CHAPTER

9 Relations

Relationships between elements of sets occur in many contexts. Every day we deal with relationships such as those between a business and its telephone number, an employee and his or her salary, a person and a relative, and so on. In mathematics we study relationships such as those between a positive integer and one that it divides, an integer and one that it is congruent to modulo 5, a real number and one that is larger than it, a real number x and the value $f(x)$ where f is a function, and so on. Relationships such as that between a program and a variable it uses, and that between a computer language and a valid statement in this language, often arise in computer science. Relationships between elements of two sets are represented using the structure called a binary relation, which is just a subset of the Cartesian product of the sets. Relations can be used to solve problems such as determining which pairs of cities are linked by airline flights in a network, or finding a viable order for the different phases of a complicated project. We will introduce a number of different properties binary relations may enjoy.

Relationships between elements of more than two sets arise in many contexts. These relationships can be represented by n-ary relations, which are collections of n-tuples. Such relations are the basis of the relational data model, the most common way to store information in computer databases. We will introduce the terminology used to study relational databases, define some important operations on them, and introduce the database query language SQL. We will conclude our brief study of n-ary relations and databases with an important application from data mining. In particular, we will show how databases of transactions, represented by n-ary relations, are used to measure the likelihood that someone buys a particular product from a store when they buy one or more other products.

Two methods for representing relations, using square matrices and using directed graphs, consisting of vertices and directed edges, will be introduced and used in later sections of the chapter. We will also study relationships that have certain collections of properties that relations may enjoy. For example, in some computer languages, only the first 31 characters of the name of a variable matter. The relation consisting of ordered pairs of strings in which the first string has the same initial 31 characters as the second string is an example of a special type of relation, known as an equivalence relation. Equivalence relations arise throughout mathematics and computer science. Finally, we will study relations called partial orderings, which generalize the notion of the less than or equal to relation. For example, the set of all pairs of strings of English letters in which the second string is the same as the first string or comes after the first in dictionary order is a partial ordering.

9.1 Relations and Their Properties

9.1.1 Introduction

Links

The most direct way to express a relationship between elements of two sets is to use ordered pairs made up of two related elements. For this reason, sets of ordered pairs are called binary relations. In this section we introduce the basic terminology used to describe binary relations. Later in this chapter we will use relations to solve problems involving communications networks, project scheduling, and identifying elements in sets with common properties.

Definition 1 Let A and B be sets. A *binary relation from A to B* is a subset of $A \times B$.

In other words, a binary relation from A to B is a set R of ordered pairs, where the first element of each ordered pair comes from A and the second element comes from B. We use the notation $a \, R \, b$ to denote that $(a, b) \in R$ and $a \not R \, b$ to denote that $(a, b) \notin R$. Moreover, when (a, b) belongs to R, a is said to be **related to** b by R.

Binary relations represent relationships between the elements of two sets. We will introduce n-ary relations, which express relationships among elements of more than two sets, later in this chapter. We will omit the word *binary* when there is no danger of confusion.

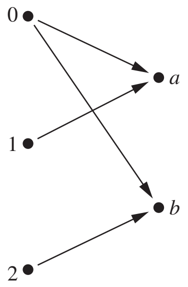
Examples 1–3 illustrate the notion of a relation.

EXAMPLE 1 Let A be the set of students in your school, and let B be the set of courses. Let R be the relation that consists of those pairs (a, b), where a is a student enrolled in course b. For instance, if Jason Goodfriend and Deborah Sherman are enrolled in CS518, the pairs (Jason Goodfriend, CS518) and (Deborah Sherman, CS518) belong to R. If Jason Goodfriend is also enrolled in CS510, then the pair (Jason Goodfriend, CS510) is also in R. However, if Deborah Sherman is not enrolled in CS510, then the pair (Deborah Sherman, CS510) is not in R.

Note that if a student is not currently enrolled in any courses there will be no pairs in R that have this student as the first element. Similarly, if a course is not currently being offered there will be no pairs in R that have this course as their second element. ◀

EXAMPLE 2 Let A be the set of cities in the U.S.A., and let B be the set of the 50 states in the U.S.A. Define the relation R by specifying that (a, b) belongs to R if a city with name a is in the state b. For instance, (Boulder, Colorado), (Bangor, Maine), (Ann Arbor, Michigan), (Middletown, New Jersey), (Middletown, New York), (Cupertino, California), and (Red Bank, New Jersey) are in R. ◀

EXAMPLE 3 Let $A = \{0, 1, 2\}$ and $B = \{a, b\}$. Then $\{(0, a), (0, b), (1, a), (2, b)\}$ is a relation from A to B. This means, for instance, that $0 \, R \, a$, but that $1 \not R \, b$. Relations can be represented graphically, as shown in Figure 1, using arrows to represent ordered pairs. Another way to represent this relation is to use a table, which is also done in Figure 1. We will discuss representations of relations in more detail in Section 9.3. ◀



R	a	b
0	×	×
1	×	
2		×

FIGURE 1 Displaying the ordered pairs in the relation R from Example 3.

9.1.2 Functions as Relations

Recall that a function f from a set A to a set B (as defined in Section 2.3) assigns exactly one element of B to each element of A. The graph of f is the set of ordered pairs (a, b) such that $b = f(a)$. Because the graph of f is a subset of $A \times B$, it is a relation from A to B. Moreover, the graph of a function has the property that every element of A is the first element of exactly one ordered pair of the graph.

Conversely, if R is a relation from A to B such that every element in A is the first element of exactly one ordered pair of R, then a function can be defined with R as its graph. This can be done by assigning to an element a of A the unique element $b \in B$ such that $(a, b) \in R$. (Note that the relation R in Example 2 is not the graph of a function because Middletown occurs more than once as the first element of an ordered pair in R.)

A relation can be used to express a one-to-many relationship between the elements of the sets A and B (as in Example 2), where an element of A may be related to more than one element of B. A function represents a relation where exactly one element of B is related to each element of A.

Relations are a generalization of graphs of functions; they can be used to express a much wider class of relationships between sets. (Recall that the graph of the function f from A to B is the set of ordered pairs $(a, f(a))$ for $a \in A$.)

9.1.3 Relations on a Set

Relations from a set A to itself are of special interest.

Definition 2 A *relation on a set A* is a relation from A to A.

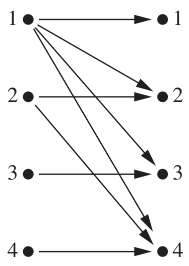
In other words, a relation on a set A is a subset of $A \times A$.

EXAMPLE 4 Let A be the set $\{1, 2, 3, 4\}$. Which ordered pairs are in the relation $R = \{(a, b) \mid a \text{ divides } b\}$?

Solution: Because (a, b) is in R if and only if a and b are positive integers not exceeding 4 such that a divides b, we see that

$$R = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 4), (3, 3), (4, 4)\}.$$

The pairs in this relation are displayed both graphically and in tabular form in Figure 2. ◀



R	1	2	3	4
1	×	×	×	×
2		×		×
3			×	
4				×

FIGURE 2 **Displaying the ordered pairs in the relation R from Example 4.**

Next, some examples of relations on the set of integers will be given in Example 5.

EXAMPLE 5 Consider these relations on the set of integers:

$$
\begin{aligned}
R_1 &= \{(a, b) \mid a \leq b\},\\
R_2 &= \{(a, b) \mid a > b\},\\
R_3 &= \{(a, b) \mid a = b \text{ or } a = -b\},\\
R_4 &= \{(a, b) \mid a = b\},\\
R_5 &= \{(a, b) \mid a = b + 1\},\\
R_6 &= \{(a, b) \mid a + b \leq 3\}.
\end{aligned}
$$

Which of these relations contain each of the pairs $(1, 1)$, $(1, 2)$, $(2, 1)$, $(1, -1)$, and $(2, 2)$?

Remark: Unlike the relations in Examples 1–4, these are relations on an infinite set.

Solution: The pair $(1, 1)$ is in R_1, R_3, R_4, and R_6; $(1, 2)$ is in R_1 and R_6; $(2, 1)$ is in R_2, R_5, and R_6; $(1, -1)$ is in R_2, R_3, and R_6; and finally, $(2, 2)$ is in R_1, R_3, and R_4. ◀

It is not hard to determine the number of relations on a finite set, because a relation on a set A is simply a subset of $A \times A$.

EXAMPLE 6 How many relations are there on a set with n elements?

Solution: A relation on a set A is a subset of $A \times A$. Because $A \times A$ has n^2 elements when A has n elements, and a set with m elements has 2^m subsets, there are 2^{n^2} subsets of $A \times A$. Thus, there are 2^{n^2} relations on a set with n elements. For example, there are $2^{3^2} = 2^9 = 512$ relations on the set $\{a, b, c\}$. ◀

9.1.4 Properties of Relations

There are several properties that are used to classify relations on a set. We will introduce the most important of these here. (You may find it instructive to study this material with the contents of Section 9.3. In that section, several methods for representing relations will be introduced that can help you understand each of the properties that we introduce here.)

In some relations an element is always related to itself. For instance, let R be the relation on the set of all people consisting of pairs (x, y) where x and y have the same mother and the same father. Then xRx for every person x.

Definition 3 A relation R on a set A is called *reflexive* if $(a, a) \in R$ for every element $a \in A$.

Remark: Using quantifiers we see that the relation R on the set A is reflexive if $\forall a((a, a) \in R)$, where the universe of discourse is the set of all elements in A.

We see that a relation on A is reflexive if every element of A is related to itself. Examples 7–9 illustrate the concept of a reflexive relation.

EXAMPLE 7 Consider the following relations on $\{1, 2, 3, 4\}$:

$$
\begin{aligned}
R_1 &= \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 4), (4, 1), (4, 4)\},\\
R_2 &= \{(1, 1), (1, 2), (2, 1)\},\\
R_3 &= \{(1, 1), (1, 2), (1, 4), (2, 1), (2, 2), (3, 3), (4, 1), (4, 4)\},\\
R_4 &= \{(2, 1), (3, 1), (3, 2), (4, 1), (4, 2), (4, 3)\},\\
R_5 &= \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 3), (3, 4), (4, 4)\},\\
R_6 &= \{(3, 4)\}.
\end{aligned}
$$

Which of these relations are reflexive?

Solution: The relations R_3 and R_5 are reflexive because they both contain all pairs of the form (a, a), namely, $(1, 1)$, $(2, 2)$, $(3, 3)$, and $(4, 4)$. The other relations are not reflexive because they do not contain all of these ordered pairs. In particular, R_1, R_2, R_4, and R_6 are not reflexive because $(3, 3)$ is not in any of these relations. ◀

EXAMPLE 8 Which of the relations from Example 5 are reflexive?

Solution: The reflexive relations from Example 5 are R_1 (because $a \leq a$ for every integer a), R_3, and R_4. For each of the other relations in this example it is easy to find a pair of the form (a, a) that is not in the relation. (This is left as an exercise for the reader.) ◀

EXAMPLE 9 Is the “divides” relation on the set of positive integers reflexive?

Solution: Because $a \mid a$ whenever a is a positive integer, the “divides” relation is reflexive. (Note that if we replace the set of positive integers with the set of all integers the relation is not reflexive because by definition 0 does not divide 0.) ◀

In some relations an element is related to a second element if and only if the second element is also related to the first element. The relation consisting of pairs (x, y), where x and y are students at your school with at least one common class has this property. Other relations have the property that if an element is related to a second element, then this second element is not related to the first. The relation consisting of the pairs (x, y), where x and y are students at your school, where x has a higher grade point average than y has this property.

Definition 4

A relation R on a set A is called *symmetric* if $(b, a) \in R$ whenever $(a, b) \in R$, for all $a, b \in A$. A relation R on a set A such that for all $a, b \in A$, if $(a, b) \in R$ and $(b, a) \in R$, then $a = b$ is called *antisymmetric*.

Remark: Using quantifiers, we see that the relation R on the set A is symmetric if $\forall a \forall b((a, b) \in R \rightarrow (b, a) \in R)$. Similarly, the relation R on the set A is antisymmetric if $\forall a \forall b(((a, b) \in R \wedge (b, a) \in R) \rightarrow (a = b))$.

In other words, a relation is symmetric if and only if a is related to b always implies that b is related to a. For instance, the equality relation is symmetric because $a = b$ if and only if $b = a$. A relation is antisymmetric if and only if there are no pairs of distinct elements a and b with a related to b and b related to a. That is, the only way to have a related to b and b related to a is for a and b to be the same element. For instance, the less than or equal to relation is antisymmetric. To see this, note that $a \leq b$ and $b \leq a$ implies that $a = b$. The terms *symmetric* and *antisymmetric* are not opposites, because a relation can have both of these properties or may lack both of them (see Exercise 10). A relation cannot be both symmetric and antisymmetric if it contains some pair of the form (a, b) in which $a \neq b$.

Remark: Although relatively few of the 2^{n^2} relations on a set with n elements are symmetric or antisymmetric, as counting arguments can show, many important relations have one of these properties. (See Exercise 49.)

EXAMPLE 10 Which of the relations from Example 7 are symmetric and which are antisymmetric?

Extra Examples

Solution: The relations R_2 and R_3 are symmetric, because in each case (b, a) belongs to the relation whenever (a, b) does. For R_2, the only thing to check is that both $(2, 1)$ and $(1, 2)$ are in the relation. For R_3, it is necessary to check that both $(1, 2)$ and $(2, 1)$ belong to the relation, and $(1, 4)$ and $(4, 1)$ belong to the relation. The reader should verify that none of the other relations is symmetric. This is done by finding a pair (a, b) such that it is in the relation but (b, a) is not.

R_4, R_5, and R_6 are all antisymmetric. For each of these relations there is no pair of elements a and b with $a \neq b$ such that both (a, b) and (b, a) belong to the relation. The reader should verify that none of the other relations is antisymmetric. This is done by finding a pair (a, b) with $a \neq b$ such that (a, b) and (b, a) are both in the relation. ◀

EXAMPLE 11 Which of the relations from Example 5 are symmetric and which are antisymmetric?

Solution: The relations R_3, R_4, and R_6 are symmetric. R_3 is symmetric, for if $a = b$ or $a = -b$, then $b = a$ or $b = -a$. R_4 is symmetric because $a = b$ implies that $b = a$. R_6 is symmetric because $a + b \leq 3$ implies that $b + a \leq 3$. The reader should verify that none of the other relations is symmetric.

The relations R_1, R_2, R_4, and R_5 are antisymmetric. R_1 is antisymmetric because the inequalities $a \leq b$ and $b \leq a$ imply that $a = b$. R_2 is antisymmetric because it is impossible that $a > b$ and $b > a$. R_4 is antisymmetric, because two elements are related with respect to R_4 if and only if they are equal. R_5 is antisymmetric because it is impossible that $a = b + 1$ and $b = a + 1$. The reader should verify that none of the other relations is antisymmetric. ◀

EXAMPLE 12 Is the "divides" relation on the set of positive integers symmetric? Is it antisymmetric?

Solution: This relation is not symmetric because $1 \mid 2$, but $2 \nmid 1$. However, it is antisymmetric. To see this, note that if a and b are positive integers with $a \mid b$ and $b \mid a$, then $a = b$ (the verification of this is left as an exercise for the reader). ◀

Let R be the relation consisting of all pairs (x, y) of students at your school, where x has taken more credits than y. Suppose that x is related to y and y is related to z. This means that x has taken more credits than y and y has taken more credits than z. We can conclude that x has taken more credits than z, so that x is related to z. What we have shown is that R has the transitive property, which is defined as follows.

Definition 5 A relation R on a set A is called *transitive* if whenever $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$, for all $a, b, c \in A$.

Remark: Using quantifiers we see that the relation R on a set A is transitive if we have $\forall a \forall b \forall c(((a, b) \in R \wedge (b, c) \in R) \rightarrow (a, c) \in R)$.

EXAMPLE 13 Which of the relations in Example 7 are transitive?

Extra Examples

Solution: R_4, R_5, and R_6 are transitive. For each of these relations, we can show that it is transitive by verifying that if (a, b) and (b, c) belong to this relation, then (a, c) also does. For instance, R_4 is transitive, because (3, 2) and (2, 1), (4, 2) and (2, 1), (4, 3) and (3, 1), and (4, 3) and (3, 2) are the only such sets of pairs, and (3, 1), (4, 1), and (4, 2) belong to R_4. The reader should verify that R_5 and R_6 are transitive.

R_1 is not transitive because (3, 4) and (4, 1) belong to R_1, but (3, 1) does not. R_2 is not transitive because (2, 1) and (1, 2) belong to R_2, but (2, 2) does not. R_3 is not transitive because (4, 1) and (1, 2) belong to R_3, but (4, 2) does not. ◀

EXAMPLE 14 Which of the relations in Example 5 are transitive?

Solution: The relations R_1, R_2, R_3, and R_4 are transitive. R_1 is transitive because $a \leq b$ and $b \leq c$ imply that $a \leq c$. R_2 is transitive because $a > b$ and $b > c$ imply that $a > c$. R_3 is transitive because $a = \pm b$ and $b = \pm c$ imply that $a = \pm c$. R_4 is clearly transitive, as the reader should verify. R_5 is not transitive because (2, 1) and (1, 0) belong to R_5, but (2, 0) does not. R_6 is not transitive because (2, 1) and (1, 2) belong to R_6, but (2, 2) does not. ◀

EXAMPLE 15 Is the "divides" relation on the set of positive integers transitive?

Solution: Suppose that a divides b and b divides c. Then there are positive integers k and l such that $b = ak$ and $c = bl$. Hence, $c = a(kl)$, so a divides c. It follows that this relation is transitive. ◀

We can use counting techniques to determine the number of relations with specific properties. Finding the number of relations with a particular property provides information about how common this property is in the set of all relations on a set with n elements.

EXAMPLE 16 How many reflexive relations are there on a set with n elements?

Solution: A relation R on a set A is a subset of $A \times A$. Consequently, a relation is determined by specifying whether each of the n^2 ordered pairs in $A \times A$ is in R. However, if R is reflexive, each of the n ordered pairs (a, a) for $a \in A$ must be in R. Each of the other $n(n-1)$ ordered pairs of the form (a, b), where $a \neq b$, may or may not be in R. Hence, by the product rule for counting, there are $2^{n(n-1)}$ reflexive relations [this is the number of ways to choose whether each element (a, b), with $a \neq b$, belongs to R]. ◀

Formulas for the number of symmetric relations and the number of antisymmetric relations on a set with n elements can be found using reasoning similar to that in Example 16 (see Exercise 49). However, no general formula is known that counts the transitive relations on a set with n elements. Currently, $T(n)$, the number of transitive relations on a set with n elements, is known only for $0 \leq n \leq 18$. For example, $T(4) = 3{,}994$, $T(5) = 154{,}303$, and $T(6) = 9{,}415{,}189$. (The values of $T(n)$ for $n = 0, 1, 2, \ldots, 18$, are the terms of the sequence A006905 in the OEIS, which is discussed in Section 2.4.)

9.1.5 Combining Relations

Because relations from A to B are subsets of $A \times B$, two relations from A to B can be combined in any way two sets can be combined. Consider Examples 17–19.

EXAMPLE 17 Let $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4\}$. The relations $R_1 = \{(1, 1), (2, 2), (3, 3)\}$ and $R_2 = \{(1, 1), (1, 2), (1, 3), (1, 4)\}$ can be combined to obtain

$$\begin{aligned}
R_1 \cup R_2 &= \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (3, 3)\},\\
R_1 \cap R_2 &= \{(1, 1)\},\\
R_1 - R_2 &= \{(2, 2), (3, 3)\},\\
R_2 - R_1 &= \{(1, 2), (1, 3), (1, 4)\}.
\end{aligned}$$

◀

EXAMPLE 18 Let A and B be the set of all students and the set of all courses at a school, respectively. Suppose that R_1 consists of all ordered pairs (a, b), where a is a student who has taken course b, and R_2 consists of all ordered pairs (a, b), where a is a student who requires course b to graduate. What are the relations $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 \oplus R_2$, $R_1 - R_2$, and $R_2 - R_1$?

Solution: The relation $R_1 \cup R_2$ consists of all ordered pairs (a, b), where a is a student who either has taken course b or needs course b to graduate, and $R_1 \cap R_2$ is the set of all ordered pairs (a, b), where a is a student who has taken course b and needs this course to graduate. Also, $R_1 \oplus R_2$ consists of all ordered pairs (a, b), where student a has taken course b but does not need it to graduate or needs course b to graduate but has not taken it. $R_1 - R_2$ is the set of ordered pairs (a, b), where a has taken course b but does not need it to graduate; that is, b is an elective course that a has taken. $R_2 - R_1$ is the set of all ordered pairs (a, b), where b is a course that a needs to graduate but has not taken. ◀

EXAMPLE 19 Let R_1 be the less than relation on the set of real numbers and let R_2 be the greater than relation on the set of real numbers, that is, $R_1 = \{(x, y) \mid x < y\}$ and $R_2 = \{(x, y) \mid x > y\}$. What are $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 - R_2$, $R_2 - R_1$, and $R_1 \oplus R_2$?

Solution: We note that $(x, y) \in R_1 \cup R_2$ if and only if $(x, y) \in R_1$ or $(x, y) \in R_2$. Hence, $(x, y) \in R_1 \cup R_2$ if and only if $x < y$ or $x > y$. Because the condition $x < y$ or $x > y$ is the same as the condition $x \neq y$, it follows that $R_1 \cup R_2 = \{(x, y) \mid x \neq y\}$. In other words, the union of the less than relation and the greater than relation is the not equals relation.

Next, note that it is impossible for a pair (x, y) to belong to both R_1 and R_2 because it is impossible that $x < y$ and $x > y$. It follows that $R_1 \cap R_2 = \emptyset$. We also see that $R_1 - R_2 = R_1$, $R_2 - R_1 = R_2$, and $R_1 \oplus R_2 = R_1 \cup R_2 - R_1 \cap R_2 = \{(x, y) \mid x \neq y\}$. ◀

There is another way that relations are combined that is analogous to the composition of functions.

Definition 6 Let R be a relation from a set A to a set B and S a relation from B to a set C. The *composite* of R and S is the relation consisting of ordered pairs (a, c), where $a \in A$, $c \in C$, and for which there exists an element $b \in B$ such that $(a, b) \in R$ and $(b, c) \in S$. We denote the composite of R and S by $S \circ R$.

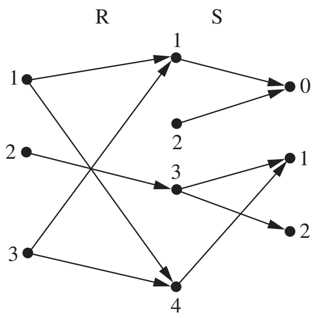
Computing the composite of two relations requires that we find elements that are the second element of ordered pairs in the first relation and the first element of ordered pairs in the second relation, as Examples 20 and 21 illustrate.

EXAMPLE 20 What is the composite of the relations R and S, where R is the relation from $\{1, 2, 3\}$ to $\{1, 2, 3, 4\}$ with $R = \{(1, 1), (1, 4), (2, 3), (3, 1), (3, 4)\}$ and S is the relation from $\{1, 2, 3, 4\}$ to $\{0, 1, 2\}$ with $S = \{(1, 0), (2, 0), (3, 1), (3, 2), (4, 1)\}$?

Solution: $S \circ R$ is constructed using all ordered pairs in R and ordered pairs in S, where the second element of the ordered pair in R agrees with the first element of the ordered pair in S. For example, the ordered pairs $(2, 3)$ in R and $(3, 1)$ in S produce the ordered pair $(2, 1)$ in $S \circ R$. Computing all the ordered pairs in the composite, we find

$$S \circ R = \{(1, 0), (1, 1), (2, 1), (2, 2), (3, 0), (3, 1)\}.$$

Figure 3 illustrates how this composition is found. In the figure, we examine all paths that travel via two directed edges from the leftmost elements to the rightmost elements via an element in the middle.



1 → 1 → 0	(1, 0)
1 → 4 → 1	(1, 1)
2 → 3 → 1	(2, 1)
2 → 3 → 2	(2, 2)
3 → 1 → 0	(3, 0)
3 → 4 → 1	(3, 1)

FIGURE 3 **Constructing $S \circ R$.** ◀

EXAMPLE 21 **Composing the Parent Relation with Itself** Let R be the relation on the set of all people such that $(a, b) \in R$ if person a is a parent of person b. Then $(a, c) \in R \circ R$ if and only if there is a person b such that $(a, b) \in R$ and $(b, c) \in R$, that is, if and only if there is a person b such that a is a parent of b and b is a parent of c. In other words, $(a, c) \in R \circ R$ if and only if a is a grandparent of c. ◀

The powers of a relation R can be recursively defined from the definition of a composite of two relations.

