $\mathbf{A}$.

EXAMPLE 9 Let $\mathbf{A} = \begin{bmatrix} 2 & -3 & 5 \\ 6 & 1 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 4 & 5 \\ 2 & -1 & 0 \\ 1 & 6 & -2 \end{bmatrix}$. Then

$$\mathbf{A}^T = \begin{bmatrix} 2 & 6 \\ -3 & 1 \\ 5 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B}^T = \begin{bmatrix} 3 & 2 & 1 \\ 4 & -1 & 6 \\ 5 & 0 & -2 \end{bmatrix}.$$

■

The following theorem summarizes the basic properties of the transpose operation.

Theorem 3 If $\mathbf{A}$ and $\mathbf{B}$ are matrices, then
(a) $(\mathbf{A}^T)^T = \mathbf{A}$
(b) $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
(c) $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ ■

A matrix $\mathbf{A} = [a_{ij}]$ is called **symmetric** if $\mathbf{A}^T = \mathbf{A}$. Thus, if $\mathbf{A}$ is symmetric, it must be a square matrix. It is easy to show that $\mathbf{A}$ is symmetric if and only if $a_{ij} = a_{ji}$. That is, $\mathbf{A}$ is symmetric if and only if the entries of $\mathbf{A}$ are symmetric with respect to the main diagonal of $\mathbf{A}$.

EXAMPLE 10 If $\mathbf{A} = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & 5 \\ -3 & 5 & 6 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & 0 \\ 3 & 2 & 1 \end{bmatrix}$, then $\mathbf{A}$ is symmetric and $\mathbf{B}$ is not symmetric. ■

If x is a nonzero number, there is a number y such that $xy = 1$. The number y is called the **multiplicative inverse** of x. When we have a multiplication for objects other than numbers, it is reasonable to ask if a multiplicative inverse exists. One example is matrix multiplication, described in this section.

If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, then we say that $\mathbf{B}$ is an ***inverse*** of $\mathbf{A}$ if $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix defined earlier. The identity matrix behaves like the number 1, in that $\mathbf{I}_n\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$ for any $n \times n$ matrix $\mathbf{A}$, so the matrix inverse is analogous to the reciprocal of a nonzero number. However, it is not clear how to construct inverses, or even when they exist.

EXAMPLE 11 An inverse of the matrix $\begin{bmatrix} 1 & 3 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ is the matrix $\begin{bmatrix} -2 & 3 & -3 \\ 1 & -1 & 1 \\ 2 & -3 & 4 \end{bmatrix}$. This can be verified by checking that

$$\begin{bmatrix} 1 & 3 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 & 3 & -3 \\ 1 & -1 & 1 \\ 2 & -3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} -2 & 3 & -3 \\ 1 & -1 & 1 \\ 2 & -3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 3 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

■

There are tests to determine if an $n \times n$ matrix has an inverse, and formulas for computing an inverse if it exists. Many of these are programmed into calculators and computer algebra systems. We will be content to give the results for the case of 2×2 matrices.

Suppose that the 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has an inverse $\mathbf{B} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$. Then

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and we have two pairs of equations

$$\begin{cases} ae + bg = 1 \\ ce + dg = 0 \end{cases} \quad \text{and} \quad \begin{cases} af + bh = 0 \\ cf + dh = 1. \end{cases}$$

When we solve the first pair for e and g, we find that we must divide by $ad - bc$. This can only be done if $ad - bc \neq 0$. The same condition is needed when solving the second pair for f and h. The results of solving for e, f, g, and h are summarized in Theorem 4.

Theorem 4 A matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has an inverse if and only if $ad - bc \neq 0$. In this case we have

$$\mathbf{A}^{-1} = \begin{bmatrix} \dfrac{d}{ad-bc} & -\dfrac{b}{ad-bc} \\ -\dfrac{c}{ad-bc} & \dfrac{a}{ad-bc} \end{bmatrix}.$$

■

■ Boolean Matrix Operations

A **Boolean matrix** (also called a **bit matrix**) is an $m \times n$ matrix whose entries are either zero or one. We shall now define three operations on Boolean matrices that have useful applications in Chapter 4.

Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $m \times n$ Boolean matrices. We define $\mathbf{A} \vee \mathbf{B} = \mathbf{C} = [c_{ij}]$, the **join** of $\mathbf{A}$ and $\mathbf{B}$, by

$$c_{ij} = \begin{cases} 1 & \text{if } a_{ij} = 1 \text{ or } b_{ij} = 1 \\ 0 & \text{if } a_{ij} \text{ and } b_{ij} \text{ are both } 0 \end{cases}$$

and $\mathbf{A} \wedge \mathbf{B} = \mathbf{D} = [d_{ij}]$, the **meet** of $\mathbf{A}$ and $\mathbf{B}$, by

$$d_{ij} = \begin{cases} 1 & \text{if } a_{ij} \text{ and } b_{ij} \text{ are both } 1 \\ 0 & \text{if } a_{ij} = 0 \text{ or } b_{ij} = 0. \end{cases}$$

Note that these operations are only possible when $\mathbf{A}$ and $\mathbf{B}$ have the same size, just as in the case of matrix addition. Instead of adding corresponding elements in $\mathbf{A}$ and $\mathbf{B}$, to compute the entries of the result, we simply examine the corresponding elements for particular patterns.

EXAMPLE 12 Let $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

(a) Compute $\mathbf{A} \vee \mathbf{B}$. (b) Compute $\mathbf{A} \wedge \mathbf{B}$.

Solution

(a) Let $\mathbf{A} \vee \mathbf{B} = [c_{ij}]$. Then, since a_{43} and b_{43} are both 0, we see that $c_{43} = 0$. In all other cases, either a_{ij} or b_{ij} is 1, so c_{ij} is also 1. Thus

$$\mathbf{A} \vee \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

(b) Let $\mathbf{A} \wedge \mathbf{B} = [d_{ij}]$. Then, since a_{11} and b_{11} are both 1, $d_{11} = 1$, and since a_{23} and b_{23} are both 1, $d_{23} = 1$. In all other cases, either a_{ij} or b_{ij} is 0, so $d_{ij} = 0$. Thus

$$A \wedge \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

■

Finally, suppose that $\mathbf{A} = [a_{ij}]$ is an $m \times p$ Boolean matrix and $\mathbf{B} = [b_{ij}]$ is a $p \times n$ Boolean matrix. Notice that the condition on the sizes of $\mathbf{A}$ and $\mathbf{B}$ is exactly the condition needed to form the matrix product $\mathbf{AB}$. We now define another kind of product.

