CHAPTER

2 Basic Structures: Sets, Functions, Sequences, Sums, and Matrices

Much of discrete mathematics is devoted to the study of discrete structures, used to represent discrete objects. Many important discrete structures are built using sets, which are collections of objects. Among the discrete structures built from sets are combinations, unordered collections of objects used extensively in counting; relations, sets of ordered pairs that represent relationships between objects; graphs, sets of vertices and edges that connect vertices; and finite state machines, used to model computing machines. These are some of the topics we will study in later chapters.

The concept of a function is extremely important in discrete mathematics. A function assigns to each element of a first set exactly one element of a second set, where the two sets are not necessarily distinct. Functions play important roles throughout discrete mathematics. They are used to represent the computational complexity of algorithms, to study the size of sets, to count objects, and in a myriad of other ways. Useful structures such as sequences and strings are special types of functions. In this chapter, we will introduce the notion of a sequence, which represents ordered lists of elements. Furthermore, we will introduce some important types of sequences and we will show how to define the terms of a sequence using earlier terms. We will also address the problem of identifying a sequence from its first few terms.

In our study of discrete mathematics, we will often add consecutive terms of a sequence of numbers. Because adding terms from a sequence, as well as other indexed sets of numbers, is such a common occurrence, a special notation has been developed for adding such terms. In this chapter, we will introduce the notation used to express summations. We will develop formulae for certain types of summations that appear throughout the study of discrete mathematics. For instance, we will encounter such summations in the analysis of the number of steps used by an algorithm to sort a list of numbers so that its terms are in increasing order.

The relative sizes of infinite sets can be studied by introducing the notion of the size, or cardinality, of a set. We say that a set is countable when it is finite or has the same size as the set of positive integers. In this chapter we will establish the surprising result that the set of rational numbers is countable, while the set of real numbers is not. We will also show how the concepts we discuss can be used to show that there are functions that cannot be computed using a computer program in any programming language.

Matrices are used in discrete mathematics to represent a variety of discrete structures. We will review the basic material about matrices and matrix arithmetic needed to represent relations and graphs. The matrix arithmetic we study will be used to solve a variety of problems involving these structures.

2.1 Sets

Introduction

In this section, we study the fundamental discrete structure on which all other discrete structures are built, namely, the set. Sets are used to group objects together. Often, but not always, the objects in a set have similar properties. For instance, all the students who are currently enrolled in your school make up a set. Likewise, all the students currently taking a course in discrete mathematics at any school make up a set. In addition, those students enrolled in your school who are taking a course in discrete mathematics form a set that can be obtained by taking the elements common to the first two collections. The language of sets is a means to study such collections in an organized fashion. We now provide a definition of a set. This definition is an intuitive definition, which is not part of a formal theory of sets.

DEFINITION 1 A *set* is an unordered collection of objects, called *elements* or *members* of the set. A set is said to *contain* its elements. We write $a \in A$ to denote that a is an element of the set A. The notation $a \notin A$ denotes that a is not an element of the set A.

It is common for sets to be denoted using uppercase letters. Lowercase letters are usually used to denote elements of sets.

There are several ways to describe a set. One way is to list all the members of a set, when this is possible. We use a notation where all members of the set are listed between braces. For example, the notation $\{a, b, c, d\}$ represents the set with the four elements a, b, c, and d. This way of describing a set is known as the **roster method**.

EXAMPLE 1 The set V of all vowels in the English alphabet can be written as $V = \{a, e, i, o, u\}$. ◀

EXAMPLE 2 The set O of odd positive integers less than 10 can be expressed by $O = \{1, 3, 5, 7, 9\}$. ◀

EXAMPLE 3 Although sets are usually used to group together elements with common properties, there is nothing that prevents a set from having seemingly unrelated elements. For instance, $\{a$, 2, Fred, New Jersey$\}$ is the set containing the four elements a, 2, Fred, and New Jersey. ◀

Sometimes the roster method is used to describe a set without listing all its members. Some members of the set are listed, and then *ellipses* (...) are used when the general pattern of the elements is obvious.

EXAMPLE 4 The set of positive integers less than 100 can be denoted by $\{1, 2, 3, \ldots, 99\}$. ◀

Extra Examples

Another way to describe a set is to use **set builder** notation. We characterize all those elements in the set by stating the property or properties they must have to be members. For instance, the set O of all odd positive integers less than 10 can be written as

$$O = \{x \mid x \text{ is an odd positive integer less than } 10\},$$

or, specifying the universe as the set of positive integers, as

$$O = \{x \in \mathbf{Z}^+ \mid x \text{ is odd and } x < 10\}.$$

We often use this type of notation to describe sets when it is impossible to list all the elements of the set. For instance, the set $\mathbf{Q}^+$ of all positive rational numbers can be written as

$$\mathbf{Q}^+ = \{x \in \mathbf{R} \mid x = \tfrac{p}{q}, \text{ for some positive integers } p \text{ and } q\}.$$

Beware that mathematicians disagree whether 0 is a natural number. We consider it quite natural.

These sets, each denoted using a boldface letter, play an important role in discrete mathematics:

$\mathbf{N} = \{0, 1, 2, 3, \ldots\}$, the set of **natural numbers**
$\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the set of **integers**
$\mathbf{Z}^+ = \{1, 2, 3, \ldots\}$, the set of **positive integers**
$\mathbf{Q} = \{p/q \mid p \in \mathbf{Z}, q \in \mathbf{Z}, \text{ and } q \neq 0\}$, the set of **rational numbers**
$\mathbf{R}$, the set of **real numbers**
$\mathbf{R}^+$, the set of **positive real numbers**
$\mathbf{C}$, the set of **complex numbers**.

(Note that some people do not consider 0 a natural number, so be careful to check how the term *natural numbers* is used when you read other books.)

Recall the notation for **intervals** of real numbers. When a and b are real numbers with $a < b$, we write

$$[a, b] = \{x \mid a \leq x \leq b\}$$
$$[a, b) = \{x \mid a \leq x < b\}$$
$$(a, b] = \{x \mid a < x \leq b\}$$
$$(a, b) = \{x \mid a < x < b\}$$

Note that $[a, b]$ is called the **closed interval** from a to b and (a, b) is called the **open interval** from a to b.

Sets can have other sets as members, as Example 5 illustrates.

EXAMPLE 5 The set $\{\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}\}$ is a set containing four elements, each of which is a set. The four elements of this set are **N**, the set of natural numbers; **Z**, the set of integers; **Q**, the set of rational numbers; and **R**, the set of real numbers. ◀

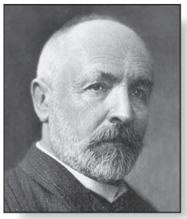
Remark: Note that the concept of a datatype, or type, in computer science is built upon the concept of a set. In particular, a **datatype** or **type** is the name of a set, together with a set of operations that can be performed on objects from that set. For example, *boolean* is the name of the set {0, 1} together with operators on one or more elements of this set, such as AND, OR, and NOT.

Because many mathematical statements assert that two differently specified collections of objects are really the same set, we need to understand what it means for two sets to be equal.

DEFINITION 2 Two sets are *equal* if and only if they have the same elements. Therefore, if A and B are sets, then A and B are equal if and only if $\forall x(x \in A \leftrightarrow x \in B)$. We write $A = B$ if A and B are equal sets.

EXAMPLE 6 The sets {1, 3, 5} and {3, 5, 1} are equal, because they have the same elements. Note that the order in which the elements of a set are listed does not matter. Note also that it does not matter if an element of a set is listed more than once, so {1, 3, 3, 3, 5, 5, 5, 5} is the same as the set {1, 3, 5} because they have the same elements. ◀

Links



GEORG CANTOR (1845–1918) Georg Cantor was born in St. Petersburg, Russia, where his father was a successful merchant. Cantor developed his interest in mathematics in his teens. He began his university studies in Zurich in 1862, but when his father died he left Zurich. He continued his university studies at the University of Berlin in 1863, where he studied under the eminent mathematicians Weierstrass, Kummer, and Kronecker. He received his doctor's degree in 1867, after having written a dissertation on number theory. Cantor assumed a position at the University of Halle in 1869, where he continued working until his death.

Cantor is considered the founder of set theory. His contributions in this area include the discovery that the set of real numbers is uncountable. He is also noted for his many important contributions to analysis. Cantor also was interested in philosophy and wrote papers relating his theory of sets with metaphysics.

Cantor married in 1874 and had five children. His melancholy temperament was balanced by his wife's happy disposition. Although he received a large inheritance from his father, he was poorly paid as a professor. To mitigate this, he tried to obtain a better-paying position at the University of Berlin. His appointment there was blocked by Kronecker, who did not agree with Cantor's views on set theory. Cantor suffered from mental illness throughout the later years of his life. He died in 1918 from a heart attack.

THE EMPTY SET There is a special set that has no elements. This set is called the **empty set**, or **null set**, and is denoted by ∅. The empty set can also be denoted by { } (that is, we represent the empty set with a pair of braces that encloses all the elements in this set). Often, a set of elements with certain properties turns out to be the null set. For instance, the set of all positive integers that are greater than their squares is the null set.

{∅} has one more element than ∅.

A set with one element is called a **singleton set**. A common error is to confuse the empty set ∅ with the set {∅}, which is a singleton set. The single element of the set {∅} is the empty set itself! A useful analogy for remembering this difference is to think of folders in a computer file system. The empty set can be thought of as an empty folder and the set consisting of just the empty set can be thought of as a folder with exactly one folder inside, namely, the empty folder.

Links

NAIVE SET THEORY Note that the term *object* has been used in the definition of a set, Definition 1, without specifying what an object is. This description of a set as a collection of objects, based on the intuitive notion of an object, was first stated in 1895 by the German mathematician Georg Cantor. The theory that results from this intuitive definition of a set, and the use of the intuitive notion that for any property whatever, there is a set consisting of exactly the objects with this property, leads to **paradoxes**, or logical inconsistencies. This was shown by the English philosopher Bertrand Russell in 1902 (see Exercise 29 for a description of one of these paradoxes). These logical inconsistencies can be avoided by building set theory beginning with axioms. However, we will use Cantor's original version of set theory, known as **naive set theory**, in this book because all sets considered in this book can be treated consistently using Cantor's original theory. Students will find familiarity with naive set theory helpful if they go on to learn about axiomatic set theory. They will also find the development of axiomatic set theory much more abstract than the material in this text. We refer the interested reader to [Su72] to learn more about axiomatic set theory.

Venn Diagrams

Sets can be represented graphically using Venn diagrams, named after the English mathematician John Venn, who introduced their use in 1881. In Venn diagrams the **universal set** U, which contains all the objects under consideration, is represented by a rectangle. (Note that the universal set varies depending on which objects are of interest.) Inside this rectangle, circles or other geometrical figures are used to represent sets. Sometimes points are used to represent the particular elements of the set. Venn diagrams are often used to indicate the relationships between sets. We show how a Venn diagram can be used in Example 7.

Assessment

EXAMPLE 7 Draw a Venn diagram that represents V, the set of vowels in the English alphabet.

Solution: We draw a rectangle to indicate the universal set U, which is the set of the 26 letters of the English alphabet. Inside this rectangle we draw a circle to represent V. Inside this circle we indicate the elements of V with points (see Figure 1). ◀

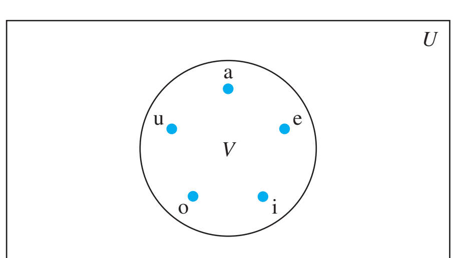


FIGURE 1 Venn Diagram for the Set of Vowels.

Subsets

It is common to encounter situations where the elements of one set are also the elements of a second set. We now introduce some terminology and notation to express such relationships between sets.

DEFINITION 3

The set A is a *subset* of B if and only if every element of A is also an element of B. We use the notation $A \subseteq B$ to indicate that A is a subset of the set B.

We see that $A \subseteq B$ if and only if the quantification

$$\forall x(x \in A \to x \in B)$$

is true. Note that to show that A is not a subset of B we need only find one element $x \in A$ with $x \notin B$. Such an x is a counterexample to the claim that $x \in A$ implies $x \in B$.

We have these useful rules for determining whether one set is a subset of another:

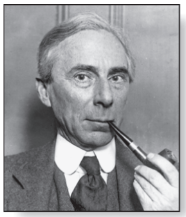
Showing that A is a Subset of B To show that $A \subseteq B$, show that if x belongs to A then x also belongs to B.

Showing that A is Not a Subset of B To show that $A \not\subseteq B$, find a single $x \in A$ such that $x \notin B$.

EXAMPLE 8 The set of all odd positive integers less than 10 is a subset of the set of all positive integers less than 10, the set of rational numbers is a subset of the set of real numbers, the set of all computer science majors at your school is a subset of the set of all students at your school, and the set of all people in China is a subset of the set of all people in China (that is, it is a subset of itself). Each of these facts follows immediately by noting that an element that belongs to the first set in each pair of sets also belongs to the second set in that pair. ◀

EXAMPLE 9 The set of integers with squares less than 100 is not a subset of the set of nonnegative integers because -1 is in the former set [as $(-1)^2 < 100$], but not the later set. The set of people who have taken discrete mathematics at your school is not a subset of the set of all computer science majors at your school if there is at least one student who has taken discrete mathematics who is not a computer science major. ◀

Links



BERTRAND RUSSELL (1872–1970) Bertrand Russell was born into a prominent English family active in the progressive movement and having a strong commitment to liberty. He became an orphan at an early age and was placed in the care of his father's parents, who had him educated at home. He entered Trinity College, Cambridge, in 1890, where he excelled in mathematics and in moral science. He won a fellowship on the basis of his work on the foundations of geometry. In 1910 Trinity College appointed him to a lectureship in logic and the philosophy of mathematics.

Russell fought for progressive causes throughout his life. He held strong pacifist views, and his protests against World War I led to dismissal from his position at Trinity College. He was imprisoned for 6 months in 1918 because of an article he wrote that was branded as seditious. Russell fought for women's suffrage in Great Britain. In 1961, at the age of 89, he was imprisoned for the second time for his protests advocating nuclear disarmament.

Russell's greatest work was in his development of principles that could be used as a foundation for all of mathematics. His most famous work is *Principia Mathematica,* written with Alfred North Whitehead, which attempts to deduce all of mathematics using a set of primitive axioms. He wrote many books on philosophy, physics, and his political ideas. Russell won the Nobel Prize for literature in 1950.

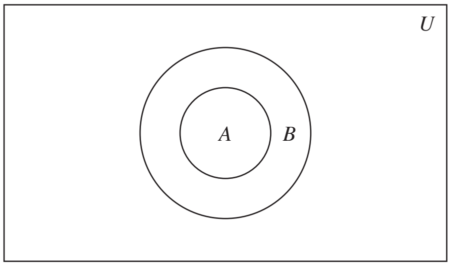


FIGURE 2 **Venn Diagram Showing that A Is a Subset of B.**

Theorem 1 shows that every nonempty set S is guaranteed to have at least two subsets, the empty set and the set S itself, that is, $\emptyset \subseteq S$ and $S \subseteq S$.

THEOREM 1 For every set S, $(i)\ \emptyset \subseteq S$ and $(ii)\ S \subseteq S$.

Proof: We will prove (i) and leave the proof of (ii) as an exercise.

Let S be a set. To show that $\emptyset \subseteq S$, we must show that $\forall x(x \in \emptyset \rightarrow x \in S)$ is true. Because the empty set contains no elements, it follows that $x \in \emptyset$ is always false. It follows that the conditional statement $x \in \emptyset \rightarrow x \in S$ is always true, because its hypothesis is always false and a conditional statement with a false hypothesis is true. Therefore, $\forall x(x \in \emptyset \rightarrow x \in S)$ is true. This completes the proof of (i). Note that this is an example of a vacuous proof. ◁

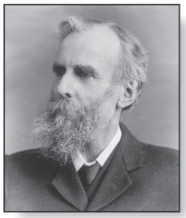
When we wish to emphasize that a set A is a subset of a set B but that $A \neq B$, we write $A \subset B$ and say that A is a **proper subset** of B. For $A \subset B$ to be true, it must be the case that $A \subseteq B$ and there must exist an element x of B that is not an element of A. That is, A is a proper subset of B if and only if

$$\forall x(x \in A \rightarrow x \in B) \wedge \exists x(x \in B \wedge x \notin A)$$

is true. Venn diagrams can be used to illustrate that a set A is a subset of a set B. We draw the universal set U as a rectangle. Within this rectangle we draw a circle for B. Because A is a subset of B, we draw the circle for A within the circle for B. This relationship is shown in Figure 2.

A useful way to show that two sets have the same elements is to show that each set is a subset of the other. In other words, we can show that if A and B are sets with $A \subseteq B$ and $B \subseteq A$, then $A = B$. That is, $A = B$ if and only if $\forall x(x \in A \rightarrow x \in B)$ and $\forall x(x \in B \rightarrow x \in A)$ or equivalently if and only if $\forall x(x \in A \leftrightarrow x \in B)$, which is what it means for the A and B to be equal. Because this method of showing two sets are equal is so useful, we highlight it here.

Links



JOHN VENN (1834–1923) John Venn was born into a London suburban family noted for its philanthropy. He attended London schools and got his mathematics degree from Caius College, Cambridge, in 1857. He was elected a fellow of this college and held his fellowship there until his death. He took holy orders in 1859 and, after a brief stint of religious work, returned to Cambridge, where he developed programs in the moral sciences. Besides his mathematical work, Venn had an interest in history and wrote extensively about his college and family.

Venn's book *Symbolic Logic* clarifies ideas originally presented by Boole. In this book, Venn presents a systematic development of a method that uses geometric figures, known now as *Venn diagrams*. Today these diagrams are primarily used to analyze logical arguments and to illustrate relationships between sets. In addition to his work on symbolic logic, Venn made contributions to probability theory described in his widely used textbook on that subject.

Showing Two Sets are Equal To show that two sets A and B are equal, show that $A \subseteq B$ and $B \subseteq A$.

Sets may have other sets as members. For instance, we have the sets

$$A = \{\emptyset, \{a\}, \{b\}, \{a, b\}\} \quad \text{and} \quad B = \{x \mid x \text{ is a subset of the set } \{a, b\}\}.$$

Note that these two sets are equal, that is, $A = B$. Also note that $\{a\} \in A$, but $a \notin A$.

The Size of a Set

Sets are used extensively in counting problems, and for such applications we need to discuss the sizes of sets.

DEFINITION 4 Let S be a set. If there are exactly n distinct elements in S where n is a nonnegative integer, we say that S is a *finite set* and that n is the *cardinality* of S. The cardinality of S is denoted by $|S|$.

Remark: The term *cardinality* comes from the common usage of the term *cardinal number* as the size of a finite set.

EXAMPLE 10 Let A be the set of odd positive integers less than 10. Then $|A| = 5$. ◀

EXAMPLE 11 Let S be the set of letters in the English alphabet. Then $|S| = 26$. ◀

EXAMPLE 12 Because the null set has no elements, it follows that $|\emptyset| = 0$. ◀

We will also be interested in sets that are not finite.

DEFINITION 5 A set is said to be *infinite* if it is not finite.

EXAMPLE 13 The set of positive integers is infinite. ◀

Extra Examples

We will extend the notion of cardinality to infinite sets in Section 2.5, a challenging topic full of surprising results.

Power Sets

Many problems involve testing all combinations of elements of a set to see if they satisfy some property. To consider all such combinations of elements of a set S, we build a new set that has as its members all the subsets of S.

DEFINITION 6 Given a set S, the *power set* of S is the set of all subsets of the set S. The power set of S is denoted by $\mathcal{P}(S)$.

EXAMPLE 14 What is the power set of the set $\{0, 1, 2\}$?

Extra Examples

Solution: The power set $\mathcal{P}(\{0, 1, 2\})$ is the set of all subsets of $\{0, 1, 2\}$. Hence,

$$\mathcal{P}(\{0, 1, 2\}) = \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

Note that the empty set and the set itself are members of this set of subsets. ◀

EXAMPLE 15 What is the power set of the empty set? What is the power set of the set $\{\emptyset\}$?

Solution: The empty set has exactly one subset, namely, itself. Consequently,

$$\mathcal{P}(\emptyset) = \{\emptyset\}.$$

The set $\{\emptyset\}$ has exactly two subsets, namely, $\emptyset$ and the set $\{\emptyset\}$ itself. Therefore,

$$\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}.$$ ◀

If a set has n elements, then its power set has 2^n elements. We will demonstrate this fact in several ways in subsequent sections of the text.

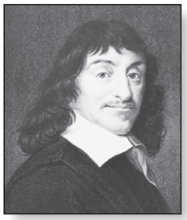
Cartesian Products

The order of elements in a collection is often important. Because sets are unordered, a different structure is needed to represent ordered collections. This is provided by **ordered *n*-tuples**.

DEFINITION 7 The *ordered n-tuple* $(a_1, a_2, \ldots, a_n)$ is the ordered collection that has a_1 as its first element, a_2 as its second element, ..., and a_n as its nth element.

We say that two ordered n-tuples are equal if and only if each corresponding pair of their elements is equal. In other words, $(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n)$ if and only if $a_i = b_i$, for $i = 1, 2, \ldots, n$. In particular, ordered 2-tuples are called **ordered pairs**. The ordered pairs (a, b) and (c, d) are equal if and only if $a = c$ and $b = d$. Note that (a, b) and (b, a) are not equal unless $a = b$.

Links



RENÉ DESCARTES (1596–1650) René Descartes was born into a noble family near Tours, France, about 200 miles southwest of Paris. He was the third child of his father's first wife; she died several days after his birth. Because of René's poor health, his father, a provincial judge, let his son's formal lessons slide until, at the age of 8, René entered the Jesuit college at La Flèche. The rector of the school took a liking to him and permitted him to stay in bed until late in the morning because of his frail health. From then on, Descartes spent his mornings in bed; he considered these times his most productive hours for thinking.

Descartes left school in 1612, moving to Paris, where he spent 2 years studying mathematics. He earned a law degree in 1616 from the University of Poitiers. At 18 Descartes became disgusted with studying and decided to see the world. He moved to Paris and became a successful gambler. However, he grew tired of bawdy living and moved to the suburb of Saint-Germain, where he devoted himself to mathematical study. When his gambling friends found him, he decided to leave France and undertake a military career. However, he never did any fighting. One day, while escaping the cold in an overheated room at a military encampment, he had several feverish dreams, which revealed his future career as a mathematician and philosopher.

After ending his military career, he traveled throughout Europe. He then spent several years in Paris, where he studied mathematics and philosophy and constructed optical instruments. Descartes decided to move to Holland, where he spent 20 years wandering around the country, accomplishing his most important work. During this time he wrote several books, including the *Discours,* which contains his contributions to analytic geometry, for which he is best known. He also made fundamental contributions to philosophy.

In 1649 Descartes was invited by Queen Christina to visit her court in Sweden to tutor her in philosophy. Although he was reluctant to live in what he called "the land of bears amongst rocks and ice," he finally accepted the invitation and moved to Sweden. Unfortunately, the winter of 1649–1650 was extremely bitter. Descartes caught pneumonia and died in mid-February.

Many of the discrete structures we will study in later chapters are based on the notion of the *Cartesian product* of sets (named after René Descartes). We first define the Cartesian product of two sets.

DEFINITION 8 Let A and B be sets. The *Cartesian product* of A and B, denoted by $A \times B$, is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$. Hence,

$$A \times B = \{(a, b) \mid a \in A \wedge b \in B\}.$$

EXAMPLE 16 Let A represent the set of all students at a university, and let B represent the set of all courses offered at the university. What is the Cartesian product $A \times B$ and how can it be used?

Extra Examples

Solution: The Cartesian product $A \times B$ consists of all the ordered pairs of the form (a, b), where a is a student at the university and b is a course offered at the university. One way to use the set $A \times B$ is to represent all possible enrollments of students in courses at the university.

EXAMPLE 17 What is the Cartesian product of $A = \{1, 2\}$ and $B = \{a, b, c\}$?

Solution: The Cartesian product $A \times B$ is

$$A \times B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

Note that the Cartesian products $A \times B$ and $B \times A$ are not equal, unless $A = \emptyset$ or $B = \emptyset$ (so that $A \times B = \emptyset$) or $A = B$ (see Exercise 24). This is illustrated in Example 18.

EXAMPLE 18 Show that the Cartesian product $B \times A$ is not equal to the Cartesian product $A \times B$, where A and B are as in Example 17.

Solution: The Cartesian product $B \times A$ is

$$B \times A = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}.$$

This is not equal to $A \times B$, which was found in Example 17.

The Cartesian product of more than two sets can also be defined.

DEFINITION 9 The *Cartesian product* of the sets $A_1, A_2, \ldots, A_n$, denoted by $A_1 \times A_2 \times \cdots \times A_n$, is the set of ordered n-tuples $(a_1, a_2, \ldots, a_n)$, where a_i belongs to A_i for $i = 1, 2, \ldots, n$. In other words,

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i \text{ for } i = 1, 2, \ldots, n\}.$$

EXAMPLE 19 What is the Cartesian product $A \times B \times C$, where $A = \{0, 1\}$, $B = \{1, 2\}$, and $C = \{0, 1, 2\}$?

Solution: The Cartesian product $A \times B \times C$ consists of all ordered triples (a, b, c), where $a \in A$, $b \in B$, and $c \in C$. Hence,

$$A \times B \times C = \{(0, 1, 0), (0, 1, 1), (0, 1, 2), (0, 2, 0), (0, 2, 1), (0, 2, 2), (1, 1, 0), (1, 1, 1), (1, 1, 2), (1, 2, 0), (1, 2, 1), (1, 2, 2)\}.$$ ◀

Remark: Note that when A, B, and C are sets, $(A \times B) \times C$ is not the same as $A \times B \times C$ (see Exercise 25).

We use the notation A^2 to denote $A \times A$, the Cartesian product of the set A with itself. Similarly, $A^3 = A \times A \times A$, $A^4 = A \times A \times A \times A$, and so on. More generally,

$$A^n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in A \text{ for } i = 1, 2, \ldots, n\}.$$

EXAMPLE 20 Suppose that $A = \{1, 2\}$. It follows that $A^2 = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ and $A^3 = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)\}$. ◀

A subset R of the Cartesian product $A \times B$ is called a **relation** from the set A to the set B. The elements of R are ordered pairs, where the first element belongs to A and the second to B. For example, $R = \{(a, 0), (a, 1), (a, 3), (b, 1), (b, 2), (c, 0), (c, 3)\}$ is a relation from the set $\{a, b, c\}$ to the set $\{0, 1, 2, 3\}$. A relation from a set A to itself is called a relation on A.

EXAMPLE 21 What are the ordered pairs in the less than or equal to relation, which contains (a, b) if $a \leq b$, on the set $\{0, 1, 2, 3\}$?

Solution: The ordered pair (a, b) belongs to R if and only if both a and b belong to $\{0, 1, 2, 3\}$ and $a \leq b$. Consequently, the ordered pairs in R are (0,0), (0,1), (0,2), (0,3), (1,1), (1,2), (1,3), (2,2), (2, 3), and (3, 3). ◀

We will study relations and their properties at length in Chapter 9.

Using Set Notation with Quantifiers

Sometimes we restrict the domain of a quantified statement explicitly by making use of a particular notation. For example, $\forall x \in S(P(x))$ denotes the universal quantification of $P(x)$ over all elements in the set S. In other words, $\forall x \in S(P(x))$ is shorthand for $\forall x(x \in S \rightarrow P(x))$. Similarly, $\exists x \in S(P(x))$ denotes the existential quantification of $P(x)$ over all elements in S. That is, $\exists x \in S(P(x))$ is shorthand for $\exists x(x \in S \wedge P(x))$.

EXAMPLE 22 What do the statements $\forall x \in \mathbf{R}\ (x^2 \geq 0)$ and $\exists x \in \mathbf{Z}\ (x^2 = 1)$ mean?

Solution: The statement $\forall x \in \mathbf{R}(x^2 \geq 0)$ states that for every real number x, $x^2 \geq 0$. This statement can be expressed as "The square of every real number is nonnegative." This is a true statement.

The statement $\exists x \in \mathbf{Z}(x^2 = 1)$ states that there exists an integer x such that $x^2 = 1$. This statement can be expressed as "There is an integer whose square is 1." This is also a true statement because $x = 1$ is such an integer (as is -1). ◀

Truth Sets and Quantifiers

We will now tie together concepts from set theory and from predicate logic. Given a predicate P, and a domain D, we define the **truth set** of P to be the set of elements x in D for which $P(x)$ is true. The truth set of $P(x)$ is denoted by $\{x \in D \mid P(x)\}$.

EXAMPLE 23 What are the truth sets of the predicates $P(x)$, $Q(x)$, and $R(x)$, where the domain is the set of integers and $P(x)$ is "$|x| = 1$," $Q(x)$ is "$x^2 = 2$," and $R(x)$ is "$|x| = x$."

Solution: The truth set of P, $\{x \in \mathbf{Z} \mid |x| = 1\}$, is the set of integers for which $|x| = 1$. Because $|x| = 1$ when $x = 1$ or $x = -1$, and for no other integers x, we see that the truth set of P is the set $\{-1, 1\}$.

The truth set of Q, $\{x \in \mathbf{Z} \mid x^2 = 2\}$, is the set of integers for which $x^2 = 2$. This is the empty set because there are no integers x for which $x^2 = 2$.

The truth set of R, $\{x \in \mathbf{Z} \mid |x| = x\}$, is the set of integers for which $|x| = x$. Because $|x| = x$ if and only if $x \geq 0$, it follows that the truth set of R is $\mathbf{N}$, the set of nonnegative integers. ◀

Note that $\forall x P(x)$ is true over the domain U if and only if the truth set of P is the set U. Likewise, $\exists x P(x)$ is true over the domain U if and only if the truth set of P is nonempty.