Definition 7 Let R be a relation on the set A. The powers R^n, $n = 1, 2, 3, \ldots$, are defined recursively by

$$R^1 = R \qquad \text{and} \qquad R^{n+1} = R^n \circ R.$$

The definition shows that $R^2 = R \circ R$, $R^3 = R^2 \circ R = (R \circ R) \circ R$, and so on.

EXAMPLE 22 Let $R = \{(1, 1), (2, 1), (3, 2), (4, 3)\}$. Find the powers R^n, $n = 2, 3, 4, \ldots$.

Solution: Because $R^2 = R \circ R$, we find that $R^2 = \{(1, 1), (2, 1), (3, 1), (4, 2)\}$. Furthermore, because $R^3 = R^2 \circ R$, $R^3 = \{(1, 1), (2, 1), (3, 1), (4, 1)\}$. Additional computation shows that R^4

is the same as R^3, so $R^4 = \{(1, 1), (2, 1), (3, 1), (4, 1)\}$. It also follows that $R^n = R^3$ for $n = 5, 6, 7, \ldots$. The reader should verify this. ◀

The following theorem shows that the powers of a transitive relation are subsets of this relation. It will be used in Section 9.4.

THEOREM 1 The relation R on a set A is transitive if and only if $R^n \subseteq R$ for $n = 1, 2, 3, \ldots$.

Proof: We first prove the "if" part of the theorem. We suppose that $R^n \subseteq R$ for $n = 1, 2, 3, \ldots$. In particular, $R^2 \subseteq R$. To see that this implies R is transitive, note that if $(a, b) \in R$ and $(b, c) \in R$, then by the definition of composition, $(a, c) \in R^2$. Because $R^2 \subseteq R$, this means that $(a, c) \in R$. Hence, R is transitive.

We will use mathematical induction to prove the only if part of the theorem. Note that this part of the theorem is trivially true for $n = 1$.

Assume that $R^n \subseteq R$, where n is a positive integer. This is the inductive hypothesis. To complete the inductive step we must show that this implies that R^{n+1} is also a subset of R. To show this, assume that $(a, b) \in R^{n+1}$. Then, because $R^{n+1} = R^n \circ R$, there is an element x with $x \in A$ such that $(a, x) \in R$ and $(x, b) \in R^n$. The inductive hypothesis, namely, that $R^n \subseteq R$, implies that $(x, b) \in R$. Furthermore, because R is transitive, and $(a, x) \in R$ and $(x, b) \in R$, it follows that $(a, b) \in R$. This shows that $R^{n+1} \subseteq R$, completing the proof. ◁

Exercises

1. List the ordered pairs in the relation R from $A = \{0, 1, 2, 3, 4\}$ to $B = \{0, 1, 2, 3\}$, where $(a, b) \in R$ if and only if

a) $a = b$. **b)** $a + b = 4$.
c) $a > b$. **d)** $a \mid b$.
e) $\gcd(a, b) = 1$. **f)** $\operatorname{lcm}(a, b) = 2$.

2. a) List all the ordered pairs in the relation $R = \{(a, b) \mid a \text{ divides } b\}$ on the set $\{1, 2, 3, 4, 5, 6\}$.

b) Display this relation graphically, as was done in Example 4.

c) Display this relation in tabular form, as was done in Example 4.

3. For each of these relations on the set $\{1, 2, 3, 4\}$, decide whether it is reflexive, whether it is symmetric, whether it is antisymmetric, and whether it is transitive.

a) $\{(2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4)\}$
b) $\{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3), (4, 4)\}$
c) $\{(2, 4), (4, 2)\}$
d) $\{(1, 2), (2, 3), (3, 4)\}$
e) $\{(1, 1), (2, 2), (3, 3), (4, 4)\}$
f) $\{(1, 3), (1, 4), (2, 3), (2, 4), (3, 1), (3, 4)\}$

4. Determine whether the relation R on the set of all people is reflexive, symmetric, antisymmetric, and/or transitive, where $(a, b) \in R$ if and only if

a) a is taller than b.
b) a and b were born on the same day.
c) a has the same first name as b.
d) a and b have a common grandparent.

5. Determine whether the relation R on the set of all Web pages is reflexive, symmetric, antisymmetric, and/or transitive, where $(a, b) \in R$ if and only if

a) everyone who has visited Web page a has also visited Web page b.
b) there are no common links found on both Web page a and Web page b.
c) there is at least one common link on Web page a and Web page b.
d) there is a Web page that includes links to both Web page a and Web page b.

6. Determine whether the relation R on the set of all real numbers is reflexive, symmetric, antisymmetric, and/or transitive, where $(x, y) \in R$ if and only if

a) $x + y = 0$. **b)** $x = \pm y$.
c) $x - y$ is a rational number.
d) $x = 2y$. **e)** $xy \geq 0$.
f) $xy = 0$. **g)** $x = 1$.
h) $x = 1$ or $y = 1$.

7. Determine whether the relation R on the set of all integers is reflexive, symmetric, antisymmetric, and/or transitive, where $(x, y) \in R$ if and only if

a) $x \neq y$. **b)** $xy \geq 1$.
c) $x = y + 1$ or $x = y - 1$.
d) $x \equiv y \pmod{7}$. **e)** x is a multiple of y.
f) x and y are both negative or both nonnegative.
g) $x = y^2$. **h)** $x \geq y^2$.

8. Show that the relation $R = \emptyset$ on a nonempty set S is symmetric and transitive, but not reflexive.

9. Show that the relation $R = \emptyset$ on the empty set $S = \emptyset$ is reflexive, symmetric, and transitive.

10. Give an example of a relation on a set that is

a) both symmetric and antisymmetric.

b) neither symmetric nor antisymmetric.

A relation R on the set A is **irreflexive** if for every $a \in A, (a, a) \notin R$. That is, R is irreflexive if no element in A is related to itself.

11. Which relations in Exercise 3 are irreflexive?

12. Which relations in Exercise 4 are irreflexive?

13. Which relations in Exercise 5 are irreflexive?

14. Which relations in Exercise 6 are irreflexive?

15. Can a relation on a set be neither reflexive nor irreflexive?

16. Use quantifiers to express what it means for a relation to be irreflexive.

17. Give an example of an irreflexive relation on the set of all people.

A relation R is called **asymmetric** if $(a, b) \in R$ implies that $(b, a) \notin R$. Exercises 18–24 explore the notion of an asymmetric relation. Exercise 22 focuses on the difference between asymmetry and antisymmetry.

18. Which relations in Exercise 3 are asymmetric?

19. Which relations in Exercise 4 are asymmetric?

20. Which relations in Exercise 5 are asymmetric?

21. Which relations in Exercise 6 are asymmetric?

22. Must an asymmetric relation also be antisymmetric? Must an antisymmetric relation be asymmetric? Give reasons for your answers.

23. Use quantifiers to express what it means for a relation to be asymmetric.

24. Give an example of an asymmetric relation on the set of all people.

25. How many different relations are there from a set with m elements to a set with n elements?

Let R be a relation from a set A to a set B. The **inverse relation** from B to A, denoted by R^{-1}, is the set of ordered pairs $\{(b, a) \mid (a, b) \in R\}$. The **complementary relation** $\overline{R}$ is the set of ordered pairs $\{(a, b) \mid (a, b) \notin R\}$.

26. Let R be the relation $R = \{(a, b) \mid a < b\}$ on the set of integers. Find

a) R^{-1}.

b) $\overline{R}$.

27. Let R be the relation $R = \{(a, b) \mid a$ divides $b\}$ on the set of positive integers. Find

a) R^{-1}.

b) $\overline{R}$.

28. Let R be the relation on the set of all states in the United States consisting of pairs (a, b) where state a borders state b. Find

a) R^{-1}.

b) $\overline{R}$.

29. Suppose that the function f from A to B is a one-to-one correspondence. Let R be the relation that equals the graph of f. That is, $R = \{(a, f(a)) \mid a \in A\}$. What is the inverse relation R^{-1}?

30. Let $R_1 = \{(1, 2), (2, 3), (3, 4)\}$ and $R_2 = \{(1, 1), (1, 2), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3), (3, 4)\}$ be relations from $\{1, 2, 3\}$ to $\{1, 2, 3, 4\}$. Find

a) $R_1 \cup R_2$.

b) $R_1 \cap R_2$.

c) $R_1 - R_2$.

d) $R_2 - R_1$.

31. Let A be the set of students at your school and B the set of books in the school library. Let R_1 and R_2 be the relations consisting of all ordered pairs (a, b), where student a is required to read book b in a course, and where student a has read book b, respectively. Describe the ordered pairs in each of these relations.

a) $R_1 \cup R_2$

b) $R_1 \cap R_2$

c) $R_1 \oplus R_2$

d) $R_1 - R_2$

e) $R_2 - R_1$

32. Let R be the relation $\{(1, 2), (1, 3), (2, 3), (2, 4), (3, 1)\}$, and let S be the relation $\{(2, 1), (3, 1), (3, 2), (4, 2)\}$. Find $S \circ R$.

33. Let R be the relation on the set of people consisting of pairs (a, b), where a is a parent of b. Let S be the relation on the set of people consisting of pairs (a, b), where a and b are siblings (brothers or sisters). What are $S \circ R$ and $R \circ S$?

Exercises 34–38 deal with these relations on the set of real numbers:

$R_1 = \{(a, b) \in \mathbf{R}^2 \mid a > b\}$, the greater than relation,
$R_2 = \{(a, b) \in \mathbf{R}^2 \mid a \geq b\}$, the greater than or equal to relation,
$R_3 = \{(a, b) \in \mathbf{R}^2 \mid a < b\}$, the less than relation,
$R_4 = \{(a, b) \in \mathbf{R}^2 \mid a \leq b\}$, the less than or equal to relation,
$R_5 = \{(a, b) \in \mathbf{R}^2 \mid a = b\}$, the equal to relation,
$R_6 = \{(a, b) \in \mathbf{R}^2 \mid a \neq b\}$, the unequal to relation.

34. Find

a) $R_1 \cup R_3$.

b) $R_1 \cup R_5$.

c) $R_2 \cap R_4$.

d) $R_3 \cap R_5$.

e) $R_1 - R_2$.

f) $R_2 - R_1$.

g) $R_1 \oplus R_3$.

h) $R_2 \oplus R_4$.

35. Find

a) $R_2 \cup R_4$.

b) $R_3 \cup R_6$.

c) $R_3 \cap R_6$.

d) $R_4 \cap R_6$.

e) $R_3 - R_6$.

f) $R_6 - R_3$.

g) $R_2 \oplus R_6$.

h) $R_3 \oplus R_5$.

36. Find

a) $R_1 \circ R_1$.

b) $R_1 \circ R_2$.

c) $R_1 \circ R_3$.

d) $R_1 \circ R_4$.

e) $R_1 \circ R_5$.

f) $R_1 \circ R_6$.

g) $R_2 \circ R_3$.

h) $R_3 \circ R_3$.

37. Find

a) $R_2 \circ R_1$.

b) $R_2 \circ R_2$.

c) $R_3 \circ R_5$.

d) $R_4 \circ R_1$.

e) $R_5 \circ R_3$.

f) $R_3 \circ R_6$.

g) $R_4 \circ R_6$.

h) $R_6 \circ R_6$.

38. Find the relations R_i^2 for $i = 1, 2, 3, 4, 5, 6$.

39. Find the relations S_i^2 for $i = 1, 2, 3, 4, 5, 6$ where

$S_1 = \{(a, b) \in \mathbf{Z}^2 \mid a > b\}$, the greater than relation,
$S_2 = \{(a, b) \in \mathbf{Z}^2 \mid a \geq b\}$, the greater than or equal to relation,
$S_3 = \{(a, b) \in \mathbf{Z}^2 \mid a < b\}$, the less than relation,
$S_4 = \{(a, b) \in \mathbf{Z}^2 \mid a \leq b\}$, the less than or equal to relation,
$S_5 = \{(a, b) \in \mathbf{Z}^2 \mid a = b\}$, the equal to relation,
$S_6 = \{(a, b) \in \mathbf{Z}^2 \mid a \neq b\}$, the unequal to relation.

40. Let R be the parent relation on the set of all people (see Example 21). When is an ordered pair in the relation R^3?

41. Let R be the relation on the set of people with doctorates such that $(a, b) \in R$ if and only if a was the thesis advisor of b. When is an ordered pair (a, b) in R^2? When is an ordered pair (a, b) in R^n, when n is a positive integer? (Assume that every person with a doctorate has a thesis advisor.)

42. Let R_1 and R_2 be the "divides" and "is a multiple of" relations on the set of all positive integers, respectively. That is, $R_1 = \{(a, b) \mid a \text{ divides } b\}$ and $R_2 = \{(a, b) \mid a \text{ is a multiple of } b\}$. Find

a) $R_1 \cup R_2$.
b) $R_1 \cap R_2$.
c) $R_1 - R_2$.
d) $R_2 - R_1$.
e) $R_1 \oplus R_2$.

43. Let R_1 and R_2 be the "congruent modulo 3" and the "congruent modulo 4" relations, respectively, on the set of integers. That is, $R_1 = \{(a, b) \mid a \equiv b \pmod 3\}$ and $R_2 = \{(a, b) \mid a \equiv b \pmod 4\}$. Find

a) $R_1 \cup R_2$.
b) $R_1 \cap R_2$.
c) $R_1 - R_2$.
d) $R_2 - R_1$.
e) $R_1 \oplus R_2$.

44. List the 16 different relations on the set $\{0, 1\}$.

45. How many of the 16 different relations on $\{0, 1\}$ contain the pair $(0, 1)$?

46. Which of the 16 relations on $\{0, 1\}$, which you listed in Exercise 44, are

a) reflexive?
b) irreflexive?
c) symmetric?
d) antisymmetric?
e) asymmetric?
f) transitive?

47. a) How many relations are there on the set $\{a, b, c, d\}$?
b) How many relations are there on the set $\{a, b, c, d\}$ that contain the pair (a, a)?

48. Let S be a set with n elements and let a and b be distinct elements of S. How many relations R are there on S such that

a) $(a, b) \in R$?
b) $(a, b) \notin R$?
c) no ordered pair in R has a as its first element?
d) at least one ordered pair in R has a as its first element?
e) no ordered pair in R has a as its first element or b as its second element?
f) at least one ordered pair in R either has a as its first element or has b as its second element?

***49.** How many relations are there on a set with n elements that are

a) symmetric?
b) antisymmetric?
c) asymmetric?
d) irreflexive?
e) reflexive and symmetric?
f) neither reflexive nor irreflexive?

***50.** How many transitive relations are there on a set with n elements if

a) $n = 1$?
b) $n = 2$?
c) $n = 3$?

51. Find the error in the "proof" of the following "theorem."

"*Theorem*": Let R be a relation on a set A that is symmetric and transitive. Then R is reflexive.

"*Proof*": Let $a \in A$. Take an element $b \in A$ such that $(a, b) \in R$. Because R is symmetric, we also have $(b, a) \in R$. Now using the transitive property, we can conclude that $(a, a) \in R$ because $(a, b) \in R$ and $(b, a) \in R$.

52. Suppose that R and S are reflexive relations on a set A. Prove or disprove each of these statements.

a) $R \cup S$ is reflexive.
b) $R \cap S$ is reflexive.
c) $R \oplus S$ is irreflexive.
d) $R - S$ is irreflexive.
e) $S \circ R$ is reflexive.

53. Show that the relation R on a set A is symmetric if and only if $R = R^{-1}$, where R^{-1} is the inverse relation.

54. Show that the relation R on a set A is antisymmetric if and only if $R \cap R^{-1}$ is a subset of the diagonal relation $\Delta = \{(a, a) \mid a \in A\}$.

55. Show that the relation R on a set A is reflexive if and only if the inverse relation R^{-1} is reflexive.

56. Show that the relation R on a set A is reflexive if and only if the complementary relation $\overline{R}$ is irreflexive.

57. Let R be a relation that is reflexive and transitive. Prove that $R^n = R$ for all positive integers n.

58. Let R be the relation on the set $\{1, 2, 3, 4, 5\}$ containing the ordered pairs $(1, 1)$, $(1, 2)$, $(1, 3)$, $(2, 3)$, $(2, 4)$, $(3, 1)$, $(3, 4)$, $(3, 5)$, $(4, 2)$, $(4, 5)$, $(5, 1)$, $(5, 2)$, and $(5, 4)$. Find

a) R^2.
b) R^3.
c) R^4.
d) R^5.

59. Let R be a reflexive relation on a set A. Show that R^n is reflexive for all positive integers n.

***60.** Let R be a symmetric relation. Show that R^n is symmetric for all positive integers n.

61. Suppose that the relation R is irreflexive. Is R^2 necessarily irreflexive? Give a reason for your answer.

62. Derive a big-O estimate for the number of integer comparisons needed to count all transitive relations on a set with n elements using the brute force approach of checking every relation of this set for transitivity.

9.2 *n*-ary Relations and Their Applications

9.2.1 Introduction

Relationships among elements of more than two sets often arise. For instance, there is a relationship involving the name of a student, the student's major, and the student's grade point average. Similarly, there is a relationship involving the airline, flight number, starting point, destination, departure time, and arrival time of a flight. An example of such a relationship in mathematics involves three integers, where the first integer is larger than the second integer, which is larger than the third. Another example is the betweenness relationship involving points on a line, such that three points are related when the second point is between the first and the third.

We will study relationships among elements from more than two sets in this section. These relationships are called ***n*-ary relations**. These relations are used to represent computer databases. These representations help us answer queries about the information stored in databases, such as: Which flights land at O'Hare Airport between 3 A.M. and 4 A.M.? Which students at your school are sophomores majoring in mathematics or computer science and have greater than a 3.0 average? Which employees of a company have worked for the company less than 5 years and make more than $50,000?

9.2.2 *n*-ary Relations

We begin with the basic definition on which the theory of relational databases rests.

Definition 1 Let $A_1, A_2, \dots, A_n$ be sets. An *n-ary relation* on these sets is a subset of $A_1 \times A_2 \times \cdots \times A_n$. The sets $A_1, A_2, \dots, A_n$ are called the *domains* of the relation, and n is called its *degree*.

EXAMPLE 1 Let R be the relation on $\mathbf{N} \times \mathbf{N} \times \mathbf{N}$ consisting of triples (a, b, c), where a, b, and c are integers with $a < b < c$. Then $(1, 2, 3) \in R$, but $(2, 4, 3) \notin R$. The degree of this relation is 3. Its domains are all equal to the set of natural numbers. ◀

EXAMPLE 2 Let R be the relation on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$ consisting of all triples of integers (a, b, c) in which a, b, and c form an arithmetic progression. That is, $(a, b, c) \in R$ if and only if there is an integer k such that $b = a + k$ and $c = a + 2k$, or equivalently, such that $b - a = k$ and $c - b = k$. Note that $(1, 3, 5) \in R$ because $3 = 1 + 2$ and $5 = 1 + 2 \cdot 2$, but $(2, 5, 9) \notin R$ because $5 - 2 = 3$ while $9 - 5 = 4$. This relation has degree 3 and its domains are all equal to the set of integers. ◀

EXAMPLE 3 Let R be the relation on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}^+$ consisting of triples (a, b, m), where a, b, and m are integers with $m \geq 1$ and $a \equiv b \pmod{m}$. Then $(8, 2, 3)$, $(-1, 9, 5)$, and $(14, 0, 7)$ all belong to R, but $(7, 2, 3)$, $(-2, -8, 5)$, and $(11, 0, 6)$ do not belong to R because $8 \equiv 2 \pmod{3}$, $-1 \equiv 9 \pmod{5}$, and $14 \equiv 0 \pmod{7}$, but $7 \not\equiv 2 \pmod{3}$, $-2 \not\equiv -8 \pmod{5}$, and $11 \not\equiv 0 \pmod{6}$. This relation has degree 3 and its first two domains are the set of all integers and its third domain is the set of positive integers. ◀

EXAMPLE 4 Let R be the relation consisting of 5-tuples (A, N, S, D, T) representing airplane flights, where A is the airline, N is the flight number, S is the starting point, D is the destination, and T is the departure time. For instance, if Nadir Express Airlines has flight 963 from Newark to Bangor at 15:00, then (Nadir, 963, Newark, Bangor, 15:00) belongs to R. The degree of this relation is 5, and its domains are the set of all airlines, the set of flight numbers, the set of cities, the set of cities (again), and the set of times. ◀

9.2.3 Databases and Relations

Links

The time required to manipulate information in a database depends on how this information is stored. The operations of adding and deleting records, updating records, searching for records, and combining records from overlapping databases are performed millions of times each day in a large database. Because of the importance of these operations, various methods for representing databases have been developed. We will discuss one of these methods, called the **relational data model**, based on the concept of a relation.

A database consists of **records**, which are n-tuples, made up of **fields**. The fields are the entries of the n-tuples. For instance, a database of student records may be made up of fields containing the name, student number, major, and grade point average of the student. The relational data model represents a database of records as an n-ary relation. Thus, student records are represented as 4-tuples of the form (*Student_name, ID_number, Major, GPA*). A sample database of six such records is

(Ackermann, 231455, Computer Science, 3.88)
(Adams, 888323, Physics, 3.45)
(Chou, 102147, Computer Science, 3.49)
(Goodfriend, 453876, Mathematics, 3.45)
(Rao, 678543, Mathematics, 3.90)
(Stevens, 786576, Psychology, 2.99).

Relations used to represent databases are also called **tables**, because these relations are often displayed as tables. Each column of the table corresponds to an *attribute* of the database. For instance, the same database of students is displayed in Table 1. The attributes of this database are Student Name, ID Number, Major, and GPA.

A domain of an n-ary relation is called a **primary key** when the value of the n-tuple from this domain determines the n-tuple. That is, a domain is a primary key when no two n-tuples in the relation have the same value from this domain.

Records are often added to or deleted from databases. Because of this, the property that a domain is a primary key is time-dependent. Consequently, a primary key should be chosen that remains one whenever the database is changed. The current collection of n-tuples in a relation is called the **extension** of the relation. The more permanent part of a database, including the name and attributes of the database, is called its **intension**. When selecting a primary key, the goal should be to select a key that can serve as a primary key for all possible extensions of the database. To do this, it is necessary to examine the intension of the database to understand the set of possible n-tuples that can occur in an extension.

TABLE 1 **Students.**

Student_name	***ID_number***	***Major***	***GPA***
Ackermann	231455	Computer Science	3.88
Adams	888323	Physics	3.45
Chou	102147	Computer Science	3.49
Goodfriend	453876	Mathematics	3.45
Rao	678543	Mathematics	3.90
Stevens	786576	Psychology	2.99

EXAMPLE 5 Which domains are primary keys for the n-ary relation displayed in Table 1, assuming that no n-tuples will be added in the future?

Extra Examples ❯

Solution: Because there is only one 4-tuple in this table for each student name, the domain of student names is a primary key. Similarly, the ID numbers in this table are unique, so the domain of ID numbers is also a primary key. However, the domain of major fields of study is not a primary key, because more than one 4-tuple contains the same major field of study. The domain of grade point averages is also not a primary key, because there are two 4-tuples containing the same GPA. ◀

Combinations of domains can also uniquely identify n-tuples in an n-ary relation. When the values of a set of domains determine an n-tuple in a relation, the Cartesian product of these domains is called a **composite key**.

EXAMPLE 6 Is the Cartesian product of the domain of major fields of study and the domain of GPAs a composite key for the n-ary relation from Table 1, assuming that no n-tuples are ever added?

Solution: Because no two 4-tuples from this table have both the same major and the same GPA, this Cartesian product is a composite key. ◀

Because primary and composite keys are used to identify records uniquely in a database, it is important that keys remain valid when new records are added to the database. Hence, checks should be made to ensure that every new record has values that are different in the appropriate field, or fields, from all other records in this table. For instance, it makes sense to use the student identification number as a key for student records if no two students ever have the same student identification number. A university should not use the name field as a key, because two students may have the same name (such as John Smith).

9.2.4 Operations on *n*-ary Relations

There are a variety of operations on n-ary relations that can be used to form new n-ary relations. Applied together, these operations can answer queries on databases that ask for all n-tuples that satisfy certain conditions.

The most basic operation on an n-ary relation is determining all n-tuples in the n-ary relation that satisfy certain conditions. For example, we may want to find all the records of all computer science majors in a database of student records. We may want to find all students who have a grade point average above 3.5. We may want to find the records of all computer science majors who have a grade point average above 3.5. To perform such tasks we use the selection operator.

Definition 2 Let R be an n-ary relation and C a condition that elements in R may satisfy. Then the *selection operator* s_C maps the n-ary relation R to the n-ary relation of all n-tuples from R that satisfy the condition C.