The **Boolean product** of $\mathbf{A}$ and $\mathbf{B}$, denoted $\mathbf{A} \odot \mathbf{B}$, is the $m \times n$ Boolean matrix $\mathbf{C} = [c_{ij}]$ defined by

$$c_{ij} = \begin{cases} 1 & \text{if } a_{ik} = 1 \text{ and } b_{kj} = 1 \text{ for some } k,\ 1 \leq k \leq p \\ 0 & \text{otherwise.} \end{cases}$$

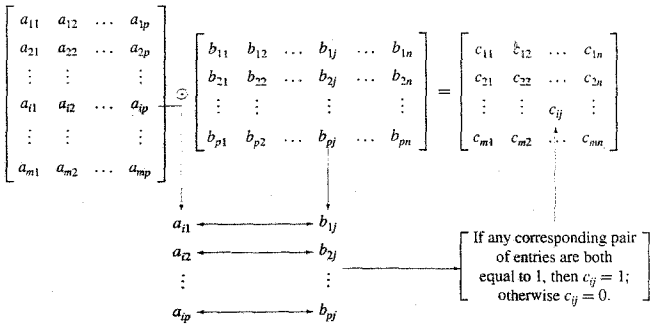


Figure 1.21

This multiplication is similar to ordinary matrix multiplication. The preceding formula states that for any i and j the element c_{ij} of $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ can be computed in the following way, as illustrated in Figure 1.21. (Compare this with Figure 1.20.)

1. Select row i of $\mathbf{A}$ and column j of $\mathbf{B}$, and arrange them side by side.
2. Compare corresponding entries. If even a single pair of corresponding entries consists of two 1's, then $c_{ij} = 1$. If this is not the case, then $c_{ij} = 0$.

We can easily perform the indicated comparisons and checks for each position of the Boolean product. Thus, at least for human beings, the computation of elements in $\mathbf{A} \odot \mathbf{B}$ is considerably easier than the computation of elements in $\mathbf{AB}$.

EXAMPLE 13 Let $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}$. Compute $\mathbf{A} \odot \mathbf{B}$.

Solution

Let $\mathbf{A} \odot \mathbf{B} = [e_{ij}]$. Then $e_{11} = 1$, since row 1 of $\mathbf{A}$ and column 1 of $\mathbf{B}$ each have a 1 as the first entry. Similarly, $e_{12} = 1$, since $a_{12} = 1$ and $b_{22} = 1$; that is, the first row of $\mathbf{A}$ and the second column of $\mathbf{B}$ have a 1 in the second position. In a similar way we see that $e_{13} = 1$. On the other hand, $e_{14} = 0$, since row 1 of $\mathbf{A}$ and column 4 of $\mathbf{B}$ do not have common 1's in any position. Proceeding in this way, we obtain

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

■

The following theorem, whose proof is left as an exercise, summarizes the basic properties of the Boolean matrix operations just defined.

Theorem 5 If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are Boolean matrices of compatible sizes, then

1. (a) $\mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A}$
   (b) $\mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}$

2. (a) $(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C} = \mathbf{A} \vee (\mathbf{B} \vee \mathbf{C})$
   (b) $(\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C} = \mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C})$
3. (a) $\mathbf{A} \wedge (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \vee (\mathbf{A} \wedge \mathbf{C})$
   (b) $\mathbf{A} \vee (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \vee \mathbf{B}) \wedge (\mathbf{A} \vee \mathbf{C})$
4. $(\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$

■

1.5 Exercises

1. Let $\mathbf{A} = \begin{bmatrix} 3 & -2 & 5 \\ 4 & 1 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}$, and $\mathbf{C} = \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & -1 \\ 2 & 0 & 8 \end{bmatrix}$.
   (a) What is a_{12}, a_{22}, a_{23}?
   (b) What is b_{11}, b_{31}?
   (c) What is c_{13}, c_{23}, c_{33}?
   (d) List the elements on the main diagonal of C.
2. Which of the following are diagonal matrices?
   (a) $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 0 & 0 \end{bmatrix}$ (b) $\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$
   (c) $\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
   (d) $\mathbf{D} = \begin{bmatrix} 2 & 6 & -2 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$
   (e) $\mathbf{E} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$
3. If $\begin{bmatrix} a+b & c+d \\ c-d & a-b \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 10 & 2 \end{bmatrix}$, find a, b, c, and d.
4. If $\begin{bmatrix} a+2b & 2a-b \\ 2c+d & c-2d \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 4 & -3 \end{bmatrix}$, find a, b, c, and d.

In Exercises 5 through 10, let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 1 & -2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 2 & 5 \\ 3 & 1 & 2 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} -3 & 2 \\ 4 & 1 \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} 3 & 2 & -1 \\ 5 & 4 & -3 \\ 0 & 1 & 2 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} -2 & 3 \\ 4 & 5 \end{bmatrix}.$$