Exercises

1. List the members of these sets.
- **a)** $\{x \mid x \text{ is a real number such that } x^2 = 1\}$
- **b)** $\{x \mid x \text{ is a positive integer less than } 12\}$
- **c)** $\{x \mid x \text{ is the square of an integer and } x < 100\}$
- **d)** $\{x \mid x \text{ is an integer such that } x^2 = 2\}$

2. Use set builder notation to give a description of each of these sets.
- **a)** $\{0, 3, 6, 9, 12\}$
- **b)** $\{-3, -2, -1, 0, 1, 2, 3\}$
- **c)** $\{m, n, o, p\}$

3. For each of these pairs of sets, determine whether the first is a subset of the second, the second is a subset of the first, or neither is a subset of the other.
- **a)** the set of airline flights from New York to New Delhi, the set of nonstop airline flights from New York to New Delhi
- **b)** the set of people who speak English, the set of people who speak Chinese
- **c)** the set of flying squirrels, the set of living creatures that can fly

4. Determine whether each of these pairs of sets are equal.
- **a)** $\{1, 3, 3, 3, 5, 5, 5, 5, 5\}, \{5, 3, 1\}$
- **b)** $\{\{1\}\}, \{1, \{1\}\}$ **c)** $\emptyset, \{\emptyset\}$

5. Suppose that $A = \{2, 4, 6\}$, $B = \{2, 6\}$, $C = \{4, 6\}$, and $D = \{4, 6, 8\}$. Determine which of these sets are subsets of which other of these sets.

6. For each of the following sets, determine whether 2 is an element of that set.
- **a)** $\{x \in \mathbf{R} \mid x \text{ is an integer greater than } 1\}$
- **b)** $\{x \in \mathbf{R} \mid x \text{ is the square of an integer}\}$
- **c)** $\{2, \{2\}\}$ **d)** $\{\{2\}, \{\{2\}\}\}$
- **e)** $\{\{2\}, \{2, \{2\}\}\}$ **f)** $\{\{\{2\}\}\}$

7. For each of the sets in Exercise 7, determine whether $\{2\}$ is an element of that set.

8. Determine whether each of these statements is true or false.
- **a)** $0 \in \emptyset$ **b)** $\emptyset \in \{0\}$
- **c)** $\{0\} \subset \emptyset$ **d)** $\emptyset \subset \{0\}$
- **e)** $x \in \{x\}$ **f)** $\{x\} \subseteq \{x\}$
- **g)** $\{x\} \in \{x\}$ **h)** $\{x\} \in \{\{x\}\}$
- **i)** $\emptyset \subseteq \{x\}$

9. Use a Venn diagram to illustrate the subset of odd integers in the set of all positive integers not exceeding 10.

10. Use a Venn diagram to illustrate the set of all months of the year whose names do not contain the letter R in the set of all months of the year.

11. Use a Venn diagram to illustrate the relationships $A \subset B$ and $B \subset C$.

12. Suppose that A, B, and C are sets such that $A \subseteq B$ and $B \subseteq C$. Show that $A \subseteq C$.

13. Find two sets A and B such that $A \in B$ and $A \subseteq B$.

14. What is the cardinality of each of these sets?
- **a)** $\{a\}$ **b)** $\{\{a\}\}$
- **c)** $\{a, \{a\}\}$ **d)** $\{a, \{a\}, \{a, \{a\}\}\}$

15. Find the power set of each of these sets, where a and b are distinct elements.
- **a)** $\{a\}$ **b)** $\{a, b\}$ **c)** $\{\emptyset, \{\emptyset\}\}$

16. Can you conclude that $A = B$ if A and B are two sets with the same power set?

17. How many elements does each of these sets have where a and b are distinct elements?

a) $\mathcal{P}(\{a, b, \{a, b\}\})$

b) $\mathcal{P}(\{\emptyset, a, \{a\}, \{\{a\}\}\})$

c) $\mathcal{P}(\mathcal{P}(\emptyset))$

18. Determine whether each of these sets is the power set of a set, where a and b are distinct elements.

a) $\emptyset$ **b)** $\{\emptyset, \{a\}\}$

c) $\{\emptyset, \{a\}, \{\emptyset, a\}\}$ **d)** $\{\emptyset, \{a\}, \{b\}, \{a, b\}\}$

19. Let $A = \{a, b, c, d\}$ and $B = \{y, z\}$. Find

a) $A \times B$. **b)** $B \times A$.

20. What is the Cartesian product $A \times B$, where A is the set of courses offered by the mathematics department at a university and B is the set of mathematics professors at this university? Give an example of how this Cartesian product can be used.

21. Suppose that $A \times B = \emptyset$, where A and B are sets. What can you conclude?

22. Let $A = \{a, b, c\}$, $B = \{x, y\}$, and $C = \{0, 1\}$. Find

a) $A \times B \times C$. **b)** $C \times B \times A$.

c) $C \times A \times B$. **d)** $B \times B \times B$.

23. How many different elements does $A \times B$ have if A has m elements and B has n elements?

24. Show that $A \times B \neq B \times A$, when A and B are nonempty, unless $A = B$.

25. Explain why $A \times B \times C$ and $(A \times B) \times C$ are not the same.

26. Translate each of these quantifications into English and determine its truth value.

a) $\forall x \in \mathbf{R}\ (x^2 \neq -1)$ **b)** $\exists x \in \mathbf{Z}\ (x^2 = 2)$

c) $\forall x \in \mathbf{Z}\ (x^2 > 0)$ **d)** $\exists x \in \mathbf{R}\ (x^2 = x)$

27. Find the truth set of each of these predicates where the domain is the set of integers.

a) $P(x)$: $x^2 < 3$ **b)** $Q(x)$: $x^2 > x$

c) $R(x)$: $2x + 1 = 0$

***28.** The defining property of an ordered pair is that two ordered pairs are equal if and only if their first elements are equal and their second elements are equal. Surprisingly, instead of taking the ordered pair as a primitive concept, we can construct ordered pairs using basic notions from set theory. Show that if we define the ordered pair (a, b) to be $\{\{a\}, \{a, b\}\}$, then $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. [*Hint:* First show that $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ if and only if $a = c$ and $b = d$.]

***29.** This exercise presents **Russell's paradox**. Let S be the set that contains a set x if the set x does not belong to itself, so that $S = \{x \mid x \notin x\}$.

a) Show the assumption that S is a member of S leads to a contradiction.

b) Show the assumption that S is not a member of S leads to a contradiction.

By parts (a) and (b) it follows that the set S cannot be defined as it was. This paradox can be avoided by restricting the types of elements that sets can have.

2.2 Set Operations

Introduction

Two, or more, sets can be combined in many different ways. For instance, starting with the set of mathematics majors at your school and the set of computer science majors at your school, we can form the set of students who are mathematics majors or computer science majors, the set of students who are joint majors in mathematics and computer science, the set of all students not majoring in mathematics, and so on.

Links

DEFINITION 1 Let A and B be sets. The *union* of the sets A and B, denoted by $A \cup B$, is the set that contains those elements that are either in A or in B, or in both.

An element x belongs to the union of the sets A and B if and only if x belongs to A or x belongs to B. This tells us that

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

The Venn diagram shown in Figure 1 represents the union of two sets A and B. The area that represents $A \cup B$ is the shaded area within either the circle representing A or the circle representing B.

We will give some examples of the union of sets.

EXAMPLE 1 The union of the sets $\{1, 3, 5\}$ and $\{1, 2, 3\}$ is the set $\{1, 2, 3, 5\}$; that is, $\{1, 3, 5\} \cup \{1, 2, 3\} = \{1, 2, 3, 5\}$. ◀

EXAMPLE 2 The union of the set of all computer science majors at your school and the set of all mathematics majors at your school is the set of students at your school who are majoring either in mathematics or in computer science (or in both). ◀

DEFINITION 2 Let A and B be sets. The *intersection* of the sets A and B, denoted by $A \cap B$, is the set containing those elements in both A and B.

An element x belongs to the intersection of the sets A and B if and only if x belongs to A and x belongs to B. This tells us that

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

The Venn diagram shown in Figure 2 represents the intersection of two sets A and B. The shaded area that is within both the circles representing the sets A and B is the area that represents the intersection of A and B.

We give some examples of the intersection of sets.

EXAMPLE 3 The intersection of the sets $\{1, 3, 5\}$ and $\{1, 2, 3\}$ is the set $\{1, 3\}$; that is, $\{1, 3, 5\} \cap \{1, 2, 3\} = \{1, 3\}$. ◀

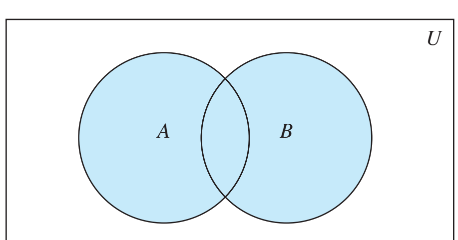
EXAMPLE 4 The intersection of the set of all computer science majors at your school and the set of all mathematics majors is the set of all students who are joint majors in mathematics and computer science. ◀

DEFINITION 3 Two sets are called *disjoint* if their intersection is the empty set.

EXAMPLE 5 Let $A = \{1, 3, 5, 7, 9\}$ and $B = \{2, 4, 6, 8, 10\}$. Because $A \cap B = \emptyset$, A and B are disjoint. ◀

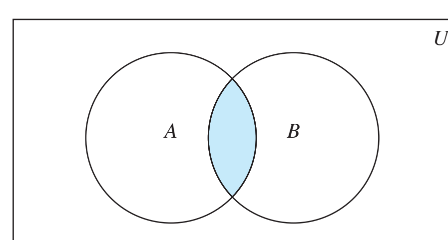
Be careful not to overcount!

We are often interested in finding the cardinality of a union of two finite sets A and B. Note that $|A| + |B|$ counts each element that is in A but not in B or in B but not in A exactly once, and each element that is in both A and B exactly twice. Thus, if the number of elements that



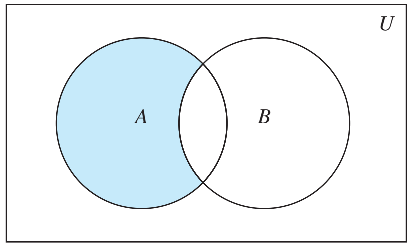
$A \cup B$ is shaded.

FIGURE 1 Venn Diagram of the Union of A and B.



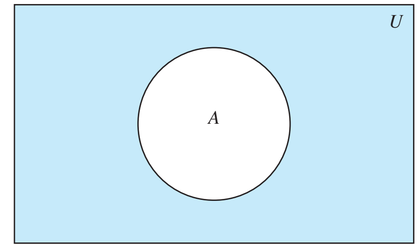
$A \cap B$ is shaded.

FIGURE 2 Venn Diagram of the Intersection of A and B.



$A - B$ is shaded.

FIGURE 3 Venn Diagram for the Difference of A and B.



$\overline{A}$ is shaded.

FIGURE 4 Venn Diagram for the Complement of the Set A.

are in both A and B is subtracted from $|A| + |B|$, elements in $A \cap B$ will be counted only once. Hence,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

The generalization of this result to unions of an arbitrary number of sets is called the **principle of inclusion–exclusion**. The principle of inclusion–exclusion is an important technique used in enumeration. We will discuss this principle and other counting techniques in detail in Chapters 6 and 8.

There are other important ways to combine sets.

DEFINITION 4 Let A and B be sets. The *difference* of A and B, denoted by $A - B$, is the set containing those elements that are in A but not in B. The difference of A and B is also called the *complement of B with respect to A*.

Remark: The difference of sets A and B is sometimes denoted by $A \backslash B$.

An element x belongs to the difference of A and B if and only if $x \in A$ and $x \notin B$. This tells us that

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

The Venn diagram shown in Figure 3 represents the difference of the sets A and B. The shaded area inside the circle that represents A and outside the circle that represents B is the area that represents $A - B$.

We give some examples of differences of sets.

EXAMPLE 6 The difference of $\{1, 3, 5\}$ and $\{1, 2, 3\}$ is the set $\{5\}$; that is, $\{1, 3, 5\} - \{1, 2, 3\} = \{5\}$. This is different from the difference of $\{1, 2, 3\}$ and $\{1, 3, 5\}$, which is the set $\{2\}$. ◀

EXAMPLE 7 The difference of the set of computer science majors at your school and the set of mathematics majors at your school is the set of all computer science majors at your school who are not also mathematics majors. ◀

Once the universal set U has been specified, the **complement** of a set can be defined.

DEFINITION 5 Let U be the universal set. The *complement* of the set A, denoted by $\overline{A}$, is the complement of A with respect to U. Therefore, the complement of the set A is $U - A$.

An element belongs to $\overline{A}$ if and only if $x \notin A$. This tells us that

$$\overline{A} = \{x \in U \mid x \notin A\}.$$

In Figure 4 the shaded area outside the circle representing A is the area representing $\overline{A}$.

We give some examples of the complement of a set.

EXAMPLE 8 Let $A = \{a, e, i, o, u\}$ (where the universal set is the set of letters of the English alphabet). Then $\overline{A} = \{b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z\}$. ◀

EXAMPLE 9 Let A be the set of positive integers greater than 10 (with universal set the set of all positive integers). Then $\overline{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. ◀

It is left to the reader (Exercise 11) to show that we can express the difference of A and B as the intersection of A and the complement of B. That is,

$$A - B = A \cap \overline{B}.$$

Set Identities

Set identities and propositional equivalences are just special cases of identities for Boolean algebra.

Table 1 lists the most important set identities. We will prove several of these identities here, using three different methods. These methods are presented to illustrate that there are often many different approaches to the solution of a problem. The proofs of the remaining identities will be left as exercises. The reader should note the similarity between these set identities and the logical equivalences discussed in Section 1.3. (Compare Table 6 of Section 1.6 and Table 1.) In fact, the set identities given can be proved directly from the corresponding logical equivalences. Furthermore, both are special cases of identities that hold for Boolean algebra (discussed in Chapter 12).

One way to show that two sets are equal is to show that each is a subset of the other. Recall that to show that one set is a subset of a second set, we can show that if an element belongs to the first set, then it must also belong to the second set. We generally use a direct proof to do this. We illustrate this type of proof by establishing the first of De Morgan's laws.

TABLE 1 Set Identities.

Identity	*Name*
$A \cap U = A$ $A \cup \emptyset = A$	Identity laws
$A \cup U = U$ $A \cap \emptyset = \emptyset$	Domination laws
$A \cup A = A$ $A \cap A = A$	Idempotent laws
$\overline{(\overline{A})} = A$	Complementation law
$A \cup B = B \cup A$ $A \cap B = B \cap A$	Commutative laws
$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$	Associative laws
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributive laws
$\overline{A \cap B} = \overline{A} \cup \overline{B}$ $\overline{A \cup B} = \overline{A} \cap \overline{B}$	De Morgan's laws
$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$	Absorption laws
$A \cup \overline{A} = U$ $A \cap \overline{A} = \emptyset$	Complement laws

EXAMPLE 10 Prove that $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

This identity says that the complement of the intersection of two sets is the union of their complements.

Solution: We will prove that the two sets $\overline{A \cap B}$ and $\overline{A} \cup \overline{B}$ are equal by showing that each set is a subset of the other.

First, we will show that $\overline{A \cap B} \subseteq \overline{A} \cup \overline{B}$. We do this by showing that if x is in $\overline{A \cap B}$, then it must also be in $\overline{A} \cup \overline{B}$. Now suppose that $x \in \overline{A \cap B}$. By the definition of complement, $x \notin A \cap B$. Using the definition of intersection, we see that the proposition $\neg((x \in A) \wedge (x \in B))$ is true.

By applying De Morgan's law for propositions, we see that $\neg(x \in A)$ or $\neg(x \in B)$. Using the definition of negation of propositions, we have $x \notin A$ or $x \notin B$. Using the definition of the complement of a set, we see that this implies that $x \in \overline{A}$ or $x \in \overline{B}$. Consequently, by the definition of union, we see that $x \in \overline{A} \cup \overline{B}$. We have now shown that $\overline{A \cap B} \subseteq \overline{A} \cup \overline{B}$.

Extra Examples

Next, we will show that $\overline{A} \cup \overline{B} \subseteq \overline{A \cap B}$. We do this by showing that if x is in $\overline{A} \cup \overline{B}$, then it must also be in $\overline{A \cap B}$. Now suppose that $x \in \overline{A} \cup \overline{B}$. By the definition of union, we know that $x \in \overline{A}$ or $x \in \overline{B}$. Using the definition of complement, we see that $x \notin A$ or $x \notin B$. Consequently, the proposition $\neg(x \in A) \vee \neg(x \in B)$ is true.

By De Morgan's law for propositions, we conclude that $\neg((x \in A) \wedge (x \in B))$ is true. By the definition of intersection, it follows that $\neg(x \in A \cap B)$. We now use the definition of complement to conclude that $x \in \overline{A \cap B}$. This shows that $\overline{A} \cup \overline{B} \subseteq \overline{A \cap B}$.

Because we have shown that each set is a subset of the other, the two sets are equal, and the identity is proved. ◀

We can more succinctly express the reasoning used in Example 10 using set builder notation, as Example 11 illustrates.

EXAMPLE 11 Use set builder notation and logical equivalences to establish the first De Morgan law $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Solution: We can prove this identity with the following steps.

$\overline{A \cap B}$	$= \{x \mid x \notin A \cap B\}$	by definition of complement
	$= \{x \mid \neg(x \in (A \cap B))\}$	by definition of does not belong symbol
	$= \{x \mid \neg(x \in A \wedge x \in B)\}$	by definition of intersection
	$= \{x \mid \neg(x \in A) \vee \neg(x \in B)\}$	by the first De Morgan law for logical equivalences
	$= \{x \mid x \notin A \vee x \notin B\}$	by definition of does not belong symbol
	$= \{x \mid x \in \overline{A} \vee x \in \overline{B}\}$	by definition of complement
	$= \{x \mid x \in \overline{A} \cup \overline{B}\}$	by definition of union
	$= \overline{A} \cup \overline{B}$	by meaning of set builder notation

Note that besides the definitions of complement, union, set membership, and set builder notation, this proof uses the second De Morgan law for logical equivalences. ◀

Proving a set identity involving more than two sets by showing each side of the identity is a subset of the other often requires that we keep track of different cases, as illustrated by the proof in Example 12 of one of the distributive laws for sets.

EXAMPLE 12 Prove the second distributive law from Table 1, which states that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ for all sets A, B, and C.

Solution: We will prove this identity by showing that each side is a subset of the other side.

Suppose that $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. By the definition of union, it follows that $x \in A$, and $x \in B$ or $x \in C$ (or both). In other words, we know that the compound proposition $(x \in A) \wedge ((x \in B) \vee (x \in C))$ is true. By the distributive law for conjunction over disjunction, it follows that $((x \in A) \wedge (x \in B)) \vee ((x \in A) \wedge (x \in C))$. We conclude that either $x \in A$ and $x \in B$, or $x \in A$ and $x \in C$. By the definition of intersection, it follows that $x \in A \cap B$ or $x \in A \cap C$. Using the definition of union, we conclude that $x \in (A \cap B) \cup (A \cap C)$. We conclude that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Now suppose that $x \in (A \cap B) \cup (A \cap C)$. Then, by the definition of union, $x \in A \cap B$ or $x \in A \cap C$. By the definition of intersection, it follows that $x \in A$ and $x \in B$ or that $x \in A$ and $x \in C$. From this we see that $x \in A$, and $x \in B$ or $x \in C$. Consequently, by the definition of union we see that $x \in A$ and $x \in B \cup C$. Furthermore, by the definition of intersection, it follows that $x \in A \cap (B \cup C)$. We conclude that $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. This completes the proof of the identity. ◀

Set identities can also be proved using **membership tables**. We consider each combination of sets that an element can belong to and verify that elements in the same combinations of sets belong to both the sets in the identity. To indicate that an element is in a set, a 1 is used; to indicate that an element is not in a set, a 0 is used. (The reader should note the similarity between membership tables and truth tables.)

EXAMPLE 13 Use a membership table to show that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Solution: The membership table for these combinations of sets is shown in Table 2. This table has eight rows. Because the columns for $A \cap (B \cup C)$ and $(A \cap B) \cup (A \cap C)$ are the same, the identity is valid. ◀

Additional set identities can be established using those that we have already proved. Consider Example 14.

TABLE 2 **A Membership Table for the Distributive Property.**

A	B	C	$B \cup C$	$A \cap (B \cup C)$	$A \cap B$	$A \cap C$	$(A \cap B) \cup (A \cap C)$
1	1	1	1	1	1	1	1
1	1	0	1	1	1	0	1
1	0	1	1	1	0	1	1
1	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0
0	1	0	1	0	0	0	0
0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0

EXAMPLE 14 Let A, B, and C be sets. Show that

$$\overline{A \cup (B \cap C)} = (\overline{C} \cup \overline{B}) \cap \overline{A}.$$

Solution: We have

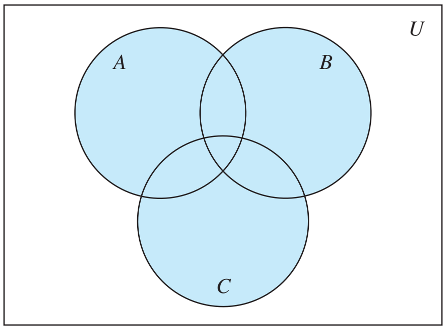
$$\begin{aligned}
\overline{A \cup (B \cap C)} &= \overline{A} \cap (\overline{B \cap C}) && \text{by the first De Morgan law} \\
&= \overline{A} \cap (\overline{B} \cup \overline{C}) && \text{by the second De Morgan law} \\
&= (\overline{B} \cup \overline{C}) \cap \overline{A} && \text{by the commutative law for intersections} \\
&= (\overline{C} \cup \overline{B}) \cap \overline{A} && \text{by the commutative law for unions.}
\end{aligned}$$

◀

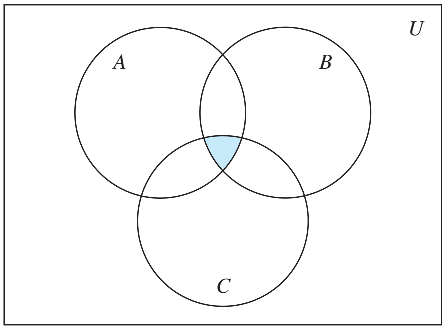
Generalized Unions and Intersections

Because unions and intersections of sets satisfy associative laws, the sets $A \cup B \cup C$ and $A \cap B \cap C$ are well defined; that is, the meaning of this notation is unambiguous when A, B, and C are sets. That is, we do not have to use parentheses to indicate which operation comes first because $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$. Note that $A \cup B \cup C$ contains those elements that are in at least one of the sets A, B, and C, and that $A \cap B \cap C$ contains those elements that are in all of A, B, and C. These combinations of the three sets, A, B, and C, are shown in Figure 5.

EXAMPLE 15 Let $A = \{0, 2, 4, 6, 8\}$, $B = \{0, 1, 2, 3, 4\}$, and $C = \{0, 3, 6, 9\}$. What are $A \cup B \cup C$ and $A \cap B \cap C$?



(a) $A \cup B \cup C$ is shaded.



(b) $A \cap B \cap C$ is shaded.

FIGURE 5 **The Union and Intersection of A, B, and C.**

Solution: The set $A \cup B \cup C$ contains those elements in at least one of A, B, and C. Hence,

$$A \cup B \cup C = \{0, 1, 2, 3, 4, 6, 8, 9\}.$$

The set $A \cap B \cap C$ contains those elements in all three of A, B, and C. Thus,

$$A \cap B \cap C = \{0\}.$$

◀

We can also consider unions and intersections of an arbitrary number of sets. We introduce these definitions.

DEFINITION 6 The *union* of a collection of sets is the set that contains those elements that are members of at least one set in the collection.

We use the notation

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$$

to denote the union of the sets $A_1, A_2, \ldots, A_n$.

DEFINITION 7 The *intersection* of a collection of sets is the set that contains those elements that are members of all the sets in the collection.

We use the notation

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$$

to denote the intersection of the sets $A_1, A_2, \ldots, A_n$. We illustrate generalized unions and intersections with Example 16.

EXAMPLE 16 For $i = 1, 2, \ldots$, let $A_i = \{i, i+1, i+2, \ldots\}$. Then,

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} \{i, i+1, i+2, \ldots\} = \{1, 2, 3, \ldots\},$$

and

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \{i, i+1, i+2, \ldots\} = \{n, n+1, n+2, \ldots\} = A_n.$$

◀

We can extend the notation we have introduced for unions and intersections to other families of sets. In particular, we use the notation

$$A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots = \bigcup_{i=1}^{\infty} A_i$$

to denote the union of the sets $A_1, A_2, \ldots, A_n, \ldots$. Similarly, the intersection of these sets is denoted by

$$A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots = \bigcap_{i=1}^{\infty} A_i.$$

More generally, when I is a set, the notations $\bigcap_{i \in I} A_i$ and $\bigcup_{i \in I} A_i$ are used to denote the intersection and union of the sets A_i for $i \in I$, respectively. Note that we have $\bigcap_{i \in I} A_i = \{x \mid \forall i \in I\ (x \in A_i)\}$ and $\bigcup_{i \in I} A_i = \{x \mid \exists i \in I\ (x \in A_i)\}$.

EXAMPLE 17 Suppose that $A_i = \{1, 2, 3, \ldots, i\}$ for $i = 1, 2, 3, \ldots$. Then,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \{1, 2, 3, \ldots, i\} = \{1, 2, 3, \ldots\} = \mathbf{Z}^+$$

and

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \{1, 2, 3, \ldots, i\} = \{1\}.$$

To see that the union of these sets is the set of positive integers, note that every positive integer n is in at least one of the sets, because it belongs to $A_n = \{1, 2, \ldots, n\}$, and every element of the sets in the union is a positive integer. To see that the intersection of these sets is the set $\{1\}$, note that the only element that belongs to all the sets $A_1, A_2, \ldots$ is 1. To see this note that $A_1 = \{1\}$ and $1 \in A_i$ for $i = 1, 2, \ldots$. ◀

Computer Representation of Sets

There are various ways to represent sets using a computer. One method is to store the elements of the set in an unordered fashion. However, if this is done, the operations of computing the union, intersection, or difference of two sets would be time-consuming, because each of these operations would require a large amount of searching for elements. We will present a method for storing elements using an arbitrary ordering of the elements of the universal set. This method of representing sets makes computing combinations of sets easy.

Assume that the universal set U is finite (and of reasonable size so that the number of elements of U is not larger than the memory size of the computer being used). First, specify an arbitrary ordering of the elements of U, for instance $a_1, a_2, \ldots, a_n$. Represent a subset A of U with the bit string of length n, where the ith bit in this string is 1 if a_i belongs to A and is 0 if a_i does not belong to A. Example 18 illustrates this technique.

EXAMPLE 18 Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and the ordering of elements of U has the elements in increasing order; that is, $a_i = i$. What bit strings represent the subset of all odd integers in U, the subset of all even integers in U, and the subset of integers not exceeding 5 in U?

Solution: The bit string that represents the set of odd integers in U, namely, $\{1, 3, 5, 7, 9\}$, has a one bit in the first, third, fifth, seventh, and ninth positions, and a zero elsewhere. It is

$$10\ 1010\ 1010.$$

(We have split this bit string of length ten into blocks of length four for easy reading.) Similarly, we represent the subset of all even integers in U, namely, $\{2, 4, 6, 8, 10\}$, by the string

$$01\ 0101\ 0101.$$

The set of all integers in U that do not exceed 5, namely, $\{1, 2, 3, 4, 5\}$, is represented by the string

$$11\ 1110\ 0000.$$ ◀

Using bit strings to represent sets, it is easy to find complements of sets and unions, intersections, and differences of sets. To find the bit string for the complement of a set from the bit string for that set, we simply change each 1 to a 0 and each 0 to 1, because $x \in A$ if and only if $x \notin \overline{A}$. Note that this operation corresponds to taking the negation of each bit when we associate a bit with a truth value—with 1 representing true and 0 representing false.

EXAMPLE 19 We have seen that the bit string for the set $\{1, 3, 5, 7, 9\}$ (with universal set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$) is

$$10\ 1010\ 1010.$$

What is the bit string for the complement of this set?

Solution: The bit string for the complement of this set is obtained by replacing 0s with 1s and vice versa. This yields the string

$$01\ 0101\ 0101,$$

which corresponds to the set $\{2, 4, 6, 8, 10\}$. ◀

To obtain the bit string for the union and intersection of two sets we perform bitwise Boolean operations on the bit strings representing the two sets. The bit in the ith position of the bit string of the union is 1 if either of the bits in the ith position in the two strings is 1 (or both are 1), and is 0 when both bits are 0. Hence, the bit string for the union is the bitwise *OR* of the bit strings for the two sets. The bit in the ith position of the bit string of the intersection is 1 when the bits in the corresponding position in the two strings are both 1, and is 0 when either of the two bits is 0 (or both are). Hence, the bit string for the intersection is the bitwise *AND* of the bit strings for the two sets.

EXAMPLE 20 The bit strings for the sets $\{1, 2, 3, 4, 5\}$ and $\{1, 3, 5, 7, 9\}$ are 11 1110 0000 and 10 1010 1010, respectively. Use bit strings to find the union and intersection of these sets.

Solution: The bit string for the union of these sets is

$$11\ 1110\ 0000 \vee 10\ 1010\ 1010 = 11\ 1110\ 1010,$$

which corresponds to the set $\{1, 2, 3, 4, 5, 7, 9\}$. The bit string for the intersection of these sets is

$$11\ 1110\ 0000 \wedge 10\ 1010\ 1010 = 10\ 1010\ 0000,$$

which corresponds to the set $\{1, 3, 5\}$. ◀

Exercises

1. Let A be the set of students who live within one mile of school and let B be the set of students who walk to classes. Describe the students in each of these sets.

a) $A \cap B$ **b)** $A \cup B$
c) $A - B$ **d)** $B - A$

2. Let $A = \{a, b, c, d, e\}$ and $B = \{a, b, c, d, e, f, g, h\}$. Find

a) $A \cup B$. **b)** $A \cap B$.
c) $A - B$. **d)** $B - A$.