EXAMPLE 7 To find the records of computer science majors in the n-ary relation R shown in Table 1, we use the operator s_{C_1}, where C_1 is the condition Major = “Computer Science.” The result is the two 4-tuples (Ackermann, 231455, Computer Science, 3.88) and (Chou, 102147, Computer Science, 3.49). Similarly, to find the records of students who have a grade point average above 3.5 in this database, we use the operator s_{C_2}, where C_2 is the condition GPA > 3.5. The result is the two 4-tuples (Ackermann, 231455, Computer Science, 3.88) and (Rao, 678543, Mathematics,

Extra Examples ❯

3.90). Finally, to find the records of computer science majors who have a GPA above 3.5, we use the operator s_{C_3}, where C_3 is the condition (Major = "Computer Science" $\wedge$ GPA > 3.5). The result consists of the single 4-tuple (Ackermann, 231455, Computer Science, 3.88). ◀

Projections are used to form new n-ary relations by deleting the same fields in every record of the relation.

Definition 3 The *projection* $P_{i_1 i_2,\ldots,i_m}$ where $i_1 < i_2 < \cdots < i_m$, maps the n-tuple $(a_1, a_2, \ldots, a_n)$ to the m-tuple $(a_{i_1}, a_{i_2}, \ldots, a_{i_m})$, where $m \leq n$.

In other words, the projection $P_{i_1,i_2,\ldots,i_m}$ deletes $n - m$ of the components of an n-tuple, leaving the i_1th, i_2th, …, and i_mth components.

EXAMPLE 8 What results when the projection $P_{1,3}$ is applied to the 4-tuples (2, 3, 0, 4), (Jane Doe, 234111001, Geography, 3.14), and (a_1, a_2, a_3, a_4)?

Solution: The projection $P_{1,3}$ sends these 4-tuples to (2, 0), (Jane Doe, Geography), and (a_1, a_3), respectively. ◀

Example 9 illustrates how new relations are produced using projections.

EXAMPLE 9 What relation results when the projection $P_{1,4}$ is applied to the relation in Table 1?

Solution: When the projection $P_{1,4}$ is used, the second and third columns of the table are deleted, and pairs representing student names and grade point averages are obtained. Table 2 displays the results of this projection. ◀

Fewer rows may result when a projection is applied to the table for a relation. This happens when some of the n-tuples in the relation have identical values in each of the m components of the projection, and only disagree in components deleted by the projection. For instance, consider the following example.

EXAMPLE 10 What is the table obtained when the projection $P_{1,2}$ is applied to the relation in Table 3?

Solution: Table 4 displays the relation obtained when $P_{1,2}$ is applied to Table 3. Note that there are fewer rows after this projection is applied. ◀

TABLE 2 **GPAs.**

Student_name	*GPA*
Ackermann	3.88
Adams	3.45
Chou	3.49
Goodfriend	3.45
Rao	3.90
Stevens	2.99

TABLE 3 **Enrollments.**

Student	*Major*	*Course*
Glauser	Biology	BI 290
Glauser	Biology	MS 475
Glauser	Biology	PY 410
Marcus	Mathematics	MS 511
Marcus	Mathematics	MS 603
Marcus	Mathematics	CS 322
Miller	Computer Science	MS 575
Miller	Computer Science	CS 455

TABLE 4 **Majors.**

Student	*Major*
Glauser	Biology
Marcus	Mathematics
Miller	Computer Science

TABLE 5 **Teaching_assignments.**

Professor	*Department*	*Course_ number*
Cruz	Zoology	335
Cruz	Zoology	412
Farber	Psychology	501
Farber	Psychology	617
Grammer	Physics	544
Grammer	Physics	551
Rosen	Computer Science	518
Rosen	Mathematics	575

TABLE 6 **Class_schedule.**

Department	*Course_ number*	*Room*	*Time*
Computer Science	518	N521	2:00 P.M.
Mathematics	575	N502	3:00 P.M.
Mathematics	611	N521	4:00 P.M.
Physics	544	B505	4:00 P.M.
Psychology	501	A100	3:00 P.M.
Psychology	617	A110	11:00 A.M.
Zoology	335	A100	9:00 A.M.
Zoology	412	A100	8:00 A.M.

The **join** operation is used to combine two tables into one when these tables share some identical fields. For instance, a table containing fields for airline, flight number, and gate, and another table containing fields for flight number, gate, and departure time can be combined into a table containing fields for airline, flight number, gate, and departure time.

Definition 4 Let R be a relation of degree m and S a relation of degree n. The *join* $J_p(R, S)$, where $p \leq m$ and $p \leq n$, is a relation of degree $m + n - p$ that consists of all $(m + n - p)$-tuples $(a_1, a_2, \ldots, a_{m-p}, c_1, c_2, \ldots, c_p, b_1, b_2, \ldots, b_{n-p})$, where the m-tuple $(a_1, a_2, \ldots, a_{m-p}, c_1, c_2, \ldots, c_p)$ belongs to R and the n-tuple $(c_1, c_2, \ldots, c_p, b_1, b_2, \ldots, b_{n-p})$ belongs to S.

In other words, the join operator J_p produces a new relation from two relations by combining all m-tuples of the first relation with all n-tuples of the second relation, where the last p components of the m-tuples agree with the first p components of the n-tuples.

EXAMPLE 11 What relation results when the join operator J_2 is used to combine the relation displayed in Tables 5 and 6?

Solution: The join J_2 produces the relation shown in Table 7. ◀

There are other operators besides projections and joins that produce new relations from existing relations. A description of these operations can be found in books on database theory.

TABLE 7 **Teaching_schedule.**

Professor	*Department*	*Course_number*	*Room*	*Time*
Cruz	Zoology	335	A100	9:00 A.M.
Cruz	Zoology	412	A100	8:00 A.M.
Farber	Psychology	501	A100	3:00 P.M.
Farber	Psychology	617	A110	11:00 A.M.
Grammer	Physics	544	B505	4:00 P.M.
Rosen	Computer Science	518	N521	2:00 P.M.
Rosen	Mathematics	575	N502	3:00 P.M.

TABLE 8 **Flights.**

Airline	*Flight_number*	*Gate*	*Destination*	*Departure_time*
Nadir	122	34	Detroit	08:10
Acme	221	22	Denver	08:17
Acme	122	33	Anchorage	08:22
Acme	323	34	Honolulu	08:30
Nadir	199	13	Detroit	08:47
Acme	222	22	Denver	09:10
Nadir	322	34	Detroit	09:44

9.2.5 SQL

Links

The database query language SQL (short for Structured Query Language) can be used to carry out the operations we have described in this section. Example 12 illustrates how SQL commands are related to operations on *n*-ary relations.

EXAMPLE 12 We will illustrate how SQL is used to express queries by showing how SQL can be employed to make a query about airline flights using Table 8. The SQL statement

```
SELECT Departure_time
FROM Flights
WHERE Destination='Detroit'
```

is used to find the projection P_5 (on the Departure_time attribute) of the selection of 5-tuples in the Flights database that satisfy the condition: Destination = 'Detroit'. The output would be a list containing the times of flights that have Detroit as their destination, namely, 08:10, 08:47, and 09:44. SQL uses the FROM clause to identify the *n*-ary relation the query is applied to, the WHERE clause to specify the condition of the selection operation, and the SELECT clause to specify the projection operation that is to be applied. (*Beware:* SQL uses SELECT to represent a projection, rather than a selection operation. This is an unfortunate example of conflicting terminology.) ◀

Example 13 shows how SQL queries can be made involving more than one table.

EXAMPLE 13 The SQL statement

```
SELECT Professor, Time
FROM Teaching_assignments, Class_schedule
WHERE Department='Mathematics'
```

is used to find the projection $P_{1,5}$ of the 5-tuples in the database (shown in Table 7), which is the join J_2 of the Teaching_assignments and Class_schedule databases in Tables 5 and 6, respectively, which satisfy the condition: Department = Mathematics. The output would consist of the single 2-tuple (Rosen, 3:00 P.M.). The SQL FROM clause is used here to find the join of two different databases. ◀

We have only touched on the basic concepts of relational databases in this section. More information can be found in [AhUl95].

9.2.6 Association Rules from Data Mining

Links

We will now introduce concepts from **data mining**, the discipline with the goal of gaining useful information from data. In particular, we will discuss information that can be gleaned from databases of transactions. We will focus on supermarket transactions, but the ideas we present are relevant in a wide range of applications.

By a **transaction** we mean a set of items bought by a customer during a visit to the store, such as {milk, eggs, bread} or {orange juice, bananas, yogurt, cream}. Stores collect large databases of transactions that can be used to help them manage their businesses. We will discuss how these databases can be used to address the question: How likely is it that a customer buys a product given that they also buy a collection of one or more specified products?

We call each product in the store an **item**. A collection of items is known as an **itemset**. A ***k*-itemset** is an itemset that contains exactly k items. The terms **transaction** and **basket** are used synonymously with the word itemset. When a store has n items, $a_1, a_2, \ldots, a_n$, for sale, each transaction can be represented by an n-tuple $b_1, b_2, \ldots, b_n$, where b_i is a binary variable that tells us whether a_i occurs in this transaction. That is, $b_i = 1$ if a_i is in this transaction and $b_i = 0$ otherwise. (Note that we only care whether an item occurs in a transaction and not how many times it occurs.) We can represent a transaction by an $(n + 1)$-tuple of the form (*transaction number*, $b_1, b_2, \ldots, b_n$). The collection of all these $(n + 1)$-tuples forms a database of transactions.

We now define additional terms used in the study of questions relating to the purchase of particular itemsets.

Definition 5

The *count* of an itemset I, denoted by $\sigma(I)$, in a set of transactions $T = \{t_1, t_2, \ldots, t_k\}$, where k is a positive integer, is the number of transactions that contain this itemset. That is,

$$\sigma(I) = |\{t_i \in T \mid I \subseteq t_i\}|.$$

The *support* of an itemset I is the probability that I is included in a randomly selected transaction from T. That is,

$$\text{support}(I) = \frac{\sigma(I)}{|T|}.$$

The **support threshold** s is specified for a particular application. **Frequent itemset mining** is the process of finding itemsets I with support greater than or equal to s. Such itemsets are said to be **frequent**.

EXAMPLE 14

Extra Examples

The morning transactions at a market stand that sells apples, pears, cider, donuts, and mangos are shown in Table 9. In Table 10 we display the corresponding binary database, where each record is an $(n + 1)$-tuple consisting of the transaction number followed by binary entries that represent this itemset. Because apples occurs in five of the eight transactions, we see that $\sigma(\{\text{apples}\}) = 5$ and $\text{support}(\{\text{apples}\}) = 5/8$. Similarly, because the itemset {apples, cider} is a subset of four of the eight transactions, we have $\sigma(\{\text{apples, cider}\}) = 4$ and $\text{support}(\{\text{cider}\}) = 4/8 = 1/2$.

If we set the support threshold to be 0.5, an item is frequent if it occurs in at least four of the eight transactions. Consequently, with this support threshold, apples, pears, donuts and cider are the frequent items. The itemset {apples, cider} is a frequent itemset, but the itemset {donuts pears} is not a frequent itemset. ◀

TABLE 9 **A Set of Transactions.**

Transaction Number	*Items*
1	{apples, pears, mangos}
2	{pears, cider}
3	{apples, cider, donuts, mangos}
4	{apples, pears, cider, donuts}
5	{apples, cider, donuts}
6	{pears, cider, donuts}
7	{pears, donuts}
8	{apples, pears, cider}

TABLE 10 **Binary Database for the Transactions in Table 9.**

Transaction Number	*Apples*	*Pears*	*Cider*	*Donuts*	*Mangos*
1	1	1	0	0	1
2	0	1	1	0	0
3	1	0	1	1	1
4	1	1	1	1	0
5	1	0	1	1	0
6	0	1	1	1	0
7	0	1	0	1	0
8	1	1	1	0	0

We can use sets of transactions to help us predict whether a customer will buy a particular item given that they also buy all the items in an itemset (which might just be one item). Before we address a question of this type, we introduce some terminology. If S is a set of items and T is a set of transactions, an **association rule** is an implication of the form $I \to J$, where I and J are disjoint subsets of S. Although this notation borrows the notation for logical implications, its meaning is not entirely analagous. The association rule $I \to J$ is not the statement that whenever I is a subset of a transaction, then J must also be one. Instead, the strength of an association rule is measured in terms of its **support**, the frequency of transactions that contain both I and J, and its **confidence**, the frequency of transactions that contain J when they also contain I.

Definition 6 If I and J are subsets of a set T of transactions, then

$$\text{support}(I \to J) = \frac{\sigma(I \cup J)}{|T|}$$

and

$$\text{confidence}(I \to J) = \frac{\sigma(I \cup J)}{\sigma(I)}.$$

The support of the association rule $I \to J$, the fraction of transactions that contain both I and J, is a useful measure because a low support value tells us that the basket containing all items in

I and all items in J is seldom purchased, whereas a high value tells us that they are purchased together in a large fraction of transactions. Note that the confidence of the association rule $I \to J$ is the conditional probability that a transaction will contain all the items in I and in J given that it contains all the items in I. So, the larger the confidence of $I \to J$, the more likely it is for J to be a subset of a transaction that contains I.

EXAMPLE 15 What are the support and the confidence of the association rule {cider, donuts} → {apples} for the set of transactions in Example 14?

Extra Examples

Solution: The support of this association rule is $\sigma(\{\text{cider, donuts}\} \cup \{\text{apples}\})/|T|$. Because $\sigma(\{\text{cider, donuts}\} \cup \{\text{apples}\}) = \sigma(\{\text{cider, donuts, apples}\}) = 3$ and $|T| = 8$, we see that the support of this rule is $3/8 = 0.375$.

The confidence of this rule is $\sigma(\{\text{cider, donuts}\} \cup \{\text{apples}\})/\sigma(\{\text{cider, donuts}\}) = 3/4 = 0.75$. ◀

An important problem in data mining is to find **strong association rules**, which have support greater than or equal to a minimum support level and confidence greater than or equal to a minimum confidence level. It is important to have efficient algorithms to find strong association rules because the number of available items can be extremely large. For instance, a supermarket may have tens of thousands, or even hundreds of thousands, of items in stock. The brute-force approach of finding association rules with sufficiently large support and confidence by computing the support and confidence of all possible association rules is infeasible because there are an exponential number of such association rules (see Exercise 41). Several widely used algorithms have been developed to solve this problem much more efficiently than brute force. Such algorithms first find frequent itemsets and then turn their attention to finding all the association rules with high confidence from the frequent itemsets that have been found. Consult data mining texts such as [Ag15] for details.

Although we have presented association rules in the context of market baskets, they are useful in a surprisingly wide variety of applications. For instance, association rules can be used to improve medical diagnoses, in which itemsets are collections of test results or symptoms and transactions are the collections of test results and symptoms found on patient records. Association rules, in which itemsets are baskets of key words and transactions are the collections of words on web pages, are used by search engines. Cases of plagiarism can be found using association rules, in which itemsets are collections of sentences and transactions are the contents of documents. Association rules also play helpful roles in various aspects of computer security, including intrusion detection, in which the itemsets are collections of patterns and transactions are the strings transmitted during network attacks. The interested reader will be able to find many more such applications by searching the web.

Exercises

1. List the triples in the relation $\{(a, b, c) \mid a, b,$ and c are integers with $0 < a < b < c < 5\}$.

2. Which 4-tuples are in the relation $\{(a, b, c, d) \mid a, b, c,$ and d are positive integers with $abcd = 6\}$?

3. List the 5-tuples in the relation in Table 8.

4. Assuming that no new n-tuples are added, find all the primary keys for the relations displayed in

a) Table 3. **b)** Table 5.

c) Table 6. **d)** Table 8.

5. Assuming that no new n-tuples are added, find a composite key with two fields containing the *Airline* field for the database in Table 8.

6. Assuming that no new n-tuples are added, find a composite key with two fields containing the Professor field for the database in Table 7.

7. The 3-tuples in a 3-ary relation represent the following attributes of a student database: student ID number, name, phone number.

a) Is student ID number likely to be a primary key?

b) Is name likely to be a primary key?

c) Is phone number likely to be a primary key?

8. The 4-tuples in a 4-ary relation represent these attributes of published books: title, ISBN, publication date, number of pages.

a) What is a likely primary key for this relation?
b) Under what conditions would (title, publication date) be a composite key?
c) Under what conditions would (title, number of pages) be a composite key?

9. The 5-tuples in a 5-ary relation represent these attributes of all people in the United States: name, Social Security number, street address, city, state.

a) Determine a primary key for this relation.
b) Under what conditions would (name, street address) be a composite key?
c) Under what conditions would (name, street address, city) be a composite key?

10. What do you obtain when you apply the selection operator s_C, where C is the condition Room = A100, to the database in Table 7?

11. What do you obtain when you apply the selection operator s_C, where C is the condition Destination = Detroit, to the database in Table 8?

12. What do you obtain when you apply the selection operator s_C, where C is the condition $(\text{Project} = 2) \wedge (\text{Quantity} \geq 50)$, to the database in Table 10?

13. What do you obtain when you apply the selection operator s_C, where C is the condition $(\text{Airline} = \text{Nadir}) \vee (\text{Destination} = \text{Denver})$, to the database in Table 8?

14. What do you obtain when you apply the projection $P_{2,3,5}$ to the 5-tuple (a, b, c, d, e)?

15. Which projection mapping is used to delete the first, second, and fourth components of a 6-tuple?

16. Display the table produced by applying the projection $P_{1,2,4}$ to Table 8.

17. Display the table produced by applying the projection $P_{1,4}$ to Table 8.

18. How many components are there in the n-tuples in the table obtained by applying the join operator J_3 to two tables with 5-tuples and 8-tuples, respectively?

19. Construct the table obtained by applying the join operator J_2 to the relations in Tables 11 and 12.

20. Show that if C_1 and C_2 are conditions that elements of the n-ary relation R may satisfy, then $s_{C_1 \wedge C_2}(R) = s_{C_1}(s_{C_2}(R))$.

21. Show that if C_1 and C_2 are conditions that elements of the n-ary relation R may satisfy, then $s_{C_1}(s_{C_2}(R)) = s_{C_2}(s_{C_1}(R))$.

22. Show that if C is a condition that elements of the n-ary relations R and S may satisfy, then $s_C(R \cup S) = s_C(R) \cup s_C(S)$.

23. Show that if C is a condition that elements of the n-ary relations R and S may satisfy, then $s_C(R \cap S) = s_C(R) \cap s_C(S)$.

24. Show that if C is a condition that elements of the n-ary relations R and S may satisfy, then $s_C(R - S) = s_C(R) - s_C(S)$.

25. Show that if R and S are both n-ary relations, then $P_{i_1,i_2,\ldots,i_m}(R \cup S) = P_{i_1,i_2,\ldots,i_m}(R) \cup P_{i_1,i_2,\ldots,i_m}(S)$.

26. Give an example to show that if R and S are both n-ary relations, then $P_{i_1,i_2,\ldots,i_m}(R \cap S)$ may be different from $P_{i_1,i_2,\ldots,i_m}(R) \cap P_{i_1,i_2,\ldots,i_m}(S)$.

27. Give an example to show that if R and S are both n-ary relations, then $P_{i_1,i_2,\ldots,i_m}(R - S)$ may be different from $P_{i_1,i_2,\ldots,i_m}(R) - P_{i_1,i_2,\ldots,i_m}(S)$.

28. a) What are the operations that correspond to the query expressed using this SQL statement?

```
SELECT Supplier
FROM Part_needs
WHERE 1000 ≤ Part_number ≤ 5000
```

b) What is the output of this query given the database in Table 11 as input?

29. a) What are the operations that correspond to the query expressed using this SQL statement?

```
SELECT Supplier, Project
FROM Part_needs, Parts_inventory
WHERE Quantity ≤ 10
```

b) What is the output of this query given the databases in Tables 11 and 12 as input?

30. Determine whether there is a primary key for the relation in Example 2.

31. Determine whether there is a primary key for the relation in Example 3.

32. Show that an n-ary relation with a primary key can be thought of as the graph of a function that maps values of the primary key to $(n - 1)$-tuples formed from values of the other domains.

33. Suppose that the transactions at a convenience store during an evening are {bread, milk, diapers, juice}, {bread, milk, diapers, eggs}, {milk, diapers, beer, eggs}, {bread, beer}, {milk, diapers, eggs, juice}, and {milk, diapers, beer}.

a) Find the count and support of diapers.
b) Find all frequent itemsets if the threshold level is 0.6.

34. Suppose that the key words on eight different web pages are {evolution, primate, Human, Neanderthal, DNA, fossil}, {evolution, Neanderthal, Denisovan, Human, DNA}, {cave, fossil, primate}, {Human, Neanderthal, Denisovan, evolution}, {DNA, genome, evolution, fossil}, {DNA, Human, Neanderthal, Denisovan, genome}, {evolution, primate, cave, fossil}, and {Human, Neanderthal, genome}.

a) Find the count and support of Neanderthal.
b) Find all frequent itemsets if the threshold level is 0.6.

35. Find the support and confidence of the association rule {beer} → {diapers} for the set of transactions in Exercise 33. (This association rule has played an important role in the development of the subject.)

36. Find the support and confidence of the association rule {human, DNA} → {Neanderthal} for the set of transactions in Exercise 34.

TABLE 11 **Part_needs.**

Supplier	*Part_number*	*Project*
23	1092	1
23	1101	3
23	9048	4
31	4975	3
31	3477	2
32	6984	4
32	9191	2
33	1001	1

TABLE 12 **Parts_inventory.**

Part_number	*Project*	*Quantity*	*Color_code*
1001	1	14	8
1092	1	2	2
1101	3	1	1
3477	2	25	2
4975	3	6	2
6984	4	10	1
9048	4	12	2
9191	2	80	4

37. Suppose that I is an itemset with positive count in a set of transactions. Find the confidence of the association rule $I \to \emptyset$.

38. Suppose that I, J, and K are itemsets. Show that the six association rules $\{I, J\} \to K$, $\{J, K\} \to I$, $\{I, K\} \to J$, $I \to \{J, K\}$, $J \to \{I, K\}$, and $K \to \{I, J\}$ all have the same support.

39. The **lift** of the association rule $I \to J$, where I and J are itemsets with positive support in a set of transactions, equals support($I \cup J$)/(support(I)support(J)).

a) Show that the lift of $I \to J$, when support(I) and support(J) are both positive, equals 1 if and only if the occurrence of I in a transaction and the occurrence of J in a transaction are independent events.

b) Find the lift of the association rule {beer} → {diapers} for the set of transactions in Exercises 33.

c) Find the lift of the association rule {evolution} → {Neanderthals, Denisovans} for the set of transactions in Exercise 34.

40. Show that if an itemset is frequent in a set of transactions, then all its subsets are also frequent itemsets in this set of transactions.

41. Given n unique items, show that there are 3^n possible association rules of the form $I \to J$, where I and J are disjoint subsets of the set of all items. Be sure to allow the association rules where I or J, or both, are empty.

9.3 Representing Relations

9.3.1 Introduction

In this section, and in the remainder of this chapter, all relations we study will be binary relations. Because of this, in this section and in the rest of this chapter, the word relation will always refer to a binary relation. There are many ways to represent a relation between finite sets. As we have seen in Section 9.1, one way is to list its ordered pairs. Another way to represent a relation is to use a table, as we did in Example 3 in Section 9.1. In this section we will discuss two alternative methods for representing relations. One method uses zero–one matrices. The other method uses pictorial representations called directed graphs, which we will discuss later in this section.

Generally, matrices are appropriate for the representation of relations in computer programs. On the other hand, people often find the representation of relations using directed graphs useful for understanding the properties of these relations.

9.3.2 Representing Relations Using Matrices

A relation between finite sets can be represented using a zero–one matrix. Suppose that R is a relation from $A = \{a_1, a_2, \ldots, a_m\}$ to $B = \{b_1, b_2, \ldots, b_n\}$. (Here the elements of the sets A and B have been listed in a particular, but arbitrary, order. Furthermore, when $A = B$ we use the same ordering for A and B.) The relation R can be represented by the matrix $\mathbf{M}_R = [m_{ij}]$, where

$$m_{ij} = \begin{cases} 1 \text{ if } (a_i, b_j) \in R, \\ 0 \text{ if } (a_i, b_j) \notin R. \end{cases}$$

In other words, the zero–one matrix representing R has a 1 as its (i, j) entry when a_i is related to b_j, and a 0 in this position if a_i is not related to b_j. (Such a representation depends on the orderings used for A and B.)

The use of matrices to represent relations is illustrated in Examples 1–6.