5. If possible, compute each of the following.
   (a) $\mathbf{C} + \mathbf{E}$ (b) $\mathbf{AB}$
   (c) $\mathbf{CB} + \mathbf{F}$ (d) $\mathbf{AB} + \mathbf{DF}$
6. If possible, compute each of the following.
   (a) $\mathbf{A}(\mathbf{BD})$ and $(\mathbf{AB})\mathbf{D}$
   (b) $\mathbf{A}(\mathbf{C} + \mathbf{E})$ and $\mathbf{AC} + \mathbf{AE}$
   (c) $\mathbf{FD} + \mathbf{AB}$
7. If possible, compute each of the following.
   (a) $\mathbf{EB} + \mathbf{FA}$
   (b) $\mathbf{A}(\mathbf{B} + \mathbf{D})$ and $\mathbf{AB} + \mathbf{AD}$
   (c) $(\mathbf{F} + \mathbf{D})\mathbf{A}$ (d) $\mathbf{AC} + \mathbf{DE}$
8. If possible, compute each of the following.
   (a) $\mathbf{A}^T$ and $(\mathbf{A}^T)^T$
   (b) $(\mathbf{C} + \mathbf{E})^T$ and $\mathbf{C}^T + \mathbf{E}^T$
   (c) $(\mathbf{AB})^T$ and $\mathbf{B}^T\mathbf{A}^T$ (d) $(\mathbf{B}^T\mathbf{C}) + \mathbf{A}$
9. If possible, compute each of the following.
   (a) $\mathbf{A}^T(\mathbf{D} + \mathbf{F})$ (b) $(\mathbf{BC})^T$ and $\mathbf{C}^T\mathbf{B}^T$
   (c) $(\mathbf{B}^T + \mathbf{A})\mathbf{C}$ (d) $(\mathbf{D}^T + \mathbf{E})\mathbf{F}$
10. Compute $\mathbf{D}^3$.
11. Let $\mathbf{A}$ be an $m \times n$ matrix. Show that $\mathbf{I}_m\mathbf{A} = \mathbf{AI}_n = \mathbf{A}$. (*Hint*: Choose a generic element of $\mathbf{I}_m\mathbf{A}$.)
12. Let $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & -2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix}$. Show that $\mathbf{AB} \neq \mathbf{BA}$.
13. Let $\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$.
   (a) Compute $\mathbf{A}^3$. (b) What is $\mathbf{A}^k$?
14. Show that $\mathbf{A0} = \mathbf{0}$ for any matrix $\mathbf{A}$.
15. Show that $\mathbf{I}_n^T = \mathbf{I}_n$.
16. (a) Show that if $\mathbf{A}$ has a row of zeros, then $\mathbf{AB}$ has a corresponding row of zeros. (*Hint*: Use the generic element definition of $\mathbf{AB}$ given in this section.)
   (b) Show that if $\mathbf{B}$ has a column of zeros, then $\mathbf{AB}$ has a corresponding column of zeros.
17. Show that the jth column of the matrix product $\mathbf{AB}$ is equal to the matrix product $\mathbf{AB}_j$, where $\mathbf{B}_j$ is the jth column of $\mathbf{B}$.
18. If $\mathbf{0}$ is the 2×2 zero matrix, find two 2×2 matrices $\mathbf{A}$ and $\mathbf{B}$, with $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$, such that $\mathbf{AB} = \mathbf{0}$.
19. If $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, show that $\mathbf{A}^2 = \mathbf{I}_2$.

20. Determine all 2×2 matrices $\mathbf{A} = \begin{bmatrix} 0 & a \\ b & c \end{bmatrix}$ such that $\mathbf{A}^2 = \mathbf{I}_2$.

21. Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices.

(a) Show that $\mathbf{A} + \mathbf{B}$ is also symmetric.

(b) Is $\mathbf{AB}$ also symmetric?

22. Let $\mathbf{A}$ be an $n \times n$ matrix.

(a) Show that $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric.

(b) Show that $\mathbf{A} + \mathbf{A}^T$ is symmetric.

23. Prove Theorem 3. [*Hint*: For part (c), show that the i, jth element of $(\mathbf{AB})^T$ equals the i, jth element of $\mathbf{B}^T\mathbf{A}^T$.]

24. Let $\mathbf{A}$ be a symmetric 2×2 matrix that has an inverse. Must $\mathbf{A}^{-1}$ also be symmetric? Explain your reasoning.

25. Find the inverse of each matrix.

(a) $\begin{bmatrix} 2 & 1 \\ 5 & 7 \end{bmatrix}$ **(b)** $\begin{bmatrix} -3 & 4 \\ 0 & 9 \end{bmatrix}$

(c) $\begin{bmatrix} 6 & 5 \\ 4 & -2 \end{bmatrix}$

26. Find the inverse of each matrix.

(a) $\begin{bmatrix} -8 & 3 \\ 9 & -2 \end{bmatrix}$ **(b)** $\begin{bmatrix} 3 & 4 \\ -1 & 10 \end{bmatrix}$

(c) $\begin{bmatrix} -2 & -9 \\ -6 & -4 \end{bmatrix}$

For Exercises 27 and 28, let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.2 & 0.4 & 0.2 \\ -0.4 & 0.2 & 0.6 \\ 0.2 & -0.6 & 0.2 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \textit{and} \quad \mathbf{D} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}.$$

27. **(a)** Verify that $\mathbf{C} = \mathbf{A}^{-1}$.

(b) Verify that $\mathbf{D} = \mathbf{B}^{-1}$.

28. Determine whether $\mathbf{CD}$ is the inverse of $\mathbf{AB}$. Explain your reasoning.

29. Show that if $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices and $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ both exist, then $(\mathbf{AB})^{-1} = (\mathbf{B}^{-1}\mathbf{A}^{-1})$.

In Exercises 30 and 31, compute $\mathbf{A} \vee \mathbf{B}$, $\mathbf{A} \wedge \mathbf{B}$, *and* $\mathbf{A} \odot \mathbf{B}$ *for the given matrices* $\mathbf{A}$ *and* $\mathbf{B}$.

30. **(a)** $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

(b) $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$

(c) $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$

31. **(a)** $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

(b) $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

(c) $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

32. Complete the following proofs.

(a) $\mathbf{A} \vee \mathbf{A} = \mathbf{A}$. Proof: Let b_{ij} be an element of $\mathbf{A} \vee \mathbf{A}$. If $b_{ij} = 0$, then $a_{ij} =$ ____, because ____. If $b_{ij} = 1$, then $a_{ij} =$ ____ because ____. Hence $b_{ij} = a_{ij}$ for each i, j pair.

(b) $\mathbf{A} \wedge \mathbf{A} = \mathbf{A}$. Proof: Let b_{ij} be an element of $\mathbf{A} \wedge \mathbf{A}$. If $b_{ij} = 0$, then ____. If $b_{ij} = 1$, then ____. (Explain.) Hence $b_{ij} = a_{ij}$ for each i, j pair.

33. Show that $\mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A}$.

34. Show that $\mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}$.

35. Show that $\mathbf{A} \vee (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$.

36. Show that $\mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C}$.

37. Show that $\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$.

38. Show that $A \wedge (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \vee (\mathbf{A} \wedge \mathbf{C})$.

39. Show that $\mathbf{A} \vee (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \vee \mathbf{B}) \wedge (\mathbf{A} \vee \mathbf{C})$.

40. What fact does Example 8 illustrate?

41. Let $\mathbf{A} = [\,a_{ij}\,]$ and $\mathbf{B} = [\,b_{ij}\,]$ be two $n \times n$ matrices and let $\mathbf{C} = [\,c_{ij}\,]$ represent $\mathbf{AB}$. Prove that if k is an integer and $k \mid a_{ij}$ for all i, j, then $k \mid c_{ij}$ for all i, j.