In Exercises 3–4 assume that A is a subset of some underlying universal set U.

3. Prove the complementation law in Table 1 by showing that $\overline{\overline{A}} = A$.

4. Prove the complement laws in Table 1 by showing that

a) $A \cup \overline{A} = U$. **b)** $A \cap \overline{A} = \emptyset$.

5. Prove the first absorption law from Table 1 by showing that if A and B are sets, then $A \cup (A \cap B) = A$.

6. Find the sets A and B if $A - B = \{1, 5, 7, 8\}$, $B - A = \{2, 10\}$, and $A \cap B = \{3, 6, 9\}$.

7. Prove the second De Morgan law in Table 1 by showing that if A and B are sets, then $\overline{A \cup B} = \overline{A} \cap \overline{B}$

a) by showing each side is a subset of the other side.
b) using a membership table.

8. Let A and B be sets. Show that

a) $(A \cap B) \subseteq A$. **b)** $A \subseteq (A \cup B)$.
c) $A - B \subseteq A$. **d)** $A \cap (B - A) = \emptyset$.
e) $A \cup (B - A) = A \cup B$.

9. Show that if A, B, and C are sets, then $\overline{A \cap B \cap C} = \overline{A} \cup \overline{B} \cup \overline{C}$

a) by showing each side is a subset of the other side.
b) using a membership table.

10. Let A, B, and C be sets. Show that

a) $(A \cup B) \subseteq (A \cup B \cup C)$.
b) $(A \cap B \cap C) \subseteq (A \cap B)$.
c) $(A - B) - C \subseteq A - C$.
d) $(A - C) \cap (C - B) = \emptyset$.
e) $(B - A) \cup (C - A) = (B \cup C) - A$.

11. Show that if A and B are sets, then

a) $A - B = A \cap \overline{B}$.
b) $(A \cap B) \cup (A \cap \overline{B}) = A$.

12. Show that if A and B are sets with $A \subseteq B$, then

a) $A \cup B = B$.
b) $A \cap B = A$.

13. Let $A = \{0, 2, 4, 6, 8, 10\}$, $B = \{0, 1, 2, 3, 4, 5, 6\}$, and $C = \{4, 5, 6, 7, 8, 9, 10\}$. Find

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

14. Draw the Venn diagrams for each of these combinations of the sets A, B, and C.

a) $A \cap (B \cup C)$ **b)** $\overline{A} \cap \overline{B} \cap \overline{C}$
c) $(A - B) \cup (A - C) \cup (B - C)$

15. What can you say about the sets A and B if we know that

a) $A \cup B = A$? **b)** $A \cap B = A$?
c) $A - B = A$? **d)** $A \cap B = B \cap A$?
e) $A - B = B - A$?

16. Can you conclude that $A = B$ if A, B, and C are sets such that

a) $A \cup C = B \cup C$? **b)** $A \cap C = B \cap C$?
c) $A \cup C = B \cup C$ and $A \cap C = B \cap C$?

17. Let A and B be subsets of a universal set U. Show that $A \subseteq B$ if and only if $\overline{B} \subseteq \overline{A}$.

The **symmetric difference** of A and B, denoted by $A \oplus B$, is the set containing those elements in either A or B, but not in both A and B.

18. Find the symmetric difference of $\{1, 3, 5\}$ and $\{1, 2, 3\}$.

19. Find the symmetric difference of the set of computer science majors at a school and the set of mathematics majors at this school.

20. Draw a Venn diagram for the symmetric difference of the sets A and B.

21. What can you say about the sets A and B if $A \oplus B = A$?

***22.** Determine whether the symmetric difference is associative; that is, if A, B, and C are sets, does it follow that $A \oplus (B \oplus C) = (A \oplus B) \oplus C$?

***23.** Suppose that A, B, and C are sets such that $A \oplus C = B \oplus C$. Must it be the case that $A = B$?

24. Show that if A and B are finite sets, then $A \cup B$ is a finite set.

25. Show that if A is an infinite set, then whenever B is a set, $A \cup B$ is also an infinite set.

***26.** Show that if A, B, and C are sets, then

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

(This is a special case of the inclusion–exclusion principle, which will be studied in Chapter 8.)

27. Let $A_i = \{1, 2, 3, \ldots, i\}$ for $i = 1, 2, 3, \ldots$. Find

a) $\bigcup_{i=1}^{n} A_i$. **b)** $\bigcap_{i=1}^{n} A_i$.

28. Let $A_i = \{\ldots, -2, -1, 0, 1, \ldots, i\}$. Find

a) $\bigcup_{i=1}^{n} A_i$. **b)** $\bigcap_{i=1}^{n} A_i$.

29. Let A_i be the set of all nonempty bit strings (that is, bit strings of length at least one) of length not exceeding i. Find

a) $\bigcup_{i=1}^{n} A_i$. **b)** $\bigcap_{i=1}^{n} A_i$.

30. Find $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$ if for every positive integer i,

a) $A_i = \{i, i+1, i+2, \ldots\}$.

b) $A_i = \{0, i\}$.

c) $A_i = (0, i)$, that is, the set of real numbers x with $0 < x < i$.

d) $A_i = (i, \infty)$, that is, the set of real numbers x with $x > i$.

31. Find $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$ if for every positive integer i,

a) $A_i = \{-i, -i+1, \ldots, -1, 0, 1, \ldots, i-1, i\}$.

b) $A_i = \{-i, i\}$.

c) $A_i = [-i, i]$, that is, the set of real numbers x with $-i \leq x \leq i$.

d) $A_i = [i, \infty)$, that is, the set of real numbers x with $x \geq i$.

32. Suppose that the universal set is $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Express each of these sets with bit strings where the ith bit in the string is 1 if i is in the set and 0 otherwise.

a) $\{3, 4, 5\}$

b) $\{1, 3, 6, 10\}$

c) $\{2, 3, 4, 7, 8, 9\}$

33. Using the same universal set as in the last problem, find the set specified by each of these bit strings.

a) 11 1100 1111

b) 01 0111 1000

c) 10 0000 0001

34. What is the bit string corresponding to the symmetric difference of two sets?

35. Show how bitwise operations on bit strings can be used to find these combinations of $A = \{a, b, c, d, e\}$, $B = \{b, c, d, g, p, t, v\}$, $C = \{c, e, i, o, u, x, y, z\}$, and $D = \{d, e, h, i, n, o, t, u, x, y\}$.

a) $A \cup B$ **b)** $A \cap B$

c) $(A \cup D) \cap (B \cup C)$ **d)** $A \cup B \cup C \cup D$

36. How can the union and intersection of n sets that all are subsets of the universal set U be found using bit strings?

The **successor** of the set A is the set $A \cup \{A\}$.

37. Find the successors of the following sets.

a) $\{1, 2, 3\}$ **b)** $\emptyset$

c) $\{\emptyset\}$ **d)** $\{\emptyset, \{\emptyset\}\}$

38. How many elements does the successor of a set with n elements have?

Sometimes the number of times that an element occurs in an unordered collection matters. **Multisets** are unordered collections of elements where an element can occur as a member more than once. The notation $\{m_1 \cdot a_1, m_2 \cdot a_2, \ldots, m_r \cdot a_r\}$ denotes the multiset with element a_1 occurring m_1 times, element a_2 occurring m_2 times, and so on. The numbers m_i, $i = 1, 2, \ldots, r$ are called the **multiplicities** of the elements a_i, $i = 1, 2, \ldots, r$.

Let P and Q be multisets. The **union** of the multisets P and Q is the multiset where the multiplicity of an element is the maximum of its multiplicities in P and Q. The **intersection** of P and Q is the multiset where the multiplicity of an element is the minimum of its multiplicities in P and Q. The **difference** of P and Q is the multiset where the multiplicity of an element is the multiplicity of the element in P less its multiplicity in Q unless this difference is negative, in which case the multiplicity is 0. The **sum** of P and Q is the multiset where the multiplicity of an element is the sum of multiplicities in P and Q. The union, intersection, and difference of P and Q are denoted by $P \cup Q$, $P \cap Q$, and $P - Q$, respectively (where these operations should not be confused with the analogous operations for sets). The sum of P and Q is denoted by $P + Q$.

39. Let A and B be the multisets $\{3 \cdot a, 2 \cdot b, 1 \cdot c\}$ and $\{2 \cdot a, 3 \cdot b, 4 \cdot d\}$, respectively. Find

a) $A \cup B$. **b)** $A \cap B$. **c)** $A - B$.
d) $B - A$. **e)** $A + B$.

Fuzzy sets are used in artificial intelligence. Each element in the universal set U has a **degree of membership**, which is a real number between 0 and 1 (including 0 and 1), in a fuzzy set S. The fuzzy set S is denoted by listing the elements with their degrees of membership (elements with 0 degree of membership are not listed). For instance, we write {0.6 Alice, 0.9 Brian, 0.4 Fred, 0.1 Oscar, 0.5 Rita} for the set F (of famous people) to indicate that Alice has a 0.6 degree of membership in F, Brian has a 0.9 degree of membership in F, Fred has a 0.4 degree of membership in F, Oscar has a 0.1 degree of membership in F, and Rita has a 0.5 degree of membership in F (so that Brian is the most famous and Oscar is the least famous of these people). Also suppose that R is the set of rich people with $R =$ {0.4 Alice, 0.8 Brian, 0.2 Fred, 0.9 Oscar, 0.7 Rita}.

40. The **complement** of a fuzzy set S is the set $\overline{S}$, with the degree of the membership of an element in $\overline{S}$ equal to 1 minus the degree of membership of this element in S. Find $\overline{F}$ (the fuzzy set of people who are not famous) and $\overline{R}$ (the fuzzy set of people who are not rich).

41. The **union** of two fuzzy sets S and T is the fuzzy set $S \cup T$, where the degree of membership of an element in $S \cup T$ is the maximum of the degrees of membership of this element in S and in T. Find the fuzzy set $F \cup R$ of rich or famous people.

42. The **intersection** of two fuzzy sets S and T is the fuzzy set $S \cap T$, where the degree of membership of an element in $S \cap T$ is the minimum of the degrees of membership of this element in S and in T. Find the fuzzy set $F \cap R$ of rich and famous people.

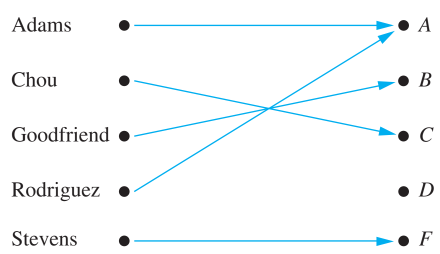


FIGURE 1 Assignment of Grades in a Discrete Mathematics Class.

2.3 Functions

Introduction

In many instances we assign to each element of a set a particular element of a second set (which may be the same as the first). For example, suppose that each student in a discrete mathematics class is assigned a letter grade from the set $\{A, B, C, D, F\}$. And suppose that the grades are A for Adams, C for Chou, B for Goodfriend, A for Rodriguez, and F for Stevens. This assignment of grades is illustrated in Figure 1.

This assignment is an example of a function. The concept of a function is extremely important in mathematics and computer science. For example, in discrete mathematics functions are used in the definition of such discrete structures as sequences and strings. Functions are also used to represent how long it takes a computer to solve problems of a given size. Many computer programs and subroutines are designed to calculate values of functions. Recursive functions, which are functions defined in terms of themselves, are used throughout computer science; they will be studied in Chapter 5. This section reviews the basic concepts involving functions needed in discrete mathematics.

DEFINITION 1

Let A and B be nonempty sets. A *function* f from A to B is an assignment of exactly one element of B to each element of A. We write $f(a) = b$ if b is the unique element of B assigned by the function f to the element a of A. If f is a function from A to B, we write $f : A \to B$.

Remark: Functions are sometimes also called **mappings** or **transformations**.

Assessment

Functions are specified in many different ways. Sometimes we explicitly state the assignments, as in Figure 1. Often we give a formula, such as $f(x) = x + 1$, to define a function. Other times we use a computer program to specify a function.

A function $f : A \to B$ can also be defined in terms of a relation from A to B. Recall from Section 2.1 that a relation from A to B is just a subset of $A \times B$. A relation from A to B that contains one, and only one, ordered pair (a, b) for every element $a \in A$, defines a function f from A to B. This function is defined by the assignment $f(a) = b$, where (a, b) is the unique ordered pair in the relation that has a as its first element.

DEFINITION 2

If f is a function from A to B, we say that A is the *domain* of f and B is the *codomain* of f. If $f(a) = b$, we say that b is the *image* of a and a is a *preimage* of b. The *range*, or *image*, of f is the set of all images of elements of A. Also, if f is a function from A to B, we say that f *maps* A to B.

Figure 2 represents a function f from A to B.

When we define a function we specify its domain, its codomain, and the mapping of elements of the domain to elements in the codomain. Two functions are **equal** when they have the same domain, have the same codomain, and map each element of their common domain to the same element in their common codomain. Note that if we change either the domain or the codomain of a function, then we obtain a different function. If we change the mapping of elements, then we also obtain a different function.

Examples 1–5 provide examples of functions. In each case, we describe the domain, the codomain, the range, and the assignment of values to elements of the domain.

EXAMPLE 1 What are the domain, codomain, and range of the function that assigns grades to students described in the first paragraph of the introduction of this section?

Solution: Let G be the function that assigns a grade to a student in our discrete mathematics class. Note that $G(\text{Adams}) = A$, for instance. The domain of G is the set {Adams, Chou, Goodfriend, Rodriguez, Stevens}, and the codomain is the set $\{A, B, C, D, F\}$. The range of G is the set $\{A, B, C, F\}$, because each grade except D is assigned to some student. ◀

EXAMPLE 2 Let R be the relation with ordered pairs (Abdul, 22), (Brenda, 24), (Carla, 21), (Desire, 22), (Eddie, 24), and (Felicia, 22). Here each pair consists of a graduate student and this student's age. Specify a function determined by this relation.

Solution: If f is a function specified by R, then $f(\text{Abdul}) = 22$, $f(\text{Brenda}) = 24$, $f(\text{Carla}) = 21$, $f(\text{Desire}) = 22$, $f(\text{Eddie}) = 24$, and $f(\text{Felicia}) = 22$. (Here, $f(x)$ is the age of x, where x is a student.) For the domain, we take the set {Abdul, Brenda, Carla, Desire, Eddie, Felicia}. We also need to specify a codomain, which needs to contain all possible ages of students. Because it is highly likely that all students are less than 100 years old, we can take the set of positive integers less than 100 as the codomain. (Note that we could choose a different codomain, such as the set of all positive integers or the set of positive integers between 10 and 90, but that would change the function. Using this codomain will also allow us to extend the function by adding the names and ages of more students later.) The range of the function we have specified is the set of different ages of these students, which is the set {21, 22, 24}. ◀

EXAMPLE 3 Extra Examples

Let f be the function that assigns the last two bits of a bit string of length 2 or greater to that string. For example, $f(11010) = 10$. Then, the domain of f is the set of all bit strings of length 2 or greater, and both the codomain and range are the set {00, 01, 10, 11}. ◀

EXAMPLE 4 Let $f: \mathbf{Z} \to \mathbf{Z}$ assign the square of an integer to this integer. Then, $f(x) = x^2$, where the domain of f is the set of all integers, the codomain of f is the set of all integers, and the range of f is the set of all integers that are perfect squares, namely, $\{0, 1, 4, 9, \ldots\}$. ◀

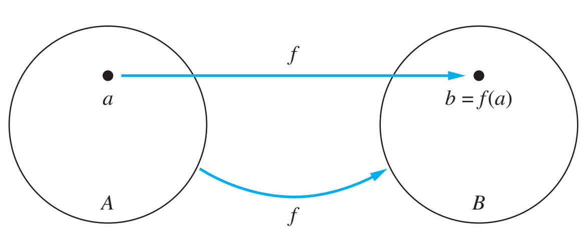


FIGURE 2 **The Function *f* Maps *A* to *B*.**

EXAMPLE 5 The domain and codomain of functions are often specified in programming languages. For instance, the Java statement

int **floor**(float real){. . .}

and the C++ function statement

int **function** (float x){. . .}

both tell us that the domain of the floor function is the set of real numbers (represented by floating point numbers) and its codomain is the set of integers. ◀

A function is called **real-valued** if its codomain is the set of real numbers, and it is called **integer-valued** if its codomain is the set of integers. Two real-valued functions or two integer-valued functions with the same domain can be added, as well as multiplied.

DEFINITION 3 Let f_1 and f_2 be functions from A to $\mathbf{R}$. Then $f_1 + f_2$ and $f_1 f_2$ are also functions from A to $\mathbf{R}$ defined for all $x \in A$ by

$$(f_1 + f_2)(x) = f_1(x) + f_2(x),$$
$$(f_1 f_2)(x) = f_1(x) f_2(x).$$

Note that the functions $f_1 + f_2$ and $f_1 f_2$ have been defined by specifying their values at x in terms of the values of f_1 and f_2 at x.

EXAMPLE 6 Let f_1 and f_2 be functions from $\mathbf{R}$ to $\mathbf{R}$ such that $f_1(x) = x^2$ and $f_2(x) = x - x^2$. What are the functions $f_1 + f_2$ and $f_1 f_2$?

Solution: From the definition of the sum and product of functions, it follows that

$$(f_1 + f_2)(x) = f_1(x) + f_2(x) = x^2 + (x - x^2) = x$$

and

$$(f_1 f_2)(x) = x^2(x - x^2) = x^3 - x^4.$$ ◀

When f is a function from A to B, the image of a subset of A can also be defined.

DEFINITION 4 Let f be a function from A to B and let S be a subset of A. The *image* of S under the function f is the subset of B that consists of the images of the elements of S. We denote the image of S by $f(S)$, so

$$f(S) = \{t \mid \exists s \in S \ (t = f(s))\}.$$

We also use the shorthand $\{f(s) \mid s \in S\}$ to denote this set.

Remark: The notation $f(S)$ for the image of the set S under the function f is potentially ambiguous. Here, $f(S)$ denotes a set, and not the value of the function f for the set S.

EXAMPLE 7 Let $A = \{a, b, c, d, e\}$ and $B = \{1, 2, 3, 4\}$ with $f(a) = 2$, $f(b) = 1$, $f(c) = 4$, $f(d) = 1$, and $f(e) = 1$. The image of the subset $S = \{b, c, d\}$ is the set $f(S) = \{1, 4\}$. ◀

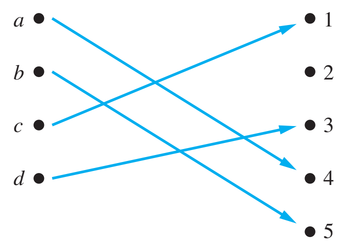


FIGURE 3 **A One-to-One Function.**

One-to-One and Onto Functions

Some functions never assign the same value to two different domain elements. These functions are said to be **one-to-one**.

DEFINITION 5 A function f is said to be *one-to-one,* or an *injunction,* if and only if $f(a) = f(b)$ implies that $a = b$ for all a and b in the domain of f. A function is said to be *injective* if it is one-to-one.

Note that a function f is one-to-one if and only if $f(a) \neq f(b)$ whenever $a \neq b$. This way of expressing that f is one-to-one is obtained by taking the contrapositive of the implication in the definition.

Remark: We can express that f is one-to-one using quantifiers as $\forall a \forall b (f(a) = f(b) \rightarrow a = b)$ or equivalently $\forall a \forall b (a \neq b \rightarrow f(a) \neq f(b))$, where the universe of discourse is the domain of the function.

Assessment

We illustrate this concept by giving examples of functions that are one-to-one and other functions that are not one-to-one.

EXAMPLE 8 Determine whether the function f from $\{a, b, c, d\}$ to $\{1, 2, 3, 4, 5\}$ with $f(a) = 4$, $f(b) = 5$, $f(c) = 1$, and $f(d) = 3$ is one-to-one.

Extra Examples

Solution: The function f is one-to-one because f takes on different values at the four elements of its domain. This is illustrated in Figure 3. ◀

EXAMPLE 9 Determine whether the function $f(x) = x^2$ from the set of integers to the set of integers is one-to-one.

Solution: The function $f(x) = x^2$ is not one-to-one because, for instance, $f(1) = f(-1) = 1$, but $1 \neq -1$.

Note that the function $f(x) = x^2$ with its domain restricted to $\mathbf{Z}^+$ is one-to-one. (Technically, when we restrict the domain of a function, we obtain a new function whose values agree with those of the original function for the elements of the restricted domain. The restricted function is not defined for elements of the original domain outside of the restricted domain.) ◀

EXAMPLE 10 Determine whether the function $f(x) = x + 1$ from the set of real numbers to itself is one-to-one.

Solution: The function $f(x) = x + 1$ is a one-to-one function. To demonstrate this, note that $x + 1 \neq y + 1$ when $x \neq y$. ◀

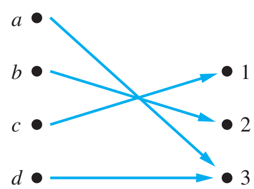


FIGURE 4 **An Onto Function.**

EXAMPLE 11 Suppose that each worker in a group of employees is assigned a job from a set of possible jobs, each to be done by a single worker. In this situation, the function f that assigns a job to each worker is one-to-one. To see this, note that if x and y are two different workers, then $f(x) \neq f(y)$ because the two workers x and y must be assigned different jobs. ◀

We now give some conditions that guarantee that a function is one-to-one.

DEFINITION 6 A function f whose domain and codomain are subsets of the set of real numbers is called *increasing* if $f(x) \leq f(y)$, and *strictly increasing* if $f(x) < f(y)$, whenever $x < y$ and x and y are in the domain of f. Similarly, f is called *decreasing* if $f(x) \geq f(y)$, and *strictly decreasing* if $f(x) > f(y)$, whenever $x < y$ and x and y are in the domain of f. (The word *strictly* in this definition indicates a strict inequality.)

Remark: A function f is increasing if $\forall x \forall y (x < y \to f(x) \leq f(y))$, strictly increasing if $\forall x \forall y (x < y \to f(x) < f(y))$, decreasing if $\forall x \forall y (x < y \to f(x) \geq f(y))$, and strictly decreasing if $\forall x \forall y (x < y \to f(x) > f(y))$, where the universe of discourse is the domain of f.

From these definitions, it can be shown that a function that is either strictly increasing or strictly decreasing must be one-to-one. However, a function that is increasing, but not strictly increasing, or decreasing, but not strictly decreasing, is not one-to-one.

For some functions the range and the codomain are equal. That is, every member of the codomain is the image of some element of the domain. Functions with this property are called **onto** functions.

DEFINITION 7 A function f from A to B is called *onto,* or a *surjection,* if and only if for every element $b \in B$ there is an element $a \in A$ with $f(a) = b$. A function f is called *surjective* if it is onto.

Remark: A function f is onto if $\forall y \exists x (f(x) = y)$, where the domain for x is the domain of the function and the domain for y is the codomain of the function.

We now give examples of onto functions and functions that are not onto.

EXAMPLE 12 Let f be the function from $\{a, b, c, d\}$ to $\{1, 2, 3\}$ defined by $f(a) = 3$, $f(b) = 2$, $f(c) = 1$, and $f(d) = 3$. Is f an onto function?

Extra Examples

Solution: Because all three elements of the codomain are images of elements in the domain, we see that f is onto. This is illustrated in Figure 4. Note that if the codomain were $\{1, 2, 3, 4\}$, then f would not be onto. ◀

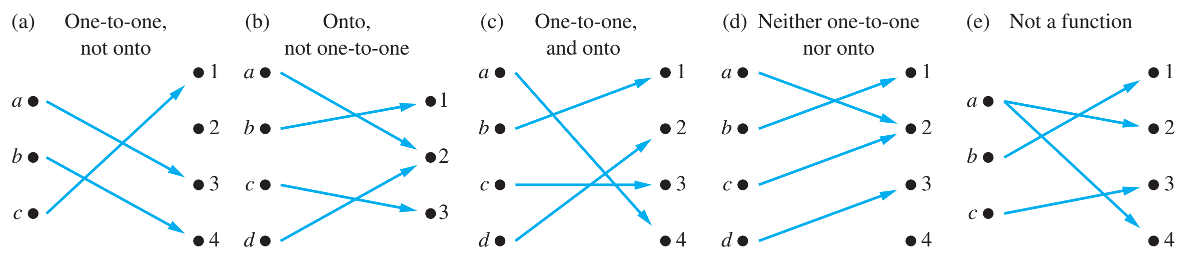


FIGURE 5 Examples of Different Types of Correspondences.

EXAMPLE 13 Is the function $f(x) = x^2$ from the set of integers to the set of integers onto?

Solution: The function f is not onto because there is no integer x with $x^2 = -1$, for instance. ◀

EXAMPLE 14 Is the function $f(x) = x + 1$ from the set of integers to the set of integers onto?

Solution: This function is onto, because for every integer y there is an integer x such that $f(x) = y$. To see this, note that $f(x) = y$ if and only if $x + 1 = y$, which holds if and only if $x = y - 1$. ◀

EXAMPLE 15 Consider the function f in Example 11 that assigns jobs to workers. The function f is onto if for every job there is a worker assigned this job. The function f is not onto when there is at least one job that has no worker assigned it. ◀

DEFINITION 8 The function f is a *one-to-one correspondence,* or a *bijection,* if it is both one-to-one and onto. We also say that such a function is *bijective*.

Examples 16 and 17 illustrate the concept of a bijection.

EXAMPLE 16 Let f be the function from $\{a, b, c, d\}$ to $\{1, 2, 3, 4\}$ with $f(a) = 4$, $f(b) = 2$, $f(c) = 1$, and $f(d) = 3$. Is f a bijection?

Solution: The function f is one-to-one and onto. It is one-to-one because no two values in the domain are assigned the same function value. It is onto because all four elements of the codomain are images of elements in the domain. Hence, f is a bijection. ◀

Figure 5 displays four functions where the first is one-to-one but not onto, the second is onto but not one-to-one, the third is both one-to-one and onto, and the fourth is neither one-to-one nor onto. The fifth correspondence in Figure 5 is not a function, because it sends an element to two different elements.

Suppose that f is a function from a set A to itself. If A is finite, then f is one-to-one if and only if it is onto. (This follows from the result in Exercise 48.) This is not necessarily the case if A is infinite (as will be shown in Section 2.5).

EXAMPLE 17 Let A be a set. The *identity function* on A is the function $\iota_A : A \to A$, where

$$\iota_A(x) = x$$

for all $x \in A$. In other words, the identity function ι_A is the function that assigns each element to itself. The function ι_A is one-to-one and onto, so it is a bijection. (Note that ι is the Greek letter iota.) ◀

For future reference, we summarize what needs be to shown to establish whether a function is one-to-one and whether it is onto. It is instructive to review Examples 8–17 in light of this summary.

To show that f is injective Show that if $f(x) = f(y)$ for arbitrary $x, y \in A$ with $x \neq y$, then $x = y$.

To show that f is not injective Find particular elements $x, y \in A$ such that $x \neq y$ and $f(x) = f(y)$.

To show that f is surjective Consider an arbitrary element $y \in B$ and find an element $x \in A$ such that $f(x) = y$.

To show that f is not surjective Find a particular $y \in B$ such that $f(x) \neq y$ for all $x \in A$.

Inverse Functions and Compositions of Functions

Now consider a one-to-one correspondence f from the set A to the set B. Because f is an onto function, every element of B is the image of some element in A. Furthermore, because f is also a one-to-one function, every element of B is the image of a *unique* element of A. Consequently, we can define a new function from B to A that reverses the correspondence given by f. This leads to Definition 9.

DEFINITION 9 Let f be a one-to-one correspondence from the set A to the set B. The *inverse function* of f is the function that assigns to an element b belonging to B the unique element a in A such that $f(a) = b$. The inverse function of f is denoted by f^{-1}. Hence, $f^{-1}(b) = a$ when $f(a) = b$.

Remark: Be sure not to confuse the function f^{-1} with the function $1/f$, which is the function that assigns to each x in the domain the value $1/f(x)$. Notice that the latter makes sense only when $f(x)$ is a non-zero real number.

Figure 6 illustrates the concept of an inverse function.

If a function f is not a one-to-one correspondence, we cannot define an inverse function of f. When f is not a one-to-one correspondence, either it is not one-to-one or it is not onto. If

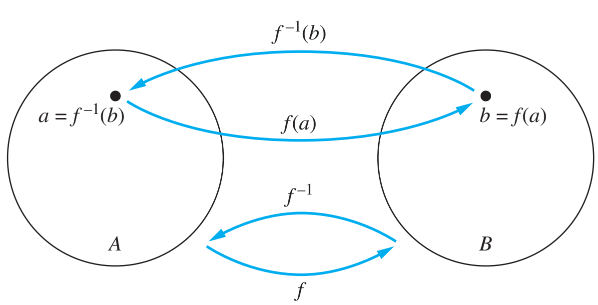


FIGURE 6 The Function f^{-1} Is the Inverse of Function f.

f is not one-to-one, some element b in the codomain is the image of more than one element in the domain. If f is not onto, for some element b in the codomain, no element a in the domain exists for which $f(a) = b$. Consequently, if f is not a one-to-one correspondence, we cannot assign to each element b in the codomain a unique element a in the domain such that $f(a) = b$ (because for some b there is either more than one such a or no such a).

A one-to-one correspondence is called **invertible** because we can define an inverse of this function. A function is **not invertible** if it is not a one-to-one correspondence, because the inverse of such a function does not exist.

EXAMPLE 18 Let f be the function from $\{a, b, c\}$ to $\{1, 2, 3\}$ such that $f(a) = 2$, $f(b) = 3$, and $f(c) = 1$. Is f invertible, and if it is, what is its inverse?