EXAMPLE 1 Suppose that $A = \{1, 2, 3\}$ and $B = \{1, 2\}$. Let R be the relation from A to B containing (a, b) if $a \in A$, $b \in B$, and $a > b$. What is the matrix representing R if $a_1 = 1$, $a_2 = 2$, and $a_3 = 3$, and $b_1 = 1$ and $b_2 = 2$?

Solution: Because $R = \{(2, 1), (3, 1), (3, 2)\}$, the matrix for R is

$$\mathbf{M}_R = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

The 1s in $\mathbf{M}_R$ show that the pairs (2, 1), (3, 1), and (3, 2) belong to R. The 0s show that no other pairs belong to R. ◀

EXAMPLE 2 Let $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3, b_4, b_5\}$. Which ordered pairs are in the relation R represented by the matrix

$$\mathbf{M}_R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}?$$

Solution: Because R consists of those ordered pairs (a_i, b_j) with $m_{ij} = 1$, it follows that

$$R = \{(a_1, b_2), (a_2, b_1), (a_2, b_3), (a_2, b_4), (a_3, b_1), (a_3, b_3), (a_3, b_5)\}.$$ ◀

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

FIGURE 1 The zero–one matrix for a reflexive relation. (Off diagonal elements can be 0 or 1.)

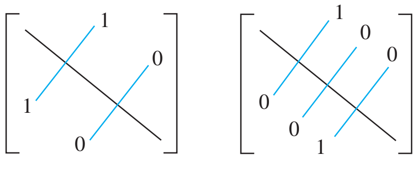
The matrix of a relation on a set, which is a square matrix, can be used to determine whether the relation has certain properties. Recall that a relation R on A is reflexive if $(a, a) \in R$ whenever $a \in A$. Thus, R is reflexive if and only if $(a_i, a_i) \in R$ for $i = 1, 2, \ldots, n$. Hence, R is reflexive if and only if $m_{ii} = 1$, for $i = 1, 2, \ldots, n$. In other words, R is reflexive if all the elements on the main diagonal of $\mathbf{M}_R$ are equal to 1, as shown in Figure 1. Note that the elements off the main diagonal can be either 0 or 1.

The relation R is symmetric if $(a, b) \in R$ implies that $(b, a) \in R$. Consequently, the relation R on the set $A = \{a_1, a_2, \ldots, a_n\}$ is symmetric if and only if $(a_j, a_i) \in R$ whenever $(a_i, a_j) \in R$. In terms of the entries of $\mathbf{M}_R$, R is symmetric if and only if $m_{ji} = 1$ whenever $m_{ij} = 1$. This also means $m_{ji} = 0$ whenever $m_{ij} = 0$. Consequently, R is symmetric if and only if $m_{ij} = m_{ji}$, for all pairs of integers i and j with $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$. Recalling the definition of the transpose of a matrix from Section 2.6, we see that R is symmetric if and only if

$$\mathbf{M}_R = (\mathbf{M}_R)^t,$$

that is, if $\mathbf{M}_R$ is a symmetric matrix. The form of the matrix for a symmetric relation is illustrated in Figure 2(a).

The relation R is antisymmetric if and only if $(a, b) \in R$ and $(b, a) \in R$ imply that $a = b$. Consequently, the matrix of an antisymmetric relation has the property that if $m_{ij} = 1$ with $i \neq j$, then $m_{ji} = 0$. Or, in other words, either $m_{ij} = 0$ or $m_{ji} = 0$ when $i \neq j$. The form of the matrix for an antisymmetric relation is illustrated in Figure 2(b).



(a) Symmetric (b) Antisymmetric

FIGURE 2 The zero–one matrices for symmetric and antisymmetric relations.

EXAMPLE 3 Suppose that the relation R on a set is represented by the matrix

$$\mathbf{M}_R = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Is R reflexive, symmetric, and/or antisymmetric?

Solution: Because all the diagonal elements of this matrix are equal to 1, R is reflexive. Moreover, because $\mathbf{M}_R$ is symmetric, it follows that R is symmetric. It is also easy to see that R is not antisymmetric. ◀

The Boolean operations join and meet (discussed in Section 2.6) can be used to find the matrices representing the union and the intersection of two relations. Suppose that R_1 and R_2 are relations on a set A represented by the matrices $\mathbf{M}_{R_1}$ and $\mathbf{M}_{R_2}$, respectively. The matrix representing the union of these relations has a 1 in the positions where either $\mathbf{M}_{R_1}$ or $\mathbf{M}_{R_2}$ has a 1. The matrix representing the intersection of these relations has a 1 in the positions where both $\mathbf{M}_{R_1}$ and $\mathbf{M}_{R_2}$ have a 1. Thus, the matrices representing the union and intersection of these relations are

$$\mathbf{M}_{R_1 \cup R_2} = \mathbf{M}_{R_1} \vee \mathbf{M}_{R_2} \qquad \text{and} \qquad \mathbf{M}_{R_1 \cap R_2} = \mathbf{M}_{R_1} \wedge \mathbf{M}_{R_2}.$$

EXAMPLE 4 Suppose that the relations R_1 and R_2 on a set A are represented by the matrices

$$\mathbf{M}_{R_1} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{M}_{R_2} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

What are the matrices representing $R_1 \cup R_2$ and $R_1 \cap R_2$?

Solution: The matrices of these relations are

$$\mathbf{M}_{R_1 \cup R_2} = \mathbf{M}_{R_1} \vee \mathbf{M}_{R_2} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

$$\mathbf{M}_{R_1 \cap R_2} = \mathbf{M}_{R_1} \wedge \mathbf{M}_{R_2} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

◀

We now turn our attention to determining the matrix for the composite of relations. This matrix can be found using the Boolean product of the matrices (discussed in Section 2.6) for these relations. In particular, suppose that R is a relation from A to B and S is a relation from B to C. Suppose that A, B, and C have m, n, and p elements, respectively. Let the zero–one matrices for $S \circ R$, R, and S be $\mathbf{M}_{S \circ R} = [t_{ij}]$, $\mathbf{M}_R = [r_{ij}]$, and $\mathbf{M}_S = [s_{ij}]$, respectively (these matrices have sizes $m \times p$, $m \times n$, and $n \times p$, respectively). The ordered pair (a_i, c_j) belongs to $S \circ R$ if and only if there is an element b_k such that (a_i, b_k) belongs to R and (b_k, c_j) belongs to S. It follows that $t_{ij} = 1$ if and only if $r_{ik} = s_{kj} = 1$ for some k. From the definition of the Boolean product, this means that

$$\mathbf{M}_{S \circ R} = \mathbf{M}_R \odot \mathbf{M}_S.$$

EXAMPLE 5 Find the matrix representing the relations $S \circ R$, where the matrices representing R and S are

$$\mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{M}_S = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Solution: The matrix for $S \circ R$ is

$$\mathbf{M}_{S \circ R} = \mathbf{M}_R \odot \mathbf{M}_S = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

◀

The matrix representing the composite of two relations can be used to find the matrix for $\mathbf{M}_{R^n}$. In particular,

$$\mathbf{M}_{R^n} = \mathbf{M}_R^{[n]},$$

from the definition of Boolean powers. Exercise 35 asks for a proof of this formula.

EXAMPLE 6 Find the matrix representing the relation R^2, where the matrix representing R is

$$\mathbf{M}_R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Solution: The matrix for R^2 is

$$\mathbf{M}_{R^2} = \mathbf{M}_R^{[2]} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

◀

9.3.3 Representing Relations Using Digraphs

We have shown that a relation can be represented by listing all of its ordered pairs or by using a zero–one matrix. There is another important way of representing a relation using a pictorial representation. Each element of the set is represented by a point, and each ordered pair is represented using an arc with its direction indicated by an arrow. We use such pictorial representations when we think of relations on a finite set as **directed graphs**, or **digraphs**.

Definition 1 A *directed graph*, or *digraph*, consists of a set V of *vertices* (or *nodes*) together with a set E of ordered pairs of elements of V called *edges* (or *arcs*). The vertex a is called the *initial vertex* of the edge (a, b), and the vertex b is called the *terminal vertex* of this edge.

An edge of the form (a, a) is represented using an arc from the vertex a back to itself. Such an edge is called a **loop**.

EXAMPLE 7 The directed graph with vertices a, b, c, and d, and edges (a, b), (a, d), (b, b), (b, d), (c, a), (c, b), and (d, b) is displayed in Figure 3. ◀

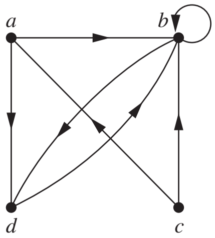


FIGURE 3
A directed graph.

The relation R on a set A is represented by the directed graph that has the elements of A as its vertices and the ordered pairs (a, b), where $(a, b) \in R$, as edges. This assignment sets up a one-to-one correspondence between the relations on a set A and the directed graphs with A as their set of vertices. Thus, every statement about relations corresponds to a statement about directed graphs, and vice versa. Directed graphs give a visual display of information about relations. As such, they are often used to study relations and their properties. (Note that relations from a set A to a set B can be represented by a directed graph where there is a vertex for each element of A and a vertex for each element of B, as shown in Section 9.1. However, when $A = B$, such representation provides much less insight than the digraph representations described here.) The use of directed graphs to represent relations on a set is illustrated in Examples 8–10.

EXAMPLE 8 The directed graph of the relation

$$R_1 = \{(1, 1), (1, 3), (2, 1), (2, 3), (2, 4), (3, 1), (3, 2), (4, 1)\}$$

on the set $\{1, 2, 3, 4\}$ is shown in Figure 4. ◀

EXAMPLE 9 What are the ordered pairs in the relation R_2 represented by the directed graph shown in Figure 5?

Solution: The ordered pairs (x, y) in the relation are

$$R_2 = \{(1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 3), (4, 1), (4, 3)\}.$$

Each of these pairs corresponds to an edge of the directed graph, with $(2, 2)$ and $(3, 3)$ corresponding to loops. ◀

We will study directed graphs extensively in Chapter 10.

The directed graph representing a relation can be used to determine whether the relation has various properties. For instance, a relation is reflexive if and only if there is a loop at every vertex of the directed graph, so that every ordered pair of the form (x, x) occurs in the relation. A relation is symmetric if and only if for every edge between distinct vertices in its digraph there is an edge in the opposite direction, so that (y, x) is in the relation whenever (x, y) is in the relation. Similarly, a relation is antisymmetric if and only if there are never two edges in opposite directions between distinct vertices. Finally, a relation is transitive if and only if whenever there is an edge from a vertex x to a vertex y and an edge from a vertex y to a vertex z, there is an edge from x to z (completing a triangle where each side is a directed edge with the correct direction).

Remark: Note that a symmetric relation can be represented by an undirected graph, which is a graph where edges do not have directions. We will study undirected graphs in Chapter 10.

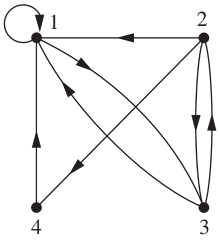


FIGURE 4 **The directed graph of the relation R_1.**

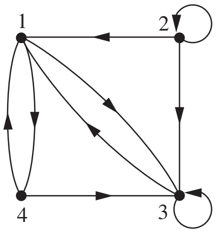
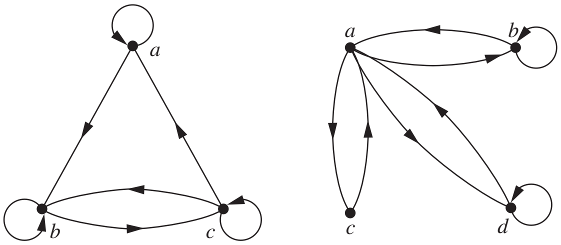


FIGURE 5 **The directed graph of the relation R_2.**



(a) Directed graph of S_1

(b) Directed graph of S_2

FIGURE 6 **The directed graphs of the relations S_1 and S_2.**

EXAMPLE 10 Determine whether the relations for the directed graphs shown in Figure 6 are reflexive, symmetric, antisymmetric, and/or transitive.

Solution: Because there are loops at every vertex of the directed graph of S_1, it is reflexive. The relation S_1 is neither symmetric nor antisymmetric because there is an edge from a to b but not one from b to a, but there are edges in both directions connecting b and c. Finally, S_1 is not transitive because there is an edge from a to b and an edge from b to c, but no edge from a to c.

Because loops are not present at all the vertices of the directed graph of S_2, this relation is not reflexive. It is symmetric and not antisymmetric, because every edge between distinct vertices is accompanied by an edge in the opposite direction. It is also not hard to see from the directed graph that S_2 is not transitive, because (c, a) and (a, b) belong to S_2, but (c, b) does not belong to S_2. ◀

Exercises

1. Represent each of these relations on $\{1, 2, 3\}$ with a matrix (with the elements of this set listed in increasing order).

a) $\{(1, 1), (1, 2), (1, 3)\}$
b) $\{(1, 2), (2, 1), (2, 2), (3, 3)\}$
c) $\{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$
d) $\{(1, 3), (3, 1)\}$

2. Represent each of these relations on $\{1, 2, 3, 4\}$ with a matrix (with the elements of this set listed in increasing order).

a) $\{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$
b) $\{(1, 1), (1, 4), (2, 2), (3, 3), (4, 1)\}$
c) $\{(1, 2), (1, 3), (1, 4), (2, 1), (2, 3), (2, 4), (3, 1), (3, 2), (3, 4), (4, 1), (4, 2), (4, 3)\}$
d) $\{(2, 4), (3, 1), (3, 2), (3, 4)\}$

3. List the ordered pairs in the relations on $\{1, 2, 3\}$ corresponding to these matrices (where the rows and columns correspond to the integers listed in increasing order).

a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

4. List the ordered pairs in the relations on $\{1, 2, 3, 4\}$ corresponding to these matrices (where the rows and columns correspond to the integers listed in increasing order).

a) $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$

5. How can the matrix representing a relation R on a set A be used to determine whether the relation is irreflexive?

6. How can the matrix representing a relation R on a set A be used to determine whether the relation is asymmetric?

7. Determine whether the relations represented by the matrices in Exercise 3 are reflexive, irreflexive, symmetric, antisymmetric, and/or transitive.

8. Determine whether the relations represented by the matrices in Exercise 4 are reflexive, irreflexive, symmetric, antisymmetric, and/or transitive.

9. How many nonzero entries does the matrix representing the relation R on $A = \{1, 2, 3, \ldots, 100\}$ consisting of the first 100 positive integers have if R is

a) $\{(a, b) \mid a > b\}$? **b)** $\{(a, b) \mid a \neq b\}$?
c) $\{(a, b) \mid a = b + 1\}$? **d)** $\{(a, b) \mid a = 1\}$?
e) $\{(a, b) \mid ab = 1\}$?

10. How many nonzero entries does the matrix representing the relation R on $A = \{1, 2, 3, \ldots, 1000\}$ consisting of the first 1000 positive integers have if R is

a) $\{(a, b) \mid a \leq b\}$?
b) $\{(a, b) \mid a = b \pm 1\}$?
c) $\{(a, b) \mid a + b = 1000\}$?
d) $\{(a, b) \mid a + b \leq 1001\}$?
e) $\{(a, b) \mid a \neq 0\}$?

11. How can the matrix for $\overline{R}$, the complement of the relation R, be found from the matrix representing R, when R is a relation on a finite set A?

12. How can the matrix for R^{-1}, the inverse of the relation R, be found from the matrix representing R, when R is a relation on a finite set A?

13. Let R be the relation represented by the matrix

$$\mathbf{M}_R = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Find the matrix representing

a) R^{-1}. **b)** $\overline{R}$. **c)** R^2.

14. Let R_1 and R_2 be relations on a set A represented by the matrices

$$\mathbf{M}_{R_1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{M}_{R_2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Find the matrices that represent

a) $R_1 \cup R_2$. **b)** $R_1 \cap R_2$. **c)** $R_2 \circ R_1$.
d) $R_1 \circ R_1$. **e)** $R_1 \oplus R_2$.

15. Let R be the relation represented by the matrix

$$\mathbf{M}_R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Find the matrices that represent

a) R^2. **b)** R^3. **c)** R^4.

16. Let R be a relation on a set A with n elements. If there are k nonzero entries in $\mathbf{M}_R$, the matrix representing R, how many nonzero entries are there in $\mathbf{M}_{R^{-1}}$, the matrix representing R^{-1}, the inverse of R?

17. Let R be a relation on a set A with n elements. If there are k nonzero entries in $\mathbf{M}_R$, the matrix representing R, how many nonzero entries are there in $\mathbf{M}_{\overline{R}}$, the matrix representing $\overline{R}$, the complement of R?

18. Draw the directed graphs representing each of the relations from Exercise 1.

19. Draw the directed graphs representing each of the relations from Exercise 2.

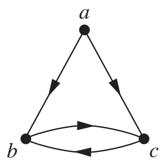
20. Draw the directed graph representing each of the relations from Exercise 3.

21. Draw the directed graph representing each of the relations from Exercise 4.

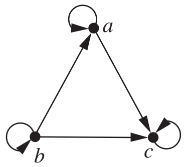
22. Draw the directed graph that represents the relation $\{(a, a), (a, b), (b, c), (c, b), (c, d), (d, a), (d, b)\}$.

In Exercises 23–28 list the ordered pairs in the relations represented by the directed graphs.

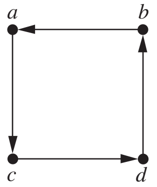
23.



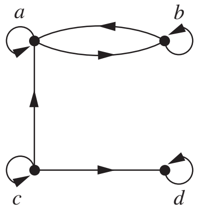
24.



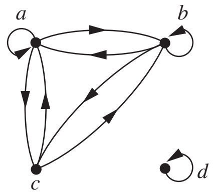
25.



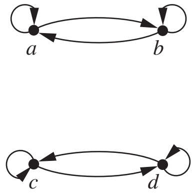
26.



27.



28.



29. How can the directed graph of a relation R on a finite set A be used to determine whether a relation is asymmetric?

30. How can the directed graph of a relation R on a finite set A be used to determine whether a relation is irreflexive?

31. Determine whether the relations represented by the directed graphs shown in Exercises 23–25 are reflexive, irreflexive, symmetric, antisymmetric, and/or transitive.

32. Determine whether the relations represented by the directed graphs shown in Exercises 26–28 are reflexive, irreflexive, symmetric, antisymmetric, asymmetric, and/or transitive.

33. Let R be a relation on a set A. Explain how to use the directed graph representing R to obtain the directed graph representing the inverse relation R^{-1}.

34. Let R be a relation on a set A. Explain how to use the directed graph representing R to obtain the directed graph representing the complementary relation $\overline{R}$.

35. Show that if $\mathbf{M}_R$ is the matrix representing the relation R, then $\mathbf{M}_R^{[n]}$ is the matrix representing the relation R^n.

36. Given the directed graphs representing two relations, how can the directed graph of the union, intersection, symmetric difference, difference, and composition of these relations be found?

9.4 Closures of Relations

9.4.1 Introduction

A computer network has data centers in Boston, Chicago, Denver, Detroit, New York, and San Diego. There are direct, one-way telephone lines from Boston to Chicago, from Boston to Detroit, from Chicago to Detroit, from Detroit to Denver, and from New York to San Diego. Let R be the relation containing (a, b) if there is a telephone line from the data center in a to that in b. How can we determine if there is some (possibly indirect) link composed of one or more telephone lines from one center to another? Because not all links are direct, such as the link from Boston to Denver that goes through Detroit, R cannot be used directly to answer this. In the language of relations, R is not transitive, so it does not contain all the pairs that can be linked. As we will show in this section, we can find all pairs of data centers that have a link by constructing a transitive relation S containing R such that S is a subset of every transitive relation containing R. Here, S is the smallest transitive relation that contains R. This relation is called the **transitive closure** of R.

9.4.2 Different Types of Closures

If R is a relation on a set A, it may or may not have some property **P**, such as reflexivity, symmetry, or transitivity. When R does not enjoy property **P**, we would like to find the smallest relation S on A with property **P** that contains R.

Definition 1 If R is a relation on a set A, then the **closure** of R with respect to **P**, if it exists, is the relation S on A with property **P** that contains R and is a subset of every subset of $A \times A$ containing R with property **P**.

If there is a relation S that is a subset of every relation containing R with property **P**, it must be unique. To see this, suppose that relations S and T both have property **P** and are subsets of every relation with property **P** that contains R. Then, S and T are subsets of each other, and so are equal. Such a relation, if it exists, is the smallest relation with property **P** that contains R because it is a subset of every relation with property **P** that contains R.

We will show how reflexive, symmetric, and transitive closures of relations can be found. In Exercises 15 and 35 we give properties **P** for which the closure of a relation with respect to **P** may not exist.

The relation $R = \{(1, 1), (1, 2), (2, 1), (3, 2)\}$ on the set $A = \{1, 2, 3\}$ is not reflexive. How can we produce a reflexive relation containing R that is as small as possible? This can be done by adding $(2, 2)$ and $(3, 3)$ to R, because these are the only pairs of the form (a, a) that are not in R. This new relation contains R. Furthermore, *any* reflexive relation that contains R must also contain $(2, 2)$ and $(3, 3)$. Because this relation contains R, is reflexive, and is contained within every reflexive relation that contains R, it is called the **reflexive closure** of R.

As this example illustrates, given a relation R on a set A, the reflexive closure of R can be formed by adding to R all pairs of the form (a, a) with $a \in A$, not already in R. The addition of these pairs produces a new relation that is reflexive, contains R, and is contained within any reflexive relation containing R. We see that the reflexive closure of R equals $R \cup \Delta$, where $\Delta = \{(a, a) \mid a \in A\}$ is the **diagonal relation** on A. (The reader should verify this.)

EXAMPLE 1 What is the reflexive closure of the relation $R = \{(a, b) \mid a < b\}$ on the set of integers?

Solution: The reflexive closure of R is

$$R \cup \Delta = \{(a, b) \mid a < b\} \cup \{(a, a) \mid a \in \mathbf{Z}\} = \{(a, b) \mid a \leq b\}.$$

◀

The relation $\{(1, 1), (1, 2), (2, 2), (2, 3), (3, 1), (3, 2)\}$ on $\{1, 2, 3\}$ is not symmetric. How can we produce a symmetric relation that is as small as possible and contains R? To do this, we need only add $(2, 1)$ and $(1, 3)$, because these are the only pairs of the form (b, a) with $(a, b) \in R$ that are not in R. This new relation is symmetric and contains R. Furthermore, *any* symmetric relation that contains R must contain this new relation, because a symmetric relation that contains R must contain $(2, 1)$ and $(1, 3)$. Consequently, this new relation is called the **symmetric closure** of R.

As this example illustrates, the symmetric closure of a relation R can be constructed by adding all ordered pairs of the form (b, a), where (a, b) is in the relation, that are not already present in R. Adding these pairs produces a relation that is symmetric, that contains R, and that is contained in any symmetric relation that contains R. The symmetric closure of a relation can be constructed by taking the union of a relation with its inverse (defined in the preamble of Exercise 26 in Section 9.1); that is, $R \cup R^{-1}$ is the symmetric closure of R, where $R^{-1} = \{(b, a) \mid (a, b) \in R\}$. The reader should verify this statement.

EXAMPLE 2 What is the symmetric closure of the relation $R = \{(a, b) \mid a > b\}$ on the set of positive integers?

Extra Examples

Solution: The symmetric closure of R is the relation

$$R \cup R^{-1} = \{(a, b) \mid a > b\} \cup \{(b, a) \mid a > b\} = \{(a, b) \mid a \neq b\}.$$

This last equality follows because R contains all ordered pairs of positive integers, where the first element is greater than the second element, and R^{-1} contains all ordered pairs of positive integers, where the first element is less than the second. ◀

Suppose that a relation R is not transitive. How can we produce a transitive relation that contains R such that this new relation is contained within any transitive relation that contains R? Can the transitive closure of a relation R be produced by adding all the pairs of the form (a, c), where (a, b) and (b, c) are already in the relation? Consider the relation $R = \{(1, 3), (1, 4), (2, 1), (3, 2)\}$ on the set $\{1, 2, 3, 4\}$. This relation is not transitive because it does not contain all pairs of the form (a, c) where (a, b) and (b, c) are in R. The pairs of this form not in R are $(1, 2)$, $(2, 3)$, $(2, 4)$, and $(3, 1)$. Adding these pairs does *not* produce a transitive relation, because the resulting relation contains $(3, 1)$ and $(1, 4)$ but does not contain $(3, 4)$. This shows that constructing the transitive closure of a relation is more complicated than constructing either the reflexive or symmetric closure. The rest of this section develops algorithms for constructing transitive closures. As will be shown later in this section, the transitive closure of a relation can be found by adding new ordered pairs that must be present and then repeating this process until no new ordered pairs are needed.

9.4.3 Paths in Directed Graphs

We will see that representing relations by directed graphs helps in the construction of transitive closures. We now introduce some terminology that we will use for this purpose.