42. Let p be a prime number with $p > 2$, and let $\mathbf{A}$ and $\mathbf{B}$ be matrices all of whose entries are integers. Suppose that p divides all the entries of $\mathbf{A}+\mathbf{B}$ and all the entries of $\mathbf{A}-\mathbf{B}$. Prove that p divides all the entries of $\mathbf{A}$ and all the entries of $\mathbf{B}$.

Another operation on matrices is ***scalar multiplication****. Let* k *be a real number and* $\mathbf{A} = [\,a_{ij}\,]$ *be an* $m \times n$ *matrix. The result of multiplying* $\mathbf{A}$ *by the scalar* k *is the matrix* $k\mathbf{A} = [\,ka_{ij}\,]$. *For Exercises 43 through 47, use the definition of scalar multiplication and the matrices given.*

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & -1 \\ 0 & 5 & 2 \\ 4 & -4 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 4 & -6 \\ 4 & 0 & 9 \\ 7 & -1 & 3 \end{bmatrix},$$

and

$$\mathbf{C} = \begin{bmatrix} 4 & 0 \\ 3 & 1 \\ -2 & 5 \end{bmatrix}$$

43. Compute each of the following.

(a) $3\mathbf{A}$ **(b)** $5\mathbf{B}$ **(c)** $-1\mathbf{C}$

44. Compute each of the following.

(a) $3(\mathbf{A} + \mathbf{B})$ **(b)** $3\mathbf{A} + 3\mathbf{B}$

(c) $-2(\mathbf{AC})$ **(d)** $\mathbf{A}(-2\mathbf{C})$

45. Show that scalar multiplication has the following property.

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$

46. Show that scalar multiplication has the following property.

$$k(\mathbf{AB}) = (k\mathbf{A})\mathbf{B} = \mathbf{A}(k\mathbf{B})$$

47. Let $\mathbf{A}$ be an $m \times n$ matrix. Find a matrix $\mathbf{K}$ such that $\mathbf{KA} = k\mathbf{A}$, for a fixed k.

1.6 Mathematical Structures

Several times in this chapter, we have defined a new kind of mathematical object; for example, a set or a matrix. Then notation was introduced for representing the new type of object and a way to determine whether two objects are the same was described. Next we classified objects of the new type; for example, finite or infinite for sets, and Boolean or symmetric for matrices. And then operations were defined for the objects and the properties of these operations were examined.

Such a collection of objects with operations defined on them and the accompanying properties form a **mathematical structure** or **system**. In this book we deal only with discrete mathematical structures.

EXAMPLE 1 The collection of sets with the operations of union, intersection, and complement and their accompanying properties is a (discrete) mathematical structure. We denote this structure by $(\text{sets}, \cup, \cap, \bar{\ })$. ■

EXAMPLE 2 The collection of 3×3 matrices with the operations of addition, multiplication, and transpose is a mathematical structure denoted by $(3 \times 3 \text{ matrices}, +, *, {}^T)$. ■

An important property we have not identified before is closure. A structure is **closed with respect to** an operation if that operation always produces another member of the collection of objects.

EXAMPLE 3 The structure $(5 \times 5 \text{ matrices}, +, *, {}^T)$ is closed with respect to addition because the sum of two 5×5 matrices is another 5×5 matrix. ■

EXAMPLE 4 The structure (odd integers, $+$, $*$) is not closed with respect to addition. The sum of two odd integers is an even integer. This structure does have the closure property for multiplication, since the product of two odd numbers is an odd number. ■

An operation that combines two objects is a **binary operation**. An operation that requires only one object is a **unary operation**. Binary operations often have similar properties, as we have seen earlier.

EXAMPLE 5

(a) Set intersection is a binary operation since it combines two sets to produce a new set.

(b) Producing the transpose of a matrix is a unary operation. ■

Common properties have been given names. For example, if the order of the objects does not affect the outcome of a binary operation, we say that the operation is **commutative**. That is, if $x \square y = y \square x$, where $\square$ is some binary operation, $\square$ is commutative.

EXAMPLE 6

(a) Join and meet for Boolean matrices are commutative operations.

$$\mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A} \quad \text{and} \quad \mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}.$$

(b) Ordinary matrix multiplication is not a commutative operation. $\mathbf{AB} \neq \mathbf{BA}$. ■

Note that when we say an operation has a property, this means that the statement of the property is true when the operation is used with any objects in the structure. If there is even one case when the statement is not true, the operation does not have that property. If $\square$ is a binary operation, then $\square$ is **associative** or **has the associative property** if

$$(x \square y) \square z = x \square (y \square z).$$

EXAMPLE 7 Set union is an associative operation, since $(A \cup B) \cup C = A \cup (B \cup C)$ is always true. ■

If a mathematical structure has two binary operations, say $\square$ and $\triangledown$, a **distributive property** has the following pattern:

$$x \square (y \triangledown z) = (x \square y) \triangledown (x \square z).$$

We say that "$\square$ distributes over $\triangledown$."

EXAMPLE 8

(a) We are familiar with the distributive property for real numbers; if a, b, and c are real numbers, then $a \cdot (b + c) = a \cdot b + a \cdot c$. Note that because we have an agreement about real number arithmetic to multiply before adding, parentheses are not needed on the right-hand side.

(b) The structure (sets, $\cup$, $\cap$, $^-$) has two distributive properties:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

and

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$ ■

Several of the structures we have seen have a unary operation and two binary operations. For such structures we can ask whether De Morgan's laws are properties of the system. If the unary operation is $\circ$ and the binary operations are $\square$ and $\triangledown$, then **De Morgan's laws** are

$$(x \square y)^\circ = x^\circ \triangledown y^\circ \quad \text{and} \quad (x \triangledown y)^\circ = x^\circ \square y^\circ.$$

EXAMPLE 9

(a) As we saw in Section 1.2, sets satisfy De Morgan's laws for union, intersection, and complement: $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$ and $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.

(b) The structure (real numbers, $+$, $*$, $\sqrt{\ }$) does not satisfy De Morgan's laws, since $\sqrt{x + y} \neq \sqrt{x} * \sqrt{y}$. ■

A structure with a binary operation $\square$ may contain a distinguished object e, with the property $x \square e = e \square x = x$ for all x in the collection. We call e an **identity for** $\square$. In fact, an identity for an operation must be unique.