Solution: The function f is invertible because it is a one-to-one correspondence. The inverse function f^{-1} reverses the correspondence given by f, so $f^{-1}(1) = c$, $f^{-1}(2) = a$, and $f^{-1}(3) = b$. ◀

EXAMPLE 19 Let $f : \mathbf{Z} \to \mathbf{Z}$ be such that $f(x) = x + 1$. Is f invertible, and if it is, what is its inverse?

Solution: The function f has an inverse because it is a one-to-one correspondence, as follows from Examples 10 and 14. To reverse the correspondence, suppose that y is the image of x, so that $y = x + 1$. Then $x = y - 1$. This means that $y - 1$ is the unique element of $\mathbf{Z}$ that is sent to y by f. Consequently, $f^{-1}(y) = y - 1$. ◀

EXAMPLE 20 Let f be the function from $\mathbf{R}$ to $\mathbf{R}$ with $f(x) = x^2$. Is f invertible?

Solution: Because $f(-2) = f(2) = 4$, f is not one-to-one. If an inverse function were defined, it would have to assign two elements to 4. Hence, f is not invertible. (Note we can also show that f is not invertible because it is not onto.) ◀

Sometimes we can restrict the domain or the codomain of a function, or both, to obtain an invertible function, as Example 21 illustrates.

EXAMPLE 21 Show that if we restrict the function $f(x) = x^2$ in Example 20 to a function from the set of all nonnegative real numbers to the set of all nonnegative real numbers, then f is invertible.

Solution: The function $f(x) = x^2$ from the set of nonnegative real numbers to the set of nonnegative real numbers is one-to-one. To see this, note that if $f(x) = f(y)$, then $x^2 = y^2$, so $x^2 - y^2 = (x + y)(x - y) = 0$. This means that $x + y = 0$ or $x - y = 0$, so $x = -y$ or $x = y$. Because both x and y are nonnegative, we must have $x = y$. So, this function is one-to-one. Furthermore, $f(x) = x^2$ is onto when the codomain is the set of all nonnegative real numbers, because each nonnegative real number has a square root. That is, if y is a nonnegative real number, there exists a nonnegative real number x such that $x = \sqrt{y}$, which means that $x^2 = y$. Because the function $f(x) = x^2$ from the set of nonnegative real numbers to the set of nonnegative real numbers is one-to-one and onto, it is invertible. Its inverse is given by the rule $f^{-1}(y) = \sqrt{y}$. ◀

DEFINITION 10 Let g be a function from the set A to the set B and let f be a function from the set B to the set C. The *composition* of the functions f and g, denoted for all $a \in A$ by $f \circ g$, is defined by

$$(f \circ g)(a) = f(g(a)).$$

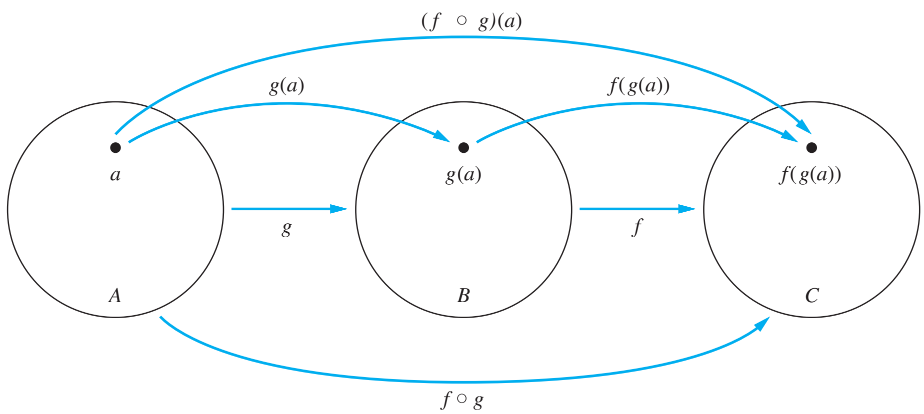


FIGURE 7 The Composition of the Functions f and g.

In other words, $f \circ g$ is the function that assigns to the element a of A the element assigned by f to $g(a)$. That is, to find $(f \circ g)(a)$ we first apply the function g to a to obtain $g(a)$ and then we apply the function f to the result $g(a)$ to obtain $(f \circ g)(a) = f(g(a))$. Note that the composition $f \circ g$ cannot be defined unless the range of g is a subset of the domain of f. In Figure 7 the composition of functions is shown.

EXAMPLE 22 Let g be the function from the set $\{a, b, c\}$ to itself such that $g(a) = b$, $g(b) = c$, and $g(c) = a$. Let f be the function from the set $\{a, b, c\}$ to the set $\{1, 2, 3\}$ such that $f(a) = 3$, $f(b) = 2$, and $f(c) = 1$. What is the composition of f and g, and what is the composition of g and f?

Solution: The composition $f \circ g$ is defined by $(f \circ g)(a) = f(g(a)) = f(b) = 2$, $(f \circ g)(b) = f(g(b)) = f(c) = 1$, and $(f \circ g)(c) = f(g(c)) = f(a) = 3$.

Note that $g \circ f$ is not defined, because the range of f is not a subset of the domain of g. ◀

EXAMPLE 23 Let f and g be the functions from the set of integers to the set of integers defined by $f(x) = 2x + 3$ and $g(x) = 3x + 2$. What is the composition of f and g? What is the composition of g and f?

Solution: Both the compositions $f \circ g$ and $g \circ f$ are defined. Moreover,

$$(f \circ g)(x) = f(g(x)) = f(3x + 2) = 2(3x + 2) + 3 = 6x + 7$$

and

$$(g \circ f)(x) = g(f(x)) = g(2x + 3) = 3(2x + 3) + 2 = 6x + 11.$$ ◀

Remark: Note that even though $f \circ g$ and $g \circ f$ are defined for the functions f and g in Example 23, $f \circ g$ and $g \circ f$ are not equal. In other words, the commutative law does not hold for the composition of functions.

When the composition of a function and its inverse is formed, in either order, an identity function is obtained. To see this, suppose that f is a one-to-one correspondence from the set A to the set B. Then the inverse function f^{-1} exists and is a one-to-one correspondence from B to A. The inverse function reverses the correspondence of the original function, so $f^{-1}(b) = a$ when $f(a) = b$, and $f(a) = b$ when $f^{-1}(b) = a$. Hence,

$$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a,$$

and

$$(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b.$$

Consequently $f^{-1} \circ f = \iota_A$ and $f \circ f^{-1} = \iota_B$, where ι_A and ι_B are the identity functions on the sets A and B, respectively. That is, $(f^{-1})^{-1} = f$.

The Graphs of Functions

We can associate a set of pairs in $A \times B$ to each function from A to B. This set of pairs is called the **graph** of the function and is often displayed pictorially to aid in understanding the behavior of the function.

DEFINITION 11 Let f be a function from the set A to the set B. The *graph* of the function f is the set of ordered pairs $\{(a, b) \mid a \in A \text{ and } f(a) = b\}$.

From the definition, the graph of a function f from A to B is the subset of $A \times B$ containing the ordered pairs with the second entry equal to the element of B assigned by f to the first entry. Also, note that the graph of a function f from A to B is the same as the relation from A to B determined by the function f, as described on page 139.

EXAMPLE 24 Display the graph of the function $f(n) = 2n + 1$ from the set of integers to the set of integers.

Solution: The graph of f is the set of ordered pairs of the form $(n, 2n + 1)$, where n is an integer. This graph is displayed in Figure 8. ◀

EXAMPLE 25 Display the graph of the function $f(x) = x^2$ from the set of integers to the set of integers.

Solution: The graph of f is the set of ordered pairs of the form $(x, f(x)) = (x, x^2)$, where x is an integer. This graph is displayed in Figure 9. ◀

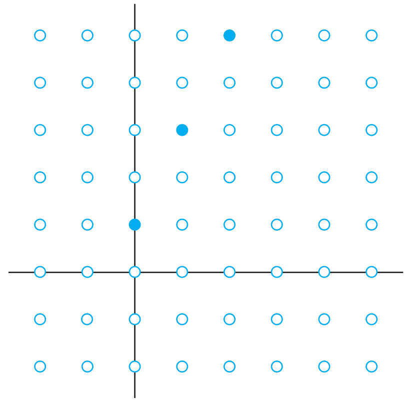


FIGURE 8 **The Graph of $f(n) = 2n + 1$ from Z to Z.**

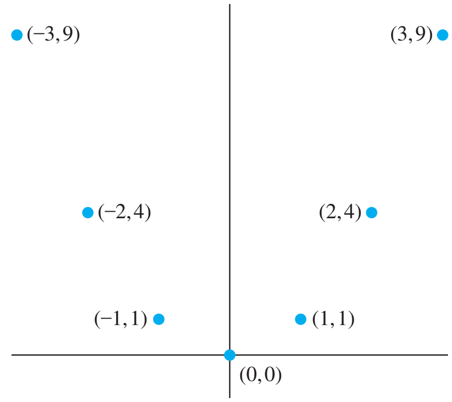


FIGURE 9 **The Graph of $f(x) = x^2$ from Z to Z.**

Some Important Functions

Next, we introduce two important functions in discrete mathematics, namely, the floor and ceiling functions. Let x be a real number. The floor function rounds x down to the closest integer less than or equal to x, and the ceiling function rounds x up to the closest integer greater than or equal to x. These functions are often used when objects are counted. They play an important role in the analysis of the number of steps used by procedures to solve problems of a particular size.

DEFINITION 12 The *floor function* assigns to the real number x the largest integer that is less than or equal to x. The value of the floor function at x is denoted by $\lfloor x \rfloor$. The *ceiling function* assigns to the real number x the smallest integer that is greater than or equal to x. The value of the ceiling function at x is denoted by $\lceil x \rceil$.

Remark: The floor function is often also called the *greatest integer function*. It is often denoted by $[x]$.

EXAMPLE 26 These are some values of the floor and ceiling functions:

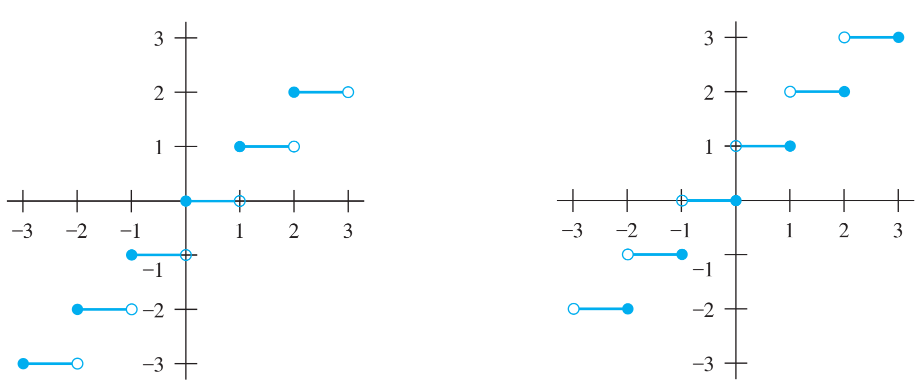
$$\lfloor \tfrac{1}{2} \rfloor = 0,\ \lceil \tfrac{1}{2} \rceil = 1,\ \lfloor -\tfrac{1}{2} \rfloor = -1,\ \lceil -\tfrac{1}{2} \rceil = 0,\ \lfloor 3.1 \rfloor = 3,\ \lceil 3.1 \rceil = 4,\ \lfloor 7 \rfloor = 7,\ \lceil 7 \rceil = 7.$$ ◀

We display the graphs of the floor and ceiling functions in Figure 10. In Figure 10(a) we display the graph of the floor function $\lfloor x \rfloor$. Note that this function has the same value throughout the interval $[n, n+1)$, namely n, and then it jumps up to $n+1$ when $x = n+1$. In Figure 10(b) we display the graph of the ceiling function $\lceil x \rceil$. Note that this function has the same value throughout the interval $(n, n+1]$, namely $n+1$, and then jumps to $n+2$ when x is a little larger than $n+1$.

The floor and ceiling functions are useful in a wide variety of applications, including those involving data storage and data transmission. Consider Examples 27 and 28, typical of basic calculations done when database and data communications problems are studied.

EXAMPLE 27 Data stored on a computer disk or transmitted over a data network are usually represented as a string of bytes. Each byte is made up of 8 bits. How many bytes are required to encode 100 bits of data?

Solution: To determine the number of bytes needed, we determine the smallest integer that is at least as large as the quotient when 100 is divided by 8, the number of bits in a byte. Consequently, $\lceil 100/8 \rceil = \lceil 12.5 \rceil = 13$ bytes are required. ◀



(a) $y = \lfloor x \rfloor$ (b) $y = \lceil x \rceil$

FIGURE 10 **Graphs of the (a) Floor and (b) Ceiling Functions.**

TABLE 1 Useful Properties of the Floor and Ceiling Functions.
(n is an integer, x is a real number)

(1a) $\lfloor x \rfloor = n$ if and only if $n \leq x < n+1$
(1b) $\lceil x \rceil = n$ if and only if $n-1 < x \leq n$
(1c) $\lfloor x \rfloor = n$ if and only if $x-1 < n \leq x$
(1d) $\lceil x \rceil = n$ if and only if $x \leq n < x+1$
(2) $x-1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x+1$
(3a) $\lfloor -x \rfloor = -\lceil x \rceil$
(3b) $\lceil -x \rceil = -\lfloor x \rfloor$
(4a) $\lfloor x+n \rfloor = \lfloor x \rfloor + n$
(4b) $\lceil x+n \rceil = \lceil x \rceil + n$

EXAMPLE 28 In asynchronous transfer mode (ATM) (a communications protocol used on backbone networks), data are organized into cells of 53 bytes. How many ATM cells can be transmitted in 1 minute over a connection that transmits data at the rate of 500 kilobits per second?

Solution: In 1 minute, this connection can transmit $500{,}000 \cdot 60 = 30{,}000{,}000$ bits. Each ATM cell is 53 bytes long, which means that it is $53 \cdot 8 = 424$ bits long. To determine the number of cells that can be transmitted in 1 minute, we determine the largest integer not exceeding the quotient when 30,000,000 is divided by 424. Consequently, $\lfloor 30{,}000{,}000/424 \rfloor = 70{,}754$ ATM cells can be transmitted in 1 minute over a 500 kilobit per second connection. ◀

Table 1, with x denoting a real number, displays some simple but important properties of the floor and ceiling functions. Because these functions appear so frequently in discrete mathematics, it is useful to look over these identities. Each property in this table can be established using the definitions of the floor and ceiling functions. Properties (1a), (1b), (1c), and (1d) follow directly from these definitions. For example, (1a) states that $\lfloor x \rfloor = n$ if and only if the integer n is less than or equal to x and $n+1$ is larger than x. This is precisely what it means for n to be the greatest integer not exceeding x, which is the definition of $\lfloor x \rfloor = n$. Properties (1b), (1c), and (1d) can be established similarly. We will prove property (4a) using a direct proof.

Proof: Suppose that $\lfloor x \rfloor = m$, where m is a positive integer. By property (1a), it follows that $m \leq x < m+1$. Adding n to all three quantities in this chain of two inequalities shows that $m + n \leq x + n < m + n + 1$. Using property (1a) again, we see that $\lfloor x+n \rfloor = m+n = \lfloor x \rfloor + n$. This completes the proof. Proofs of the other properties are left as exercises. ◁

The floor and ceiling functions enjoy many other useful properties besides those displayed in Table 1. There are also many statements about these functions that may appear to be correct, but actually are not. We will consider statements about the floor and ceiling functions in Examples 29 and 30.

A useful approach for considering statements about the floor function is to let $x = n + \epsilon$, where $n = \lfloor x \rfloor$ is an integer, and ϵ, the fractional part of x, satisfies the inequality $0 \leq \epsilon < 1$. Similarly, when considering statements about the ceiling function, it is useful to write $x = n - \epsilon$, where $n = \lceil x \rceil$ is an integer and $0 \leq \epsilon < 1$.

EXAMPLE 29 Prove that if x is a real number, then $\lfloor 2x \rfloor = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor$.

Extra Examples

Solution: To prove this statement we let $x = n + \epsilon$, where n is an integer and $0 \le \epsilon < 1$. There are two cases to consider, depending on whether ϵ is less than, or greater than or equal to $\frac{1}{2}$. (The reason we choose these two cases will be made clear in the proof.)

We first consider the case when $0 \le \epsilon < \frac{1}{2}$. In this case, $2x = 2n + 2\epsilon$ and $\lfloor 2x \rfloor = 2n$ because $0 \le 2\epsilon < 1$. Similarly, $x + \frac{1}{2} = n + (\frac{1}{2} + \epsilon)$, so $\lfloor x + \frac{1}{2} \rfloor = n$, because $0 < \frac{1}{2} + \epsilon < 1$. Consequently, $\lfloor 2x \rfloor = 2n$ and $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = n + n = 2n$.

Next, we consider the case when $\frac{1}{2} \le \epsilon < 1$. In this case, $2x = 2n + 2\epsilon = (2n + 1) + (2\epsilon - 1)$. Because $0 \le 2\epsilon - 1 < 1$, it follows that $\lfloor 2x \rfloor = 2n + 1$. Because $\lfloor x + \frac{1}{2} \rfloor = \lfloor n + (\frac{1}{2} + \epsilon) \rfloor = \lfloor n + 1 + (\epsilon - \frac{1}{2}) \rfloor$ and $0 \le \epsilon - \frac{1}{2} < 1$, it follows that $\lfloor x + \frac{1}{2} \rfloor = n + 1$. Consequently, $\lfloor 2x \rfloor = 2n + 1$ and $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = n + (n + 1) = 2n + 1$. This concludes the proof. ◀

EXAMPLE 30 Prove or disprove that $\lceil x + y \rceil = \lceil x \rceil + \lceil y \rceil$ for all real numbers x and y.

Solution: Although this statement may appear reasonable, it is false. A counterexample is supplied by $x = \frac{1}{2}$ and $y = \frac{1}{2}$. With these values we find that $\lceil x + y \rceil = \lceil \frac{1}{2} + \frac{1}{2} \rceil = \lceil 1 \rceil = 1$, but $\lceil x \rceil + \lceil y \rceil = \lceil \frac{1}{2} \rceil + \lceil \frac{1}{2} \rceil = 1 + 1 = 2$. ◀

There are certain types of functions that will be used throughout the text. These include polynomial, logarithmic, and exponential functions. A brief review of the properties of these functions needed in this text is given in Appendix 2. In this book the notation $\log x$ will be used to denote the logarithm to the base 2 of x, because 2 is the base that we will usually use for logarithms. We will denote logarithms to the base b, where b is any real number greater than 1, by $\log_b x$, and the natural logarithm by $\ln x$.

Another function we will use throughout this text is the **factorial function** $f : \mathbf{N} \to \mathbf{Z}^+$, denoted by $f(n) = n!$. The value of $f(n) = n!$ is the product of the first n positive integers, so $f(n) = 1 \cdot 2 \cdots (n - 1) \cdot n$ [and $f(0) = 0! = 1$].

EXAMPLE 31 We have $f(1) = 1! = 1$, $f(2) = 2! = 1 \cdot 2 = 2$, $f(6) = 6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$, and $f(20) = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 13 \cdot 14 \cdot 15 \cdot 16 \cdot 17 \cdot 18 \cdot 19 \cdot 20 =$ 2,432,902,008,176,640,000. ◀

Example 31 illustrates that the factorial function grows extremely rapidly as n grows. The rapid growth of the factorial function is made clearer by Stirling's formula, a result from

Links

JAMES STIRLING (1692–1770) James Stirling was born near the town of Stirling, Scotland. His family strongly supported the Jacobite cause of the Stuarts as an alternative to the British crown. The first information known about James is that he entered Balliol College, Oxford, on a scholarship in 1711. However, he later lost his scholarship when he refused to pledge his allegiance to the British crown. The first Jacobean rebellion took place in 1715, and Stirling was accused of communicating with rebels. He was charged with cursing King George, but he was acquitted of these charges. Even though he could not graduate from Oxford because of his politics, he remained there for several years. Stirling published his first work, which extended Newton's work on plane curves, in 1717. He traveled to Venice, where a chair of mathematics had been promised to him, an appointment that unfortunately fell through. Nevertheless, Stirling stayed in Venice, continuing his mathematical work. He attended the University of Padua in 1721, and in 1722 he returned to Glasgow. Stirling apparently fled Italy after learning the secrets of the Italian glass industry, avoiding the efforts of Italian glass makers to assassinate him to protect their secrets.

In late 1724 Stirling moved to London, staying there 10 years teaching mathematics and actively engaging in research. In 1730 he published *Methodus Differentialis*, his most important work, presenting results on infinite series, summations, interpolation, and quadrature. It is in this book that his asymptotic formula for $n!$ appears. Stirling also worked on gravitation and the shape of the earth; he stated, but did not prove, that the earth is an oblate spheroid. Stirling returned to Scotland in 1735, when he was appointed manager of a Scottish mining company. He was very successful in this role and even published a paper on the ventilation of mine shafts. He continued his mathematical research, but at a reduced pace, during his years in the mining industry. Stirling is also noted for surveying the River Clyde with the goal of creating a series of locks to make it navigable. In 1752 the citizens of Glasgow presented him with a silver teakettle as a reward for this work.

higher mathematics that tell us that $n! \sim \sqrt{2\pi n}(n/e)^n$. Here, we have used the notation $f(n) \sim g(n)$, which means that the ratio $f(n)/g(n)$ approaches 1 as n grows without bound (that is, $\lim_{n\to\infty} f(n)/g(n) = 1$). The symbol $\sim$ is read "is asymptotic to." Stirling's formula is named after James Stirling, a Scottish mathematician of the eighteenth century.

Partial Functions

A program designed to evaluate a function may not produce the correct value of the function for all elements in the domain of this function. For example, a program may not produce a correct value because evaluating the function may lead to an infinite loop or an overflow. Similarly, in abstract mathematics, we often want to discuss functions that are defined only for a subset of the real numbers, such as $1/x$, $\sqrt{x}$, and $\arcsin(x)$. Also, we may want to use such notions as the "youngest child" function, which is undefined for a couple having no children, or the "time of sunrise," which is undefined for some days above the Arctic Circle. To study such situations, we use the concept of a partial function.

DEFINITION 13 A *partial function* f from a set A to a set B is an assignment to each element a in a subset of A, called the *domain of definition* of f, of a unique element b in B. The sets A and B are called the *domain* and *codomain* of f, respectively. We say that f is *undefined* for elements in A that are not in the domain of definition of f. When the domain of definition of f equals A, we say that f is a *total function.*

Remark: We write $f : A \to B$ to denote that f is a partial function from A to B. Note that this is the same notation as is used for functions. The context in which the notation is used determines whether f is a partial function or a total function.

EXAMPLE 32 The function $f : \mathbf{Z} \to \mathbf{R}$ where $f(n) = \sqrt{n}$ is a partial function from $\mathbf{Z}$ to $\mathbf{R}$ where the domain of definition is the set of nonnegative integers. Note that f is undefined for negative integers. ◀

Exercises

1. Why is f not a function from $\mathbf{R}$ to $\mathbf{R}$ if
- **a)** $f(x) = 1/x$?
- **b)** $f(x) = \sqrt{x}$?
- **c)** $f(x) = \pm\sqrt{(x^2+1)}$?

2. Determine whether f is a function from $\mathbf{Z}$ to $\mathbf{R}$ if
- **a)** $f(n) = \pm n$.
- **b)** $f(n) = \sqrt{n^2+1}$.
- **c)** $f(n) = 1/(n^2-4)$.

3. Determine whether f is a function from the set of all bit strings to the set of integers if
- **a)** $f(S)$ is the position of a 0 bit in S.
- **b)** $f(S)$ is the number of 1 bits in S.
- **c)** $f(S)$ is the smallest integer i such that the ith bit of S is 1 and $f(S) = 0$ when S is the empty string, the string with no bits.

4. Find the domain and range of these functions. Note that in each case, to find the domain, determine the set of elements assigned values by the function.
- **a)** the function that assigns to each nonnegative integer its last digit
- **b)** the function that assigns the next largest integer to a positive integer
- **c)** the function that assigns to a bit string the number of one bits in the string
- **d)** the function that assigns to a bit string the number of bits in the string

5. Find these values.
- **a)** $\lceil \frac{3}{4} \rceil$
- **b)** $\lfloor \frac{7}{8} \rfloor$
- **c)** $\lceil -\frac{3}{4} \rceil$
- **d)** $\lfloor -\frac{7}{8} \rfloor$
- **e)** $\lceil 3 \rceil$
- **f)** $\lfloor -1 \rfloor$
- **g)** $\lfloor \frac{1}{2} + \lceil \frac{3}{2} \rceil \rfloor$
- **h)** $\lfloor \frac{1}{2} \cdot \lfloor \frac{5}{2} \rfloor \rfloor$

6. Determine whether each of these functions from $\{a, b, c, d\}$ to itself is one-to-one.
- **a)** $f(a) = b$, $f(b) = a$, $f(c) = c$, $f(d) = d$
- **b)** $f(a) = b$, $f(b) = b$, $f(c) = d$, $f(d) = c$
- **c)** $f(a) = d$, $f(b) = b$, $f(c) = c$, $f(d) = d$

7. Which functions in Exercise 6 are onto?

8. Determine whether each of these functions from $\mathbf{Z}$ to $\mathbf{Z}$ is one-to-one.
- **a)** $f(n) = n - 1$
- **b)** $f(n) = n^2 + 1$
- **c)** $f(n) = n^3$
- **d)** $f(n) = \lceil n/2 \rceil$

9. Which functions in Exercise 8 are onto?

10. Consider these functions from the set of students in a discrete mathematics class. Under what conditions is the function one-to-one if it assigns to a student his or her

a) mobile phone number.

b) student identification number.

c) final grade in the class.

d) home town.

11. Consider these functions from the set of teachers in a school. Under what conditions is the function one-to-one if it assigns to a teacher his or her

a) office.

b) assigned bus to chaperone in a group of buses taking students on a field trip.

c) salary.

d) social security number.

12. Specify a codomain for each of the functions in Exercise 16. Under what conditions is each of these functions with the codomain you specified onto?

13. Specify a codomain for each of the functions in Exercise 11. Under what conditions is each of the functions with the codomain you specified onto?

14. Give an example of a function from **N** to **N** that is

a) one-to-one but not onto.

b) onto but not one-to-one.

c) both onto and one-to-one (but different from the identity function).

d) neither one-to-one nor onto.

15. Determine whether each of these functions is a bijection from **R** to **R**.

a) $f(x) = 2x + 1$

b) $f(x) = x^2 + 1$

c) $f(x) = x^3$

d) $f(x) = (x^2 + 1)/(x^2 + 2)$

16. Show that the function $f(x) = e^x$ from the set of real numbers to the set of real numbers is not invertible, but if the codomain is restricted to the set of positive real numbers, the resulting function is invertible.

17. Show that the function $f(x) = |x|$ from the set of real numbers to the set of nonnegative real numbers is not invertible, but if the domain is restricted to the set of nonnegative real numbers, the resulting function is invertible.

18. Let $S = \{-1, 0, 2, 4, 7\}$. Find $f(S)$ if

a) $f(x) = 1$.

b) $f(x) = 2x + 1$.

c) $f(x) = \lceil x/5 \rceil$.

d) $f(x) = \lfloor (x^2 + 1)/3 \rfloor$.

19. Suppose that g is a function from A to B and f is a function from B to C.

a) Show that if both f and g are one-to-one functions, then $f \circ g$ is also one-to-one.

b) Show that if both f and g are onto functions, then $f \circ g$ is also onto.

***20.** If f and $f \circ g$ are one-to-one, does it follow that g is one-to-one? Justify your answer.

***21.** If f and $f \circ g$ are onto, does it follow that g is onto? Justify your answer.

22. Find $f \circ g$ and $g \circ f$, where $f(x) = x^2 + 1$ and $g(x) = x + 2$, are functions from **R** to **R**.

23. Find $f + g$ and fg for the functions f and g given in Exercise 22.

24. Let $f(x) = ax + b$ and $g(x) = cx + d$, where a, b, c, and d are constants. Determine necessary and sufficient conditions on the constants a, b, c, and d so that $f \circ g = g \circ f$.

25. Show that the function $f(x) = ax + b$ from **R** to **R** is invertible, where a and b are constants, with $a \neq 0$, and find the inverse of f.

26. Let f be a function from the set A to the set B. Let S and T be subsets of A. Show that

a) $f(S \cup T) = f(S) \cup f(T)$.

b) $f(S \cap T) \subseteq f(S) \cap f(T)$.

27. a) Give an example to show that the inclusion in part (b) in Exercise 26 may be proper.

b) Show that if f is one-to-one, the inclusion in part (b) in Exercise 26 is an equality.

Let f be a function from the set A to the set B. Let S be a subset of B. We define the **inverse image** of S to be the subset of A whose elements are precisely all pre-images of all elements of S. We denote the inverse image of S by $f^{-1}(S)$, so $f^{-1}(S) = \{a \in A \mid f(a) \in S\}$. (*Beware:* The notation f^{-1} is used in two different ways. Do not confuse the notation introduced here with the notation $f^{-1}(y)$ for the value at y of the inverse of the invertible function f. Notice also that $f^{-1}(S)$, the inverse image of the set S, makes sense for all functions f, not just invertible functions.)

28. Let f be the function from **R** to **R** defined by $f(x) = x^2$. Find

a) $f^{-1}(\{1\})$.

b) $f^{-1}(\{x \mid 0 < x < 1\})$.

c) $f^{-1}(\{x \mid x > 4\})$.

29. Show that $\lceil x - \frac{1}{2} \rceil$ is the closest integer to the number x, except when x is midway between two integers, when it is the smaller of these two integers.

30. Show that if x is a real number, then $\lceil x \rceil - \lfloor x \rfloor = 1$ if x is not an integer and $\lceil x \rceil - \lfloor x \rfloor = 0$ if x is an integer.