A path in a directed graph is obtained by traversing along edges (in the same direction as indicated by the arrow on the edge).

Definition 2

A *path* from a to b in the directed graph G is a sequence of edges (x_0, x_1), (x_1, x_2), $(x_2, x_3), \ldots, (x_{n-1}, x_n)$ in G, where n is a nonnegative integer, and $x_0 = a$ and $x_n = b$, that is, a sequence of edges where the terminal vertex of an edge is the same as the initial vertex in the next edge in the path. This path is denoted by $x_0, x_1, x_2, \ldots, x_{n-1}, x_n$ and has *length* n. We view the empty set of edges as a path of length zero from a to a. A path of length $n \geq 1$ that begins and ends at the same vertex is called a *circuit* or *cycle*.

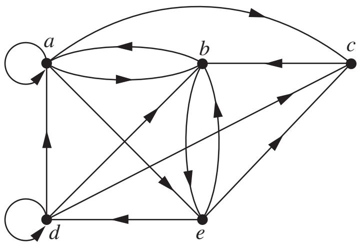


FIGURE 1 **A directed graph.**

A path in a directed graph can pass through a vertex more than once. Moreover, an edge in a directed graph can occur more than once in a path.

EXAMPLE 3 Which of the following are paths in the directed graph shown in Figure 1: a, b, e, d; a, e, c, d, b; b, a, c, b, a, a, b; d, c; c, b, a; e, b, a, b, a, b, e? What are the lengths of those that are paths? Which of the paths in this list are circuits?

Solution: Because each of (a, b), (b, e), and (e, d) is an edge, a, b, e, d is a path of length three. Because (c, d) is not an edge, a, e, c, d, b is not a path. Also, b, a, c, b, a, a, b is a path of length six because (b, a), (a, c), (c, b), (b, a), (a, a), and (a, b) are all edges. We see that d, c is a path of length one, because (d, c) is an edge. Also c, b, a is a path of length two, because (c, b) and (b, a) are edges. All of (e, b), (b, a), (a, b), (b, a), (a, b), and (b, e) are edges, so e, b, a, b, a, b, e is a path of length six.

The two paths b, a, c, b, a, a, b and e, b, a, b, a, b, e are circuits because they begin and end at the same vertex. The paths a, b, e, d; c, b, a; and d, c are not circuits. ◀

The term *path* also applies to relations. Carrying over the definition from directed graphs to relations, there is a **path** from a to b in R if there is a sequence of elements $a, x_1, x_2, \ldots, x_{n-1}, b$ with $(a, x_1) \in R$, $(x_1, x_2) \in R$, …, and $(x_{n-1}, b) \in R$. Theorem 1 can be obtained from the definition of a path in a relation.

THEOREM 1 Let R be a relation on a set A. There is a path of length n, where n is a positive integer, from a to b if and only if $(a, b) \in R^n$.

Proof: We will use mathematical induction. By definition, there is a path from a to b of length one if and only if $(a, b) \in R$, so the theorem is true when $n = 1$.

Assume that the theorem is true for the positive integer n. This is the inductive hypothesis. There is a path of length $n + 1$ from a to b if and only if there is an element $c \in A$ such that there is a path of length one from a to c, so $(a, c) \in R$, and a path of length n from c to b, that is, $(c, b) \in R^n$. Consequently, by the inductive hypothesis, there is a path of length $n + 1$ from a to b if and only if there is an element c with $(a, c) \in R$ and $(c, b) \in R^n$. But there is such an element if and only if $(a, b) \in R^{n+1}$. Therefore, there is a path of length $n + 1$ from a to b if and only if $(a, b) \in R^{n+1}$. This completes the proof. ◁

9.4.4 Transitive Closures

We now show that finding the transitive closure of a relation is equivalent to determining which pairs of vertices in the associated directed graph are connected by a path. With this in mind, we define a new relation.

Definition 3 Let R be a relation on a set A. The *connectivity relation* R^* consists of the pairs (a, b) such that there is a path of length at least one from a to b in R.

Because R^n consists of the pairs (a, b) such that there is a path of length n from a to b, it follows that R^* is the union of all the sets R^n. In other words,

$$R^* = \bigcup_{n=1}^{\infty} R^n.$$

The connectivity relation is useful in many models.

EXAMPLE 4 Let R be the relation on the set of all people in the world that contains (a, b) if a has met b. What is R^n, where n is a positive integer greater than one? What is R^*?

Solution: The relation R^2 contains (a, b) if there is a person c such that $(a, c) \in R$ and $(c, b) \in R$, that is, if there is a person c such that a has met c and c has met b. Similarly, R^n consists of those pairs (a, b) such that there are people $x_1, x_2, \ldots, x_{n-1}$ such that a has met x_1, x_1 has met $x_2, \ldots,$ and x_{n-1} has met b.

The relation R^* contains (a, b) if there is a sequence of people, starting with a and ending with b, such that each person in the sequence has met the next person in the sequence. (There are many interesting conjectures about R^*. Do you think that this connectivity relation includes the pair with you as the first element and the president of Mongolia as the second element? We will use graphs to model this application in Chapter 10.) ◀

EXAMPLE 5 Let R be the relation on the set of all subway stops in New York City that contains (a, b) if it is possible to travel from stop a to stop b without changing trains. What is R^n when n is a positive integer? What is R^*?

Solution: The relation R^n contains (a, b) if it is possible to travel from stop a to stop b by making at most $n - 1$ changes of trains. The relation R^* consists of the ordered pairs (a, b) where it is possible to travel from stop a to stop b making as many changes of trains as necessary. (The reader should verify these statements.) ◀

EXAMPLE 6 Let R be the relation on the set of all states in the United States that contains (a, b) if state a and state b have a common border. What is R^n, where n is a positive integer? What is R^*?

Solution: The relation R^n consists of the pairs (a, b), where it is possible to go from state a to state b by crossing exactly n state borders. R^* consists of the ordered pairs (a, b), where it is possible to go from state a to state b crossing as many borders as necessary. (The reader should verify these statements.) The only ordered pairs not in R^* are those containing states that are not connected to the continental United States (that is, those pairs containing Alaska or Hawaii). ◀

Theorem 2 shows that the transitive closure of a relation and the associated connectivity relation are the same.

THEOREM 2 The transitive closure of a relation R equals the connectivity relation R^*.

Proof: Note that R^* contains R by definition. To show that R^* is the transitive closure of R we must also show that R^* is transitive and that $R^* \subseteq S$ whenever S is a transitive relation that contains R.

First, we show that R^* is transitive. If $(a, b) \in R^*$ and $(b, c) \in R^*$, then there are paths from a to b and from b to c in R. We obtain a path from a to c by starting with the path from a to b and following it with the path from b to c. Hence, $(a, c) \in R^*$. It follows that R^* is transitive.

Now suppose that S is a transitive relation containing R. Because S is transitive, S^n also is transitive (the reader should verify this) and $S^n \subseteq S$ (by Theorem 1 of Section 9.1). Furthermore, because

$$S^* = \bigcup_{k=1}^{\infty} S^k$$

and $S^k \subseteq S$, it follows that $S^* \subseteq S$. Now note that if $R \subseteq S$, then $R^* \subseteq S^*$, because any path in R is also a path in S. Consequently, $R^* \subseteq S^* \subseteq S$. Thus, any transitive relation that contains R must also contain R^*. Therefore, R^* is the transitive closure of R. ◁

Now that we know that the transitive closure equals the connectivity relation, we turn our attention to the problem of computing this relation. We do not need to examine arbitrarily long paths to determine whether there is a path between two vertices in a finite directed graph. As Lemma 1 shows, it is sufficient to examine paths containing no more than n edges, where n is the number of elements in the set.

LEMMA 1 Let A be a set with n elements, and let R be a relation on A. If there is a path of length at least one in R from a to b, then there is such a path with length not exceeding n. Moreover, when $a \neq b$, if there is a path of length at least one in R from a to b, then there is such a path with length not exceeding $n - 1$.

Proof: Suppose there is a path from a to b in R. Let m be the length of the shortest such path. Suppose that $x_0, x_1, x_2, \ldots, x_{m-1}, x_m$, where $x_0 = a$ and $x_m = b$, is such a path.

Suppose that $a = b$ and that $m > n$, so that $m \geq n + 1$. By the pigeonhole principle, because there are n vertices in A, among the m vertices $x_0, x_1, \ldots, x_{m-1}$, at least two are equal (see Figure 2).

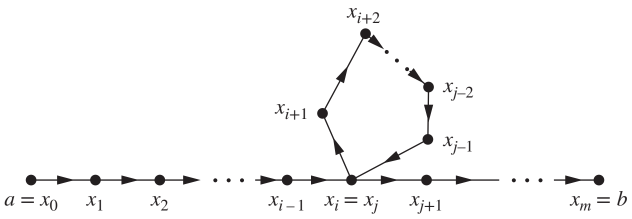


FIGURE 2 Producing a path with length not exceeding *n*.

Suppose that $x_i = x_j$ with $0 \leq i < j \leq m - 1$. Then the path contains a circuit from x_i to itself. This circuit can be deleted from the path from a to b, leaving a path, namely, $x_0, x_1, \ldots, x_i, x_{j+1}, \ldots, x_{m-1}, x_m$, from a to b of shorter length. Hence, the path of shortest length must have length less than or equal to n.

The case where $a \neq b$ is left as an exercise for the reader. ◁

From Lemma 1, we see that the transitive closure of R is the union of $R, R^2, R^3, \ldots,$ and R^n. This follows because there is a path in R^* between two vertices if and only if there is a path between these vertices in R^i, for some positive integer i with $i \leq n$. Because

$$R^* = R \cup R^2 \cup R^3 \cup \cdots \cup R^n$$

and the zero–one matrix representing a union of relations is the join of the zero–one matrices of these relations, the zero–one matrix for the transitive closure is the join of the zero–one matrices of the first n powers of the zero–one matrix of R.

THEOREM 3 Let $\mathbf{M}_R$ be the zero–one matrix of the relation R on a set with n elements. Then the zero–one matrix of the transitive closure R^* is

$$\mathbf{M}_{R^*} = \mathbf{M}_R \vee \mathbf{M}_R^{[2]} \vee \mathbf{M}_R^{[3]} \vee \cdots \vee \mathbf{M}_R^{[n]}.$$

EXAMPLE 7 Find the zero–one matrix of the transitive closure of the relation R where

$$\mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

Solution: By Theorem 3, it follows that the zero–one matrix of R^* is

$$\mathbf{M}_{R^*} = \mathbf{M}_R \vee \mathbf{M}_R^{[2]} \vee \mathbf{M}_R^{[3]}.$$

Because

$$\mathbf{M}_R^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{M}_R^{[3]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix},$$

it follows that

$$\mathbf{M}_{R^*} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

◀

Links

Theorem 3 can be used as a basis for an algorithm for computing the matrix of the relation R^*. To find this matrix, the successive Boolean powers of $\mathbf{M}_R$, up to the nth power, are computed. As each power is calculated, its join with the join of all smaller powers is formed. When this is done with the nth power, the matrix for R^* has been found. This procedure is displayed as Algorithm 1.

ALGORITHM 1 A Procedure for Computing the Transitive Closure.

```
procedure transitive closure ($\mathbf{M}_R$ : zero–one $n \times n$ matrix)
$\mathbf{A} := \mathbf{M}_R$
$\mathbf{B} := \mathbf{A}$
for $i := 2$ to $n$
    $\mathbf{A} := \mathbf{A} \odot \mathbf{M}_R$
    $\mathbf{B} := \mathbf{B} \vee \mathbf{A}$
return $\mathbf{B}$ {$\mathbf{B}$ is the zero–one matrix for $R^*$}
```

We can easily find the number of bit operations used by Algorithm 1 to determine the transitive closure of a relation. Computing the Boolean powers $\mathbf{M}_R, \mathbf{M}_R^{[2]}, \ldots, \mathbf{M}_R^{[n]}$ requires that $n-1$ Boolean products of $n \times n$ zero–one matrices be found. Each of these Boolean products can be found using $n^2(2n-1)$ bit operations. Hence, these products can be computed using $n^2(2n-1)(n-1)$ bit operations.

To find $\mathbf{M}_{R^*}$ from the n Boolean powers of $\mathbf{M}_R$, $n-1$ joins of zero–one matrices need to be found. Computing each of these joins uses n^2 bit operations. Hence, $(n-1)n^2$ bit operations are used in this part of the computation. Therefore, when Algorithm 1 is used, the matrix of the transitive closure of a relation on a set with n elements can be found using $n^2(2n-1)(n-1) + (n-1)n^2 = 2n^3(n-1)$, which is $O(n^4)$ bit operations. The remainder of this section describes a more efficient algorithm for finding transitive closures.

9.4.5 Warshall's Algorithm

Links

Warshall's algorithm, named after Stephen Warshall, who described this algorithm in 1960, is an efficient method for computing the transitive closure of a relation. Algorithm 1 can find the transitive closure of a relation on a set with n elements using $2n^3(n-1)$ bit operations. However, the transitive closure can be found by Warshall's algorithm using only $2n^3$ bit operations.

Remark: Warshall's algorithm is sometimes called the Roy–Warshall algorithm, because Bernard Roy described this algorithm in 1959.

Suppose that R is a relation on a set with n elements. Let $v_1, v_2, \ldots, v_n$ be an arbitrary listing of these n elements. The concept of the **interior vertices** of a path is used in Warshall's algorithm. If $a, x_1, x_2, \ldots, x_{m-1}, b$ is a path, its interior vertices are $x_1, x_2, \ldots, x_{m-1}$, that is, all the vertices of the path that occur somewhere other than as the first and last vertices in the path. For instance, the interior vertices of a path a, c, d, f, g, h, b, j in a directed graph are c, d, f, g, h, and b. The interior vertices of a, c, d, a, f, b are c, d, a, and f. (Note that the first vertex in the path is not an interior vertex unless it is visited again by the path, except as the last vertex. Similarly, the last vertex in the path is not an interior vertex unless it was visited previously by the path, except as the first vertex.)

Warshall's algorithm is based on the construction of a sequence of zero–one matrices. These matrices are $\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_n$, where $\mathbf{W}_0 = \mathbf{M}_R$ is the zero–one matrix of this relation, and $\mathbf{W}_k = [w_{ij}^{(k)}]$, where $w_{ij}^{(k)} = 1$ if there is a path from v_i to v_j such that all the interior vertices of this path are in the set $\{v_1, v_2, \ldots, v_k\}$ (the first k vertices in the list) and is 0 otherwise. (The first and last vertices in the path may be outside the set of the first k vertices in the list.) Note that $\mathbf{W}_n = \mathbf{M}_{R^*}$, because the (i, j)th entry of $\mathbf{M}_{R^*}$ is 1 if and only if there is a path from v_i to v_j, with all interior vertices in the set $\{v_1, v_2, \ldots, v_n\}$ (but these are the only vertices in the directed graph). Example 8 illustrates what the matrix $\mathbf{W}_k$ represents.

EXAMPLE 8 Let R be the relation with directed graph shown in Figure 3. Let a, b, c, d be a listing of the elements of the set. Find the matrices $\mathbf{W}_0$, $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$, and $\mathbf{W}_4$. The matrix $\mathbf{W}_4$ is the transitive closure of R.

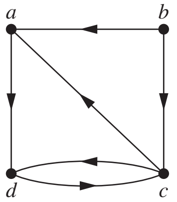


FIGURE 3 **The directed graph of the relation *R*.**

Solution: Let $v_1 = a$, $v_2 = b$, $v_3 = c$, and $v_4 = d$. $\mathbf{W}_0$ is the matrix of the relation. Hence,

$$\mathbf{W}_0 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

$\mathbf{W}_1$ has 1 as its (i, j)th entry if there is a path from v_i to v_j that has only $v_1 = a$ as an interior vertex. Note that all paths of length one can still be used because they have no interior vertices. Also, there is now an allowable path from b to d, namely, b, a, d. Hence,

$$\mathbf{W}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

$\mathbf{W}_2$ has 1 as its (i, j)th entry if there is a path from v_i to v_j that has only $v_1 = a$ and/or $v_2 = b$ as its interior vertices, if any. Because there are no edges that have b as a terminal vertex, no new paths are obtained when we permit b to be an interior vertex. Hence, $\mathbf{W}_2 = \mathbf{W}_1$.

$\mathbf{W}_3$ has 1 as its (i, j)th entry if there is a path from v_i to v_j that has only $v_1 = a$, $v_2 = b$, and/or $v_3 = c$ as its interior vertices, if any. We now have paths from d to a, namely, d, c, a, and from d to d, namely, d, c, d. Hence,

$$\mathbf{W}_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Finally, $\mathbf{W}_4$ has 1 as its (i, j)th entry if there is a path from v_i to v_j that has $v_1 = a$, $v_2 = b$, $v_3 = c$, and/or $v_4 = d$ as interior vertices, if any. Because these are all the vertices of the graph, this entry is 1 if and only if there is a path from v_i to v_j. Hence,

$$\mathbf{W}_4 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

This last matrix, $\mathbf{W}_4$, is the matrix of the transitive closure. ◀

Warshall's algorithm computes $\mathbf{M}_{R^*}$ by efficiently computing $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n = \mathbf{M}_{R^*}$. This observation shows that we can compute $\mathbf{W}_k$ directly from $\mathbf{W}_{k-1}$: There is a path from v_i to v_j with no vertices other than $v_1, v_2, \ldots, v_k$ as interior vertices if and only if either there is a path from v_i to v_j with its interior vertices among the first $k - 1$ vertices in the list, or there are paths from v_i to v_k and from v_k to v_j that have interior vertices only among the first $k - 1$ vertices in the list. That is, either a path from v_i to v_j already existed before v_k was permitted as an interior vertex, or allowing v_k as an interior vertex produces a path that goes from v_i to v_k and then from v_k to v_j. These two cases are shown in Figure 4.

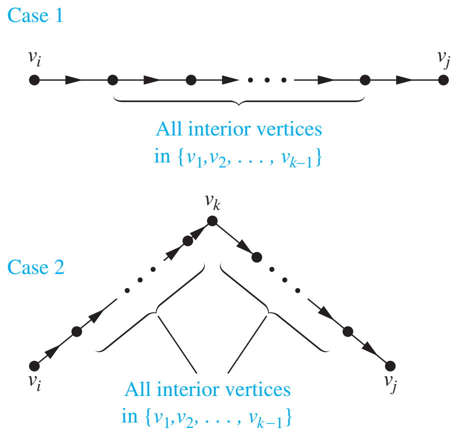

Case 1
v_i
v_j
All interior vertices in $\{v_1, v_2, \ldots, v_{k-1}\}$
v_k
Case 2
v_i
v_j
All interior vertices in $\{v_1, v_2, \ldots, v_{k-1}\}$

FIGURE 4 **Adding v_k to the set of allowable interior vertices.**

The first type of path exists if and only if $w_{ij}^{[k-1]} = 1$, and the second type of path exists if and only if both $w_{ik}^{[k-1]}$ and $w_{kj}^{[k-1]}$ are 1. Hence, $w_{ij}^{[k]}$ is 1 if and only if either $w_{ij}^{[k-1]}$ is 1 or both $w_{ik}^{[k-1]}$ and $w_{kj}^{[k-1]}$ are 1. This gives us Lemma 2.

LEMMA 2

Let $\mathbf{W}_k = [w_{ij}^{[k]}]$ be the zero–one matrix that has a 1 in its (i, j)th position if and only if there is a path from v_i to v_j with interior vertices from the set $\{v_1, v_2, \ldots, v_k\}$. Then

$$w_{ij}^{[k]} = w_{ij}^{[k-1]} \vee (w_{ik}^{[k-1]} \wedge w_{kj}^{[k-1]}),$$

whenever i, j, and k are positive integers not exceeding n.

Lemma 2 gives us the means to compute efficiently the matrices $\mathbf{W}_k$, $k = 1, 2, \ldots, n$. We display the pseudocode for Warshall's algorithm, using Lemma 2, as Algorithm 2.

Links

Courtesy of Stephen Warshall

STEPHEN WARSHALL (1935–2006) Stephen Warshall, born in New York City, went to public school in Brooklyn. He attended Harvard University, receiving his degree in mathematics in 1956. He never received an advanced degree, because at that time no programs were available in his areas of interest. However, he took graduate courses at several different universities and contributed to the development of computer science and software engineering.

After graduating from Harvard, Warshall worked at ORO (Operation Research Office), which was set up by Johns Hopkins to do research and development for the U.S. Army. In 1958 he left ORO to take a position at a company called Technical Operations, where he helped build a research and development laboratory for military software projects. In 1961 he left Technical Operations to found Massachusetts Computer Associates. Later, this company became part of Applied Data Research (ADR). After the merger, Warshall sat on the board of directors of ADR and managed a variety of projects and organizations. He retired from ADR in 1982.

During his career Warshall carried out research and development in operating systems, compiler design, language design, and operations research. In the 1971–1972 academic year he presented lectures on software engineering at French universities. There is an interesting anecdote about his proof that the transitive closure algorithm, now known as Warshall's algorithm, is correct. He and a colleague at Technical Operations bet a bottle of rum on who first could determine whether this algorithm always works. Warshall came up with his proof overnight, winning the bet and the rum, which he shared with the loser of the bet. Because Warshall did not like sitting at a desk, he did much of his creative work in unconventional places, such as on a sailboat in the Indian Ocean or in a Greek lemon orchard.

ALGORITHM 2 Warshall Algorithm.

```
procedure Warshall ($\mathbf{M}_R$ : $n \times n$ zero–one matrix)
$\mathbf{W} := \mathbf{M}_R$
for $k := 1$ to $n$
    for $i := 1$ to $n$
        for $j := 1$ to $n$
            $w_{ij} := w_{ij} \vee (w_{ik} \wedge w_{kj})$
return $\mathbf{W}$ {$\mathbf{W} = [w_{ij}]$ is $\mathbf{M}_{R^*}$}
```

The computational complexity of Warshall's algorithm can easily be computed in terms of bit operations. To find the entry $w_{ij}^{[k]}$ from the entries $w_{ij}^{[k-1]}$, $w_{ik}^{[k-1]}$, and $w_{kj}^{[k-1]}$ using Lemma 2 requires two bit operations. To find all n^2 entries of $\mathbf{W}_k$ from those of $\mathbf{W}_{k-1}$ requires $2n^2$ bit operations. Because Warshall's algorithm begins with $\mathbf{W}_0 = \mathbf{M}_R$ and computes the sequence of n zero–one matrices $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n = \mathbf{M}_{R^*}$, the total number of bit operations used is $n \cdot 2n^2 = 2n^3$.

Exercises

1. Let R be the relation on the set $\{0, 1, 2, 3\}$ containing the ordered pairs $(0, 1)$, $(1, 1)$, $(1, 2)$, $(2, 0)$, $(2, 2)$, and $(3, 0)$. Find the

a) reflexive closure of R. **b)** symmetric closure of R.

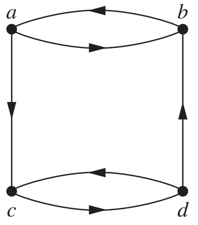
2. Let R be the relation $\{(a, b) \mid a \neq b\}$ on the set of integers. What is the reflexive closure of R?

3. Let R be the relation $\{(a, b) \mid a$ divides $b\}$ on the set of integers. What is the symmetric closure of R?

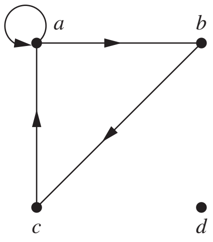
4. How can the directed graph representing the reflexive closure of a relation on a finite set be constructed from the directed graph of the relation?

In Exercises 5–7 draw the directed graph of the reflexive closure of the relations with the directed graph shown.

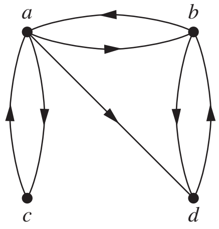
5.



6.



7.



8. How can the directed graph representing the symmetric closure of a relation on a finite set be constructed from the directed graph for this relation?

9. Find the directed graphs of the symmetric closures of the relations with directed graphs shown in Exercises 5–7.

10. Find the smallest relation containing the relation in Example 2 that is both reflexive and symmetric.

11. Find the directed graph of the smallest relation that is both reflexive and symmetric that contains each of the relations with directed graphs shown in Exercises 5–7.

12. Suppose that the relation R on the finite set A is represented by the matrix $\mathbf{M}_R$. Show that the matrix that represents the reflexive closure of R is $\mathbf{M}_R \vee \mathbf{I}_n$.