Theorem 1 If e is an identity for a binary operation $\square$, then e is unique.

Proof

Assume another object i also has the identity property, so $x \square i = i \square x = x$. Then $e \square i = e$, but since e is an identity for $\square$, $i \square e = e \square i = i$. Thus, $i = e$. Therefore there is at most one object with the identity property for $\square$. ■

This is one of our first examples of a proof that does not proceed directly. We assumed that there were two identity elements and showed that they were in fact the same element.

EXAMPLE 10 For ($n \times n$ matrices, $+$, $*$, ^{T}), $\mathbf{I}_n$ is the identity for matrix multiplication and the $n \times n$ zero matrix is the identity for matrix addition. ■

If a binary operation $\square$ has an identity e, we say y is a **$\square$-inverse** of x if $x \square y = y \square x = e$.

Theorem 2 If $\square$ is an associative operation and x has a $\square$-inverse y, then y is unique.

Proof

Assume there is another $\square$-inverse for x, say z. Then $(z \square x) \square y = e \square y = y$ and $z \square (x \square y) = z \square e = z$. Since $\square$ is associative, $(z \square x) \square y = z \square (x \square y)$ and so $y = z$. ■

EXAMPLE 11

(a) In the structure (3×3 matrices, $+$, $*$, ^{T}), each matrix $\mathbf{A} = [a_{ij}]$ has a $+$-inverse, or additive inverse, $-\mathbf{A} = [-a_{ij}]$.

(b) In the structure (integers, $+$, $*$), only the integers 1 and -1 have multiplicative inverses. ■

EXAMPLE 12 Let $\square$, $\triangledown$, and $\circ$ be defined for the set $\{0, 1\}$ by the following tables.

$\square$	0	1
0	0	1
1	1	0

$\triangledown$	0	1
0	0	0
1	0	1

x°	x
0	1
1	0

Thus $1 \square 0 = 1$, $0 \triangledown 1 = 0$, and $1^\circ = 0$. Determine if each of the following is true for $(\{0, 1\}, \square, \triangledown, \circ)$.

(a) $\square$ is commutative. **(b)** $\triangledown$ is associative.

(c) De Morgan's laws hold.

(d) Two distributive properties hold for the structure.

Solution

(a) The statement $x \square y = y \square x$ must be true for all choices of x and y. Here there is only one case to check: Is $0 \square 1 = 1 \square 0$ true? Since both $0 \square 1$ and $1 \square 0$ are 1, $\square$ is commutative.

(b) The eight possible cases to be checked are left as an exercise. See Exercise 6(b).

(c) $(0 \square 0)^\circ = 0^\circ = 1$ $\quad 0^\circ \triangledown 0^\circ = 1 \triangledown 1 = 1$
$(0 \square 1)^\circ = 1^\circ = 0$ $\quad 0^\circ \triangledown 1^\circ = 1 \triangledown 0 = 0$
$(1 \square 1)^\circ = 0^\circ = 1$ $\quad 1^\circ \triangledown 1^\circ = 0 \triangledown 0 = 0$

The last pair shows that De Morgan's laws do not hold in this structure.

(d) One possible distributive property is $x \square (y \triangledown z) = (x \square y) \triangledown (x \square z)$. We must check all possible cases. One way to organize this is shown in a table.

x	y	z	$y \triangledown z$	$x \square (y \triangledown z)$	$x \square y$	$x \square z$	$(x \square y) \triangledown (x \square z)$
0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0
0	1	0	0	0	1	0	0
0	1	1	1	1	1	1	1
1	0	0	0	1	1	1	1
1	0	1	0	1	1	0	0
1	1	0	0	1	0	1	0
1	1	1	1	0	0	0	0
				(A)			(B)

Since columns (A) and (B) are not identical, this possible distributive property does not hold in this structure. The check for the other distributive property is Exercise 7. ■

In later sections, we will find it useful to consider mathematical structures themselves as objects and to classify them according to the properties associated with their operations.

1.6 Exercises

In Exercises 1 and 2, tell whether the structure has the closure property with respect to the operation.

1. **(a)** (sets, $\cup$, $\cap$, $^-$) union
   (b) (sets, $\cup$, $\cap$, $^-$) complement
2. **(a)** (4 × 4 matrices, $+$, $*$, ^{T}) multiplication
   (b) (3 × 5 matrices, $+$, $*$, ^{T}) transpose

In Exercises 3 and 4, tell whether the structure has the closure property with respect to the operation.

3. **(a)** (integers, $+$, $-$, $*$, $\div$) division
   (b) (A^*, catenation) catenation
4. **(a)** ($n \times n$ Boolean matrices, $\vee$, $\wedge$, ^{T}) meet
   (b) (prime numbers, $+$, $*$) addition
5. Show that $\oplus$ is a commutative operation for sets.
6. Using the definitions in Example 12, (a) show that $\square$ is associative. (b) Show that $\triangledown$ is associative.
7. Using the definitions in Example 12, determine if the other possible distributive property holds.
8. Give the identity element, if one exists, for each binary operation in the given structure.
   (a) (real numbers, $+$, $*$, $\sqrt{\ }$)
   (b) (sets, $\cup$, $\cap$, $^-$)
   (c) ($\{0, 1\}$, $\square$, $\triangledown$, $*$) as defined in Example 12
   (d) (subsets of a finite set A, $\oplus$, $^-$)
9. Give the identity element, if one exists, for each binary operation in the structure (5 × 5 Boolean matrices, $\vee$, $\wedge$, $\odot$).

In Exercises 10 through 16, use the structure S = ($n \times n$ diagonal matrices, $+$, $*$, ^{T}).

10. Show that S is closed with respect to addition.
11. Show that S is closed with respect to multiplication.
12. Show that S is closed with respect to the transpose operation.
13. Does S have an identity for addition? If so, what is it?
14. Does S have an identity for multiplication? If so, what is it?
15. Let $\mathbf{A}$ be an $n \times n$ diagonal matrix. Describe the additive inverse of $\mathbf{A}$.
16. Let $\mathbf{A}$ be an $n \times n$ diagonal matrix. Describe the multiplicative inverse of $\mathbf{A}$.

In Exercises 17 through 23, use the structure R = (M, $+$, $*$, ^{T}), where M is the set of matrices of the form $\begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix}$, where a is a real number.