31. Show that if x is a real number, then $x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1$.

32. Show that if x is a real number and m is an integer, then $\lceil x + m \rceil = \lceil x \rceil + m$.

33. Show that if x is a real number and n is an integer, then

a) $x < n$ if and only if $\lfloor x \rfloor < n$.

b) $n < x$ if and only if $n < \lceil x \rceil$.

34. Show that if x is a real number and n is an integer, then

a) $x \leq n$ if and only if $\lceil x \rceil \leq n$.

b) $n \leq x$ if and only if $n \leq \lfloor x \rfloor$.

35. Prove that if n is an integer, then $\lfloor n/2 \rfloor = n/2$ if n is even and $(n - 1)/2$ if n is odd.

36. Prove that if x is a real number, then $\lfloor -x \rfloor = -\lceil x \rceil$ and $\lceil -x \rceil = -\lfloor x \rfloor$.

37. The function INT is found on some calculators, where $\text{INT}(x) = \lfloor x \rfloor$ when x is a nonnegative real number and $\text{INT}(x) = \lceil x \rceil$ when x is a negative real number. Show that this INT function satisfies the identity $\text{INT}(-x) = -\text{INT}(x)$.

38. Let a and b be real numbers with $a < b$. Use the floor and/or ceiling functions to express the number of integers n that satisfy the inequality $a \le n \le b$.

39. Let a and b be real numbers with $a < b$. Use the floor and/or ceiling functions to express the number of integers n that satisfy the inequality $a < n < b$.

40. How many bytes are required to encode n bits of data where n equals

a) 4? **b)** 10? **c)** 500? **d)** 3000?

41. How many bytes are required to encode n bits of data where n equals

a) 7? **b)** 17? **c)** 1001? **d)** 28,800?

42. Draw the graph of the function $f(n) = 1 - n^2$ from **Z** to **Z**.

43. Draw the graph of the function $f(x) = \lfloor 2x \rfloor$ from **R** to **R**.

44. Draw graphs of each of these functions.

a) $f(x) = \lceil 3x - 2 \rceil$ **b)** $f(x) = \lceil 0.2x \rceil$
c) $f(x) = \lfloor -1/x \rfloor$ **d)** $f(x) = \lfloor x^2 \rfloor$
e) $f(x) = \lceil x/2 \rceil \lfloor x/2 \rfloor$ **f)** $f(x) = \lfloor x/2 \rfloor + \lceil x/2 \rfloor$
g) $f(x) = \lfloor 2\lceil x/2 \rceil + \frac{1}{2} \rfloor$

45. Find the inverse function of $f(x) = x^3 + 1$.

46. Suppose that f is an invertible function from Y to Z and g is an invertible function from X to Y. Show that the inverse of the composition $f \circ g$ is given by $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

47. Let S be a subset of a universal set U. The **characteristic function** f_S of S is the function from U to the set $\{0, 1\}$ such that $f_S(x) = 1$ if x belongs to S and $f_S(x) = 0$ if x does not belong to S. Let A and B be sets. Show that for all $x \in U$,

a) $f_{A\cap B}(x) = f_A(x) \cdot f_B(x)$
b) $f_{A\cup B}(x) = f_A(x) + f_B(x) - f_A(x) \cdot f_B(x)$
c) $f_{\overline{A}}(x) = 1 - f_A(x)$
d) $f_{A\oplus B}(x) = f_A(x) + f_B(x) - 2f_A(x)f_B(x)$

48. Suppose that f is a function from A to B, where A and B are finite sets with $|A| = |B|$. Show that f is one-to-one if and only if it is onto.

49. Prove or disprove each of these statements about the floor and ceiling functions.

a) $\lfloor \lceil x \rceil \rfloor = \lceil x \rceil$ for all real numbers x.
b) $\lfloor x + y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$ for all real numbers x and y.
c) $\lceil \lceil x/2 \rceil /2 \rceil = \lceil x/4 \rceil$ for all real numbers x.
d) $\lfloor \sqrt{\lceil x \rceil} \rfloor = \lfloor \sqrt{x} \rfloor$ for all positive real numbers x.
e) $\lfloor x \rfloor + \lfloor y \rfloor + \lfloor x + y \rfloor \le \lfloor 2x \rfloor + \lfloor 2y \rfloor$ for all real numbers x and y.

50. Prove that if x is a positive real number, then

a) $\lfloor \sqrt{\lfloor x \rfloor} \rfloor = \lfloor \sqrt{x} \rfloor$.
b) $\lceil \sqrt{\lceil x \rceil} \rceil = \lceil \sqrt{x} \rceil$.

51. For each of these partial functions, determine its domain, codomain, domain of definition, and the set of values for which it is undefined. Also, determine whether it is a total function.

a) $f : \mathbf{Z} \to \mathbf{R}$, $f(n) = 1/n$
b) $f : \mathbf{Z} \to \mathbf{Z}$, $f(n) = \lceil n/2 \rceil$
c) $f : \mathbf{Z} \times \mathbf{Z} \to \mathbf{Q}$, $f(m, n) = m/n$
d) $f : \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$, $f(m, n) = mn$
e) $f : \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$, $f(m, n) = m - n$ if $m > n$

52. **a)** Show that a partial function from A to B can be viewed as a function f^* from A to $B \cup \{u\}$, where u is not an element of B and

$$f^*(a) = \begin{cases} f(a) & \text{if } a \text{ belongs to the domain of definition of } f \\ u & \text{if } f \text{ is undefined at } a. \end{cases}$$

b) Using the construction in (a), find the function f^* corresponding to each partial function in Exercise 77.

53. **a)** Show that if a set S has cardinality m, where m is a positive integer, then there is a one-to-one correspondence between S and the set $\{1, 2, \ldots, m\}$.

b) Show that if S and T are two sets each with m elements, where m is a positive integer, then there is a one-to-one correspondence between S and T.

2.4 Sequences and Summations

Introduction

Sequences are ordered lists of elements, used in discrete mathematics in many ways. For example, they can be used to represent solutions to certain counting problems, as we will see in Chapter 8. They are also an important data structure in computer science. We will often need to work with sums of terms of sequences in our study of discrete mathematics. This section reviews the use of summation notation, basic properties of summations, and formulas for the sums of terms of some particular types of sequences.

The terms of a sequence can be specified by providing a formula for each term of the sequence. In this section we describe another way to specify the terms of a sequence using a recurrence relation, which expresses each term as a combination of the previous terms. We will introduce one method, known as iteration, for finding a closed formula for the terms of a sequence specified via a recurrence relation. Identifying a sequence when the first few terms are provided is a useful skill when solving problems in discrete mathematics. We will provide some tips, including a useful tool on the Web, for doing so.

Sequences

A sequence is a discrete structure used to represent an ordered list. For example, 1, 2, 3, 5, 8 is a sequence with five terms and $1, 3, 9, 27, 81, \ldots, 3^n, \ldots$ is an infinite sequence.

DEFINITION 1 A *sequence* is a function from a subset of the set of integers (usually either the set $\{0, 1, 2, \ldots\}$ or the set $\{1, 2, 3, \ldots\}$) to a set S. We use the notation a_n to denote the image of the integer n. We call a_n a *term* of the sequence.

We use the notation $\{a_n\}$ to describe the sequence. (Note that a_n represents an individual term of the sequence $\{a_n\}$. Be aware that the notation $\{a_n\}$ for a sequence conflicts with the notation for a set. However, the context in which we use this notation will always make it clear when we are dealing with sets and when we are dealing with sequences. Moreover, although we have used the letter a in the notation for a sequence, other letters or expressions may be used depending on the sequence under consideration. That is, the choice of the letter a is arbitrary.)

We describe sequences by listing the terms of the sequence in order of increasing subscripts.

EXAMPLE 1 Consider the sequence $\{a_n\}$, where

$$a_n = \frac{1}{n}.$$

The list of the terms of this sequence, beginning with a_1, namely,

$$a_1, a_2, a_3, a_4, \ldots,$$

starts with

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots.$$

◀

DEFINITION 2 A *geometric progression* is a sequence of the form

$$a, ar, ar^2, \ldots, ar^n, \ldots$$

where the *initial term a* and the *common ratio r* are real numbers.

Remark: A geometric progression is a discrete analogue of the exponential function $f(x) = ar^x$.

EXAMPLE 2 The sequences $\{b_n\}$ with $b_n = (-1)^n$, $\{c_n\}$ with $c_n = 2 \cdot 5^n$, and $\{d_n\}$ with $d_n = 6 \cdot (1/3)^n$ are geometric progressions with initial term and common ratio equal to 1 and -1; 2 and 5; and 6 and 1/3, respectively, if we start at $n = 0$. The list of terms $b_0, b_1, b_2, b_3, b_4, \ldots$ begins with

$$1, -1, 1, -1, 1, \ldots;$$

the list of terms $c_0, c_1, c_2, c_3, c_4, \ldots$ begins with

$$2, 10, 50, 250, 1250, \ldots;$$

and the list of terms $d_0, d_1, d_2, d_3, d_4, \ldots$ begins with

$$6, 2, \frac{2}{3}, \frac{2}{9}, \frac{2}{27}, \ldots.$$ ◀

DEFINITION 3 An *arithmetic progression* is a sequence of the form

$$a, a + d, a + 2d, \ldots, a + nd, \ldots$$

where the *initial term* a and the *common difference* d are real numbers.

Remark: An arithmetic progression is a discrete analogue of the linear function $f(x) = dx + a$.

EXAMPLE 3 The sequences $\{s_n\}$ with $s_n = -1 + 4n$ and $\{t_n\}$ with $t_n = 7 - 3n$ are both arithmetic progressions with initial terms and common differences equal to -1 and 4, and 7 and -3, respectively, if we start at $n = 0$. The list of terms $s_0, s_1, s_2, s_3, \ldots$ begins with

$$-1, 3, 7, 11, \ldots,$$

and the list of terms $t_0, t_1, t_2, t_3, \ldots$ begins with

$$7, 4, 1, -2, \ldots.$$ ◀

Sequences of the form $a_1, a_2, \ldots, a_n$ are often used in computer science. These finite sequences are also called **strings**. This string is also denoted by $a_1a_2 \ldots a_n$. (Recall that bit strings, which are finite sequences of bits, were introduced in Section 1.1.) The **length** of a string is the number of terms in this string. The **empty string**, denoted by λ, is the string that has no terms. The empty string has length zero.

EXAMPLE 4 The string *abcd* is a string of length four. ◀

Recurrence Relations

In Examples 1–3 we specified sequences by providing explicit formulas for their terms. There are many other ways to specify a sequence. For example, another way to specify a sequence is to provide one or more initial terms together with a rule for determining subsequent terms from those that precede them.

DEFINITION 4 A *recurrence relation* for the sequence $\{a_n\}$ is an equation that expresses a_n in terms of one or more of the previous terms of the sequence, namely, $a_0, a_1, \ldots, a_{n-1}$, for all integers n with $n \geq n_0$, where n_0 is a nonnegative integer. A sequence is called a *solution* of a recurrence relation if its terms satisfy the recurrence relation. (A recurrence relation is said to *recursively define* a sequence. We will explain this alternative terminology in Chapter 5.)

EXAMPLE 5 Let $\{a_n\}$ be a sequence that satisfies the recurrence relation $a_n = a_{n-1} + 3$ for $n = 1, 2, 3, \ldots,$ and suppose that $a_0 = 2$. What are a_1, a_2, and a_3?

Solution: We see from the recurrence relation that $a_1 = a_0 + 3 = 2 + 3 = 5$. It then follows that $a_2 = 5 + 3 = 8$ and $a_3 = 8 + 3 = 11$. ◀

EXAMPLE 6 Let $\{a_n\}$ be a sequence that satisfies the recurrence relation $a_n = a_{n-1} - a_{n-2}$ for $n = 2, 3, 4, \ldots$, and suppose that $a_0 = 3$ and $a_1 = 5$. What are a_2 and a_3?

Solution: We see from the recurrence relation that $a_2 = a_1 - a_0 = 5 - 3 = 2$ and $a_3 = a_2 - a_1 = 2 - 5 = -3$. We can find a_4, a_5, and each successive term in a similar way. ◀

The **initial conditions** for a recursively defined sequence specify the terms that precede the first term where the recurrence relation takes effect. For instance, the initial condition in Example 5 is $a_0 = 2$, and the initial conditions in Example 6 are $a_0 = 3$ and $a_1 = 5$. Using mathematical induction, a proof technique introduced in Chapter 5, it can be shown that a recurrence relation together with its initial conditions determines a unique solution.

Hop along to Chapter 8 to learn how to find a formula for the Fibonacci numbers.

Next, we define a particularly useful sequence defined by a recurrence relation, known as the **Fibonacci sequence**, after the Italian mathematician Fibonacci who was born in the 12th century (see Chapter 5 for his biography). We will study this sequence in depth in Chapters 5 and 8, where we will see why it is important for many applications, including modeling the population growth of rabbits.

DEFINITION 5

Links

The *Fibonacci sequence*, $f_0, f_1, f_2, \ldots$, is defined by the initial conditions $f_0 = 0$, $f_1 = 1$, and the recurrence relation

$$f_n = f_{n-1} + f_{n-2}$$

for $n = 2, 3, 4, \ldots$.

EXAMPLE 7 Find the Fibonacci numbers f_2, f_3, f_4, f_5, and f_6.

Solution: The recurrence relation for the Fibonacci sequence tells us that we find successive terms by adding the previous two terms. Because the initial conditions tell us that $f_0 = 0$ and $f_1 = 1$, using the recurrence relation in the definition we find that

$$\begin{aligned} f_2 &= f_1 + f_0 = 1 + 0 = 1, \\ f_3 &= f_2 + f_1 = 1 + 1 = 2, \\ f_4 &= f_3 + f_2 = 2 + 1 = 3, \\ f_5 &= f_4 + f_3 = 3 + 2 = 5, \\ f_6 &= f_5 + f_4 = 5 + 3 = 8. \end{aligned}$$ ◀

EXAMPLE 8 Suppose that $\{a_n\}$ is the sequence of integers defined by $a_n = n!$, the value of the factorial function at the integer n, where $n = 1, 2, 3, \ldots$. Because $n! = n((n-1)(n-2)\ldots 2 \cdot 1) = n(n-1)! = na_{n-1}$, we see that the sequence of factorials satisfies the recurrence relation $a_n = na_{n-1}$, together with the initial condition $a_1 = 1$. ◀

We say that we have solved the recurrence relation together with the initial conditions when we find an explicit formula, called a **closed formula**, for the terms of the sequence.

EXAMPLE 9 Determine whether the sequence $\{a_n\}$, where $a_n = 3n$ for every nonnegative integer n, is a solution of the recurrence relation $a_n = 2a_{n-1} - a_{n-2}$ for $n = 2, 3, 4, \ldots$. Answer the same question where $a_n = 2^n$ and where $a_n = 5$.

Solution: Suppose that $a_n = 3n$ for every nonnegative integer n. Then, for $n \geq 2$, we see that $2a_{n-1} - a_{n-2} = 2(3(n-1)) - 3(n-2) = 3n = a_n$. Therefore, $\{a_n\}$, where $a_n = 3n$, is a solution of the recurrence relation.

Suppose that $a_n = 2^n$ for every nonnegative integer n. Note that $a_0 = 1, a_1 = 2$, and $a_2 = 4$. Because $2a_1 - a_0 = 2 \cdot 2 - 1 = 3 \neq a_2$, we see that $\{a_n\}$, where $a_n = 2^n$, is not a solution of the recurrence relation.

Suppose that $a_n = 5$ for every nonnegative integer n. Then for $n \geq 2$, we see that $a_n = 2a_{n-1} - a_{n-2} = 2 \cdot 5 - 5 = 5 = a_n$. Therefore, $\{a_n\}$, where $a_n = 5$, is a solution of the recurrence relation. ◀

Many methods have been developed for solving recurrence relations. Here, we will introduce a straightforward method known as iteration via several examples. In Chapter 8 we will study recurrence relations in depth. In that chapter we will show how recurrence relations can be used to solve counting problems and we will introduce several powerful methods that can be used to solve many different recurrence relations.

EXAMPLE 10 Solve the recurrence relation and initial condition in Example 5.

Solution: We can successively apply the recurrence relation in Example 5, starting with the initial condition $a_1 = 2$, and working upward until we reach a_n to deduce a closed formula for the sequence. We see that

$$\begin{aligned}
a_2 &= 2 + 3\\
a_3 &= (2 + 3) + 3 = 2 + 3 \cdot 2\\
a_4 &= (2 + 2 \cdot 3) + 3 = 2 + 3 \cdot 3\\
&\vdots\\
a_n &= a_{n-1} + 3 = (2 + 3 \cdot (n-2)) + 3 = 2 + 3(n-1).
\end{aligned}$$

We can also successively apply the recurrence relation in Example 5, starting with the term a_n and working downward until we reach the initial condition $a_1 = 2$ to deduce this same formula. The steps are

$$\begin{aligned}
a_n &= a_{n-1} + 3\\
&= (a_{n-2} + 3) + 3 = a_{n-2} + 3 \cdot 2\\
&= (a_{n-3} + 3) + 3 \cdot 2 = a_{n-3} + 3 \cdot 3\\
&\vdots\\
&= a_2 + 3(n-2) = (a_1 + 3) + 3(n-2) = 2 + 3(n-1).
\end{aligned}$$

At each iteration of the recurrence relation, we obtain the next term in the sequence by adding 3 to the previous term. We obtain the nth term after $n-1$ iterations of the recurrence relation. Hence, we have added $3(n-1)$ to the initial term $a_0 = 2$ to obtain a_n. This gives us the closed formula $a_n = 2 + 3(n-1)$. Note that this sequence is an arithmetic progression. ◀

The technique used in Example 10 is called **iteration**. We have iterated, or repeatedly used, the recurrence relation. The first approach is called **forward substitution** – we found successive terms beginning with the initial condition and ending with a_n. The second approach is called **backward substitution**, because we began with a_n and iterated to express it in terms of falling terms of the sequence until we found it in terms of a_1. Note that when we use iteration, we essential guess a formula for the terms of the sequence. To prove that our guess is correct, we need to use mathematical induction, a technique we discuss in Chapter 5.

In Chapter 8 we will show that recurrence relations can be used to model a wide variety of problems. We provide one such example here, showing how to use a recurrence relation to find compound interest.

EXAMPLE 11 **Compound Interest** Suppose that a person deposits \$10,000 in a savings account at a bank yielding 11% per year with interest compounded annually. How much will be in the account after 30 years?

Extra Examples

Solution: To solve this problem, let P_n denote the amount in the account after n years. Because the amount in the account after n years equals the amount in the account after $n-1$ years plus interest for the nth year, we see that the sequence $\{P_n\}$ satisfies the recurrence relation

$$P_n = P_{n-1} + 0.11 P_{n-1} = (1.11) P_{n-1}.$$

The initial condition is $P_0 = 10{,}000$.

We can use an iterative approach to find a formula for P_n. Note that

$$\begin{aligned}
P_1 &= (1.11) P_0 \\
P_2 &= (1.11) P_1 = (1.11)^2 P_0 \\
P_3 &= (1.11) P_2 = (1.11)^3 P_0 \\
&\vdots \\
P_n &= (1.11) P_{n-1} = (1.11)^n P_0.
\end{aligned}$$

When we insert the initial condition $P_0 = 10{,}000$, the formula $P_n = (1.11)^n 10{,}000$ is obtained.

Inserting $n = 30$ into the formula $P_n = (1.11)^n 10{,}000$ shows that after 30 years the account contains

$$P_{30} = (1.11)^{30} 10{,}000 = \$228{,}922.97.$$ ◀

Special Integer Sequences

A common problem in discrete mathematics is finding a closed formula, a recurrence relation, or some other type of general rule for constructing the terms of a sequence. Sometimes only a few terms of a sequence solving a problem are known; the goal is to identify the sequence. Even though the initial terms of a sequence do not determine the entire sequence (after all, there are infinitely many different sequences that start with any finite set of initial terms), knowing the first few terms may help you make an educated conjecture about the identity of your sequence. Once you have made this conjecture, you can try to verify that you have the correct sequence.

When trying to deduce a possible formula, recurrence relation, or some other type of rule for the terms of a sequence when given the initial terms, try to find a pattern in these terms. You might also see whether you can determine how a term might have been produced from those preceding it. There are many questions you could ask, but some of the more useful are:

- Are there runs of the same value? That is, does the same value occur many times in a row?
- Are terms obtained from previous terms by adding the same amount or an amount that depends on the position in the sequence?
- Are terms obtained from previous terms by multiplying by a particular amount?
- Are terms obtained by combining previous terms in a certain way?
- Are there cycles among the terms?

EXAMPLE 12 Find formulae for the sequences with the following first five terms: (a) 1, 1/2, 1/4, 1/8, 1/16 (b) 1, 3, 5, 7, 9 (c) 1, −1, 1, −1, 1.

Extra Examples

Solution: (a) We recognize that the denominators are powers of 2. The sequence with $a_n = 1/2^n$, $n = 0, 1, 2, \ldots$ is a possible match. This proposed sequence is a geometric progression with $a = 1$ and $r = 1/2$.

(b) We note that each term is obtained by adding 2 to the previous term. The sequence with $a_n = 2n + 1$, $n = 0, 1, 2, \ldots$ is a possible match. This proposed sequence is an arithmetic progression with $a = 1$ and $d = 2$.

(c) The terms alternate between 1 and −1. The sequence with $a_n = (-1)^n$, $n = 0, 1, 2 \ldots$ is a possible match. This proposed sequence is a geometric progression with $a = 1$ and $r = -1$. ◀

Examples 13–15 illustrate how we can analyze sequences to find how the terms are constructed.

EXAMPLE 13 How can we produce the terms of a sequence if the first 10 terms are 1, 2, 2, 3, 3, 3, 4, 4, 4, 4?

Solution: In this sequence, the integer 1 appears once, the integer 2 appears twice, the integer 3 appears three times, and the integer 4 appears four times. A reasonable rule for generating this sequence is that the integer n appears exactly n times, so the next five terms of the sequence would all be 5, the following six terms would all be 6, and so on. The sequence generated this way is a possible match. ◀

EXAMPLE 14 How can we produce the terms of a sequence if the first 10 terms are 5, 11, 17, 23, 29, 35, 41, 47, 53, 59?

Solution: Note that each of the first 10 terms of this sequence after the first is obtained by adding 6 to the previous term. (We could see this by noticing that the difference between consecutive terms is 6.) Consequently, the nth term could be produced by starting with 5 and adding 6 a total of $n - 1$ times; that is, a reasonable guess is that the nth term is $5 + 6(n - 1) = 6n - 1$. (This is an arithmetic progression with $a = 5$ and $d = 6$.) ◀

EXAMPLE 15 How can we produce the terms of a sequence if the first 10 terms are 1, 3, 4, 7, 11, 18, 29, 47, 76, 123?

Solution: Observe that each successive term of this sequence, starting with the third term, is the sum of the two previous terms. That is, $4 = 3 + 1$, $7 = 4 + 3$, $11 = 7 + 4$, and so on. Consequently, if L_n is the nth term of this sequence, we guess that the sequence is determined by the recurrence relation $L_n = L_{n-1} + L_{n-2}$ with initial conditions $L_1 = 1$ and $L_2 = 3$ (the same recurrence relation as the Fibonacci sequence, but with different initial conditions). This sequence is known as the **Lucas sequence**, after the French mathematician François Édouard Lucas. Lucas studied this sequence and the Fibonacci sequence in the nineteenth century. ◀

Another useful technique for finding a rule for generating the terms of a sequence is to compare the terms of a sequence of interest with the terms of a well-known integer sequence, such as terms of an arithmetic progression, terms of a geometric progression, perfect squares, perfect cubes, and so on. The first 10 terms of some sequences you may want to keep in mind are displayed in Table 1.

EXAMPLE 16 Conjecture a simple formula for a_n if the first 10 terms of the sequence $\{a_n\}$ are 1, 7, 25, 79, 241, 727, 2185, 6559, 19681, 59047.

TABLE 1 Some Useful Sequences.

nth Term	*First 10 Terms*
n^2	1, 4, 9, 16, 25, 36, 49, 64, 81, 100, ...
n^3	1, 8, 27, 64, 125, 216, 343, 512, 729, 1000, ...
n^4	1, 16, 81, 256, 625, 1296, 2401, 4096, 6561, 10000, ...
2^n	2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, ...
3^n	3, 9, 27, 81, 243, 729, 2187, 6561, 19683, 59049, ...
$n!$	1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800, ...
f_n	1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

Solution: To attack this problem, we begin by looking at the difference of consecutive terms, but we do not see a pattern. When we form the ratio of consecutive terms to see whether each term is a multiple of the previous term, we find that this ratio, although not a constant, is close to 3. So it is reasonable to suspect that the terms of this sequence are generated by a formula involving 3^n. Comparing these terms with the corresponding terms of the sequence $\{3^n\}$, we notice that the nth term is 2 less than the corresponding power of 3. We see that $a_n = 3^n - 2$ for $1 \le n \le 10$ and conjecture that this formula holds for all n. ◀

We will see throughout this text that integer sequences appear in a wide range of contexts in discrete mathematics. Sequences we have encountered or will encounter include the sequence of prime numbers (Chapter 4), the number of ways to order n discrete objects (Chapter 6), the number of moves required to solve the famous Tower of Hanoi puzzle with n disks (Chapter 8), and the number of rabbits on an island after n months (Chapter 8).

Check out the puzzles at the OEIS site.

Links

Integer sequences appear in an amazingly wide range of subject areas besides discrete mathematics, including biology, engineering, chemistry, and physics, as well as in puzzles. An amazing database of over 200,000 different integer sequences can be found in the *On-Line Encyclopedia of Integer Sequences (OEIS)*. This database was originated by Neil Sloane in the 1960s. The last printed version of this database was published in 1995 ([SIPI95]); the current encyclopedia would occupy more than 750 volumes of the size of the 1995 book with more than 10,000 new submissions a year. There is also a program accessible via the Web that you can use to find sequences from the encyclopedia that match initial terms you provide.

Summations

Next, we consider the addition of the terms of a sequence. For this we introduce **summation notation**. We begin by describing the notation used to express the sum of the terms

$$a_m, a_{m+1}, \ldots, a_n$$

from the sequence $\{a_n\}$. We use the notation

$$\sum_{j=m}^{n} a_j, \qquad \textstyle\sum_{j=m}^{n} a_j, \qquad \text{or} \qquad \textstyle\sum_{m \le j \le n} a_j$$

(read as the sum from $j = m$ to $j = n$ of a_j) to represent

$$a_m + a_{m+1} + \cdots + a_n.$$

Here, the variable j is called the **index of summation**, and the choice of the letter j as the variable is arbitrary; that is, we could have used any other letter, such as i or k. Or, in notation,

$$\sum_{j=m}^{n} a_j = \sum_{i=m}^{n} a_i = \sum_{k=m}^{n} a_k.$$

Here, the index of summation runs through all integers starting with its **lower limit** m and ending with its **upper limit** n. A large uppercase Greek letter sigma, $\sum$, is used to denote summation.

The usual laws for arithmetic apply to summations. For example, when a and b are real numbers, we have $\sum_{j=1}^{n}(ax_j + by_j) = a\sum_{y=1}^{n} x_j + b\sum_{j=1}^{n} y_j$, where $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are real numbers. (We do not present a formal proof of this identity here. Such a proof can be constructed using mathematical induction, a proof method we introduce in Chapter 5. The proof also uses the commutative and associative laws for addition and the distributive law of multiplication over addition.)

We give some examples of summation notation.

EXAMPLE 17 Use summation notation to express the sum of the first 100 terms of the sequence $\{a_j\}$, where $a_j = 1/j$ for $j = 1, 2, 3, \ldots$.

Extra Examples

Solution: The lower limit for the index of summation is 1, and the upper limit is 100. We write this sum as

$$\sum_{j=1}^{100} \frac{1}{j}.$$

◀

EXAMPLE 18 What is the value of $\sum_{j=1}^{5} j^2$?

Solution: We have

$$\begin{aligned}\sum_{j=1}^{5} j^2 &= 1^2 + 2^2 + 3^2 + 4^2 + 5^2\\ &= 1 + 4 + 9 + 16 + 25\\ &= 55.\end{aligned}$$

◀

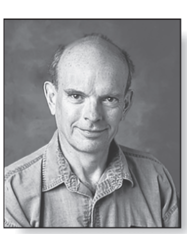
EXAMPLE 19 What is the value of $\sum_{k=4}^{8}(-1)^k$?

Solution: We have

$$\begin{aligned}\sum_{k=4}^{8}(-1)^k &= (-1)^4 + (-1)^5 + (-1)^6 + (-1)^7 + (-1)^8\\ &= 1 + (-1) + 1 + (-1) + 1\\ &= 1.\end{aligned}$$

◀

Links



NEIL SLOANE (BORN 1939) Neil Sloane studied mathematics and electrical engineering at the University of Melbourne on a scholarship from the Australian state telephone company. He mastered many telephone-related jobs, such as erecting telephone poles, in his summer work. After graduating, he designed minimal-cost telephone networks in Australia. In 1962 he came to the United States and studied electrical engineering at Cornell University. His Ph.D. thesis was on what are now called neural networks. He took a job at Bell Labs in 1969, working in many areas, including network design, coding theory, and sphere packing. He now works for AT&T Labs, moving there from Bell Labs when AT&T split up in 1996. One of his favorite problems is the **kissing problem** (a name he coined), which asks how many spheres can be arranged in n dimensions so that they all touch a central sphere of the same size. (In two dimensions the answer is 6, because 6 pennies can be placed so that they touch a central penny. In three dimensions, 12 billiard balls can be placed so that they touch a central billiard ball. Two billiard balls that just touch are said to “kiss,” giving rise to the terminology “kissing problem” and “kissing number.”) Sloane, together with Andrew Odlyzko, showed that in 8 and 24 dimensions, the optimal kissing numbers are, respectively, 240 and 196,560. The kissing number is known in dimensions 1, 2, 3, 4, 8, and 24, but not in any other dimensions. Sloane’s books include *Sphere Packings, Lattices and Groups,* 3d ed., with John Conway; *The Theory of Error-Correcting Codes* with Jessie MacWilliams; *The Encyclopedia of Integer Sequences* with Simon Plouffe (which has grown into the famous OEIS website); and *The Rock-Climbing Guide to New Jersey Crags* with Paul Nick. The last book demonstrates his interest in rock climbing; it includes more than 50 climbing sites in New Jersey.