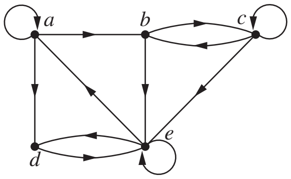
13. Suppose that the relation R on the finite set A is represented by the matrix $\mathbf{M}_R$. Show that the matrix that represents the symmetric closure of R is $\mathbf{M}_R \vee \mathbf{M}_R^t$.

14. Show that the closure of a relation R with respect to a property **P**, if it exists, is the intersection of all the relations with property **P** that contain R.

15. When is it possible to define the "irreflexive closure" of a relation R, that is, a relation that contains R, is irreflexive, and is contained in every irreflexive relation that contains R?

16. Determine whether these sequences of vertices are paths in this directed graph.

a) a, b, c, e
b) b, e, c, b, e
c) a, a, b, e, d, e
d) b, c, e, d, a, a, b
e) b, c, c, b, e, d, e, d
f) $a, a, b, b, c, c, b, e, d$



17. Find all circuits of length three in the directed graph in Exercise 16.

18. Determine whether there is a path in the directed graph in Exercise 16 beginning at the first vertex given and ending at the second vertex given.

a) a, b **b)** b, a **c)** b, b
d) a, e **e)** b, d **f)** c, d
g) d, d **h)** e, a **i)** e, c

19. Let R be the relation on the set $\{1, 2, 3, 4, 5\}$ containing the ordered pairs (1, 3), (2, 4), (3, 1), (3, 5), (4, 3), (5, 1), (5, 2), and (5, 4). Find

a) R^2. **b)** R^3. **c)** R^4.
d) R^5. **e)** R^6. **f)** R^*.

20. Let R be the relation that contains the pair (a, b) if a and b are cities such that there is a direct nonstop airline flight from a to b. When is (a, b) in

a) R^2? **b)** R^3? **c)** R^*?

21. Let R be the relation on the set of all students containing the ordered pair (a, b) if a and b are in at least one common class and $a \neq b$. When is (a, b) in

a) R^2? **b)** R^3? **c)** R^*?

22. Suppose that the relation R is reflexive. Show that R^* is reflexive.

23. Suppose that the relation R is symmetric. Show that R^* is symmetric.

24. Suppose that the relation R is irreflexive. Is the relation R^2 necessarily irreflexive?

25. Use Algorithm 1 to find the transitive closures of these relations on $\{1, 2, 3, 4\}$.

a) $\{(1, 2), (2, 1), (2, 3), (3, 4), (4, 1)\}$
b) $\{(2, 1), (2, 3), (3, 1), (3, 4), (4, 1), (4, 3)\}$
c) $\{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$
d) $\{(1, 1), (1, 4), (2, 1), (2, 3), (3, 1), (3, 2), (3, 4), (4, 2)\}$

26. Use Algorithm 1 to find the transitive closures of these relations on $\{a, b, c, d, e\}$.

a) $\{(a, c), (b, d), (c, a), (d, b), (e, d)\}$
b) $\{(b, c), (b, e), (c, e), (d, a), (e, b), (e, c)\}$
c) $\{(a, b), (a, c), (a, e), (b, a), (b, c), (c, a), (c, b), (d, a), (e, d)\}$
d) $\{(a, e), (b, a), (b, d), (c, d), (d, a), (d, c), (e, a), (e, b), (e, c), (e, e)\}$

27. Use Warshall's algorithm to find the transitive closures of the relations in Exercise 25.

28. Use Warshall's algorithm to find the transitive closures of the relations in Exercise 26.

29. Find the smallest relation containing the relation $\{(1, 2), (1, 4), (3, 3), (4, 1)\}$ that is

a) reflexive and transitive.
b) symmetric and transitive.
c) reflexive, symmetric, and transitive.

30. Finish the proof of the case when $a \neq b$ in Lemma 1.

31. Algorithms have been devised that use $O(n^{2.8})$ bit operations to compute the Boolean product of two $n \times n$ zero–one matrices. Assuming that these algorithms can be used, give big-O estimates for the number of bit operations using Algorithm 1 and using Warshall's algorithm to find the transitive closure of a relation on a set with n elements.

***32.** Devise an algorithm using the concept of interior vertices in a path to find the length of the shortest path between two vertices in a directed graph, if such a path exists.

33. Adapt Algorithm 1 to find the reflexive closure of the transitive closure of a relation on a set with n elements.

34. Adapt Warshall's algorithm to find the reflexive closure of the transitive closure of a relation on a set with n elements.

35. Show that the closure with respect to the property **P** of the relation $R = \{(0, 0), (0, 1), (1, 1), (2, 2)\}$ on the set $\{0, 1, 2\}$ does not exist if **P** is the property

a) "is not reflexive."
b) "has an odd number of elements."

36. Give an example of a relation R on the set $\{a, b, c\}$ such that the symmetric closure of the reflexive closure of the transitive closure of R is not transitive.

9.5 Equivalence Relations

9.5.1 Introduction

In some programming languages the names of variables can contain an unlimited number of characters. However, there is a limit on the number of characters that are checked when a compiler determines whether two variables are equal. For instance, in traditional C, only the first eight characters of a variable name are checked by the compiler. (These characters are uppercase or lowercase letters, digits, or underscores.) Consequently, the compiler considers strings longer than eight characters that agree in their first eight characters the same. Let R be the relation on the set of strings of characters such that sRt, where s and t are two strings, if s and t are at least eight characters long and the first eight characters of s and t agree, or $s = t$. It is easy to see that R is reflexive, symmetric, and transitive. Moreover, R divides the set of all strings into classes, where all strings in a particular class are considered the same by a compiler for traditional C.

The integers a and b are related by the "congruence modulo 4" relation when 4 divides $a - b$. We will show later that this relation is reflexive, symmetric, and transitive. It is not hard to see that a is related to b if and only if a and b have the same remainder when divided by 4. It follows that this relation splits the set of integers into four different classes.

When we care only what remainder an integer leaves when it is divided by 4, we need only know which class it is in, not its particular value.

These two relations, R and congruence modulo 4, are examples of equivalence relations, namely, relations that are reflexive, symmetric, and transitive. In this section we will show that such relations split sets into disjoint classes of equivalent elements. Equivalence relations arise whenever we care only whether an element of a set is in a certain class of elements, instead of caring about its particular identity.

9.5.2 Equivalence Relations

Links ❯

In this section we will study relations with a particular combination of properties that allows them to be used to relate objects that are similar in some way.

Definition 1

A relation on a set A is called an *equivalence relation* if it is reflexive, symmetric, and transitive.

Equivalence relations are important in every branch of mathematics!

Equivalence relations are important throughout mathematics and computer science. One reason for this is that in an equivalence relation, when two elements are related it makes sense to say they are equivalent.

Definition 2

Two elements a and b that are related by an equivalence relation are called *equivalent*. The notation $a \sim b$ is often used to denote that a and b are equivalent elements with respect to a particular equivalence relation.

For the notion of equivalent elements to make sense, every element should be equivalent to itself, as the reflexive property guarantees for an equivalence relation. It makes sense to say that a and b are related (not just that a is related to b) by an equivalence relation, because when a is related to b, by the symmetric property, b is related to a. Furthermore, because an equivalence relation is transitive, if a and b are equivalent and b and c are equivalent, it follows that a and c are equivalent.

Examples 1–5 illustrate the notion of an equivalence relation.

EXAMPLE 1 Let R be the relation on the set of integers such that aRb if and only if $a = b$ or $a = -b$. In Section 9.1 we showed that R is reflexive, symmetric, and transitive. It follows that R is an equivalence relation. ◀

EXAMPLE 2 Let R be the relation on the set of real numbers such that aRb if and only if $a - b$ is an integer. Is R an equivalence relation?

Extra Examples ❯

Solution: Because $a - a = 0$ is an integer for all real numbers a, aRa for all real numbers a. Hence, R is reflexive. Now suppose that aRb. Then $a - b$ is an integer, so $b - a$ is also an integer. Hence, bRa. It follows that R is symmetric. If aRb and bRc, then $a - b$ and $b - c$ are integers. Therefore, $a - c = (a - b) + (b - c)$ is also an integer. Hence, aRc. Thus, R is transitive. Consequently, R is an equivalence relation. ◀

One of the most widely used equivalence relations is congruence modulo m, where m is an integer greater than 1.

EXAMPLE 3 **Congruence Modulo** m Let m be an integer with $m > 1$. Show that the relation

$$R = \{(a, b) \mid a \equiv b \pmod{m}\}$$

is an equivalence relation on the set of integers.

Solution: Recall from Section 4.1 that $a \equiv b \pmod{m}$ if and only if m divides $a - b$. Note that $a - a = 0$ is divisible by m, because $0 = 0 \cdot m$. Hence, $a \equiv a \pmod{m}$, so congruence modulo m is reflexive. Now suppose that $a \equiv b \pmod{m}$. Then $a - b$ is divisible by m, so $a - b = km$, where k is an integer. It follows that $b - a = (-k)m$, so $b \equiv a \pmod{m}$. Hence, congruence modulo m is symmetric. Next, suppose that $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. Then m divides both $a - b$ and $b - c$. Therefore, there are integers k and l with $a - b = km$ and $b - c = lm$. Adding these two equations shows that $a - c = (a - b) + (b - c) = km + lm = (k + l)m$. Thus, $a \equiv c \pmod{m}$. Therefore, congruence modulo m is transitive. It follows that congruence modulo m is an equivalence relation. ◀

EXAMPLE 4 Suppose that R is the relation on the set of strings of English letters such that aRb if and only if $l(a) = l(b)$, where $l(x)$ is the length of the string x. Is R an equivalence relation?

Solution: Because $l(a) = l(a)$, it follows that aRa whenever a is a string, so that R is reflexive. Next, suppose that aRb, so that $l(a) = l(b)$. Then bRa, because $l(b) = l(a)$. Hence, R is symmetric. Finally, suppose that aRb and bRc. Then $l(a) = l(b)$ and $l(b) = l(c)$. Hence, $l(a) = l(c)$, so aRc. Consequently, R is transitive. Because R is reflexive, symmetric, and transitive, it is an equivalence relation. ◀

EXAMPLE 5 Let n be a positive integer and S a set of strings. Suppose that R_n is the relation on S such that sR_nt if and only if $s = t$, or both s and t have at least n characters and the first n characters of s and t are the same. That is, a string of fewer than n characters is related only to itself; a string s with at least n characters is related to a string t if and only if t has at least n characters and t begins with the n characters at the start of s. For example, let $n = 3$ and let S be the set of all bit strings. Then sR_3t either when $s = t$ or both s and t are bit strings of length 3 or more that begin with the same three bits. For instance, $01R_3 01$ and $00111R_3 00101$, but $01\not{R}_3 010$ and $01011\not{R}_3 01110$.

Show that for every set S of strings and every positive integer n, R_n is an equivalence relation on S.

Solution: The relation R_n is reflexive because $s = s$, so that sR_ns whenever s is a string in S. If sR_nt, then either $s = t$ or s and t are both at least n characters long that begin with the same n characters. This means that tR_ns. We conclude that R_n is symmetric.

Now suppose that sR_nt and tR_nu. Then either $s = t$ or s and t are at least n characters long and s and t begin with the same n characters, and either $t = u$ or t and u are at least n characters long and t and u begin with the same n characters. From this, we can deduce that either $s = u$ or both s and u are n characters long and s and u begin with the same n characters (because in this case we know that s, t, and u are all at least n characters long and both s and u begin with the same n characters as t does). Consequently, R_n is transitive. It follows that R_n is an equivalence relation. ◀

In Examples 6 and 7 we look at two relations that are not equivalence relations.

EXAMPLE 6 Show that the "divides" relation on the set of positive integers in not an equivalence relation.

Solution: By Examples 9 and 15 in Section 9.1, we know that the "divides" relation is reflexive and transitive. However, by Example 12 in Section 9.1, we know that this relation is not symmetric (for instance, $2 \mid 4$ but $4 \nmid 2$). We conclude that the "divides" relation on the set of positive integers is not an equivalence relation. ◀

EXAMPLE 7 Let R be the relation on the set of real numbers such that xRy if and only if x and y are real numbers that differ by less than 1, that is, $|x - y| < 1$. Show that R is not an equivalence relation.

Solution: R is reflexive because $|x - x| = 0 < 1$ whenever $x \in \mathbf{R}$. R is symmetric, for if xRy, where x and y are real numbers, then $|x - y| < 1$, which tells us that $|y - x| = |x - y| < 1$, so that yRx. However, R is not an equivalence relation because it is not transitive. Take $x = 2.8$, $y = 1.9$, and $z = 1.1$, so that $|x - y| = |2.8 - 1.9| = 0.9 < 1$, $|y - z| = |1.9 - 1.1| = 0.8 < 1$, but $|x - z| = |2.8 - 1.1| = 1.7 > 1$. That is, $2.8\,R\,1.9$, $1.9\,R\,1.1$, but $2.8 \not R\, 1.1$. ◀

9.5.3 Equivalence Classes

Let A be the set of all students in your school who graduated from high school. Consider the relation R on A that consists of all pairs (x, y), where x and y graduated from the same high school. Given a student x, we can form the set of all students equivalent to x with respect to R. This set consists of all students who graduated from the same high school as x did. This subset of A is called an equivalence class of the relation.

Definition 3 Let R be an equivalence relation on a set A. The set of all elements that are related to an element a of A is called the *equivalence class* of a. The equivalence class of a with respect to R is denoted by $[a]_R$. When only one relation is under consideration, we can delete the subscript R and write $[a]$ for this equivalence class.

In other words, if R is an equivalence relation on a set A, the equivalence class of the element a is

$$[a]_R = \{s \mid (a, s) \in R\}.$$

If $b \in [a]_R$, then b is called a **representative** of this equivalence class. Any element of a class can be used as a representative of this class. That is, there is nothing special about the particular element chosen as the representative of the class.

EXAMPLE 8 What is the equivalence class of an integer for the equivalence relation of Example 1?

Solution: Because an integer is equivalent to itself and its negative in this equivalence relation, it follows that $[a] = \{-a, a\}$. This set contains two distinct integers unless $a = 0$. For instance, $[7] = \{-7, 7\}$, $[-5] = \{-5, 5\}$, and $[0] = \{0\}$. ◀

EXAMPLE 9 What are the equivalence classes of 0, 1, 2, and 3 for congruence modulo 4?

Extra Examples ❯

Solution: The equivalence class of 0 contains all integers a such that $a \equiv 0 \pmod 4$. The integers in this class are those divisible by 4. Hence, the equivalence class of 0 for this relation is

$$[0] = \{\ldots, -8, -4, 0, 4, 8, \ldots\}.$$

The equivalence class of 1 contains all the integers a such that $a \equiv 1 \pmod 4$. The integers in this class are those that have a remainder of 1 when divided by 4. Hence, the equivalence class of 1 for this relation is

$$[1] = \{\dots, -7, -3, 1, 5, 9, \dots\}.$$

The equivalence class of 2 contains all the integers a such that $a \equiv 2 \pmod 4$. The integers in this class are those that have a remainder of 2 when divided by 4. Hence, the equivalence class of 2 for this relation is

$$[2] = \{\dots, -6, -2, 2, 6, 10, \dots\}.$$

The equivalence class of 3 contains all the integers a such that $a \equiv 3 \pmod 4$. The integers in this class are those that have a remainder of 3 when divided by 4. Hence, the equivalence class of 3 for this relation is

$$[3] = \{\dots, -5, -1, 3, 7, 11, \dots\}.$$

Note that every integer is in exactly one of the four equivalence classes and that the integer n is in the class containing n **mod** 4. ◀

In Example 9 the equivalence classes of 0, 1, 2, and 3 with respect to congruence modulo 4 were found. Example 9 can easily be generalized, replacing 4 with any positive integer m. The equivalence classes of the relation congruence modulo m are called the **congruence classes modulo** m. The congruence class of an integer a modulo m is denoted by $[a]_m$, so $[a]_m = \{\dots, a - 2m, a - m, a, a + m, a + 2m, \dots\}$. For instance, from Example 9 we have $[0]_4 = \{\dots, -8, -4, 0, 4, 8, \dots\}$, $[1]_4 = \{\dots, -7, -3, 1, 5, 9, \dots\}$, $[2]_4 = \{\dots, -6, -2, 2, 6, 10, \dots\}$, and $[3]_4 = \{\dots, -5, -1, 3, 7, 11, \dots\}$.

EXAMPLE 10 What is the equivalence class of the string 0111 with respect to the equivalence relation R_3 from Example 5 on the set of all bit strings? (Recall that sR_3t if and only if s and t are bit strings with $s = t$ or s and t are strings of at least three bits that start with the same three bits.)

Solution: The bit strings equivalent to 0111 are the bit strings with at least three bits that begin with 011. These are the bit strings 011, 0110, 0111, 01100, 01101, 01110, 01111, and so on. Consequently,

$$[011]_{R_3} = \{011, 0110, 0111, 01100, 01101, 01110, 01111, \dots\}.$$ ◀

EXAMPLE 11 **Identifiers in the C Programming Language** In the C programming language, an **identifier** is the name of a variable, a function, or another type of entity. Each identifier is a nonempty string of characters where each character is a lowercase or an uppercase English letter, a digit, or an underscore, and the first character is a lowercase or an uppercase English letter. Identifiers can be any length. This allows developers to use as many characters as they want to name an entity, such as a variable. However, for compilers for some versions of C, there is a limit on the number of characters checked when two names are compared to see whether they refer to the same thing. For example, Standard C compilers consider two identifiers the same when they agree in their first 31 characters. Consequently, developers must be careful not to use identifiers with the same initial 31 characters for different things. We see that two identifiers are considered the same when they are related by the relation R_{31} in Example 5. Using Example 5, we know that R_{31}, on the set of all identifiers in Standard C, is an equivalence relation.

What are the equivalence classes of each of the identifiers Number_of_tropical_storms, Number_of_named_tropical_storms, and Number_of_named_tropical_storms_in_the_Atlantic_in_2017?

Solution: Note that when an identifier is less than 31 characters long, by the definition of R_{31}, its equivalence class contains only itself. Because the identifier Number_of_tropical_storms is 25 characters long, its equivalence class contains exactly one element, namely, itself.

The identifier Number_of_named_tropical_storms is exactly 31 characters long. An identifier is equivalent to it when it starts with these same 31 characters. Consequently, every identifier at least 31 characters long that starts with Number_of_named_tropical_storms is equivalent to this identifier. It follows that the equivalence class of Number_of_named_tropical_storms is the set of all identifiers that begin with the 31 characters Number_of_named_tropical_storms.

An identifier is equivalent to the Number_of_named_tropical_storms_in_the_Atlantic_in_2017 if and only if it begins with its first 31 characters. Because these characters are Number_of_named_tropical_storms, we see that an identifier is equivalent to Number_of_named_tropical_storms_in_the_Atlantic_in_2017 if and only if it is equivalent to Number_of_named_tropical_storms. It follows that these last two identifiers have the same equivalence class. ◀

9.5.4 Equivalence Classes and Partitions

Let A be the set of students at your school who are majoring in exactly one subject, and let R be the relation on A consisting of pairs (x, y), where x and y are students with the same major. Then R is an equivalence relation, as the reader should verify. We can see that R splits all students in A into a collection of disjoint subsets, where each subset contains students with a specified major. For instance, one subset contains all students majoring (just) in computer science, and a second subset contains all students majoring in history. Furthermore, these subsets are equivalence classes of R. This example illustrates how the equivalence classes of an equivalence relation partition a set into disjoint, nonempty subsets. We will make these notions more precise in the following discussion.

Let R be a relation on the set A. Theorem 1 shows that the equivalence classes of two elements of A are either identical or disjoint.

THEOREM 1 Let R be an equivalence relation on a set A. These statements for elements a and b of A are equivalent:

(i) aRb $\qquad$ (ii) $[a] = [b]$ $\qquad$ (iii) $[a] \cap [b] \neq \emptyset$

Proof: We first show that (i) implies (ii). Assume that aRb. We will prove that $[a] = [b]$ by showing $[a] \subseteq [b]$ and $[b] \subseteq [a]$. Suppose $c \in [a]$. Then aRc. Because aRb and R is symmetric, we know that bRa. Furthermore, because R is transitive and bRa and aRc, it follows that bRc. Hence, $c \in [b]$. This shows that $[a] \subseteq [b]$. The proof that $[b] \subseteq [a]$ is similar; it is left as an exercise for the reader.

Second, we will show that (ii) implies (iii). Assume that $[a] = [b]$. It follows that $[a] \cap [b] \neq \emptyset$ because $[a]$ is nonempty (because $a \in [a]$ because R is reflexive).

Next, we will show that (iii) implies (i). Suppose that $[a] \cap [b] \neq \emptyset$. Then there is an element c with $c \in [a]$ and $c \in [b]$. In other words, aRc and bRc. By the symmetric property, cRb. Then by transitivity, because aRc and cRb, we have aRb.

Because (i) implies (ii), (ii) implies (iii), and (iii) implies (i), the three statements, (i), (ii), and (iii), are equivalent. ◁

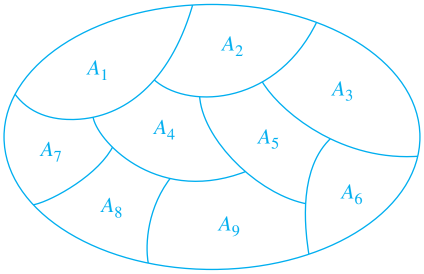


FIGURE 1 **A partition of a set.**

We are now in a position to show how an equivalence relation *partitions* a set. Let R be an equivalence relation on a set A. The union of the equivalence classes of R is all of A, because an element a of A is in its own equivalence class, namely, $[a]_R$. In other words,

$$\bigcup_{a\in A}[a]_R = A.$$

In addition, from Theorem 1, it follows that these equivalence classes are either equal or disjoint, so

$$[a]_R \cap [b]_R = \emptyset,$$

when $[a]_R \neq [b]_R$.

Recall that an *index set* is a set whose members label, or index, the elements of a set.

These two observations show that the equivalence classes form a partition of A, because they split A into disjoint subsets. More precisely, a **partition** of a set S is a collection of disjoint nonempty subsets of S that have S as their union. In other words, the collection of subsets A_i, $i \in I$ (where I is an index set) forms a partition of S if and only if

$$A_i \neq \emptyset \text{ for } i \in I,$$

$$A_i \cap A_j = \emptyset \text{ when } i \neq j,$$

and

$$\bigcup_{i\in I} A_i = S.$$

(Here the notation $\bigcup_{i\in I} A_i$ represents the union of the sets A_i for all $i \in I$.) Figure 1 illustrates the concept of a partition of a set.

EXAMPLE 12 Suppose that $S = \{1, 2, 3, 4, 5, 6\}$. The collection of sets $A_1 = \{1, 2, 3\}$, $A_2 = \{4, 5\}$, and $A_3 = \{6\}$ forms a partition of S, because these sets are disjoint and their union is S. ◀

We have seen that the equivalence classes of an equivalence relation on a set form a partition of the set. The subsets of S in this partition are the equivalence classes. Conversely, every partition of a set can be used to form an equivalence relation. Two elements are equivalent with respect to this relation if and only if they are in the same subset of S in the partition.

To see this, assume that $\{A_i \mid i \in I\}$ is a partition on S. Let R be the relation on S consisting of the pairs (x, y), where x and y belong to the same subset A_i in the partition. To show that R is an equivalence relation we must show that R is reflexive, symmetric, and transitive.

We see that $(a, a) \in R$ for every $a \in S$, because a is in the same subset of S as itself. Hence, R is reflexive. If $(a, b) \in R$, then b and a are in the same subset of S in the partition, so that

$(b, a) \in R$ as well. Hence, R is symmetric. If $(a, b) \in R$ and $(b, c) \in R$, then a and b are in the same subset X of S in the partition, and b and c are in the same subset Y of S of the partition. Because the subsets of S in the partition are disjoint and b belongs to X and Y, it follows that $X = Y$. Consequently, a and c belong to the same subset of S in the partition, so $(a, c) \in R$. Thus, R is transitive.

It follows that R is an equivalence relation. The equivalence classes of R consist of subsets of S containing related elements, and by the definition of R, these are the subsets of S in the partition. Theorem 2 summarizes the connections we have established between equivalence relations and partitions.

THEOREM 2 Let R be an equivalence relation on a set S. Then the equivalence classes of R form a partition of S. Conversely, given a partition $\{A_i \mid i \in I\}$ of the set S, there is an equivalence relation R that has the sets A_i, $i \in I$, as its equivalence classes.

Example 13 shows how to construct an equivalence relation from a partition.

EXAMPLE 13 List the ordered pairs in the equivalence relation R produced by the partition $A_1 = \{1, 2, 3\}$, $A_2 = \{4, 5\}$, and $A_3 = \{6\}$ of $S = \{1, 2, 3, 4, 5, 6\}$, given in Example 12.