17. Show that R is closed with respect to addition.
18. Show that R is closed with respect to multiplication.
19. Show that R is closed with respect to the transpose operation.
20. Does R have an identity for addition? If so, what is it?
21. Does R have an identity for multiplication? If so, what is it?
22. Let $\mathbf{A}$ be an element of M. Describe the additive inverse for $\mathbf{A}$.
23. Let $\mathbf{A}$ be an element of M. Describe the multiplicative inverse for $\mathbf{A}$.

In Exercises 24 through 28, let R = ($\mathbb{Q}$, $\square$), where $x \square y = \frac{x+y}{2}$. Determine which of the following properties hold for this structure:

24. Closure
25. Commutative
26. Associative
27. An identity element
28. An inverse for every element
29. Let R = (2 × 1 matrices, $\triangledown$), where

$$\begin{bmatrix} x \\ y \end{bmatrix} \triangledown \begin{bmatrix} w \\ z \end{bmatrix} = \begin{bmatrix} x+w \\ y+z+1 \end{bmatrix}.$$

Determine which of the following properties hold for this structure.
(a) Closure **(b)** Commutative
(c) Associative

30. Let R be as in Exercise 29. Determine which of the following properties hold for this structure.
(a) An identity element
(b) An inverse for every element
31. Let S = (1 × 2 matrices, $\square$), where $\begin{bmatrix} x & y \end{bmatrix} \square \begin{bmatrix} w & z \end{bmatrix} = \begin{bmatrix} x+w & \frac{y+z}{2} \end{bmatrix}$. Determine which of the following properties hold for this structure.
(a) Closure **(b)** Commutative
(c) Associative
32. Let S be as in Exercise 31. Determine which of the following properties hold for this structure.
(a) An identity element
(b) An inverse for every element
33. **(a)** Give a symbolic statement of the distributive property for scalar multiplication over $\triangledown$ as defined in Exercise 29.
(b) Is the distributive property in part (a) a property of R?
34. **(a)** Give a symbolic statement of the distributive property for scalar multiplication over $\square$ as defined in Exercise 31.
(b) Is the distributive property in part (a) a property of S?
35. For a Boolean matrix $\mathbf{B}$, we define comp $\mathbf{B}$ to be the matrix formed by changing each 0 entry of $\mathbf{B}$ to 1 and each 1 entry of $\mathbf{B}$ to 0. Let R = (5 × 5 Boolean matrices, $\wedge$, $\vee$, comp). Do De Morgan's laws hold for R? Justify your answer.

The properties of a mathematical structure can be used to rewrite expressions just as is done in ordinary algebra. In Exercises 36 through 39, rewrite the given expression to produce the requested result.

36. $(A \cup B) \cap (A \cup \overline{B})$ one set, no operations
37. $\overline{(A \cap B)} \cap A$ two sets, two operations
38. $\overline{(A \cup B)} \cup (\overline{A} \cap \overline{B})$ two sets, two operations
39. $\overline{(A \cup \overline{B})} \cap (\overline{A} \cup B)$ one set, no operations

Tips for Proofs

Many exercises in this chapter ask that you show, prove, or verify a statement. To show or prove a statement means to give a written explanation demonstrating that the statement is always true. To verify a statement, in this book, means to check its truth for a particular case; see, for example, Section 1.2, Exercises 16 and 22.

Most proofs required in this chapter proceed directly from the given conditions using definitions and previously proven facts; an example is Section 1.4, Theorem 2. A powerful tool for constructing a proof is to choose a generic object of the type in the statement and to see what you know about this object. Remember that you must explain why the statement is always true, so choosing a specific object will only verify the statement for that object.

The most common way to show that two sets are equal is to show each is a subset of the other (Section 1.2, Theorem 1).

In proving statements about sets or matrix operations, try to work at the level of object names rather than at the element or entry-level. For example, Section 1.5, Exercise 22 is more easily proved by using the facts that if $\mathbf{A}$ is symmetric, then $\mathbf{A}^T = \mathbf{A}$ and Theorem 3 rather than by using the fact that if $\mathbf{A} = [a_{ij}]$ is symmetric, then $a_{ij} = a_{ji}$ for each i and j.

One other style of direct proof is seen in Section 1.6, Example 12. Sometimes we show the statement is always true by examining all possible cases.