Sometimes it is useful to shift the index of summation in a sum. This is often done when two sums need to be added but their indices of summation do not match. When shifting an index of summation, it is important to make the appropriate changes in the corresponding summand. This is illustrated by Example 20.

EXAMPLE 20 Suppose we have the sum

$$\sum_{j=1}^{5} j^2$$

Extra Examples

but want the index of summation to run between 0 and 4 rather than from 1 to 5. To do this, we let $k = j - 1$. Then the new summation index runs from 0 (because $k = 1 - 0 = 0$ when $j = 1$) to 4 (because $k = 5 - 1 = 4$ when $j = 5$), and the term j^2 becomes $(k + 1)^2$. Hence,

$$\sum_{j=1}^{5} j^2 = \sum_{k=0}^{4} (k+1)^2.$$

It is easily checked that both sums are $1 + 4 + 9 + 16 + 25 = 55$. ◀

Sums of terms of geometric progressions commonly arise (such sums are called **geometric series**). Theorem 1 gives us a formula for the sum of terms of a geometric progression.

THEOREM 1 If a and r are real numbers and $r \neq 0$, then

$$\sum_{j=0}^{n} ar^j = \begin{cases} \dfrac{ar^{n+1} - a}{r - 1} & \text{if } r \neq 1 \\ (n+1)a & \text{if } r = 1. \end{cases}$$

Proof: Let

$$S_n = \sum_{j=0}^{n} ar^j.$$

To compute S, first multiply both sides of the equality by r and then manipulate the resulting sum as follows:

$$\begin{aligned} rS_n &= r\sum_{j=0}^{n} ar^j && \text{substituting summation formula for } S \\ &= \sum_{j=0}^{n} ar^{j+1} && \text{by the distributive property} \\ &= \sum_{k=1}^{n+1} ar^k && \text{shifting the index of summation, with } k = j+1 \\ &= \left(\sum_{k=0}^{n} ar^k\right) + (ar^{n+1} - a) && \text{removing } k = n+1 \text{ term and adding } k = 0 \text{ term} \\ &= S_n + (ar^{n+1} - a) && \text{substituting } S \text{ for summation formula} \end{aligned}$$

From these equalities, we see that

$$rS_n = S_n + (ar^{n+1} - a).$$

Solving for S_n shows that if $r \neq 1$, then

$$S_n = \frac{ar^{n+1} - a}{r - 1}.$$

If $r = 1$, then the $S_n = \sum_{j=0}^{n} ar^j = \sum_{j=0}^{n} a = (n+1)a$. ◁

EXAMPLE 21 Double summations arise in many contexts (as in the analysis of nested loops in computer programs). An example of a double summation is

$$\sum_{i=1}^{4}\sum_{j=1}^{3} ij.$$

To evaluate the double sum, first expand the inner summation and then continue by computing the outer summation:

$$\begin{aligned}\sum_{i=1}^{4}\sum_{j=1}^{3} ij &= \sum_{i=1}^{4}(i + 2i + 3i)\\ &= \sum_{i=1}^{4} 6i\\ &= 6 + 12 + 18 + 24 = 60.\end{aligned}$$

◀

We can also use summation notation to add all values of a function, or terms of an indexed set, where the index of summation runs over all values in a set. That is, we write

$$\sum_{s \in S} f(s)$$

to represent the sum of the values $f(s)$, for all members s of S.

EXAMPLE 22 What is the value of $\sum_{s \in \{0,2,4\}} s$?

Solution: Because $\sum_{s \in \{0,2,4\}} s$ represents the sum of the values of s for all the members of the set $\{0, 2, 4\}$, it follows that

$$\sum_{s \in \{0,2,4\}} s = 0 + 2 + 4 = 6.$$

◀

Certain sums arise repeatedly throughout discrete mathematics. Having a collection of formulae for such sums can be useful; Table 2 provides a small table of formulae for commonly occurring sums.

We derived the first formula in this table in Theorem 1. The next three formulae give us the sum of the first n positive integers, the sum of their squares, and the sum of their cubes. These three formulae can be derived in many different ways (for example, see Exercises 23 and 24). Also note that each of these formulae, once known, can easily be proved using mathematical induction, the subject of Section 5.1. The last two formulae in the table involve infinite series and will be discussed shortly.

Example 23 illustrates how the formulae in Table 2 can be useful.

TABLE 2 Some Useful Summation Formulae.

Sum	*Closed Form*
$\sum_{k=0}^{n} ar^k \ (r \neq 0)$	$\frac{ar^{n+1} - a}{r - 1}, r \neq 1$
$\sum_{k=1}^{n} k$	$\frac{n(n+1)}{2}$
$\sum_{k=1}^{n} k^2$	$\frac{n(n+1)(2n+1)}{6}$
$\sum_{k=1}^{n} k^3$	$\frac{n^2(n+1)^2}{4}$
$\sum_{k=0}^{\infty} x^k, \lvert x\rvert < 1$	$\frac{1}{1-x}$
$\sum_{k=1}^{\infty} kx^{k-1}, \lvert x\rvert < 1$	$\frac{1}{(1-x)^2}$

EXAMPLE 23 Find $\sum_{k=50}^{100} k^2$.

Solution: First note that because $\sum_{k=1}^{100} k^2 = \sum_{k=1}^{49} k^2 + \sum_{k=50}^{100} k^2$, we have

$$\sum_{k=50}^{100} k^2 = \sum_{k=1}^{100} k^2 - \sum_{k=1}^{49} k^2.$$

Using the formula $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ from Table 2 (and proved in Exercise 38), we see that

$$\sum_{k=50}^{100} k^2 = \frac{100 \cdot 101 \cdot 201}{6} - \frac{49 \cdot 50 \cdot 99}{6} = 338{,}350 - 40{,}425 = 297{,}925.$$ ◀

SOME INFINITE SERIES Although most of the summations in this book are finite sums, infinite series are important in some parts of discrete mathematics. Infinite series are usually studied in a course in calculus and even the definition of these series requires the use of calculus, but sometimes they arise in discrete mathematics, because discrete mathematics deals with infinite collections of discrete elements. In particular, in our future studies in discrete mathematics, we will find the closed forms for the infinite series in Examples 24 and 25 to be quite useful.

EXAMPLE 24 (*Requires calculus*) Let x be a real number with $|x| < 1$. Find $\sum_{n=0}^{\infty} x^n$.

Extra Examples

Solution: By Theorem 1 with $a = 1$ and $r = x$ we see that $\sum_{n=0}^{k} x^n = \frac{x^{k+1} - 1}{x - 1}$. Because $|x| < 1$, x^{k+1} approaches 0 as k approaches infinity. It follows that

$$\sum_{n=0}^{\infty} x^n = \lim_{k \to \infty} \frac{x^{k+1} - 1}{x - 1} = \frac{0 - 1}{x - 1} = \frac{1}{1 - x}.$$ ◀

We can produce new summation formulae by differentiating or integrating existing formulae.

EXAMPLE 25 (*Requires calculus*) Differentiating both sides of the equation

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},$$

from Example 24 we find that

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

(This differentiation is valid for $|x| < 1$ by a theorem about infinite series.) ◀

Exercises

1. Find these terms of the sequence $\{a_n\}$, where $a_n = 2 \cdot (-3)^n + 5^n$.
   a) a_0 **b)** a_1 **c)** a_4 **d)** a_5
2. What are the terms a_0, a_1, a_2, and a_3 of the sequence $\{a_n\}$, where a_n equals
   a) $2^n + 1$? **b)** $(n+1)^{n+1}$?
   c) $\lfloor n/2 \rfloor$? **d)** $\lfloor n/2 \rfloor + \lceil n/2 \rceil$?
3. List the first 10 terms of each of these sequences.
   a) the sequence that begins with 2 and in which each successive term is 3 more than the preceding term
   b) the sequence that lists each positive integer three times, in increasing order
   c) the sequence that lists the odd positive integers in increasing order, listing each odd integer twice
   d) the sequence whose nth term is $n! - 2^n$
   e) the sequence that begins with 3, where each succeeding term is twice the preceding term
   f) the sequence whose first term is 2, second term is 4, and each succeeding term is the sum of the two preceding terms
   g) the sequence whose nth term is the number of bits in the binary expansion of the number n (defined in Section 4.2)
   h) the sequence where the nth term is the number of letters in the English word for the index n
4. Find at least three different sequences beginning with the terms 1, 2, 4 whose terms are generated by a simple formula or rule.
5. Find at least three different sequences beginning with the terms 3, 5, 7 whose terms are generated by a simple formula or rule.
6. Find the first five terms of the sequence defined by each of these recurrence relations and initial conditions.
   a) $a_n = 6a_{n-1}, a_0 = 2$
   b) $a_n = a_{n-1}^2, a_1 = 2$
   c) $a_n = a_{n-1} + 3a_{n-2}, a_0 = 1, a_1 = 2$
   d) $a_n = na_{n-1} + n^2 a_{n-2}, a_0 = 1, a_1 = 1$
   e) $a_n = a_{n-1} + a_{n-3}, a_0 = 1, a_1 = 2, a_2 = 0$
7. Let $a_n = 2^n + 5 \cdot 3^n$ for $n = 0, 1, 2, \ldots$.
   a) Find a_0, a_1, a_2, a_3, and a_4.
   b) Show that $a_2 = 5a_1 - 6a_0$, $a_3 = 5a_2 - 6a_1$, and $a_4 = 5a_3 - 6a_2$.
   c) Show that $a_n = 5a_{n-1} - 6a_{n-2}$ for all integers n with $n \geq 2$.
8. Show that the sequence $\{a_n\}$ is a solution of the recurrence relation $a_n = -3a_{n-1} + 4a_{n-2}$ if
   a) $a_n = 0$. **b)** $a_n = 1$.
   c) $a_n = (-4)^n$. **d)** $a_n = 2(-4)^n + 3$.
9. Find the solution to each of these recurrence relations and initial conditions. Use an iterative approach such as that used in Example 10.
   a) $a_n = 3a_{n-1}, a_0 = 2$
   b) $a_n = a_{n-1} + 2, a_0 = 3$
   c) $a_n = a_{n-1} + n, a_0 = 1$
   d) $a_n = a_{n-1} + 2n + 3, a_0 = 4$
   e) $a_n = 2a_{n-1} - 1, a_0 = 1$
   f) $a_n = 3a_{n-1} + 1, a_0 = 1$
   g) $a_n = na_{n-1}, a_0 = 5$
   h) $a_n = 2na_{n-1}, a_0 = 1$
10. A person deposits \$1000 in an account that yields 9% interest compounded annually.
    a) Set up a recurrence relation for the amount in the account at the end of n years.
    b) Find an explicit formula for the amount in the account at the end of n years.
    c) How much money will the account contain after 100 years?
11. Suppose that the number of bacteria in a colony triples every hour.
    a) Set up a recurrence relation for the number of bacteria after n hours have elapsed.
    b) If 100 bacteria are used to begin a new colony, how many bacteria will be in the colony in 10 hours?
12. An employee joined a company in 2009 with a starting salary of \$50,000. Every year this employee receives a raise of \$1000 plus 5% of the salary of the previous year.
    a) Set up a recurrence relation for the salary of this employee n years after 2009.
    b) What will the salary of this employee be in 2017?
    c) Find an explicit formula for the salary of this employee n years after 2009.

13. Find a recurrence relation for the balance $B(k)$ owed at the end of k months on a loan of \$5000 at a rate of 7% if a payment of \$100 is made each month. [*Hint:* Express $B(k)$ in terms of $B(k-1)$; the monthly interest is $(0.07/12)B(k-1)$.]

14. a) Find a recurrence relation for the balance $B(k)$ owed at the end of k months on a loan at a rate of r if a payment P is made on the loan each month. [*Hint:* Express $B(k)$ in terms of $B(k-1)$ and note that the monthly interest rate is $r/12$.]

b) Determine what the monthly payment P should be so that the loan is paid off after T months.

15. For each of these lists of integers, provide a simple formula or rule that generates the terms of an integer sequence that begins with the given list. Assuming that your formula or rule is correct, determine the next three terms of the sequence.

a) 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, . . .
b) 1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 8, 8, . . .
c) 1, 0, 2, 0, 4, 0, 8, 0, 16, 0, . . .
d) 3, 6, 12, 24, 48, 96, 192, . . .
e) 15, 8, 1, −6, −13, −20, −27, . . .
f) 3, 5, 8, 12, 17, 23, 30, 38, 47, . . .
g) 2, 16, 54, 128, 250, 432, 686, . . .
h) 2, 3, 7, 25, 121, 721, 5041, 40321, . . .

****16.** Show that if a_n denotes the nth positive integer that is not a perfect square, then $a_n = n + \{\sqrt{n}\}$, where $\{x\}$ denotes the integer closest to the real number x.

17. What are the values of these sums?

a) $\sum_{k=1}^{5}(k+1)$ **b)** $\sum_{j=0}^{4}(-2)^j$
c) $\sum_{i=1}^{10} 3$ **d)** $\sum_{j=0}^{8}(2^{j+1}-2^j)$

18. What are the values of these sums, where $S = \{1, 3, 5, 7\}$?

a) $\sum_{j\in S} j$ **b)** $\sum_{j\in S} j^2$
c) $\sum_{j\in S}(1/j)$ **d)** $\sum_{j\in S} 1$

19. What is the value of each of these sums of terms of a geometric progression?

a) $\sum_{j=0}^{8} 3\cdot 2^j$ **b)** $\sum_{j=1}^{8} 2^j$
c) $\sum_{j=2}^{8}(-3)^j$ **d)** $\sum_{j=0}^{8} 2\cdot(-3)^j$

20. Compute each of these double sums.

a) $\sum_{i=1}^{2}\sum_{j=1}^{3}(i+j)$ **b)** $\sum_{i=0}^{2}\sum_{j=0}^{3}(2i+3j)$
c) $\sum_{i=1}^{3}\sum_{j=0}^{2} i$ **d)** $\sum_{i=0}^{2}\sum_{j=1}^{3} ij$

21. Show that $\sum_{j=1}^{n}(a_j - a_{j-1}) = a_n - a_0$, where $a_0, a_1, \ldots, a_n$ is a sequence of real numbers. This type of sum is called **telescoping**.

22. Use the identity $1/(k(k+1)) = 1/k - 1/(k+1)$ and Exercise 21 to compute $\sum_{k=1}^{n} 1/(k(k+1))$.

23. Sum both sides of the identity $k^2 - (k-1)^2 = 2k - 1$ from $k = 1$ to $k = n$ and use Exercise 35 to find

a) a formula for $\sum_{k=1}^{n}(2k-1)$ (the sum of the first n odd natural numbers).

b) a formula for $\sum_{k=1}^{n} k$.

***24.** Use the technique given in Exercise 35, together with the result of Exercise 37b, to derive the formula for $\sum_{k=1}^{n} k^2$ given in Table 2. [*Hint*: Take $a_k = k^3$ in the telescoping sum in Exercise 35.]

25. Find $\sum_{k=100}^{200} k$. (Use Table 2.)

26. Find $\sum_{k=99}^{200} k^3$. (Use Table 2.)

***27.** Find a formula for $\sum_{k=0}^{m}\lfloor\sqrt{k}\rfloor$, when m is a positive integer.

***28.** Find a formula for $\sum_{k=0}^{m}\lfloor\sqrt[3]{k}\rfloor$, when m is a positive integer.

There is also a special notation for products. The product of $a_m, a_{m+1}, \ldots, a_n$ is represented by $\prod_{j=m}^{n} a_j$, read as the product from $j = m$ to $j = n$ of a_j.

29. What are the values of the following products?

a) $\prod_{i=0}^{10} i$ **b)** $\prod_{i=5}^{8} i$
c) $\prod_{i=1}^{100}(-1)^i$ **d)** $\prod_{i=1}^{10} 2$

Recall that the value of the factorial function at a positive integer n, denoted by $n!$, is the product of the positive integers from 1 to n, inclusive. Also, we specify that $0! = 1$.

30. Express $n!$ using product notation.

31. Find $\sum_{j=0}^{4} j!$.

32. Find $\prod_{j=0}^{4} j!$.

2.5 Cardinality of Sets

Introduction

In Definition 4 of Section 2.1 we defined the cardinality of a finite set as the number of elements in the set. We use the cardinalities of finite sets to tell us when they have the same size, or when one is bigger than the other. In this section we extend this notion to infinite sets. That is, we will define what it means for two infinite sets to have the same cardinality, providing us with a way to measure the relative sizes of infinite sets.

We will be particularly interested in countably infinite sets, which are sets with the same cardinality as the set of positive integers. We will establish the surprising result that the set of rational numbers is countably infinite. We will also provide an example of an uncountable set when we show that the set of real numbers is not countable.

The concepts developed in this section have important applications to computer science. A function is called uncomputable if no computer program can be written to find all its values, even with unlimited time and memory. We will use the concepts in this section to explain why uncomputable functions exist.

We now define what it means for two sets to have the same size, or cardinality. In Section 2.1, we discussed the cardinality of finite sets and we defined the size, or cardinality, of such sets. In Exercise 53 of Section 2.3 we showed that there is a one-to-one correspondence between any two finite sets with the same number of elements. We use this observation to extend the concept of cardinality to all sets, both finite and infinite.

DEFINITION 1 A set S is finite with cardinality $n \in N$ if there is a bijection rom the set $\{0, 1, \ldots, n-1\}$ to S. A set is infinite if it is not finte.

THEOREM 1 The set N of natural numbers is an infinite set.

Proof: Consider the injection $f : N \to N$ deined as $f(x) = 3x$. The range of f is a subset of the domain of f.

Some facts which could easily be seen are:

1. If S' is infinite and is a subset of S, S is infinite.
2. Every subset of a finite set is finite.
3. If $F : S \to T$ be an injection and S is infinite, then T is infinite.
4. If S is an infinite set $P(S)$ is infinite.
5. If S and T are infinite sets $S \cup T$ is infinite.
6. If S is infinite and $T \neq \emptyset$ the $S \times T$ is infinite.
7. If S is infinite and $T \neq \emptyset$ the set of functions from T to S is infinite.

DEFINITION 2 The sets A and B have the same *cardinality* if and only if there is a one-to-one correspondence from A to B. When A and B have the same cardinality, we write $|A| = |B|$.

For infinite sets the definition of cardinality provides a relative measure of the sizes of two sets, rather than a measure of the size of one particular set. We can also define what it means for one set to have a smaller cardinality than another set.

DEFINITION 3 If there is a one-to-one function from A to B, the cardinality of A is less than or the same as the cardinality of B and we write $|A| \leq |B|$. Moreover, when $|A| \leq |B|$ and A and B have different cardinality, we say that the cardinality of A is less than the cardinality of B and we write $|A| < |B|$.

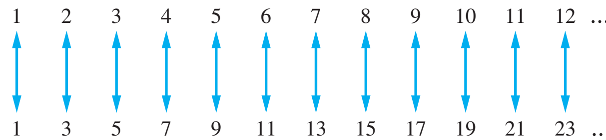


FIGURE 1 A One-to-One Correspondence Between $\mathbf{Z}^+$ and the Set of Odd Positive Integers.

Countable Sets

We will now split infinite sets into two groups, those with the same cardinality as the set of natural numbers and those with a different cardinality.

DEFINITION 4 A set that is either finite or has the same cardinality as the set of positive integers is called *countable*. A set that is not countable is called *uncountable*. When an infinite set S is countable, we denote the cardinality of S by $\aleph_0$ (where $\aleph$ is aleph, the first letter of the Hebrew alphabet). We write $|S| = \aleph_0$ and say that S has cardinality "aleph null."

We illustrate how to show a set is countable in the next example.

EXAMPLE 1 Show that the set of odd positive integers is a countable set.

Solution: To show that the set of odd positive integers is countable, we will exhibit a one-to-one correspondence between this set and the set of positive integers. Consider the function

$$f(n) = 2n - 1$$

from $\mathbf{Z}^+$ to the set of odd positive integers. We show that f is a one-to-one correspondence by showing that it is both one-to-one and onto. To see that it is one-to-one, suppose that $f(n) = f(m)$. Then $2n - 1 = 2m - 1$, so $n = m$. To see that it is onto, suppose that t is an odd positive integer. Then t is 1 less than an even integer $2k$, where k is a natural number. Hence $t = 2k - 1 = f(k)$. We display this one-to-one correspondence in Figure 1. ◀

An infinite set is countable if and only if it is possible to list the elements of the set in a sequence (indexed by the positive integers). The reason for this is that a one-to-one correspondence f from the set of positive integers to a set S can be expressed in terms of a sequence $a_1, a_2, \ldots, a_n, \ldots$, where $a_1 = f(1), a_2 = f(2), \ldots, a_n = f(n), \ldots$.

You can always get a room at Hilbert's Grand Hotel!

Links

HILBERT'S GRAND HOTEL We now describe a paradox that shows that something impossible with finite sets may be possible with infinite sets. The famous mathematician David Hilbert invented the notion of the **Grand Hotel**, which has a countably infinite number of rooms, each occupied by a guest. When a new guest arrives at a hotel with a finite number of rooms, and all rooms are occupied, this guest cannot be accommodated without evicting a current guest. However, we can always accommodate a new guest at the Grand Hotel, even when all rooms are already occupied, as we show in Example 2. Exercises 4 and 6 ask you to show that we can accommodate a finite number of new guests and a countable number of new guests, respectively, at the fully occupied Grand Hotel.

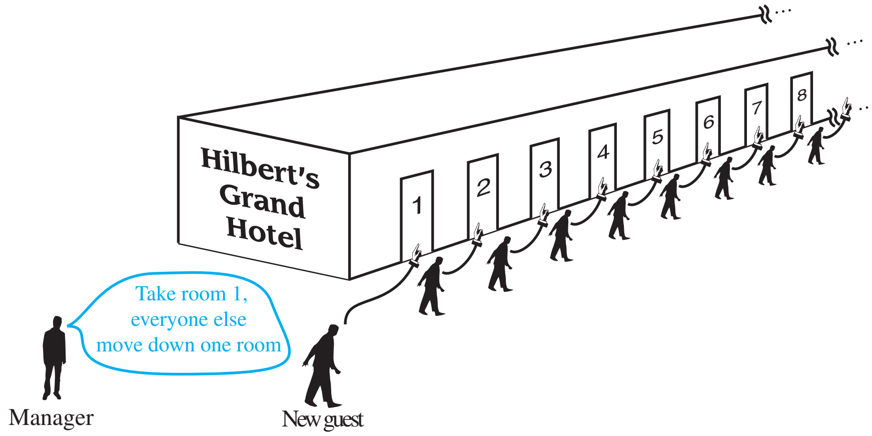


FIGURE 2 A New Guest Arrives at Hilbert's Grand Hotel.

EXAMPLE 2 How can we accommodate a new guest arriving at the fully occupied Grand Hotel without removing any of the current guests?

Solution: Because the rooms of the Grand Hotel are countable, we can list them as Room 1, Room 2, Room 3, and so on. When a new guest arrives, we move the guest in Room 1 to Room 2, the guest in Room 2 to Room 3, and in general, the guest in Room n to Room $n + 1$, for all positive integers n. This frees up Room 1, which we assign to the new guest, and all the current guests still have rooms. We illustrate this situation in Figure 2. ◀

When there are finitely many room in a hotel, the notion that all rooms are occupied is equivalent to the notion that no new guests can be accommodated. However, Hilbert's paradox of the Grand Hotel can be explained by noting that this equivalence no longer holds when there are infinitely many room.

EXAMPLES OF COUNTABLE AND UNCOUNTABLE SETS We will now show that certain sets of numbers are countable. We begin with the set of all integers. Note that we can show that the set of all integers is countable by listing its members.

EXAMPLE 3 Show that the set of all integers is countable.

Solution: We can list all integers in a sequence by starting with 0 and alternating between positive and negative integers: $0, 1, -1, 2, -2, \ldots$. Alternatively, we could find a one-to-one correspondence between the set of positive integers and the set of all integers. We leave it to the

Links

DAVID HILBERT (1862–1943) Hilbert, born in Königsberg, the city famous in mathematics for its seven bridges, was the son of a judge. During his tenure at Göttingen University, from 1892 to 1930, he made many fundamental contributions to a wide range of mathematical subjects. He almost always worked on one area of mathematics at a time, making important contributions, then moving to a new mathematical subject. Some areas in which Hilbert worked are the calculus of variations, geometry, algebra, number theory, logic, and mathematical physics. Besides his many outstanding original contributions, Hilbert is remembered for his famous list of 23 difficult problems. He described these problems at the 1900 International Congress of Mathematicians, as a challenge to mathematicians at the birth of the twentieth century. Since that time, they have spurred a tremendous amount and variety of research. Although many of these problems have now been solved, several remain open, including the Riemann hypothesis, which is part of Problem 8 on Hilbert's list. Hilbert was also the author of several important textbooks in number theory and geometry.

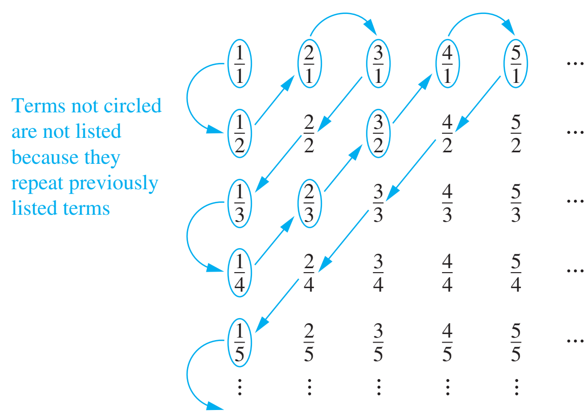


FIGURE 3 The Positive Rational Numbers Are Countable.

reader to show that the function $f(n) = n/2$ when n is even and $f(n) = -(n-1)/2$ when n is odd is such a function. Consequently, the set of all integers is countable. ◀

It is not surprising that the set of odd integers and the set of all integers are both countable sets (as shown in Examples 1 and 3). Many people are amazed to learn that the set of rational numbers is countable, as Example 4 demonstrates.

EXAMPLE 4 Show that the set of positive rational numbers is countable.

Solution: It may seem surprising that the set of positive rational numbers is countable, but we will show how we can list the positive rational numbers as a sequence $r_1, r_2, \ldots, r_n, \ldots$. First, note that every positive rational number is the quotient p/q of two positive integers. We can arrange the positive rational numbers by listing those with denominator $q = 1$ in the first row, those with denominator $q = 2$ in the second row, and so on, as displayed in Figure 3.

The key to listing the rational numbers in a sequence is to first list the positive rational numbers p/q with $p + q = 2$, followed by those with $p + q = 3$, followed by those with $p + q = 4$, and so on, following the path shown in Figure 3. Whenever we encounter a number p/q that is already listed, we do not list it again. For example, when we come to $2/2 = 1$ we do not list it because we have already listed $1/1 = 1$. The initial terms in the list of positive rational numbers we have constructed are 1, 1/2, 2, 3, 1/3, 1/4, 2/3, 3/2, 4, 5, and so on. These numbers are shown circled; the uncircled numbers in the list are those we leave out because they are already listed. Because all positive rational numbers are listed once, as the reader can verify, we have shown that the set of positive rational numbers is countable. ◀

An Uncountable Set

Not all infinite sets have the same size!

Links

We have seen that the set of positive rational numbers is a countable set. Do we have a promising candidate for an uncountable set? The first place we might look is the set of real numbers. In Example 5 we use an important proof method, introduced in 1879 by Georg Cantor and known as the **Cantor diagonalization argument**, to prove that the set of real numbers is not countable. This proof method is used extensively in mathematical logic and in the theory of computation.

EXAMPLE 5 Show that the set of real numbers is an uncountable set.

Extra Examples

Solution: To show that the set of real numbers is uncountable, we suppose that the set of real numbers is countable and arrive at a contradiction. Then, the subset of all real numbers that fall between 0 and 1 would also be countable (because any subset of a countable set is also countable; see Exercise 13). Under this assumption, the real numbers between 0 and 1 can be listed in some order, say, $r_1, r_2, r_3, \ldots$. Let the decimal representation of these real numbers be

$$\begin{aligned} r_1 &= 0.d_{11}d_{12}d_{13}d_{14} \ldots \\ r_2 &= 0.d_{21}d_{22}d_{23}d_{24} \ldots \\ r_3 &= 0.d_{31}d_{32}d_{33}d_{34} \ldots \\ r_4 &= 0.d_{41}d_{42}d_{43}d_{44} \ldots \\ &\vdots \end{aligned}$$

where $d_{ij} \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. (For example, if $r_1 = 0.23794102\ldots$, we have $d_{11} = 2$, $d_{12} = 3$, $d_{13} = 7$, and so on.) Then, form a new real number with decimal expansion $r = 0.d_1d_2d_3d_4\ldots$, where the decimal digits are determined by the following rule:

$$d_i = \begin{cases} 4 \text{ if } d_{ii} \neq 4 \\ 5 \text{ if } d_{ii} = 4. \end{cases}$$

(As an example, suppose that $r_1 = 0.23794102\ldots$, $r_2 = 0.44590138\ldots$, $r_3 = 0.09118764\ldots$, $r_4 = 0.80553900\ldots$, and so on. Then we have $r = 0.d_1d_2d_3d_4\ldots = 0.4544\ldots$, where $d_1 = 4$ because $d_{11} \neq 4$, $d_2 = 5$ because $d_{22} = 4$, $d_3 = 4$ because $d_{33} \neq 4$, $d_4 = 4$ because $d_{44} \neq 4$, and so on.)