Solution: The subsets of S in the partition are the equivalence classes of R. The pair $(a, b) \in R$ if and only if a and b are in the same subset of the S in the partition. The pairs (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), and (3, 3) belong to R because $A_1 = \{1, 2, 3\}$ is an equivalence class; the pairs (4, 4), (4, 5), (5, 4), and (5, 5) belong to R because $A_2 = \{4, 5\}$ is an equivalence class; and finally the pair (6, 6) belongs to R because $\{6\}$ is an equivalence class. No pair other than those listed belongs to R. ◀

The congruence classes modulo m provide a useful illustration of Theorem 2. There are m different congruence classes modulo m, corresponding to the m different remainders possible when an integer is divided by m. These m congruence classes are denoted by $[0]_m, [1]_m, \ldots, [m-1]_m$. They form a partition of the set of integers.

EXAMPLE 14 What are the sets in the partition of the integers arising from congruence modulo 4?

Solution: In Example 9 we found the four congruence classes, $[0]_4$, $[1]_4$, $[2]_4$, and $[3]_4$. They are the sets

$$
\begin{aligned}
[0]_4 &= \{\ldots, -8, -4, 0, 4, 8, \ldots\},\\
[1]_4 &= \{\ldots, -7, -3, 1, 5, 9, \ldots\},\\
[2]_4 &= \{\ldots, -6, -2, 2, 6, 10, \ldots\},\\
[3]_4 &= \{\ldots, -5, -1, 3, 7, 11, \ldots\}.
\end{aligned}
$$

These congruence classes are disjoint, and every integer is in exactly one of them. In other words, as Theorem 2 says, these congruence classes form a partition. ◀

We now provide an example of a partition of the set of all strings arising from an equivalence relation on this set.

EXAMPLE 15 Let R_3 be the relation from Example 5. What are the sets in the partition of the set of all bit strings arising from the relation R_3 on the set of all bit strings? (Recall that sR_3t, where s and t are bit strings, if $s = t$ or s and t are bit strings with at least three bits that agree in their first three bits.)

Solution: Note that every bit string of length less than three is equivalent only to itself. Hence $[\lambda]_{R_3} = \{\lambda\}$, $[0]_{R_3} = \{0\}$, $[1]_{R_3} = \{1\}$, $[00]_{R_3} = \{00\}$, $[01]_{R_3} = \{01\}$, $[10]_{R_3} = \{10\}$, and $[11]_{R_3} = \{11\}$. Note that every bit string of length three or more is equivalent to one of the eight bit strings 000, 001, 010, 011, 100, 101, 110, and 111. We have

$$[000]_{R_3} = \{000, 0000, 0001, 00000, 00001, 00010, 00011, \ldots\},$$

$$[001]_{R_3} = \{001, 0010, 0011, 00100, 00101, 00110, 00111, \ldots\},$$

$$[010]_{R_3} = \{010, 0100, 0101, 01000, 01001, 01010, 01011, \ldots\},$$

$$[011]_{R_3} = \{011, 0110, 0111, 01100, 01101, 01110, 01111, \ldots\},$$

$$[100]_{R_3} = \{100, 1000, 1001, 10000, 10001, 10010, 10011, \ldots\},$$

$$[101]_{R_3} = \{101, 1010, 1011, 10100, 10101, 10110, 10111, \ldots\},$$

$$[110]_{R_3} = \{110, 1100, 1101, 11000, 11001, 11010, 11011, \ldots\},$$

$$[111]_{R_3} = \{111, 1110, 1111, 11100, 11101, 11110, 11111, \ldots\}.$$

These 15 equivalence classes are disjoint and every bit string is in exactly one of them. As Theorem 2 tells us, these equivalence classes partition the set of all bit strings. ◀

Exercises

1. Which of these relations on $\{0, 1, 2, 3\}$ are equivalence relations? Determine the properties of an equivalence relation that the others lack.

a) $\{(0, 0), (1, 1), (2, 2), (3, 3)\}$

b) $\{(0, 0), (0, 2), (2, 0), (2, 2), (2, 3), (3, 2), (3, 3)\}$

c) $\{(0, 0), (1, 1), (1, 2), (2, 1), (2, 2), (3, 3)\}$

d) $\{(0, 0), (1, 1), (1, 3), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$

e) $\{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 2), (3, 3)\}$

2. Which of these relations on the set of all people are equivalence relations? Determine the properties of an equivalence relation that the others lack.

a) $\{(a, b) \mid a$ and b are the same age$\}$

b) $\{(a, b) \mid a$ and b have the same parents$\}$

c) $\{(a, b) \mid a$ and b share a common parent$\}$

d) $\{(a, b) \mid a$ and b have met$\}$

e) $\{(a, b) \mid a$ and b speak a common language$\}$

3. Which of these relations on the set of all functions from **Z** to **Z** are equivalence relations? Determine the properties of an equivalence relation that the others lack.

a) $\{(f, g) \mid f(1) = g(1)\}$

b) $\{(f, g) \mid f(0) = g(0)$ or $f(1) = g(1)\}$

c) $\{(f, g) \mid f(x) - g(x) = 1$ for all $x \in \mathbf{Z}\}$

d) $\{(f, g) \mid$ for some $C \in \mathbf{Z}$, for all $x \in \mathbf{Z}, f(x) - g(x) = C\}$

e) $\{(f, g) \mid f(0) = g(1)$ and $f(1) = g(0)\}$

4. Define three equivalence relations on the set of students in your discrete mathematics class different from the relations discussed in the text. Determine the equivalence classes for each of these equivalence relations.

5. Define three equivalence relations on the set of buildings on a college campus. Determine the equivalence classes for each of these equivalence relations.

6. Define three equivalence relations on the set of classes offered at your school. Determine the equivalence classes for each of these equivalence relations.

7. Show that the relation of logical equivalence on the set of all compound propositions is an equivalence relation. What are the equivalence classes of **F** and of **T**?

8. Let R be the relation on the set of all sets of real numbers such that $S\,R\,T$ if and only if S and T have the same cardinality. Show that R is an equivalence relation. What are the equivalence classes of the sets $\{0, 1, 2\}$ and **Z**?

9. Suppose that A is a nonempty set, and f is a function that has A as its domain. Let R be the relation on A consisting of all ordered pairs (x, y) such that $f(x) = f(y)$.

a) Show that R is an equivalence relation on A.

b) What are the equivalence classes of R?

10. Suppose that A is a nonempty set and R is an equivalence relation on A. Show that there is a function f with A as its domain such that $(x, y) \in R$ if and only if $f(x) = f(y)$.

11. Show that the relation R consisting of all pairs (x, y) such that x and y are bit strings of length three or more that agree in their first three bits is an equivalence relation on the set of all bit strings of length three or more.

12. Show that the relation R consisting of all pairs (x, y) such that x and y are bit strings of length three or more that agree except perhaps in their first three bits is an equivalence relation on the set of all bit strings of length three or more.

13. Show that the relation R consisting of all pairs (x, y) such that x and y are bit strings that agree in their first and third bits is an equivalence relation on the set of all bit strings of length three or more.

14. Let R be the relation consisting of all pairs (x, y) such that x and y are strings of uppercase and lowercase English letters with the property that for every positive integer n, the nth characters in x and y are the same letter, either uppercase or lowercase. Show that R is an equivalence relation.

15. Let R be the relation on the set of ordered pairs of positive integers such that $((a, b), (c, d)) \in R$ if and only if $a + d = b + c$. Show that R is an equivalence relation.

16. Let R be the relation on the set of ordered pairs of positive integers such that $((a, b), (c, d)) \in R$ if and only if $ad = bc$. Show that R is an equivalence relation.

17. (*Requires calculus*)

a) Show that the relation R on the set of all differentiable functions from $\mathbf{R}$ to $\mathbf{R}$ consisting of all pairs (f, g) such that $f'(x) = g'(x)$ for all real numbers x is an equivalence relation.

b) Which functions are in the same equivalence class as the function $f(x) = x^2$?

18. (*Requires calculus*)

a) Let n be a positive integer. Show that the relation R on the set of all polynomials with real-valued coefficients consisting of all pairs (f, g) such that $f^{(n)}(x) = g^{(n)}(x)$ is an equivalence relation. [Here $f^{(n)}(x)$ is the nth derivative of $f(x)$.]

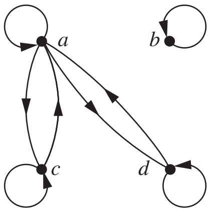
b) Which functions are in the same equivalence class as the function $f(x) = x^4$, where $n = 3$?

19. Let R be the relation on the set of all URLs (or Web addresses) such that $x\,R\,y$ if and only if the Web page at x is the same as the Web page at y. Show that R is an equivalence relation.

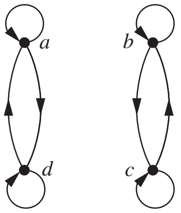
20. Let R be the relation on the set of all people who have visited a particular Web page such that $x\,R\,y$ if and only if person x and person y have followed the same set of links starting at this Web page (going from Web page to Web page until they stop using the Web). Show that R is an equivalence relation.

In Exercises 21–23 determine whether the relation with the directed graph shown is an equivalence relation.

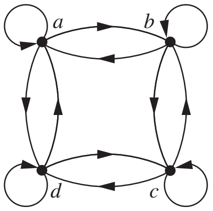
21.



22.



23.



24. Determine whether the relations represented by these zero–one matrices are equivalence relations.

a) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ **c)** $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

25. Show that the relation R on the set of all bit strings such that $s\,R\,t$ if and only if s and t contain the same number of 1s is an equivalence relation.

26. What are the equivalence classes of the equivalence relations in Exercise 1?

27. What are the equivalence classes of the equivalence relations in Exercise 2?

28. What are the equivalence classes of the equivalence relations in Exercise 3?

29. What is the equivalence class of the bit string 011 for the equivalence relation in Exercise 25?

30. What are the equivalence classes of these bit strings for the equivalence relation in Exercise 11?

a) 010 **b)** 1011 **c)** 11111 **d)** 01010101

31. What are the equivalence classes of the bit strings in Exercise 30 for the equivalence relation from Exercise 12?

32. What are the equivalence classes of the bit strings in Exercise 30 for the equivalence relation from Exercise 13?

33. What are the equivalence classes of the bit strings in Exercise 30 for the equivalence relation R_4 from Example 5 on the set of all bit strings? (Recall that bit strings s and t are equivalent under R_4 if and only if they are equal or they are both at least four bits long and agree in their first four bits.)

34. What are the equivalence classes of the bit strings in Exercise 30 for the equivalence relation R_5 from Example 5 on the set of all bit strings? (Recall that bit strings s and t are equivalent under R_5 if and only if they are equal or they are both at least five bits long and agree in their first five bits.)

35. What is the congruence class $[n]_5$ (that is, the equivalence class of n with respect to congruence modulo 5) when n is

a) 2? **b)** 3? **c)** 6? **d)** −3?

36. What is the congruence class $[4]_m$ when m is

a) 2? **b)** 3? **c)** 6? **d)** 8?

37. Give a description of each of the congruence classes modulo 6.

38. What is the equivalence class of each of these strings with respect to the equivalence relation in Exercise 14?

a) *No* **b)** *Yes* **c)** *Help*

39. a) What is the equivalence class of $(1, 2)$ with respect to the equivalence relation in Exercise 15?

b) Give an interpretation of the equivalence classes for the equivalence relation R in Exercise 15. [*Hint:* Look at the difference $a - b$ corresponding to (a, b).]

40. a) What is the equivalence class of $(1, 2)$ with respect to the equivalence relation in Exercise 16?

b) Give an interpretation of the equivalence classes for the equivalence relation R in Exercise 16. [*Hint:* Look at the ratio a/b corresponding to (a, b).]

41. Which of these collections of subsets are partitions of $\{1, 2, 3, 4, 5, 6\}$?

a) $\{1, 2\}, \{2, 3, 4\}, \{4, 5, 6\}$

b) $\{1\}, \{2, 3, 6\}, \{4\}, \{5\}$

c) $\{2, 4, 6\}, \{1, 3, 5\}$ **d)** $\{1, 4, 5\}, \{2, 6\}$

42. Which of these collections of subsets are partitions of $\{-3, -2, -1, 0, 1, 2, 3\}$?

a) $\{-3, -1, 1, 3\}, \{-2, 0, 2\}$

b) $\{-3, -2, -1, 0\}, \{0, 1, 2, 3\}$

c) $\{-3, 3\}, \{-2, 2\}, \{-1, 1\}, \{0\}$

d) $\{-3, -2, 2, 3\}, \{-1, 1\}$

43. Which of these collections of subsets are partitions of the set of bit strings of length 8?

a) the set of bit strings that begin with 1, the set of bit strings that begin with 00, and the set of bit strings that begin with 01

b) the set of bit strings that contain the string 00, the set of bit strings that contain the string 01, the set of bit strings that contain the string 10, and the set of bit strings that contain the string 11

c) the set of bit strings that end with 00, the set of bit strings that end with 01, the set of bit strings that end with 10, and the set of bit strings that end with 11

d) the set of bit strings that end with 111, the set of bit strings that end with 011, and the set of bit strings that end with 00

e) the set of bit strings that contain $3k$ ones for some nonnegative integer k, the set of bit strings that contain $3k + 1$ ones for some nonnegative integer k, and the set of bit strings that contain $3k + 2$ ones for some nonnegative integer k.

44. Which of these collections of subsets are partitions of the set of integers?

a) the set of even integers and the set of odd integers

b) the set of positive integers and the set of negative integers

c) the set of integers divisible by 3, the set of integers leaving a remainder of 1 when divided by 3, and the set of integers leaving a remainder of 2 when divided by 3

d) the set of integers less than −100, the set of integers with absolute value not exceeding 100, and the set of integers greater than 100

e) the set of integers not divisible by 3, the set of even integers, and the set of integers that leave a remainder of 3 when divided by 6

45. Which of these are partitions of the set $\mathbf{Z} \times \mathbf{Z}$ of ordered pairs of integers?

a) the set of pairs (x, y), where x or y is odd; the set of pairs (x, y), where x is even; and the set of pairs (x, y), where y is even

b) the set of pairs (x, y), where both x and y are odd; the set of pairs (x, y), where exactly one of x and y is odd; and the set of pairs (x, y), where both x and y are even

c) the set of pairs (x, y), where x is positive; the set of pairs (x, y), where y is positive; and the set of pairs (x, y), where both x and y are negative

d) the set of pairs (x, y), where $3 \mid x$ and $3 \mid y$; the set of pairs (x, y), where $3 \mid x$ and $3 \nmid y$; the set of pairs (x, y), where $3 \nmid x$ and $3 \mid y$; and the set of pairs (x, y), where $3 \nmid x$ and $3 \nmid y$

e) the set of pairs (x, y), where $x > 0$ and $y > 0$; the set of pairs (x, y), where $x > 0$ and $y \leq 0$; the set of pairs (x, y), where $x \leq 0$ and $y > 0$; and the set of pairs (x, y), where $x \leq 0$ and $y \leq 0$

f) the set of pairs (x, y), where $x \neq 0$ and $y \neq 0$; the set of pairs (x, y), where $x = 0$ and $y \neq 0$; and the set of pairs (x, y), where $x \neq 0$ and $y = 0$

46. Which of these are partitions of the set of real numbers?

a) the negative real numbers, $\{0\}$, the positive real numbers

b) the set of irrational numbers, the set of rational numbers

c) the set of intervals $[k, k + 1]$, $k = \ldots, -2, -1, 0, 1, 2, \ldots$

d) the set of intervals $(k, k + 1)$, $k = \ldots, -2, -1, 0, 1, 2, \ldots$

e) the set of intervals $(k, k + 1]$, $k = \ldots, -2, -1, 0, 1, 2, \ldots$

f) the sets $\{x + n \mid n \in \mathbf{Z}\}$ for all $x \in [0, 1)$

47. List the ordered pairs in the equivalence relations produced by these partitions of $\{0, 1, 2, 3, 4, 5\}$.

a) $\{0\}, \{1, 2\}, \{3, 4, 5\}$

b) $\{0, 1\}, \{2, 3\}, \{4, 5\}$

c) $\{0, 1, 2\}, \{3, 4, 5\}$

d) $\{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}$

48. List the ordered pairs in the equivalence relations produced by these partitions of $\{a, b, c, d, e, f, g\}$.

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

A partition P_1 is called a **refinement** of the partition P_2 if every set in P_1 is a subset of one of the sets in P_2.

49. Show that the partition formed from congruence classes modulo 6 is a refinement of the partition formed from congruence classes modulo 3.

50. Show that the partition of the set of people living in the United States consisting of subsets of people living in the same county (or parish) and same state is a refinement of the partition consisting of subsets of people living in the same state.

51. Show that the partition of the set of bit strings of length 16 formed by equivalence classes of bit strings that agree on the last eight bits is a refinement of the partition formed from the equivalence classes of bit strings that agree on the last four bits.

In Exercises 52 and 53, R_n refers to the family of equivalence relations defined in Example 5. Recall that $s\,R_n\,t$, where s and t are two strings if $s = t$ or s and t are strings with at least n characters that agree in their first n characters.

52. Show that the partition of the set of all bit strings formed by equivalence classes of bit strings with respect to the equivalence relation R_4 is a refinement of the partition formed by equivalence classes of bit strings with respect to the equivalence relation R_3.

53. Show that the partition of the set of all identifiers in C formed by the equivalence classes of identifiers with respect to the equivalence relation R_{31} is a refinement of the partition formed by equivalence classes of identifiers with respect to the equivalence relation R_8. (Compilers for "old" C consider identifiers the same when their names agree in their first eight characters, while compilers in standard C consider identifiers the same when their names agree in their first 31 characters.)

54. Suppose that R_1 and R_2 are equivalence relations on a set A. Let P_1 and P_2 be the partitions that correspond to R_1 and R_2, respectively. Show that $R_1 \subseteq R_2$ if and only if P_1 is a refinement of P_2.

55. Find the smallest equivalence relation on the set $\{a, b, c, d, e\}$ containing the relation $\{(a, b), (a, c), (d, e)\}$.

56. Suppose that R_1 and R_2 are equivalence relations on the set S. Determine whether each of these combinations of R_1 and R_2 must be an equivalence relation.

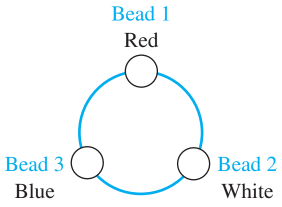
a) $R_1 \cup R_2$ **b)** $R_1 \cap R_2$ **c)** $R_1 \oplus R_2$

57. Consider the equivalence relation from Example 2, namely, $R = \{(x, y) \mid x - y \text{ is an integer}\}$.

a) What is the equivalence class of 1 for this equivalence relation?

b) What is the equivalence class of 1/2 for this equivalence relation?

***58.** Each bead on a bracelet with three beads is either red, white, or blue, as illustrated in the figure shown.



Define the relation R between bracelets as: (B_1, B_2), where B_1 and B_2 are bracelets, belongs to R if and only if B_2 can be obtained from B_1 by rotating it or rotating it and then reflecting it.

a) Show that R is an equivalence relation.

b) What are the equivalence classes of R?

***59.** Let R be the relation on the set of all colorings of the 2×2 checkerboard where each of the four squares is colored either red or blue so that (C_1, C_2), where C_1 and C_2 are 2×2 checkerboards with each of their four squares colored blue or red, belongs to R if and only if C_2 can be obtained from C_1 either by rotating the checkerboard or by rotating it and then reflecting it.

a) Show that R is an equivalence relation.

b) What are the equivalence classes of R?

60. a) Let R be the relation on the set of functions from $\mathbf{Z}^+$ to $\mathbf{Z}^+$ such that (f, g) belongs to R if and only if f is $\Theta(g)$ (see Section 3.2). Show that R is an equivalence relation.

b) Describe the equivalence class containing $f(n) = n^2$ for the equivalence relation of part (a).

61. Determine the number of different equivalence relations on a set with three elements by listing them.

62. Determine the number of different equivalence relations on a set with four elements by listing them.

***63.** Do we necessarily get an equivalence relation when we form the transitive closure of the symmetric closure of the reflexive closure of a relation?

***64.** Do we necessarily get an equivalence relation when we form the symmetric closure of the reflexive closure of the transitive closure of a relation?

65. Suppose we use Theorem 2 to form a partition P from an equivalence relation R. What is the equivalence relation R' that results if we use Theorem 2 again to form an equivalence relation from P?

66. Suppose we use Theorem 2 to form an equivalence relation R from a partition P. What is the partition P' that results if we use Theorem 2 again to form a partition from R?

67. Devise an algorithm to find the smallest equivalence relation containing a given relation.

***68.** Let $p(n)$ denote the number of different equivalence relations on a set with n elements (and by Theorem 2 the number of partitions of a set with n elements). Show that $p(n)$ satisfies the recurrence relation $p(n) = \sum_{j=0}^{n-1} C(n-1, j)p(n-j-1)$ and the initial condition $p(0) = 1$. (*Note:* The numbers $p(n)$ are called **Bell numbers** after the American mathematician E. T. Bell.)

69. Use Exercise 68 to find the number of different equivalence relations on a set with n elements, where n is a positive integer not exceeding 10.

9.6 Partial Orderings

9.6.1 Introduction

Links

We often use relations to order some or all of the elements of sets. For instance, we order words using the relation containing pairs of words (x, y), where x comes before y in the dictionary. We schedule projects using the relation consisting of pairs (x, y), where x and y are tasks in a project such that x must be completed before y begins. We order the set of integers using the relation containing the pairs (x, y), where x is less than y. When we add all of the pairs of the form (x, x) to these relations, we obtain a relation that is reflexive, antisymmetric, and transitive. These are properties that characterize relations used to order the elements of sets.

Definition 1 A relation R on a set S is called a *partial ordering* or *partial order* if it is reflexive, antisymmetric, and transitive. A set S together with a partial ordering R is called a *partially ordered set*, or *poset*, and is denoted by (S, R). Members of S are called *elements* of the poset.

We give examples of posets in Examples 1–3.

EXAMPLE 1 Show that the greater than or equal to relation ($\geq$) is a partial ordering on the set of integers.

Extra Examples

Solution: Because $a \geq a$ for every integer a, $\geq$ is reflexive. If $a \geq b$ and $b \geq a$, then $a = b$. Hence, $\geq$ is antisymmetric. Finally, $\geq$ is transitive because $a \geq b$ and $b \geq c$ imply that $a \geq c$. It follows that $\geq$ is a partial ordering on the set of integers and $(\mathbf{Z}, \geq)$ is a poset. ◀

EXAMPLE 2 The divisibility relation $|$ is a partial ordering on the set of positive integers, because it is reflexive, antisymmetric, and transitive, as was shown in Section 9.1. We see that $(\mathbf{Z}^+, |)$ is a poset. Recall that ($\mathbf{Z}^+$ denotes the set of positive integers.) ◀

EXAMPLE 3 Show that the inclusion relation $\subseteq$ is a partial ordering on the power set of a set S.

Solution: Because $A \subseteq A$ whenever A is a subset of S, $\subseteq$ is reflexive. It is antisymmetric because $A \subseteq B$ and $B \subseteq A$ imply that $A = B$. Finally, $\subseteq$ is transitive, because $A \subseteq B$ and $B \subseteq C$ imply that $A \subseteq C$. Hence, $\subseteq$ is a partial ordering on $P(S)$, and $(P(S), \subseteq)$ is a poset. ◀

Example 4 illustrates a relation that is not a partial ordering.

EXAMPLE 4 Let R be the relation on the set of people such that xRy if x and y are people and x is older than y. Show that R is not a partial ordering.

Extra Examples

Solution: Note that R is antisymmetric because if a person x is older than a person y, then y is not older than x. That is, if xRy, then $y \not R x$. The relation R is transitive because if person x is older than person y and y is older than person z, then x is older than z. That is, if xRy

and yRz, then xRz. However, R is not reflexive, because no person is older than himself or herself. That is, $x \not R x$ for all people x. It follows that R is not a partial ordering. ◀

In different posets different symbols such as $\leq$, $\subseteq$, and $|$, are used for a partial ordering. However, we need a symbol that we can use when we discuss the ordering relation in an arbitrary poset. Customarily, the notation $a \preccurlyeq b$ is used to denote that $(a, b) \in R$ in an arbitrary poset (S, R). This notation is used because the less than or equal to relation on the set of real numbers is the most familiar example of a partial ordering and the symbol $\preccurlyeq$ is similar to the $\leq$ symbol. (Note that the symbol $\preccurlyeq$ is used to denote the relation in *any* poset, not just the less than or equal to relation.) The notation $a \prec b$ denotes that $a \preccurlyeq b$, but $a \neq b$. Also, we say "a is less than b" or "b is greater than a" if $a \prec b$.