Key Ideas for Review

- Set: a well-defined collection of objects
- $\varnothing$ (empty set): the set with no elements
- Equal sets: sets with the same elements
- $A \subseteq B$ (A is a subset of B): Every element of A is an element of B.
- $|A|$ (cardinality of A): the number of elements of A
- Infinite set: see page 4
- $P(A)$ (power set of A): the set of all subsets of A
- $A \cup B$ (union of A and B): $\{x \mid x \in A \text{ or } x \in B\}$
- $A \cap B$ (intersection of A and B):$\{x \mid x \in A \text{ and } x \in B\}$
- Disjoint sets: two sets with no elements in common
- $A - B$ (complement of B with respect to A): $\{x \mid x \in A \text{ and } x \notin B\}$
- $\overline{A}$ (complement of A): $\{x \mid x \notin A\}$
- Algebraic properties of set operations: see pages 8–9
- Theorem (the addition principle): If A and B are finite sets, then $|A \cup B| = |A| + |B| - |A \cap B|$.
- Theorem (the three-set addition principle): If A, B, and C are finite sets, then $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$.
- Inclusion-exclusion principle: see page 9
- Sequence: list of objects arranged in a definite order
- Recursive formula: formula that uses previously defined terms
- Explicit formula: formula that does not use previously defined terms
- Linear array: see page 14
- Characteristic function of a set A: $f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$
- Countable set: a set that corresponds to a sequence
- Word: finite sequence of elements of A
- Regular expression: see page 17
- Theorem: If n and m are integers and $n > 0$, we can write $m = qn + r$ for integers q and r with $0 \leq r < n$. Moreover, there is just one way to do this.
- GCD(a, b): $d = \text{GCD}(a, b)$ if $d \mid a$, $d \mid b$, and d is the largest common divisor of a and b.
- Theorem: If d is GCD(a, b), then (a) $d = sa + tb$ for some integers s and t. (b) If $c \mid a$ and $c \mid b$, then $c \mid d$.
- Relatively prime: two integers a and b with GCD$(a, b) = 1$
- Euclidean algorithm: method used to find GCD(a, b); see pages 22–23
- LCM(a, b): $c = \text{LCM}(a, b)$ if $a \mid c$, $b \mid c$, and c is the smallest common multiple of a and b
- Theorem: GCD$(a, b) \cdot$ LCM$(a, b) = ab$
- Base b expansion of a number: see page 27
- Cryptology: the science of producing and deciphering secret codes
- Bacon's code: see page 29
- Steganography: the science of concealment of coded information
- mod-n function: $f_n(z) = r$, where r is the remainder when z is divided by n
- Matrix: rectangular array of numbers
- Size of a matrix: **A** is $m \times n$ if it has m rows and n columns
- Diagonal matrix: a square matrix with zero entries off the main diagonal
- Equal matrices: matrices of the same size whose corresponding entries are equal
- $\mathbf{A} + \mathbf{B}$: the matrix obtained by adding corresponding entries of **A** and **B**
- Zero matrix: a matrix all of whose entries are zero
- **AB**: see page 33
- $\mathbf{I}_n$ (identity matrix): a square matrix with ones on the diagonal and zeros elsewhere
- Array of dimension 2: see page 34
- $\mathbf{A}^T$: the matrix obtained from **A** by interchanging the rows and columns of **A**
- Symmetric matrix: $\mathbf{A}^T = \mathbf{A}$
- Inverse of a matrix: see page 36
- Boolean matrix: a matrix whose entries are either one or zero
- $\mathbf{A} \vee \mathbf{B}$: see page 37
- $\mathbf{A} \wedge \mathbf{B}$: see page 37
- $\mathbf{A} \odot \mathbf{B}$: see page 37
- Properties of Boolean matrix operations: see pages 38–39
- Mathematical structure: a collection of objects with operations defined on them and the accompanying properties
- Binary operation: an operation that combines two objects
- Unary operation: an operation that requires only one object
- Closure property: each application of the operation produces another object in the collection
- Associative property: $(x \square y) \square z = x \square (y \square z)$
- Distributive property: $x \square (y \triangledown z) = (x \square y) \triangledown (x \square z)$
- De Morgan's laws: $(x \square y)^\circ = x^\circ \triangledown y^\circ$ and $(x \triangledown y)^\circ = x^\circ \square y^\circ$
- Identity for $\square$: an element e such that $x \square e = e \square x = x$ for all x in the structure
- $\square$-inverse for x: an element y such that $x \square y = y \square x = e$, where e is the identity for $\square$

Review Questions

1. What kind of mathematical object is $P(A)$?
2. What kind of mathematical object is $|P(A)|$?
3. What kind of mathematical object is LCM(a, b)?
4. What kind of mathematical object is $k\mathbf{A}$?
5. What are the components of a mathematical structure?

Chapter 1 Self-Test

1. Let $A = \{x \mid x \text{ is a real number and } 0 < x < 1\}$, $B = \{x \mid x \text{ is a real number and } x^2 + 1 = 0\}$, $C = \{x \mid x = 4m, m \in Z\}$, $D = \{0, 2, 4, 6, \ldots\}$, and $E = \{x \mid x \in Z \text{ and } x^2 \leq 100\}$.
   (a) Identify the following as true or false.
   (i) $C \subseteq D$ (ii) $\{4, 16\} \subseteq C$
   (iii) $\{4, 16\} \subseteq E$ (iv) $D \subseteq D$
   (v) $B \subseteq \varnothing$
   (b) Identify the following as true or false.
   (i) $C \cap E \nsubseteq (C \cup E)$
   (ii) $\varnothing \subseteq (A \cap B)$ (iii) $C \cap D = D$
   (iv) $C \cup E \subseteq D$
   (v) $A \cap D \subseteq A \cap C$
2. Let $A = \{x \mid x = 2n, n \in Z^+\}$, $B = \{x \mid x = 2n + 1, n \in Z^+\}$, $C = \{x \mid x = 4n, n \in Z^+\}$, and $D = \{x \mid x(x^2 - 6x + 8) = 0, x \in Z\}$. Use Z as the universal set and find
   (a) $A \cup B$ (b) $\overline{A}$
   (c) $(A \cap D) \oplus (A \cap B)$ (d) $A \cup C$
   (e) $A - C$
3. Draw a Venn diagram to represent (a) $A \cap \overline{B}$ and (b) $\overline{A} \cap \overline{B}$.
4. Under what conditions will $A \cap B = A \cup B$?
5. Suppose that 109 of the 150 mathematics students at Verysmall College take at least one of the following computer languages: PASCAL, BASIC, C^{++}. Suppose 45 study BASIC, 61 study PASCAL, 53 study C^{++}, 18 study BASIC and PASCAL, 15 study BASIC and C^{++}, and 23 study PASCAL and C^{++}.
   (a) How many students study all three languages?
   (b) How many students study only BASIC?
   (c) How many students do not study any of the languages?
6. Define a sequence as follows: $a_0 = 0$, $a_1 = 0$, $a_n = 1 - 3a_{n-1} + 2a_{n-2}$. Compute the first six terms of this sequence.
7. Let $U = \{a, b, c, d, e, f, g, h, i, j\}$, $A = \{a, b, d, f\}$, $B = \{a, b, c, h, j\}$, $C = \{b, c, f, h, i\}$, and $D = \{g, h\}$. Represent each of the following sets by an array of zeros and ones.
   (a) $A \cup B$ (b) $A \cap B$
   (c) $A \cap (B \cup C)$ (d) $(\overline{A} \cap B) \cup D$
8. Let $I = \{a, b, c\}$. In each part that follows is listed a string in I^* and a regular expression over I. For each, state whether the string belongs to the regular set corresponding to the expression.
   (a) ab a^*bc^* (b) $acbb$ $((acb) \vee b)^*$
   (c) bc $((ab^*) \vee c)$ (d) $abaca$ $(ab)^*ac$
9. Use the Euclidean algorithm to compute GCD(4389, 7293) and write it as $s(7293) + t(4389)$.
10. Let $\mathbf{A} = \begin{bmatrix} 2 & 6 & 4 \\ -1 & 3 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 0 \\ -3 & 1 \end{bmatrix}$. Compute, if possible, each of the following.
    (a) **AB** (b) **BA** (c) $\mathbf{B}^T$
    (d) $\mathbf{A} + \mathbf{B}$ (e) $\mathbf{A}^T\mathbf{B}$ (f) $\mathbf{B}^{-1}$
    (g) $\mathbf{B}^{-1}\mathbf{A}$
11. Let $\mathbf{C} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$. Compute each of the following.
    (a) $\mathbf{C} \odot \mathbf{D}$ (b) $\mathbf{C} \vee \mathbf{D}$ (c) $\mathbf{C} \wedge \mathbf{D}$
12. Let $S = (2 \times 2$ Boolean matrices, $\wedge$, $\vee$, $\odot)$ and **A** be a 2×2 Boolean matrix. Describe the $\wedge$-inverse of **A** in S.