A number with a decimal expansion that terminates has a second decimal expansion ending with an infinite sequence of 9s because $1 = 0.999\ldots$.

Every real number has a unique decimal expansion (when the possibility that the expansion has a tail end that consists entirely of the digit 9 is excluded). Therefore, the real number r is not equal to any of $r_1, r_2, \ldots$ because the decimal expansion of r differs from the decimal expansion of r_i in the ith place to the right of the decimal point, for each i.

Because there is a real number r between 0 and 1 that is not in the list, the assumption that all the real numbers between 0 and 1 could be listed must be false. Therefore, all the real numbers between 0 and 1 cannot be listed, so the set of real numbers between 0 and 1 is uncountable. Any set with an uncountable subset is uncountable (see Exercise 13). Hence, the set of real numbers is uncountable. ◀

RESULTS ABOUT CARDINALITY We will now discuss some results about the cardinality of sets. First, we will prove that the union of two countable sets is also countable.

THEOREM 2 If A and B are countable sets, then $A \cup B$ is also countable.

Proof: Suppose that A and B are both countable sets. Without loss of generality, we can assume that A and B are disjoint. (If they are not, we can replace B by $B - A$, because $A \cap (B - A) = \emptyset$ and $A \cup (B - A) = A \cup B$.) Furthermore, without loss of generality, if one of the two sets is countably infinite and other finite, we can assume that B is the one that is finite.

This proof uses WLOG and cases.

There are three cases to consider: (i) A and B are both finite, (ii) A is infinite and B is finite, and (iii) A and B are both countably infinite.

Case (i): Note that when A and B are finite, $A \cup B$ is also finite, and therefore, countable.

Case (ii): Because A is countably infinite, its elements can be listed in an infinite sequence $a_1, a_2, a_3, \ldots, a_n, \ldots$ and because B is finite, its terms can be listed as $b_1, b_2, \ldots, b_m$ for some positive integer m. We can list the elements of $A \cup B$ as $b_1, b_2, \ldots, b_m, a_1, a_2, a_3, \ldots, a_n, \ldots$. This means that $A \cup B$ is countably infinite.

Case (iii): Because both A and B are countably infinite, we can list their elements as a_1, $a_2, a_3, \ldots, a_n, \ldots$ and $b_1, b_2, b_3, \ldots, b_n, \ldots$, respectively. By alternating terms of these two sequences we can list the elements of $A \cup B$ in the infinite sequence a_1, b_1, a_2, b_2, a_3, $b_3, \ldots, a_n, b_n, \ldots$. This means $A \cup B$ must be countably infinite.

We have completed the proof, as we have shown that $A \cup B$ is countable in all three cases.

Because of its importance, we now state a key theorem in the study of cardinality.

THEOREM 3 **SCHRÖDER-BERNSTEIN THEOREM** If A and B are sets with $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$. In other words, if there are one-to-one functions f from A to B and g from B to A, then there is a one-to-one correspondence between A and B.

Because Theorem 2 seems to be quite straightforward, we might expect that it has an easy proof. However, even though it can be proved without using advanced mathematics, no known proof is easy to explain. Consequently, we omit a proof here. We refer the interested reader to [AiZiHo09] and [Ve06] for a proof. This result is called the Schröder-Bernstein theorem after Ernst Schröder who published a flawed proof of it in 1898 and Felix Bernstein, a student of Georg Cantor, who presented a proof in 1897. However, a proof of this theorem was found in notes of Richard Dedekind dated 1887. Dedekind was a German mathematician who made important contributions to the foundations of mathematics, abstract algebra, and number theory.

We illustrate the use of Theorem 2 with an example.

EXAMPLE 6 Show that the $|(0, 1)| = |(0, 1]|$.

Solution: It is not at all obvious how to find a one-to-one correspondence between $(0, 1)$ and $(0, 1]$ to show that $|(0, 1)| = |(0, 1]|$. Fortunately, we can use the Schröder-Bernstein theorem instead. Finding a one-to-one function from $(0, 1)$ to $(0, 1]$ is simple. Because $(0, 1) \subset (0, 1]$, $f(x) = x$ is a one-to-one function from $(0, 1)$ to $(0, 1]$. Finding a one-to-one function from $(0, 1]$ to $(0, 1)$ is also not difficult. The function $g(x) = x/2$ is clearly one-to-one and maps $(0, 1]$ to $(0, 1/2] \subset (0, 1)$. As we have found one-to-one functions from $(0, 1)$ to $(0, 1]$ and from $(0, 1]$ to $(0, 1)$, the Schröder-Bernstein theorem tells us that $|(0, 1)| = |(0, 1]|$.

DEFINITION 5 Let S be a set. An enumeration of S is a surjective function f from an initial segment of N to S. If f is injective also (and hence bijective), then f is an enumeration without repetitions. If f is not injective, then f is an enumeration with repetitions.

EXAMPLE 7

1. If $S = \{\alpha, \beta, \gamma, \delta\}$, then $< \alpha, \beta, \gamma, \delta >$ is an enumeration. $< \alpha, \gamma, \beta, \beta, \delta, \alpha >$ is an enumeration with repetition. $< \gamma, \alpha, \delta, \beta >$ is an enumeration without repetition.

2. Let S be the set of natural numbers of the form $3n$, where $n \in N$. $< 0, 3, 6, 9, \ldots >$ is an enumeration defined by $f(n) = 3n$ $< 6, 3, 0, 15, 12, 9, \ldots >$ is also an enumeration defined by

$$f(n) = 3n + 6 \text{ if } n = 3k \text{ for some } k$$

$$f(n) = 3n \text{ if } n = 3k + 1 \text{ for some } k \in N$$

$$f(n) = 3n - 6 \text{ if } n = 3k + 2 \text{ for some } k \in N$$

It is straightforward to see that a set S is countable if and only if there exists an enumeration of S.

EXAMPLE 8 Let $\Sigma = \{a, b\}$. The set of strings over Σ^* is a countably infinite set. An enumeration τ of Σ^* is possible. $\tau(0) = \lambda$, the empty string. Assuming a comes before b in Σ. $\tau(1) = a$ and $\tau(2) = b$. $\tau(3)$ to $\tau(6)$ are strings of length 3 and so on. What is the 49th string in the enumeration?

$$\tau(0) \text{ –string of length } 0$$

$$\tau(1) - \tau(2) \text{ –string of length } 1$$

$$\tau(3) - \tau(6) \text{ –string of length } 2$$

$$\tau(7) - \tau(14) \text{ –string of length } 3$$

$$\tau(15) - \tau(30) \text{ –string of length } 4$$

$$\tau(31) - \tau(62) \text{ –string of length } 5$$

Hence the 49th string is of length 5. Among these $\tau(31)$ to $\tau(46)$ begin with a and $\tau(47)$ to $\tau(62)$ begin with b.

$$\tau(47) = b\ a\ a\ a$$

$$\tau(48) = b\ a\ a\ b$$

$$\tau(49) = b\ a\ b\ a$$

Hence the 49th string in the enumeration is $b\ a\ b\ a$.

THEOREM 4 Every infinite set contains a countably infinite subset.

Proof: Let S be an infinite set. Select a sequence of elements $< a_0, a_1, \ldots >$ from S as follows:

Select a_0 from S

Select a_1 from $S - \{a_0\}$

Select a_2 from $S - \{a_0, a_1\}$

and so on. a_{i+1} will be selected from $S - \{a_0, a_1, \ldots, a_i\}$. Each $S - \{a_0, \ldots, a_i\}$ is infinite. Otherwise $(S - \{a_0, \ldots, a_i\}) \cup \{a_0, \ldots, a_i\} = S$ will be finite. This set $\{a_0, \ldots, a_i, \ldots\}$ is countable infinite subset of S.

THEOREM 5 The union of a countable collection of countable sets is countable.

Proof: Let S_0, S_1, S_2, ... be a countable collection of countable sets and $S_i =< a_{i0}, a_{i1}, a_{i2}, \ldots >$

$$
\begin{array}{lcccccccc}
S_0: & (a_{00} & & a_{01} & & a_{02} & & a_{03}\ldots) \\
 & & \swarrow & & \swarrow & & \swarrow & \\
S_1: & (a_{10} & & a_{11} & & a_{12} & & a_{13}\ldots) \\
 & & \swarrow & & \swarrow & & & \\
S_2: & (a_{20} & & a_{21} & & a_{22} & & a_{23}\ldots) \\
 & & \swarrow & & & & & \\
S_3: & (a_{30} & & a_{31} & & a_{32} & & a_{33}\ldots)
\end{array}
$$

Enumerate the elements as shown in the diagram.

Let $b_0, b_1, \ldots$ be the sequence of elements.

$b_0 = a_{00}$

$b_1 = a_{01}$

$b_2 = a_{10}$

$b_3 = a_{02}$

$b_4 = a_{11}$

$b_5 = a_{20}$ and so on.

This is an enumeration of the elements of $\bigcup_{i=0}^{\infty} A_i$ and hence $\bigcup_{i=0}^{\infty} A_i$ is countable.

THEOREM 6 Let Σ be a finite alphabet and Σ^* the set of all strings over Σ. Then $\mathcal{P}(\Sigma^*)$ is uncountable.

Proof: Let $< x_0, x_1, x_2, \ldots >$ be an enumeration of strings in Σ^* and if possible let $< A_0, A_1, \ldots >$ be an enumeration of the sets in $\mathcal{P}(\Sigma^*)$. Construct a binary matrix M as follows.

$$
\begin{array}{c|cccc}
 & x_0 & x_1 & x_2 & \cdots \\
\hline
A_0 & a_{00} & a_{01} & a_{02} & \cdots \\
A_1 & a_{10} & a_{11} & a_{12} & \cdots \\
A_2 & a_{20} & a_{21} & a_{22} & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{array}
$$

$a_{ij} = 1$ if $x_j \in A_i$ and $a_{ij} = 0$ if $x_j \notin A_i$. Now we construct a set A as follows $x_i \in A$ if $a_{ii} = 0$ (i.e., $x_i \notin A_i$, $A = \{x_i | x_i \notin A_i, i \in N\}$).

Now we can see that A cannot be any A_j and hence cannot appear in the enumeration $< A_0, A_1, \ldots >$ even though $A \in \mathcal{P}(\Sigma^*)$. This is seen by the following contradiction. Suppose if possible let $A = A_j$.

A consists of strings x_j such that $a_{jj} = 0$. So if $x_j \in A$, then $a_{jj} = 0$ and by the way we constructed the matrix M, if $a_{jj} = 0 x_j \notin A_j$. Hence $A \neq A_j$. Hence we cannot have an enumeration of the sets in $\mathcal{P}(\Sigma^*)$ and so $\mathcal{P}(\Sigma^*)$ is an uncountable set.

Can we compare the cardinality of uncountable sets? We have seen that the set [0, 1] (real numbers between 0 and 1) and $\mathcal{P}(\Sigma^*)$ are uncountable sets. Can we look at them as having the same cardinality? Are there uncountable sets X_1 and X_2 such that $|X_1| \leq |X_2|$?

DEFINITION 6

A set S is of cardinality c if there is a bijection from [0, 1] to S. The label c is given to this because of the fact that the set [0, 1] is often called a continuum.

DEFINITION 7

Let S and T be sets. Then S and T are equipotent or have the same cardinality, denoted by $|S| = |T|$, if there is a bijection from S to T.

The following result follows immediately.

THEOREM 7

Equipotence is an equivalence relation over any collection of sets.

DEFINITION 8

The cardinality of S is no greater than the cardinality of T, denoted as $|S| \leq |T|$, if there is an injection from S to T. The cardinality of S is less than the cardinality of T, written $|S| < |T|$, if there exists an injection but no bijection from S to T.

It is easy to see that if S and T are sets, then exactly one of the following three conditions will hold:

1. $|S| < |T|$
2. $|S| = |T|$
3. $|T| < |S|$

Also we can see that if $|S| \leq |T|$ and $|T| \leq |S|$, then $|T| = |S|$.

THEOREM 8

Let S be a finite set. Then $|S| < \aleph_0 < c$.

Proof: Let $|S| = n$. Let $Z_n = \{0, \ldots, n-1\}$, $|S| = |Z_n|$. Then there is an injection $f : Z_n \to N$ defined as $f(x) = x$. But we cannot have a bijection from $N \to S$. Hence $|S| < |N|$ and thus $|S| < \aleph_0$.

We have also seen that there cannot be a bijection from [0, 1] to N. But we can have an injection from N to [0, 1] defined as

$$f(n) = \frac{1}{n+2}.$$

Hence $|N| < |[0, 1]|$ or $\aleph_0 < c$.

THEOREM 9 If S is an infinite set, then $\aleph_0 \leq |S|$.

Proof: We have already seen that if S is infinite, then S contain a countably infinite subset S'. Since the mapping f defined as $f: S' \to S$

$$f(x) = x \text{ for all } x \in S'$$

is an injection from A' to A, we conclude that $|S'| \leq |S|$. $|S'| = \aleph_0$. Hence $\aleph_0 \leq |S|$.

This leads us to the famous continuum hypothesis, which asserts that there is no cardinal number between $\aleph_0$ and c. In other words, the **continuum hypothesis** states that there is no set A such that $\aleph_0$, the cardinality of the set of positive integers, is less than $|A|$ and $|A|$ is less than c, the cardinality of real numbers.

THEOREM 10 Let S be a set. Then $|S| < |\mathcal{P}(S)|$.

Proof: Consider the injection $: S \to \mathcal{P}(S)$ defined as $f(a) = \{a\}$ for all $a \in S$. Since we have such an injection

$$|S| \leq |\mathcal{P}(S)|.$$

Next we show $|S| \neq |\mathcal{P}(S)|$. Let g be an arbitrary function from S to $\mathcal{P}(S)$. We can show that g is not surjective and hence not bijective.

The function g maps each element of S to a subset of S. An element x may or may not be in the subset defined by $g(x)$. Now let us define a subset A of S as follows:

$$A = \{x | x \notin g(x)\}$$

A is a subset of S and hence $A \in \mathcal{P}(S)$.

We show that for no a in S, $g(a) = A$.

Suppose for some a, $g(a) = A$, then

$$\begin{aligned} a \in A &\leftrightarrow a \in \{x | x \notin g(x)\} \\ &\leftrightarrow a \notin g(a) \\ &\leftrightarrow a \notin A \end{aligned}$$

and we arrive at a contradiction and the assumption that $g(a) = A$ for some a in S is not correct. It follows that g is not surjective and hence not bijective. We have chosen g arbitrarily and hence we can conclude that no bijection exists from S to $\mathcal{P}(S)$ and therefore $|S| \neq |\mathcal{P}(S)|$.

Hence $|S| < |\mathcal{P}(S)|$.

Using the above result, we see that we have an infinite hierarchy of uncountable sets (i.e., infinite sequence of cardinal numbers, each of which is smaller than the next in the sequence).

$$\aleph_0 = |N| < |\mathcal{P}(N)| < |\mathcal{P}(\mathcal{P}(N))| < \cdots$$

The **continuum hypothesis** says that there is no cardinal number X between X_0 and c. It was stated by Cantor in 1877. He labored unsuccessfully to prove it, becoming extremely dismayed that he could not. By 1900, settling the continuum hypothesis was considered to be among the most important unsolved problems in mathematics. It was the first problem posed by David Hilbert in his famous 1900 list of open problems in mathematics.

The continuum hypothesis is still an open question and remains an area for active research. However, it has been shown that it can be neither proved nor disproved under the standard set theory axioms in modern mathematics, the Zermelo-Fraenkel axioms. The Zermelo-Fraenkel axioms were formulated to avoid the paradoxes of naive set theory, such as Russell's paradox, but there is much controversy whether they should be replaced by some other set of axioms for set theory.

UNCOMPUTABLE FUNCTIONS We will now describe an important application of the concepts of this section to computer science. In particular, we will show that there are functions whose values cannot be computed by any computer program.

DEFINITION 9 We say that a function is **computable** if there is a computer program in some programming language that finds the values of this function. If a function is not computable we say it is **uncomputable**.

To show that there are uncomputable functions, we need to establish two results. First, we need to show that the set of all computer programs in any particular programming language is countable. This can be proved by noting that a computer programs in a particular language can be thought of as a string of characters from a finite alphabet (see Exercise 29). Next, we show that there are uncountably many different functions from a particular countably infinite set to itself. In particular, Exercise 30 shows that the set of functions from the set of positive integers to itself is uncountable. This is a consequence of the uncountability of the real numbers between 0 and 1 (see Example 5). Putting these two results together shows that there are uncomputable functions.

Exercises

1. Determine whether each of these sets is finite, countably infinite, or uncountable. For those that are countably infinite, exhibit a one-to-one correspondence between the set of positive integers and that set.

a) the negative integers

b) the even integers

c) the integers less than 100

d) the real numbers between 0 and $\frac{1}{2}$

e) the positive integers less than 1,000,000,000

f) the integers that are multiples of 7

2. Determine whether each of these sets is finite, countably infinite, or uncountable. For those that are countably infinite, exhibit a one-to-one correspondence between the set of positive integers and that set.

a) the integers greater than 10

b) the odd negative integers

c) the integers with absolute value less than 1,000,000

d) the real numbers between 0 and 2

e) the set $A \times \mathbf{Z}^+$ where $A = \{2, 3\}$

f) the integers that are multiples of 10

3. Determine whether each of these sets is countable or uncountable. For those that are countably infinite, exhibit a one-to-one correspondence between the set of positive integers and that set.

a) all bit strings not containing the bit 0

b) all positive rational numbers that cannot be written with denominators less than 4

c) the real numbers not containing 0 in their decimal representation

d) the real numbers containing only a finite number of 1s in their decimal representation

4. Show that a finite group of guests arriving at Hilbert's fully occupied Grand Hotel can be given rooms without evicting any current guest.

5. Suppose that Hilbert's Grand Hotel is fully occupied on the day the hotel expands to a second building which also contains a countably infinite number of rooms. Show that the current guests can be spread out to fill every room of the two buildings of the hotel.

6. Show that a countably infinite number of guests arriving at Hilbert's fully occupied Grand Hotel can be given rooms without evicting any current guest.

***7.** Suppose that a countably infinite number of buses, each containing a countably infinite number of guests, arrive at Hilbert's fully occupied Grand Hotel. Show that all the arriving guests can be accommodated without evicting any current guest.

8. Give an example of two uncountable sets A and B such that $A - B$ is

a) finite.

b) countably infinite.

c) uncountable.

9. Give an example of two uncountable sets A and B such that $A \cap B$ is

a) finite.

b) countably infinite.

c) uncountable.

10. Show that if A and B are sets and $A \subset B$ then $|A| \leq |B|$.

11. Explain why the set A is countable if and only if $|A| \leq |\mathbf{Z}^+|$.

12. Show that if A and B are sets with the same cardinality, then $|A| \leq |B|$ and $|B| \leq |A|$.

☞ **13.** Show that if A and B are sets, A is uncountable, and $A \subseteq B$, then B is uncountable.

☞ **14.** Show that a subset of a countable set is also countable.

15. If A is an uncountable set and B is a countable set, must $A - B$ be uncountable?

16. Show that if A and B are sets $|A| = |B|$, then $|\mathcal{P}(A)| = |\mathcal{P}(B)|$.

17. Show that if A is an infinite set, then it contains a countably infinite subset.

18. Show that there is no infinite set A such that $|A| < |\mathbf{Z}^+| = \aleph_0$.

19. Prove that if it is possible to label each element of an infinite set S with a finite string of keyboard characters, from a finite list characters, where no two elements of S have the same label, then S is a countably infinite set.

20. Use Exercise 19 to provide a proof different from that in the text that the set of rational numbers is countable. [*Hint:* Show that you can express a rational number as a string of digits with a slash and possibly a minus sign.]

***21.** Show that the union of a countable number of countable sets is countable.

***22.** Show that the set of real numbers that are solutions of quadratic equations $ax^2 + bx + c = 0$, where a, b, and c are integers, is countable.

***23.** Show that $\mathbf{Z}^+ \times \mathbf{Z}^+$ is countable by showing that the polynomial function $f : \mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ with $f(m, n) = (m + n - 2)(m + n - 1)/2 + m$ is one-to-one and onto.

***24.** Show that when you substitute $(3n + 1)^2$ for each occurrence of n and $(3m + 1)^2$ for each occurrence of m in the right-hand side of the formula for the function $f(m, n)$ in Exercise 31, you obtain a one-to-one polynomial function $\mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$. It is an open question whether there is a one-to-one polynomial function $\mathbf{Q} \times \mathbf{Q} \to \mathbf{Q}$.

25. Use the Schröder-Bernstein theorem to show that $(0, 1)$ and $[0, 1]$ have the same cardinality

26. Show that $(0, 1)$ and $\mathbf{R}$ have the same cardinality. [*Hint:* Use the Schröder-Bernstein theorem.]

27. Show that there is no one-to-one correspondence from the set of positive integers to the power set of the set of positive integers. [*Hint:* Assume that there is such a one-to-one correspondence. Represent a subset of the set of positive integers as an infinite bit string with ith bit 1 if i belongs to the subset and 0 otherwise. Suppose that you can list these infinite strings in a sequence indexed by the positive integers. Construct a new bit string with its ith bit equal to the complement of the ith bit of the ith string in the list. Show that this new bit string cannot appear in the list.]

***28.** Show that there is a one-to-one correspondence from the set of subsets of the positive integers to the set real numbers between 0 and 1. Use this result and Exercises 26 and 27 to conclude that $\aleph_0 < |\mathcal{P}(\mathbf{Z}^+)| = |\mathbf{R}|$. [*Hint:* Look at the first part of the hint for Exercise 27.]

***29.** Show that the set of all computer programs in a particular programming language is countable. [*Hint:* A computer program written in a programming language can be thought of as a string of symbols from a finite alphabet.]

***30.** Show that the set of functions from the positive integers to the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is uncountable. [*Hint:* First set up a one-to-one correspondence between the set of real numbers between 0 and 1 and a subset of these functions. Do this by associating to the real number $0.d_1d_2 \ldots d_n \ldots$ the function f with $f(n) = d_n$.]

2.6 Matrices

Introduction

Matrices are used throughout discrete mathematics to express relationships between elements in sets. In subsequent chapters we will use matrices in a wide variety of models. For instance, matrices will be used in models of communications networks and transportation systems. Many algorithms will be developed that use these matrix models. This section reviews matrix arithmetic that will be used in these algorithms.

DEFINITION 1

A *matrix* is a rectangular array of numbers. A matrix with m rows and n columns is called an $m \times n$ matrix. The plural of matrix is *matrices.* A matrix with the same number of rows as columns is called *square.* Two matrices are *equal* if they have the same number of rows and the same number of columns and the corresponding entries in every position are equal.

EXAMPLE 1 The matrix $\begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 1 & 3 \end{bmatrix}$ is a 3×2 matrix. ◀

We now introduce some terminology about matrices. Boldface uppercase letters will be used to represent matrices.

DEFINITION 2

Let m and n be positive integers and let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

The ith *row* of $\mathbf{A}$ is the $1 \times n$ matrix $[a_{i1}, a_{i2}, \dots, a_{in}]$. The jth *column* of $\mathbf{A}$ is the $m \times 1$ matrix

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ \cdot \\ a_{mj} \end{bmatrix}.$$

The (i, j)th *element* or *entry* of $\mathbf{A}$ is the element a_{ij}, that is, the number in the ith row and jth column of $\mathbf{A}$. A convenient shorthand notation for expressing the matrix $\mathbf{A}$ is to write $\mathbf{A} = [a_{ij}]$, which indicates that $\mathbf{A}$ is the matrix with its (i, j)th element equal to a_{ij}.

Matrix Arithmetic

The basic operations of matrix arithmetic will now be discussed, beginning with a definition of matrix addition.

DEFINITION 3 Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $m \times n$ matrices. The *sum* of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} + \mathbf{B}$, is the $m \times n$ matrix that has $a_{ij} + b_{ij}$ as its (i, j)th element. In other words, $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$.

The sum of two matrices of the same size is obtained by adding elements in the corresponding positions. Matrices of different sizes cannot be added, because the sum of two matrices is defined only when both matrices have the same number of rows and the same number of columns.

EXAMPLE 2 We have

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

◀

We now discuss matrix products. A product of two matrices is defined only when the number of columns in the first matrix equals the number of rows of the second matrix.

DEFINITION 4 Let $\mathbf{A}$ be an $m \times k$ matrix and $\mathbf{B}$ be a $k \times n$ matrix. The *product* of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{AB}$, is the $m \times n$ matrix with its (i, j)th entry equal to the sum of the products of the corresponding elements from the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$. In other words, if $\mathbf{AB} = [c_{ij}]$, then

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}.$$

In Figure 1 the colored row of $\mathbf{A}$ and the colored column of $\mathbf{B}$ are used to compute the element c_{ij} of $\mathbf{AB}$. The product of two matrices is not defined when the number of columns in the first matrix and the number of rows in the second matrix are not the same.

We now give some examples of matrix products.

EXAMPLE 3 Let

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

Find $\mathbf{AB}$ if it is defined.

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ik} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mk} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1j} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2j} & \ldots & b_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{k1} & b_{k2} & \ldots & b_{kj} & \ldots & b_{kn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \vdots & \vdots & c_{ij} & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mn} \end{bmatrix}$$

FIGURE 1 The Product of $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$.

Extra Examples

Solution: Because $\mathbf{A}$ is a 4×3 matrix and $\mathbf{B}$ is a 3×2 matrix, the product $\mathbf{AB}$ is defined and is a 4×2 matrix. To find the elements of $\mathbf{AB}$, the corresponding elements of the rows of $\mathbf{A}$ and the columns of $\mathbf{B}$ are first multiplied and then these products are added. For instance, the element in the $(3, 1)$th position of $\mathbf{AB}$ is the sum of the products of the corresponding elements of the third row of $\mathbf{A}$ and the first column of $\mathbf{B}$; namely, $3 \cdot 2 + 1 \cdot 1 + 0 \cdot 3 = 7$. When all the elements of $\mathbf{AB}$ are computed, we see that

$$\mathbf{AB} = \begin{bmatrix} 14 & 4 \\ 8 & 9 \\ 7 & 13 \\ 8 & 2 \end{bmatrix}.$$

◂

Matrix multiplication is *not* commutative. That is, if $\mathbf{A}$ and $\mathbf{B}$ are two matrices, it is not necessarily true that $\mathbf{AB}$ and $\mathbf{BA}$ are the same. In fact, it may be that only one of these two products is defined. For instance, if $\mathbf{A}$ is 2×3 and $\mathbf{B}$ is 3×4, then $\mathbf{AB}$ is defined and is 2×4; however, $\mathbf{BA}$ is not defined, because it is impossible to multiply a 3×4 matrix and a 2×3 matrix.

In general, suppose that $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $r \times s$ matrix. Then $\mathbf{AB}$ is defined only when $n = r$ and $\mathbf{BA}$ is defined only when $s = m$. Moreover, even when $\mathbf{AB}$ and $\mathbf{BA}$ are both defined, they will not be the same size unless $m = n = r = s$. Hence, if both $\mathbf{AB}$ and $\mathbf{BA}$ are defined and are the same size, then both $\mathbf{A}$ and $\mathbf{B}$ must be square and of the same size. Furthermore, even with $\mathbf{A}$ and $\mathbf{B}$ both $n \times n$ matrices, $\mathbf{AB}$ and $\mathbf{BA}$ are not necessarily equal, as Example 4 demonstrates.

EXAMPLE 4 Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Does $\mathbf{AB} = \mathbf{BA}$?

Solution: We find that

$$\mathbf{AB} = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}.$$

Hence, $\mathbf{AB} \neq \mathbf{BA}$. ◂

Transposes and Powers of Matrices

We now introduce an important matrix with entries that are zeros and ones.

DEFINITION 5 The *identity matrix of order* n is the $n \times n$ matrix $\mathbf{I}_n = [\delta_{ij}]$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Hence

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Multiplying a matrix by an appropriately sized identity matrix does not change this matrix. In other words, when $\mathbf{A}$ is an $m \times n$ matrix, we have

$$\mathbf{A}\mathbf{I}_n = \mathbf{I}_m\mathbf{A} = \mathbf{A}.$$

Powers of square matrices can be defined. When $\mathbf{A}$ is an $n \times n$ matrix, we have

$$\mathbf{A}^0 = \mathbf{I}_n, \qquad \mathbf{A}^r = \underbrace{\mathbf{A}\mathbf{A}\mathbf{A}\cdots\mathbf{A}}_{r \text{ times}}.$$

The operation of interchanging the rows and columns of a square matrix arises in many contexts.

DEFINITION 6 Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix. The *transpose* of $\mathbf{A}$, denoted by $\mathbf{A}^t$, is the $n \times m$ matrix obtained by interchanging the rows and columns of $\mathbf{A}$. In other words, if $\mathbf{A}^t = [b_{ij}]$, then $b_{ij} = a_{ji}$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

EXAMPLE 5 The transpose of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ is the matrix $\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$. ◀

Matrices that do not change when their rows and columns are interchanged are often important.

DEFINITION 7 A square matrix $\mathbf{A}$ is called *symmetric* if $\mathbf{A} = \mathbf{A}^t$. Thus $\mathbf{A} = [a_{ij}]$ is symmetric if $a_{ij} = a_{ji}$ for all i and j with $1 \le i \le n$ and $1 \le j \le n$.

Note that a matrix is symmetric if and only if it is square and it is symmetric with respect to its main diagonal (which consists of entries that are in the ith row and ith column for some i). This symmetry is displayed in Figure 2.

EXAMPLE 6 The matrix $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ is symmetric. ◀

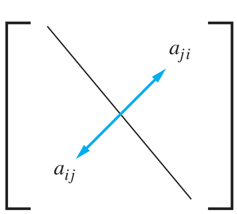


FIGURE 2 **A Symmetric Matrix.**

Zero–One Matrices

A matrix all of whose entries are either 0 or 1 is called a **zero–one matrix**. Zero–one matrices are often used to represent discrete structures, as we will see in Chapters 9 and 10. Algorithms using these structures are based on Boolean arithmetic with zero–one matrices. This arithmetic is based on the Boolean operations $\wedge$ and $\vee$, which operate on pairs of bits, defined by

$$b_1 \wedge b_2 = \begin{cases} 1 & \text{if } b_1 = b_2 = 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$b_1 \vee b_2 = \begin{cases} 1 & \text{if } b_1 = 1 \text{ or } b_2 = 1 \\ 0 & \text{otherwise.} \end{cases}$$

DEFINITION 8 Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $m \times n$ zero–one matrices. Then the *join* of $\mathbf{A}$ and $\mathbf{B}$ is the zero–one matrix with (i, j)th entry $a_{ij} \vee b_{ij}$. The join of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \vee \mathbf{B}$. The *meet* of $\mathbf{A}$ and $\mathbf{B}$ is the zero–one matrix with (i, j)th entry $a_{ij} \wedge b_{ij}$. The meet of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \wedge \mathbf{B}$.

EXAMPLE 7 Find the join and meet of the zero–one matrices

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

Solution: We find that the join of $\mathbf{A}$ and $\mathbf{B}$ is

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

The meet of $\mathbf{A}$ and $\mathbf{B}$ is

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

We now define the **Boolean product** of two matrices.

DEFINITION 9 Let $\mathbf{A} = [a_{ij}]$ be an $m \times k$ zero–one matrix and $\mathbf{B} = [b_{ij}]$ be a $k \times n$ zero–one matrix. Then the *Boolean product* of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \odot \mathbf{B}$, is the $m \times n$ matrix with (i, j)th entry c_{ij} where

$$c_{ij} = (a_{i1} \wedge b_{1j}) \vee (a_{i2} \wedge b_{2j}) \vee \cdots \vee (a_{ik} \wedge b_{kj}).$$

Note that the Boolean product of $\mathbf{A}$ and $\mathbf{B}$ is obtained in an analogous way to the ordinary product of these matrices, but with addition replaced with the operation $\vee$ and with multiplication replaced with the operation $\wedge$. We give an example of the Boolean products of matrices.

EXAMPLE 8 Find the Boolean product of $\mathbf{A}$ and $\mathbf{B}$, where

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

Solution: The Boolean product $\mathbf{A} \odot \mathbf{B}$ is given by

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} (1 \wedge 1) \vee (0 \wedge 0) & (1 \wedge 1) \vee (0 \wedge 1) & (1 \wedge 0) \vee (0 \wedge 1) \\ (0 \wedge 1) \vee (1 \wedge 0) & (0 \wedge 1) \vee (1 \wedge 1) & (0 \wedge 0) \vee (1 \wedge 1) \\ (1 \wedge 1) \vee (0 \wedge 0) & (1 \wedge 1) \vee (0 \wedge 1) & (1 \wedge 0) \vee (0 \wedge 1) \end{bmatrix}$$

$$= \begin{bmatrix} 1 \vee 0 & 1 \vee 0 & 0 \vee 0 \\ 0 \vee 0 & 0 \vee 1 & 0 \vee 1 \\ 1 \vee 0 & 1 \vee 0 & 0 \vee 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$ ◀

We can also define the Boolean powers of a square zero–one matrix. These powers will be used in our subsequent studies of paths in graphs, which are used to model such things as communications paths in computer networks.

DEFINITION 10 Let $\mathbf{A}$ be a square zero–one matrix and let r be a positive integer. The rth Boolean power of $\mathbf{A}$ is the Boolean product of r factors of $\mathbf{A}$. The rth Boolean product of $\mathbf{A}$ is denoted by $\mathbf{A}^{[r]}$.

Hence

$$\mathbf{A}^{[r]} = \underbrace{\mathbf{A} \odot \mathbf{A} \odot \mathbf{A} \odot \cdots \odot \mathbf{A}}_{r \text{ times}}.$$

(This is well defined because the Boolean product of matrices is associative.) We also define $\mathbf{A}^{[0]}$ to be $\mathbf{I}_n$.

EXAMPLE 9 Let $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$. Find $\mathbf{A}^{[n]}$ for all positive integers n.

Solution: We find that

$$\mathbf{A}^{[2]} = \mathbf{A} \odot \mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

We also find that

$$\mathbf{A}^{[3]} = \mathbf{A}^{[2]} \odot \mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \mathbf{A}^{[4]} = \mathbf{A}^{[3]} \odot \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Additional computation shows that

$$\mathbf{A}^{[5]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The reader can now see that $\mathbf{A}^{[n]} = \mathbf{A}^{[5]}$ for all positive integers n with $n \geq 5$. ◀

Exercises

1. Let $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 0 & 4 & 6 \\ 1 & 1 & 3 & 7 \end{bmatrix}$.

a) What size is $\mathbf{A}$?
b) What is the third column of $\mathbf{A}$?
c) What is the second row of $\mathbf{A}$?
d) What is the element of $\mathbf{A}$ in the (3, 2)th position?
e) What is $\mathbf{A}^t$?

2. Find $\mathbf{A} + \mathbf{B}$, where

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

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

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

$\mathbf{B} = \begin{bmatrix} -3 & 9 & -3 & 4 \\ 0 & -2 & -1 & 2 \end{bmatrix}$.

3. Find $\mathbf{AB}$ if

a) $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 4 \\ 1 & 3 \end{bmatrix}$.

b) $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & 3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 3 & -2 & -1 \\ 1 & 0 & 2 \end{bmatrix}$.

c) $\mathbf{A} = \begin{bmatrix} 4 & -3 \\ 3 & -1 \\ 0 & -2 \\ -1 & 5 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -1 & 3 & 2 & -2 \\ 0 & -1 & 4 & -3 \end{bmatrix}$.

4. Find the product **AB**, where

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

b) $\mathbf{A} = \begin{bmatrix} 1 & -3 & 0 \\ 1 & 2 & 2 \\ 2 & 1 & -1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & -1 & 2 & 3 \\ -1 & 0 & 3 & -1 \\ -3 & -2 & 0 & 2 \end{bmatrix}$.

c) $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 7 & 2 \\ -4 & -3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 4 & -1 & 2 & 3 & 0 \\ -2 & 0 & 3 & 4 & 1 \end{bmatrix}$.

5. Find a matrix **A** such that

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

[*Hint:* Finding **A** requires that you solve systems of linear equations.]

6. Find a matrix **A** such that

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 1 \\ 4 & 0 & 3 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 7 & 1 & 3 \\ 1 & 0 & 3 \\ -1 & -3 & 7 \end{bmatrix}.$$

7. Let **A** be an $m \times n$ matrix and let **0** be the $m \times n$ matrix that has all entries equal to zero. Show that $\mathbf{A} = \mathbf{0} + \mathbf{A} = \mathbf{A} + \mathbf{0}$.

8. Show that matrix addition is commutative; that is, show that if **A** and **B** are both $m \times n$ matrices, then $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

9. Show that matrix addition is associative; that is, show that if **A**, **B**, and **C** are all $m \times n$ matrices, then $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$.

10. Let **A** be a 3×4 matrix, **B** be a 4×5 matrix, and **C** be a 4×4 matrix. Determine which of the following products are defined and find the size of those that are defined.

a) **AB** **b)** **BA** **c)** **AC**
d) **CA** **e)** **BC** **f)** **CB**

11. What do we know about the sizes of the matrices **A** and **B** if both of the products **AB** and **BA** are defined?

12. In this exercise we show that matrix multiplication is distributive over matrix addition.

a) Suppose that **A** and **B** are $m \times k$ matrices and that **C** is a $k \times n$ matrix. Show that $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

b) Suppose that **C** is an $m \times k$ matrix and that **A** and **B** are $k \times n$ matrices. Show that $\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$.

13. In this exercise we show that matrix multiplication is associative. Suppose that **A** is an $m \times p$ matrix, **B** is a $p \times k$ matrix, and **C** is a $k \times n$ matrix. Show that $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

14. The $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ is called a **diagonal matrix** if $a_{ij} = 0$ when $i \neq j$. Show that the product of two $n \times n$ diagonal matrices is again a diagonal matrix. Give a simple rule for determining this product.

15. Let

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

Find a formula for $\mathbf{A}^n$, whenever n is a positive integer.

16. Show that $(\mathbf{A}^t)^t = \mathbf{A}$.

17. Let **A** and **B** be two $n \times n$ matrices. Show that

a) $(\mathbf{A} + \mathbf{B})^t = \mathbf{A}^t + \mathbf{B}^t$.

b) $(\mathbf{AB})^t = \mathbf{B}^t\mathbf{A}^t$.

If **A** and **B** are $n \times n$ matrices with $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, then **B** is called the **inverse** of **A** (this terminology is appropriate because such a matrix **B** is unique) and **A** is said to be **invertible**. The notation $\mathbf{B} = \mathbf{A}^{-1}$ denotes that **B** is the inverse of **A**.

18. Show that

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

is the inverse of

$$\begin{bmatrix} 7 & -8 & 5 \\ -4 & 5 & -3 \\ 1 & -1 & 1 \end{bmatrix}.$$

19. Let **A** be the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Show that if $ad - bc \neq 0$, then

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

20. Let

$$\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}.$$

a) Find $\mathbf{A}^{-1}$. [*Hint:* Use Exercise 19.]

b) Find $\mathbf{A}^3$.

c) Find $(\mathbf{A}^{-1})^3$.

d) Use your answers to (b) and (c) to show that $(\mathbf{A}^{-1})^3$ is the inverse of $\mathbf{A}^3$.

21. Let **A** be an invertible matrix. Show that $(\mathbf{A}^n)^{-1} = (\mathbf{A}^{-1})^n$ whenever n is a positive integer.

22. Let **A** be a matrix. Show that the matrix $\mathbf{AA}^t$ is symmetric. [*Hint:* Show that this matrix equals its transpose with the help of Exercise 17b.]

23. Suppose that $\mathbf{A}$ is an $n \times n$ matrix where n is a positive integer. Show that $\mathbf{A} + \mathbf{A}^t$ is symmetric.

24. **a)** Show that the system of simultaneous linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

in the variables $x_1, x_2, \ldots, x_n$ can be expressed as $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} = [a_{ij}]$, $\mathbf{X}$ is an $n \times 1$ matrix with x_i the entry in its ith row, and $\mathbf{B}$ is an $n \times 1$ matrix with b_i the entry in its ith row.

b) Show that if the matrix $\mathbf{A} = [a_{ij}]$ is invertible (as defined in the preamble to Exercise 18), then the solution of the system in part (a) can be found using the equation $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$.

25. Use Exercises 18 and 24 to solve the system

$$\begin{aligned} 7x_1 - 8x_2 + 5x_3 &= 5 \\ -4x_1 + 5x_2 - 3x_3 &= -3 \\ x_1 - x_2 + x_3 &= 0 \end{aligned}$$

26. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Find

a) $\mathbf{A} \vee \mathbf{B}$. **b)** $\mathbf{A} \wedge \mathbf{B}$. **c)** $\mathbf{A} \odot \mathbf{B}$.

27. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Find

a) $\mathbf{A} \vee \mathbf{B}$. **b)** $\mathbf{A} \wedge \mathbf{B}$. **c)** $\mathbf{A} \odot \mathbf{B}$.

28. Find the Boolean product of $\mathbf{A}$ and $\mathbf{B}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

29. Let

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

Find

a) $\mathbf{A}^{[2]}$. **b)** $\mathbf{A}^{[3]}$.

c) $\mathbf{A} \vee \mathbf{A}^{[2]} \vee \mathbf{A}^{[3]}$.

30. Let $\mathbf{A}$ be a zero–one matrix. Show that

a) $\mathbf{A} \vee \mathbf{A} = \mathbf{A}$. **b)** $\mathbf{A} \wedge \mathbf{A} = \mathbf{A}$.

31. In this exercise we show that the meet and join operations are commutative. Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ zero–one matrices. Show that

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

32. In this exercise we show that the meet and join operations are associative. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be $m \times n$ zero–one matrices. Show that

a) $(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C} = \mathbf{A} \vee (\mathbf{B} \vee \mathbf{C})$.

b) $(\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C} = \mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C})$.

33. We will establish distributive laws of the meet over the join operation in this exercise. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be $m \times n$ zero–one matrices. Show that

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

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

34. Let $\mathbf{A}$ be an $n \times n$ zero–one matrix. Let $\mathbf{I}$ be the $n \times n$ identity matrix. Show that $\mathbf{A} \odot \mathbf{I} = \mathbf{I} \odot \mathbf{A} = \mathbf{A}$.

35. In this exercise we will show that the Boolean product of zero–one matrices is associative. Assume that $\mathbf{A}$ is an $m \times p$ zero–one matrix, $\mathbf{B}$ is a $p \times k$ zero–one matrix, and $\mathbf{C}$ is a $k \times n$ zero–one matrix. Show that $\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$.

Key Terms and Results

TERMS

set: a collection of distinct objects

axiom: a basic assumption of a theory

paradox: a logical inconsistency

element, member of a set: an object in a set

roster method: a method that describes a set by listing its elements

set builder notation: the notation that describes a set by stating a property an element must have to be a member

∅ (empty set, null set): the set with no members

universal set: the set containing all objects under consideration

Venn diagram: a graphical representation of a set or sets

$S = T$ (set equality): S and T have the same elements

$S \subseteq T$ (S is a subset of T): every element of S is also an element of T

$S \subset T$ (S is a proper subset of T): S is a subset of T and $S \neq T$

finite set: a set with n elements, where n is a nonnegative integer

infinite set: a set that is not finite

$|S|$ (the cardinality of S): the number of elements in S

$P(S)$ (the power set of S): the set of all subsets of S

$A \cup B$ (the union of A and B): the set containing those elements that are in at least one of A and B

$A \cap B$ (the intersection of A and B): the set containing those elements that are in both A and B.

$A - B$ (the difference of A and B): the set containing those elements that are in A but not in B

$\overline{A}$ (the complement of A): the set of elements in the universal set that are not in A

$A \oplus B$ (the symmetric difference of A and B): the set containing those elements in exactly one of A and B

membership table: a table displaying the membership of elements in sets

function from A to B: an assignment of exactly one element of B to each element of A

domain of f: the set A, where f is a function from A to B

codomain of f: the set B, where f is a function from A to B

b is the image of a under f: $b = f(a)$

a is a pre-image of b under f: $f(a) = b$

range of f: the set of images of f

onto function, surjection: a function from A to B such that every element of B is the image of some element in A

one-to-one function, injection: a function such that the images of elements in its domain are distinct

one-to-one correspondence, bijection: a function that is both one-to-one and onto

inverse of f: the function that reverses the correspondence given by f (when f is a bijection)

$f \circ g$ (composition of f and g): the function that assigns $f(g(x))$ to x

$\lfloor x \rfloor$ (floor function): the largest integer not exceeding x

$\lceil x \rceil$ (ceiling function): the smallest integer greater than or equal to x

partial function: an assignment to each element in a subset of the domain a unique element in the codomain

sequence: a function with domain that is a subset of the set of integers

geometric progression: a sequence of the form $a, ar, ar^2, \ldots,$ where a and r are real numbers

arithmetic progression: a sequence of the form a, $a + d$, $a + 2d, \ldots,$ where a and d are real numbers

string: a finite sequence

empty string: a string of length zero

recurrence relation: a equation that expresses the nth term a_n of a sequence in terms of one or more of the previous terms of the sequence for all integers n greater than a particular integer

$\sum_{i=1}^{n} a_i$: the sum $a_1 + a_2 + \cdots + a_n$

$\prod_{i=1}^{n} a_i$: the product $a_1 a_2 \cdots a_n$

cardinality: two sets A and B have the same cardinality if there is a one-to-one correspondence from A to B

countable set: a set that either is finite or can be placed in one-to-one correspondence with the set of positive integers

uncountable set: a set that is not countable

$\aleph_0$ (aleph null): the cardinality of a countable set

$\mathfrak{c}$: the cardinality of the set of real numbers

Cantor diagonalization argument: a proof technique used to show that the set of real numbers is uncountable

computable function: a function for which there is a computer program in some programming language that finds its values

uncomputable function: a function for which no computer program in a programming language exists that finds its values

continuum hypothesis: the statement there no set A exists such that $\aleph_0 < |A| < \mathfrak{c}$

matrix: a rectangular array of numbers

matrix addition: see page 178

matrix multiplication: see page 179

$\mathbf{I}_n$ (identity matrix of order n): the $n \times n$ matrix that has entries equal to 1 on its diagonal and 0s elsewhere

$\mathbf{A}^t$ (transpose of A): the matrix obtained from **A** by interchanging the rows and columns

symmetric matrix: a matrix is symmetric if it equals its transpose

zero–one matrix: a matrix with each entry equal to either 0 or 1

$\mathbf{A} \vee \mathbf{B}$ (the join of A and B): see page 181

$\mathbf{A} \wedge \mathbf{B}$ (the meet of A and B): see page 181

$\mathbf{A} \odot \mathbf{B}$ (the Boolean product of A and B): see page 182

RESULTS

The set identities given in Table 1 in Section 2.2
The summation formulae in Table 2 in Section 2.4
The set of rational numbers is countable.
The set of real numbers is uncountable.

Review Questions

1. Explain what it means for one set to be a subset of another set. How do you prove that one set is a subset of another set?

2. What is the empty set? Show that the empty set is a subset of every set.

3. a) Define $|S|$, the cardinality of the set S.
   b) Give a formula for $|A \cup B|$, where A and B are sets.

4. a) Define the power set of a set S.
   b) When is the empty set in the power set of a set S?
   c) How many elements does the power set of a set S with n elements have?

5. a) Define the union, intersection, difference, and symmetric difference of two sets.
   b) What are the union, intersection, difference, and symmetric difference of the set of positive integers and the set of odd integers?

6. a) Explain what it means for two sets to be equal.
   b) Describe as many of the ways as you can to show that two sets are equal.
   c) Show in at least two different ways that the sets $A - (B \cap C)$ and $(A - B) \cup (A - C)$ are equal.

7. Explain the relationship between logical equivalences and set identities.

8. a) Define the domain, codomain, and range of a function.
   b) Let $f(n)$ be the function from the set of integers to the set of integers such that $f(n) = n^2 + 1$. What are the domain, codomain, and range of this function?
9. a) Define what it means for a function from the set of positive integers to the set of positive integers to be one-to-one.
   b) Define what it means for a function from the set of positive integers to the set of positive integers to be onto.
   c) Give an example of a function from the set of positive integers to the set of positive integers that is both one-to-one and onto.
   d) Give an example of a function from the set of positive integers to the set of positive integers that is one-to-one but not onto.
   e) Give an example of a function from the set of positive integers to the set of positive integers that is not one-to-one but is onto.
   f) Give an example of a function from the set of positive integers to the set of positive integers that is neither one-to-one nor onto.
10. a) Define the inverse of a function.
   b) When does a function have an inverse?
   c) Does the function $f(n) = 10 - n$ from the set of integers to the set of integers have an inverse? If so, what is it?
11. a) Define the floor and ceiling functions from the set of real numbers to the set of integers.
   b) For which real numbers x is it true that $\lfloor x \rfloor = \lceil x \rceil$?
12. Conjecture a formula for the terms of the sequence that begins 8, 14, 32, 86, 248 and find the next three terms of your sequence.
13. Suppose that $a_n = a_{n-1} - 5$ for $n = 1, 2, \ldots$. Find a formula for a_n.
14. What is the sum of the terms of the geometric progression $a + ar + \cdots + ar^n$ when $r \neq 1$?
15. Show that the set of odd integers is countable.
16. Give an example of an uncountable set.
17. Define the product of two matrices **A** and **B**. When is this product defined?
18. Show that matrix multiplication is not commutative.

Supplementary Exercises

1. Let A be the set of English words that contain the letter x, and let B be the set of English words that contain the letter q. Express each of these sets as a combination of A and B.
   a) The set of English words that do not contain the letter x.
   b) The set of English words that contain both an x and a q.
   c) The set of English words that contain an x but not a q.
   d) The set of English words that do not contain either an x or a q.
   e) The set of English words that contain an x or a q, but not both.
2. Show that if A is a subset of B, then the power set of A is a subset of the power set of B.
3. Suppose that A and B are sets such that the power set of A is a subset of the power set of B. Does it follow that A is a subset of B?
4. Let **E** denote the set of even integers and **O** denote the set of odd integers. As usual, let **Z** denote the set of all integers. Determine each of these sets.
   a) $\mathbf{E} \cup \mathbf{O}$ b) $\mathbf{E} \cap \mathbf{O}$ c) $\mathbf{Z} - \mathbf{E}$ d) $\mathbf{Z} - \mathbf{O}$
5. Show that if A and B are finite sets, then $|A \cap B| \leq |A \cup B|$. Determine when this relationship is an equality.
6. Let A and B be sets in a finite universal set U. List the following in order of increasing size.
   a) $|A|, |A \cup B|, |A \cap B|, |U|, |\emptyset|$
   b) $|A - B|, |A \oplus B|, |A| + |B|, |A \cup B|, |\emptyset|$
7. Let f and g be functions from $\{1, 2, 3, 4\}$ to $\{a, b, c, d\}$ and from $\{a, b, c, d\}$ to $\{1, 2, 3, 4\}$, respectively, with $f(1) = d$, $f(2) = c$, $f(3) = a$, and $f(4) = b$, and $g(a) = 2$, $g(b) = 1$, $g(c) = 3$, and $g(d) = 2$.
   a) Is f one-to-one? Is g one-to-one?
   b) Is f onto? Is g onto?
   c) Does either f or g have an inverse? If so, find this inverse.
8. Suppose that f is a function from A to B where A and B are finite sets. Explain why $|f(S)| \leq |S|$ for all subsets S of A.
9. Suppose that f is a function from A to B where A and B are finite sets. Explain why $|f(S)| = |S|$ for all subsets S of A if and only if f is one-to-one.

Suppose that f is a function from A to B. We define the function S_f from $\mathcal{P}(A)$ to $\mathcal{P}(B)$ by the rule $S_f(X) = f(X)$ for each subset X of A. Similarly, we define the function $S_{f^{-1}}$ from $\mathcal{P}(B)$ to $\mathcal{P}(A)$ by the rule $S_{f^{-1}}(Y) = f^{-1}(Y)$ for each subset Y of B. Here, we are using Definition 4, and the definition of the inverse image of a set found in the preamble to Exercise 28, both in Section 2.3.

*10. Suppose that f is a function from the set A to the set B. Prove that
   a) if f is one-to-one, then S_f is a one-to-one function from $\mathcal{P}(A)$ to $\mathcal{P}(B)$.
   b) if f is onto function, then S_f is an onto function from $\mathcal{P}(A)$ to $\mathcal{P}(B)$.
   c) if f is onto function, then $S_{f^{-1}}$ is a one-to-one function from $\mathcal{P}(B)$ to $\mathcal{P}(A)$.

d) if f is one-to-one, then $S_{f^{-1}}$ is an onto function from $\mathcal{P}(B)$ to $\mathcal{P}(A)$.

e) if f is a one-to-one correspondence, then S_f is a one-to-one correspondence from $\mathcal{P}(A)$ to $\mathcal{P}(B)$ and $S_{f^{-1}}$ is a one-to-one correspondence from $\mathcal{P}(B)$ to $\mathcal{P}(A)$. [*Hint:* Use parts (a)-(d).]

11. Prove that if f and g are functions from A to B and $S_f = S_g$ (using the definition in the preamble to Exercise 10), then $f(x) = g(x)$ for all $x \in A$.

12. Show that if n is an integer, then $n = \lceil n/2 \rceil + \lfloor n/2 \rfloor$.

13. For which real numbers x and y is it true that $\lfloor x + y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$?

14. For which real numbers x and y is it true that $\lceil x + y \rceil = \lceil x \rceil + \lceil y \rceil$?

15. For which real numbers x and y is it true that $\lceil x + y \rceil = \lceil x \rceil + \lfloor y \rfloor$?

16. Prove that $\lfloor n/2 \rfloor \lceil n/2 \rceil = \lfloor n^2/4 \rfloor$ for all integers n.

17. Prove that if m is an integer, then $\lfloor x \rfloor + \lfloor m - x \rfloor = m - 1$, unless x is an integer, in which case, it equals m.

18. Prove that if x is a real number, then $\lfloor \lfloor x/2 \rfloor /2 \rfloor = \lfloor x/4 \rfloor$.

19. Prove that if n is an odd integer, then $\lceil n^2/4 \rceil = (n^2 + 3)/4$.

***20.** We define the **Ulam numbers** by setting $u_1 = 1$ and $u_2 = 2$. Furthermore, after determining whether the integers less than n are Ulam numbers, we set n equal to the next Ulam number if it can be written uniquely as the sum of two different Ulam numbers. Note that $u_3 = 3$, $u_4 = 4$, $u_5 = 6$, and $u_6 = 8$.

a) Find the first 20 Ulam numbers.
b) Prove that there are infinitely many Ulam numbers.

21. Determine the value of $\prod_{k=1}^{100} \frac{k+1}{k}$. (The notation used here for products is defined in the preamble to Exercise 29 in Section 2.4.)

***22.** Determine a rule for generating the terms of the sequence that begins 1, 3, 4, 8, 15, 27, 50, 92, ..., and find the next four terms of the sequence.

***23.** Determine a rule for generating the terms of the sequence that begins 2, 3, 3, 5, 10, 13, 39, 43, 172, 177, 885, 891, ..., and find the next four terms of the sequence.

24. Show that the set of irrational numbers is an uncountable set.

25. Show that the set S is a countable set if there is a function f from S to the positive integers such that $f^{-1}(j)$ is countable whenever j is a positive integer.

26. Show that the set of all finite subsets of the set of positive integers is a countable set.

27. Find $\mathbf{A}^n$ if $\mathbf{A}$ is

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

28. Show that if $\mathbf{A} = c\mathbf{I}$, where c is a real number and $\mathbf{I}$ is the $n \times n$ identity matrix, then $\mathbf{AB} = \mathbf{BA}$ whenever $\mathbf{B}$ is an $n \times n$ matrix.

29. Show that if $\mathbf{A}$ is a 2×2 matrix such that $\mathbf{AB} = \mathbf{BA}$ whenever $\mathbf{B}$ is a 2×2 matrix, then $\mathbf{A} = c\mathbf{I}$, where c is a real number and $\mathbf{I}$ is the 2×2 identity matrix.

30. Show that if $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices and $\mathbf{AB}$ exists, then $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

31. Let $\mathbf{A}$ be an $n \times n$ matrix and let $\mathbf{0}$ be the $n \times n$ matrix all of whose entries are zero. Show that the following are true.

a) $\mathbf{A} \odot \mathbf{0} = \mathbf{0} \odot \mathbf{A} = \mathbf{0}$
b) $\mathbf{A} \vee \mathbf{0} = \mathbf{0} \vee \mathbf{A} = \mathbf{A}$
c) $\mathbf{A} \wedge \mathbf{0} = \mathbf{0} \wedge \mathbf{A} = \mathbf{0}$

Computer Projects

Write programs with the specified input and output.

1. Given subsets A and B of a set with n elements, use bit strings to find $\overline{A}$, $A \cup B$, $A \cap B$, $A - B$, and $A \oplus B$.

2. Given multisets A and B from the same universal set, find $A \cup B$, $A \cap B$, $A - B$, and $A + B$ (see preamble to Exercise 61 of Section 2.2).

3. Given fuzzy sets A and B, find $\overline{A}$, $A \cup B$, and $A \cap B$ (see preamble to Exercise 63 of Section 2.2).

4. Given a function f from $\{1, 2, \ldots, n\}$ to the set of integers, determine whether f is one-to-one.

5. Given a function f from $\{1, 2, \ldots, n\}$ to itself, determine whether f is onto.

6. Given a bijection f from the set $\{1, 2, \ldots, n\}$ to itself, find f^{-1}.

7. Given an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$, find $\mathbf{AB}$.

8. Given a square matrix $\mathbf{A}$ and a positive integer n, find $\mathbf{A}^n$.

9. Given a square matrix, determine whether it is symmetric.

10. Given two $m \times n$ Boolean matrices, find their meet and join.

11. Given an $m \times k$ Boolean matrix $\mathbf{A}$ and a $k \times n$ Boolean matrix $\mathbf{B}$, find the Boolean product of $\mathbf{A}$ and $\mathbf{B}$.

12. Given a square Boolean matrix $\mathbf{A}$ and a positive integer n, find $\mathbf{A}^{[n]}$.

Computations and Explorations

Use a computational program or programs you have written to do these exercises.

1. Given two finite sets, list all elements in the Cartesian product of these two sets.
2. Given a finite set, list all elements of its power set.
3. Calculate the number of one-to-one functions from a set S to a set T, where S and T are finite sets of various sizes. Can you determine a formula for the number of such functions? (We will find such a formula in Chapter 6.)
4. Calculate the number of onto functions from a set S to a set T, where S and T are finite sets of various sizes. Can you determine a formula for the number of such functions? (We will find such a formula in Chapter 8.)

Writing Projects

Respond to these with essays using outside sources.

1. Discuss how an axiomatic set theory can be developed to avoid Russell's paradox. (See Exercise 46 of Section 2.1.)
2. Research where the concept of a function first arose, and describe how this concept was first used.
3. Explain the different ways in which the *Encyclopedia of Integer Sequences* has been found useful. Also, describe a few of the more unusual sequences in this encyclopedia and how they arise.
4. Define the recently invented EKG sequence and describe some of its properties and open questions about it.
5. Look up the definition of a transcendental number. Explain how to show that such numbers exist and how such numbers can be constructed. Which famous numbers can be shown to be transcendental and for which famous numbers is it still unknown whether they are transcendental?