Coding Exercises

For each of the following, write the requested program or subroutine in pseudocode (as described in Appendix A) or in a programming language that you know. Test your code either with a paper-and-pencil trace or with a computer run.

In Exercises 1 through 3, assume that A and B are finite sets of integers. Write a subroutine to compute the specified set.

1. $A \cup B$ 2. $A \cap B$ 3. $A - B$

4. Consider the sequence recursively defined by $g(0) = 1$, $g(1) = -1$, $g(n) = 3g(n - 1) - 2g(n - 2)$.
   (a) Write a subroutine that will print the first 20 terms of the sequence.

(b) Write a subroutine that will print the first n terms of the sequence. The user should be able to supply the value of n at runtime.

5. Write a subroutine to find the least common multiple of two positive integers.

Experiment 1

In many voting procedures the rules are one person, one vote, and a simple majority is required to elect a candidate or to pass a motion. But it is not unusual to have a procedure where individual voters have more than one vote or where something other than a simple majority is required. An example of such a situation is when not all shareholders in a company own the same number of shares, and each shareholder has as many votes as shares. Does a shareholder with twice as many shares as another have twice as much control, or power, over the company? In this experiment you will investigate this question and some related ones. First, we begin with some definitions. The number of votes that a voter has is called the voter's **weight**. Here only counting numbers can be weights. The total number of votes needed to elect a candidate or to pass a motion is the **quota**. The collection of the quota and the individual weights for all voters is called a **weighted voting system**. If the voters are designated $v_1, v_2, \ldots, v_k$ with corresponding weights $w_1, w_2, \ldots, w_k$ and q is the quota, then the weighted voting system may be conveniently represented by $[q : w_1, w_2, \ldots, w_k]$. For ease of computations, the weights are usually listed from largest to smallest.

1. For the weighted voting system $[9 : 9, 4, 2, 1]$, what is the quota? How many voters are there? What is the total number of votes available?

2. In a weighted voting system $[q : w_1, w_2, \ldots, w_k]$, what are the restrictions on the possible values of q? Explain each restriction.

3. For the weighted voting system $[9 : 9, 4, 2, 1]$, describe how much power voter v_1 has. Such a voter is called a **dictator**. Why is this appropriate? Could a system have two dictators? Explain why or why not.

4. For $[8 : 5, 3, 2, 1]$, is v_1 a dictator? Describe v_1's power relative to the other voters.

More interesting cases arise when the power of each voter is not so obvious as in these first examples. One way to measure a voter's power was developed by John Banzhaf in 1965. A **coalition** is a subset of the voters in a weighted voting system. If the total number of votes controlled by the members of the coalition equals or exceeds the quota, we call the coalition a winning coalition. If not, this is a losing coalition.

5. **(a)** List all the coalitions for $[9 : 9, 4, 2, 1]$. Which of these are winning coalitions?

(b) List all the winning coalitions for $[8 : 5, 3, 2, 1]$.

Banzhaf's idea is to measure a voter's power by examining how many times removal of that voter from a coalition would change the coalition from winning to losing. Consider the system $[7 : 5, 4, 3]$. The winning coalitions are $\{v_1, v_2\}$, $\{v_1, v_3\}$, $\{v_2, v_3\}$, and $\{v_1, v_2, v_3\}$. Each member of the first three coalitions has the power to change it from winning to losing, but none have this power in the last coalition. All together there are six opportunities for change. Each of v_1, v_2, v_3 has two of these opportunities. We record this information as the **Banzhaf power distribution** for the system: $v_1 : \frac{2}{6}$, $v_2 : \frac{2}{6}$, $v_3 : \frac{2}{6}$. According to this analysis, all three voters have the same amount of power despite having different weights. The fraction of power assigned to a voter is the voter's **Banzhaf power index**.

6. Here is a test for Banzhaf's definition of power. Calculate the Banzhaf power distribution for $[9 : 9, 4, 2, 1]$. Explain how the results are consistent with the designation of v_1 as a dictator.

7. Calculate the Banzhaf power distribution for $[8 : 5, 3, 2, 1]$. A voter like v_1 that must belong to every winning coalition has **veto power** in the system.

8. Let $[q : 6, 3, 1]$ be a weighted voting system.

(a) Give values for q for which the system has a dictator and identify that voter.

(b) Give values for q for which one or more voters have veto power and identify these voters.

(c) Give values for q for which at least one player is powerless, but there is no dictator. Which player is powerless?

Banzhaf's idea is adaptable to cases where each voter has one vote, but special rules for voting apply.

9. The four partners in a company agree that each partner has one vote and a simple majority passes a motion. In the case of a tie the coalition containing the senior partner is the winning coalition. Give the Banzhaf power distribution for this system. Would the distribution change if the tie-breaking rule were changed to the coalition containing the most junior member is the losing coalition? Explain.

10. Suppose you are the voter with weight one in $[8 : 5, 3, 2, 1]$.

(a) What is your Banzhaf power index?

(b) Unhappy with this situation, you offer to buy a vote from one of the other voters. If each is willing to sell and each asks the same price, from whom should you buy a vote and why? Give the Banzhaf power distribution for this system for the resulting weighted voting system.

11. Here is another feature of Banzhaf's way of measuring power. Let $[q : w_1, w_2, \ldots, w_k]$ be a weighted voting system and n be a positive integer. Prove that the Banzhaf power distributions for $[q : w_1, w_2, \ldots, w_k]$ and $[nq : nw_1, nw_2, \ldots, nw_k]$ are the same.

12. We now return to the original question about power. Suppose we have a weighted voting system in which v_1 has weight w_1, v_2 has weight w_2, and $w_1 = 2w_2$. Construct such a system where the Banzhaf power index of v_1 is

(a) the same as that of v_2

(b) twice that of v_2

(c) more than twice that of v_2.