When a and b are elements of the poset $(S, \preccurlyeq)$, it is not necessary that either $a \preccurlyeq b$ or $b \preccurlyeq a$. For instance, in $(P(\mathbf{Z}), \subseteq)$, $\{1, 2\}$ is not related to $\{1, 3\}$, and vice versa, because neither set is contained within the other. Similarly, in $(\mathbf{Z}^+, |)$, 2 is not related to 3 and 3 is not related to 2, because $2 \nmid 3$ and $3 \nmid 2$. This leads to Definition 2.

Definition 2

The elements a and b of a poset $(S, \preccurlyeq)$ are called *comparable* if either $a \preccurlyeq b$ or $b \preccurlyeq a$. When a and b are elements of S such that neither $a \preccurlyeq b$ nor $b \preccurlyeq a$, a and b are called *incomparable*.

EXAMPLE 5 In the poset $(\mathbf{Z}^+, |)$, are the integers 3 and 9 comparable? Are 5 and 7 comparable?

Solution: The integers 3 and 9 are comparable, because $3 \mid 9$. The integers 5 and 7 are incomparable, because $5 \nmid 7$ and $7 \nmid 5$. ◀

The adjective "partial" is used to describe partial orderings because pairs of elements may be incomparable. When every two elements in the set are comparable, the relation is called a **total ordering**.

Definition 3

If $(S, \preccurlyeq)$ is a poset and every two elements of S are comparable, S is called a *totally ordered* or *linearly ordered set*, and $\preccurlyeq$ is called a *total order* or a *linear order.* A totally ordered set is also called a *chain*.

EXAMPLE 6 The poset $(\mathbf{Z}, \leq)$ is totally ordered, because $a \leq b$ or $b \leq a$ whenever a and b are integers. ◀

EXAMPLE 7 The poset $(\mathbf{Z}^+, |)$ is not totally ordered because it contains elements that are incomparable, such as 5 and 7. ◀

In Chapter 6 we noted that $(\mathbf{Z}^+, \leq)$ is well-ordered, where $\leq$ is the usual less than or equal to relation. We now define well-ordered sets.

Definition 4

$(S, \preccurlyeq)$ is a *well-ordered set* if it is a poset such that $\preccurlyeq$ is a total ordering and every nonempty subset of S has a least element.

EXAMPLE 8 The set of ordered pairs of positive integers, $\mathbf{Z}^+ \times \mathbf{Z}^+$, with $(a_1, a_2) \preccurlyeq (b_1, b_2)$ if $a_1 < b_1$, or if $a_1 = b_1$ and $a_2 \leq b_2$ (the lexicographic ordering), is a well-ordered set. The verification of this is left as Exercise 53. The set $\mathbf{Z}$, with the usual $\leq$ ordering, is not well-ordered because the set of negative integers, which is a subset of $\mathbf{Z}$, has no least element. ◀

At the end of Section 5.3 we showed how to use the principle of well-ordered induction (there called generalized induction) to prove results about a well-ordered set. We now state and prove that this proof technique is valid.

THEOREM 1 **THE PRINCIPLE OF WELL-ORDERED INDUCTION** Suppose that S is a well-ordered set. Then $P(x)$ is true for all $x \in S$, if

INDUCTIVE STEP: For every $y \in S$, if $P(x)$ is true for all $x \in S$ with $x \prec y$, then $P(y)$ is true.

Proof: Suppose it is not the case that $P(x)$ is true for all $x \in S$. Then there is an element $y \in S$ such that $P(y)$ is false. Consequently, the set $A = \{x \in S \mid P(x) \text{ is false}\}$ is nonempty. Because S is well-ordered, A has a least element a. By the choice of a as a least element of A, we know that $P(x)$ is true for all $x \in S$ with $x \prec a$. This implies by the inductive step $P(a)$ is true. This contradiction shows that $P(x)$ must be true for all $x \in S$. ◁

Remark: We do not need a basis step in a proof using the principle of well-ordered induction because if x_0 is the least element of a well-ordered set, the inductive step tells us that $P(x_0)$ is true. This follows because there are no elements $x \in S$ with $x \prec x_0$, so we know (using a vacuous proof) that $P(x)$ is true for all $x \in S$ with $x \prec x_0$.

The principle of well-ordered induction is a versatile technique for proving results about well-ordered sets. Even when it is possible to use mathematical induction for the set of positive integers to prove a theorem, it may be simpler to use the principle of well-ordered induction, as we saw in Examples 5 and 6 in Section 6.2, where we proved a result about the well-ordered set $(\mathbf{N} \times \mathbf{N}, \preccurlyeq)$ where $\preccurlyeq$ is lexicographic ordering on $\mathbf{N} \times \mathbf{N}$.

9.6.2 Lexicographic Order

The words in a dictionary are listed in alphabetic, or lexicographic, order, which is based on the ordering of the letters in the alphabet. This is a special case of an ordering of strings on a set constructed from a partial ordering on the set. We will show how this construction works in any poset.

First, we will show how to construct a partial ordering on the Cartesian product of two posets, $(A_1, \preccurlyeq_1)$ and $(A_2, \preccurlyeq_2)$. The **lexicographic ordering** $\preccurlyeq$ on $A_1 \times A_2$ is defined by specifying that one pair is less than a second pair if the first entry of the first pair is less than (in A_1) the first entry of the second pair, or if the first entries are equal, but the second entry of this pair is less than (in A_2) the second entry of the second pair. In other words, (a_1, a_2) is less than (b_1, b_2), that is,

$$(a_1, a_2) \prec (b_1, b_2),$$

either if $a_1 \prec_1 b_1$ or if both $a_1 = b_1$ and $a_2 \prec_2 b_2$.

We obtain a partial ordering $\preccurlyeq$ by adding equality to the ordering $\prec$ on $A_1 \times A_2$. The verification of this is left as an exercise.

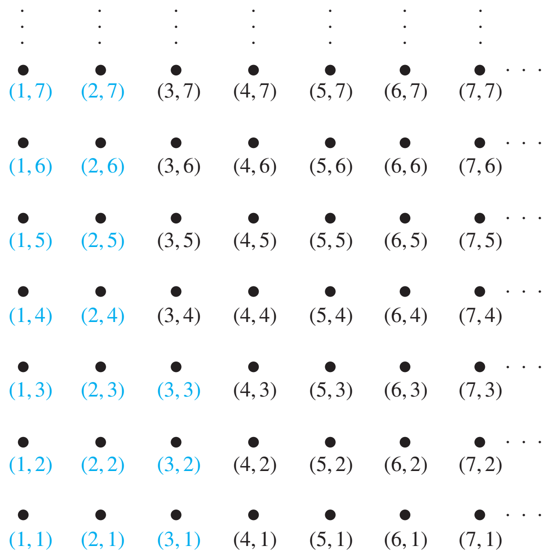


FIGURE 1 The ordered pairs less than (3, 4) in lexicographic order.

EXAMPLE 9 Determine whether $(3, 5) \prec (4, 8)$, whether $(3, 8) \prec (4, 5)$, and whether $(4, 9) \prec (4, 11)$ in the poset $(\mathbf{Z} \times \mathbf{Z}, \preccurlyeq)$, where $\preccurlyeq$ is the lexicographic ordering constructed from the usual $\leq$ relation on $\mathbf{Z}$.

Solution: Because $3 < 4$, it follows that $(3, 5) \prec (4, 8)$ and that $(3, 8) \prec (4, 5)$. We have $(4, 9) \prec (4, 11)$, because the first entries of (4, 9) and (4, 11) are the same but $9 < 11$. ◀

In Figure 1 the ordered pairs in $\mathbf{Z}^+ \times \mathbf{Z}^+$ that are less than (3, 4) are highlighted. A lexicographic ordering can be defined on the Cartesian product of n posets $(A_1, \preccurlyeq_1)$, $(A_2, \preccurlyeq_2), \ldots, (A_n, \preccurlyeq_n)$. Define the partial ordering $\preccurlyeq$ on $A_1 \times A_2 \times \cdots \times A_n$ by

$$(a_1, a_2, \ldots, a_n) \prec (b_1, b_2, \ldots, b_n)$$

if $a_1 \prec_1 b_1$, or if there is an integer $i > 0$ such that $a_1 = b_1, \ldots, a_i = b_i$, and $a_{i+1} \prec_{i+1} b_{i+1}$. In other words, one n-tuple is less than a second n-tuple if the entry of the first n-tuple in the first position where the two n-tuples disagree is less than the entry in that position in the second n-tuple.

EXAMPLE 10 Note that $(1, 2, 3, 5) \prec (1, 2, 4, 3)$, because the entries in the first two positions of these 4-tuples agree, but in the third position the entry in the first 4-tuple, 3, is less than that in the second 4-tuple, 4. (Here the ordering on 4-tuples is the lexicographic ordering that comes from the usual less than or equal to relation on the set of integers.) ◀

We can now define lexicographic ordering of strings. Consider the strings $a_1a_2 \ldots a_m$ and $b_1b_2 \ldots b_n$ on a partially ordered set S. Suppose these strings are not equal. Let t be the minimum of m and n. The definition of lexicographic ordering is that the string $a_1a_2 \ldots a_m$ is less than $b_1b_2 \ldots b_n$ if and only if

$$(a_1, a_2, \ldots, a_t) \prec (b_1, b_2, \ldots, b_t), \text{ or}$$
$$(a_1, a_2, \ldots, a_t) = (b_1, b_2, \ldots, b_t) \text{ and } m < n,$$

where $\prec$ in this inequality represents the lexicographic ordering of S^t. In other words, to determine the ordering of two different strings, the longer string is truncated to the length of the shorter string, namely, to $t = \min(m, n)$ terms. Then the t-tuples made up of the first t terms of each string are compared using the lexicographic ordering on S^t. One string is less than another string if the t-tuple corresponding to the first string is less than the t-tuple of the second string, or if these two t-tuples are the same, but the second string is longer. The verification that this is a partial ordering is left as Exercise 38 for the reader.

EXAMPLE 11 Consider the set of strings of lowercase English letters. Using the ordering of letters in the alphabet, a lexicographic ordering on the set of strings can be constructed. A string is less than a second string if the letter in the first string in the first position where the strings differ comes before the letter in the second string in this position, or if the first string and the second string agree in all positions, but the second string has more letters. This ordering is the same as that used in dictionaries. For example,

$$\textit{discreet} \prec \textit{discrete},$$

because these strings differ first in the seventh position, and $e \prec t$. Also,

$$\textit{discreet} \prec \textit{discreetness},$$

because the first eight letters agree, but the second string is longer. Furthermore,

$$\textit{discrete} \prec \textit{discretion},$$

because

$$\textit{discrete} \prec \textit{discreti}.$$

◀

9.6.3 Hasse Diagrams

Many edges in the directed graph for a finite poset do not have to be shown because they must be present. For instance, consider the directed graph for the partial ordering $\{(a, b) \mid a \leq b\}$ on the set $\{1, 2, 3, 4\}$, shown in Figure 2(a). Because this relation is a partial ordering, it is reflexive, and its directed graph has loops at all vertices. Consequently, we do not have to show these loops because they must be present; in Figure 2(b) loops are not shown. Because a partial ordering is transitive, we do not have to show those edges that must be present because of transitivity. For example, in Figure 2(c) the edges (1, 3), (1, 4), and (2, 4) are not shown because they must be present. If we assume that all edges are pointed "upward" (as they are drawn in the figure), we do not have to show the directions of the edges; Figure 2(c) does not show directions.

In general, we can represent a finite poset $(S, \preccurlyeq)$ using this procedure: Start with the directed graph for this relation. Because a partial ordering is reflexive, a loop (a, a) is present at every vertex a. Remove these loops. Next, remove all edges that must be in the partial ordering because

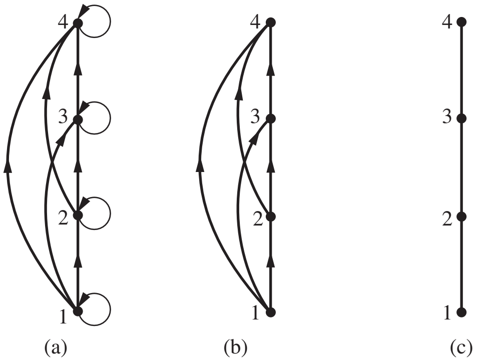
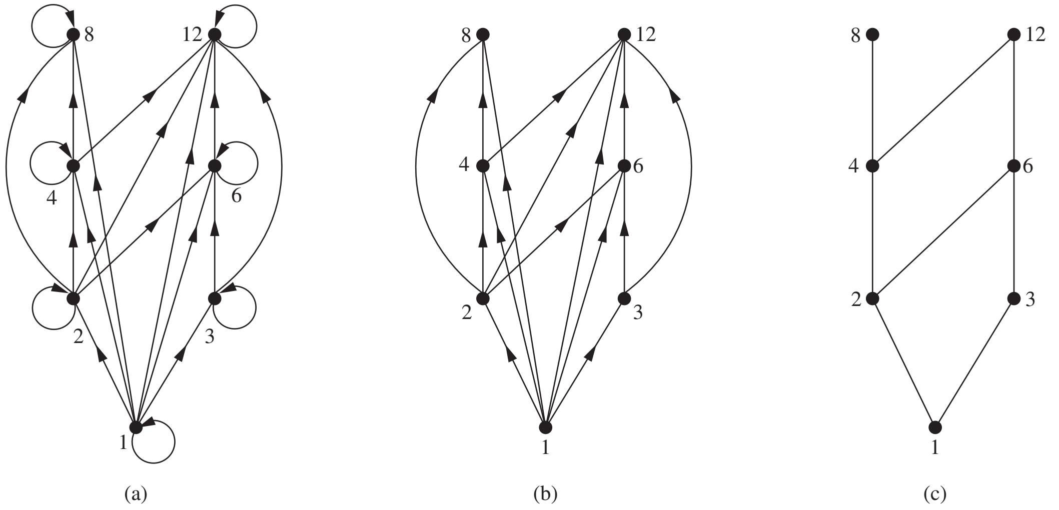


FIGURE 2 Constructing the Hasse diagram for ({1, 2, 3, 4}, ≤).



(a) (b) (c)

FIGURE 3 Constructing the Hasse diagram of ({1, 2, 3, 4, 6, 8, 12}, |).

of the presence of other edges and transitivity. That is, remove all edges (x, y) for which there is an element $z \in S$ such that $x \prec z$ and $z \prec x$. Finally, arrange each edge so that its initial vertex is below its terminal vertex (as it is drawn on paper). Remove all the arrows on the directed edges, because all edges point "upward" toward their terminal vertex.

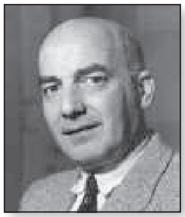
These steps are well defined, and only a finite number of steps need to be carried out for a finite poset. When all the steps have been taken, the resulting diagram contains sufficient information to find the partial ordering, as we will explain later. The resulting diagram is called the **Hasse diagram** of $(S, \preccurlyeq)$, named after the twentieth-century German mathematician Helmut Hasse who made extensive use of them.

Let $(S, \preccurlyeq)$ be a poset. We say that an element $y \in S$ **covers** an element $x \in S$ if $x \prec y$ and there is no element $z \in S$ such that $x \prec z \prec y$. The set of pairs (x, y) such that y covers x is called the **covering relation** of $(S, \preccurlyeq)$. From the description of the Hasse diagram of a poset, we see that the edges in the Hasse diagram of $(S, \preccurlyeq)$ are upwardly pointing edges corresponding to the pairs in the covering relation of $(S, \preccurlyeq)$. Furthermore, we can recover a poset from its covering relation, because it is the reflexive transitive closure of its covering relation. (Exercise 31 asks for a proof of this fact.) This tells us that we can construct a partial ordering from its Hasse diagram.

EXAMPLE 12 Draw the Hasse diagram representing the partial ordering $\{(a, b) \mid a \text{ divides } b\}$ on $\{1, 2, 3, 4, 6, 8, 12\}$.

Solution: Begin with the digraph for this partial order, as shown in Figure 3(a). Remove all loops, as shown in Figure 3(b). Then delete all the edges implied by the transitive property. These

Links



©ullstein bild/ullstein bild Dtl./Getty Images

HELMUT HASSE (1898–1979) Helmut Hasse was born in Kassel, Germany. He served in the German navy after high school. He began his university studies at Göttingen University in 1918, moving in 1920 to Marburg University to study under the number theorist Kurt Hensel. During this time, Hasse made fundamental contributions to algebraic number theory. He became Hensel's successor at Marburg, later becoming director of the renowned mathematical institute at Göttingen in 1934, and took a position at Hamburg University in 1950. Hasse served for 50 years as an editor of *Crelle's Journal*, an illustrious German mathematics periodical, taking over the job of chief editor in 1936 when the Nazis forced Hensel to resign. During World War II Hasse worked on applied mathematics research for the German navy. He was noted for the clarity and personal style of his lectures and was devoted both to number theory and to his students. (Hasse has been controversial for connections with the Nazi party. Investigations have shown he was a strong German nationalist but not an ardent Nazi.)

are (1, 4), (1, 6), (1, 8), (1, 12), (2, 8), (2, 12), and (3, 12). Arrange all edges to point upward, and delete all arrows to obtain the Hasse diagram. The resulting Hasse diagram is shown in Figure 3(c). ◀

EXAMPLE 13 Draw the Hasse diagram for the partial ordering $\{(A, B) \mid A \subseteq B\}$ on the power set $P(S)$, where $S = \{a, b, c\}$.

Solution: The Hasse diagram for this partial ordering is obtained from the associated digraph by deleting all the loops and all the edges that occur from transitivity, namely, $\big(\emptyset, \{a, b\}\big)$, $\big(\emptyset, \{a, c\}\big)$, $\big(\emptyset, \{b, c\}\big)$, $\big(\emptyset, \{a, b, c\}\big)$, $(\{a\}, \{a, b, c\})$, $(\{b\}, \{a, b, c\})$, and $(\{c\}, \{a, b, c\})$. Finally, all edges point upward, and arrows are deleted. The resulting Hasse diagram is illustrated in Figure 4. ◀

9.6.4 Maximal and Minimal Elements

Elements of posets that have certain extremal properties are important for many applications. An element of a poset is called maximal if it is not less than any element of the poset. That is, a is **maximal** in the poset $(S, \preccurlyeq)$ if there is no $b \in S$ such that $a \prec b$. Similarly, an element of a poset is called minimal if it is not greater than any element of the poset. That is, a is **minimal** if there is no element $b \in S$ such that $b \prec a$. Maximal and minimal elements are easy to spot using a Hasse diagram. They are the "top" and "bottom" elements in the diagram.

EXAMPLE 14 Which elements of the poset $(\{2, 4, 5, 10, 12, 20, 25\}, \mid)$ are maximal, and which are minimal?

Solution: The Hasse diagram in Figure 5 for this poset shows that the maximal elements are 12, 20, and 25, and the minimal elements are 2 and 5. As this example shows, a poset can have more than one maximal element and more than one minimal element. ◀

Sometimes there is an element in a poset that is greater than every other element. Such an element is called the greatest element. That is, a is the **greatest element** of the poset $(S, \preccurlyeq)$ if $b \preccurlyeq a$ for all $b \in S$. The greatest element is unique when it exists [see Exercise 40(a)]. Likewise, an element is called the least element if it is less than all the other elements in the poset. That is, a is the **least element** of $(S, \preccurlyeq)$ if $a \preccurlyeq b$ for all $b \in S$. The least element is unique when it exists [see Exercise 40(b)].

EXAMPLE 15 Determine whether the posets represented by each of the Hasse diagrams in Figure 6 have a greatest element and a least element.

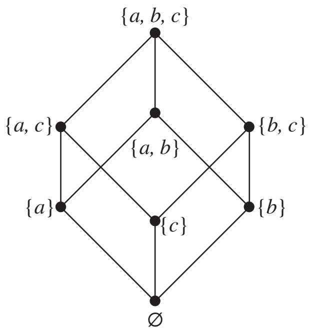


FIGURE 4 The Hasse diagram of $(P(\{a, b, c\}), \subseteq)$.

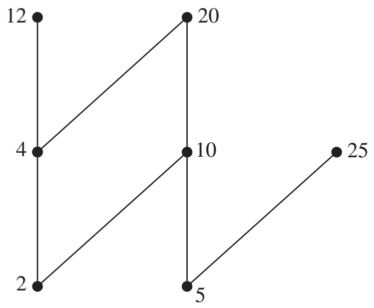


FIGURE 5 The Hasse diagram of a poset.

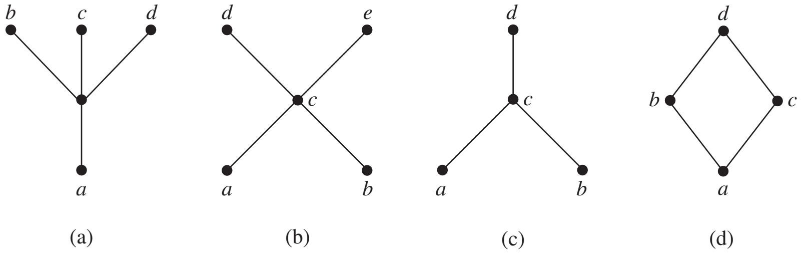


FIGURE 6 **Hasse diagrams of four posets.**

Solution: The least element of the poset with Hasse diagram (a) is a. This poset has no greatest element. The poset with Hasse diagram (b) has neither a least nor a greatest element. The poset with Hasse diagram (c) has no least element. Its greatest element is d. The poset with Hasse diagram (d) has least element a and greatest element d. ◀

EXAMPLE 16 Let S be a set. Determine whether there is a greatest element and a least element in the poset $(P(S), \subseteq)$.

Solution: The least element is the empty set, because $\emptyset \subseteq T$ for any subset T of S. The set S is the greatest element in this poset, because $T \subseteq S$ whenever T is a subset of S. ◀

EXAMPLE 17 Is there a greatest element and a least element in the poset $(\mathbf{Z}^+, |)$?

Solution: The integer 1 is the least element because $1|n$ whenever n is a positive integer. Because there is no integer that is divisible by all positive integers, there is no greatest element. ◀

Sometimes it is possible to find an element that is greater than or equal to all the elements in a subset A of a poset $(S, \preccurlyeq)$. If u is an element of S such that $a \preccurlyeq u$ for all elements $a \in A$, then u is called an **upper bound** of A. Likewise, there may be an element less than or equal to all the elements in A. If l is an element of S such that $l \preccurlyeq a$ for all elements $a \in A$, then l is called a **lower bound** of A.

EXAMPLE 18 Find the lower and upper bounds of the subsets $\{a, b, c\}$, $\{j, h\}$, and $\{a, c, d, f\}$ in the poset with the Hasse diagram shown in Figure 7.

Solution: The upper bounds of $\{a, b, c\}$ are e, f, j, and h, and its only lower bound is a. There are no upper bounds of $\{j, h\}$, and its lower bounds are a, b, c, d, e, and f. The upper bounds of $\{a, c, d, f\}$ are f, h, and j, and its lower bound is a. ◀

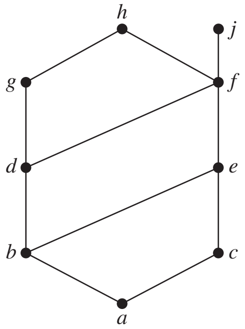


FIGURE 7 **The Hasse diagram of a poset.**

The element x is called the **least upper bound** of the subset A if x is an upper bound that is less than every other upper bound of A. Because there is only one such element, if it exists, it makes sense to call this element *the* least upper bound [see Exercise 42(a)]. That is, x is the least upper bound of A if $a \preccurlyeq x$ whenever $a \in A$, and $x \preccurlyeq z$ whenever z is an upper bound of A. Similarly, the element y is called the **greatest lower bound** of A if y is a lower bound of A and $z \preccurlyeq y$ whenever z is a lower bound of A. The greatest lower bound of A is unique if it exists [see Exercise 42(b)]. The greatest lower bound and least upper bound of a subset A are denoted by glb(A) and lub(A), respectively.

EXAMPLE 19 Find the greatest lower bound and the least upper bound of $\{b, d, g\}$, if they exist, in the poset shown in Figure 7.

Solution: The upper bounds of $\{b, d, g\}$ are g and h. Because $g \prec h$, g is the least upper bound. The lower bounds of $\{b, d, g\}$ are a and b. Because $a \prec b$, b is the greatest lower bound. ◀

EXAMPLE 20 Find the greatest lower bound and the least upper bound of the sets $\{3, 9, 12\}$ and $\{1, 2, 4, 5, 10\}$, if they exist, in the poset $(\mathbf{Z}^+, |)$.

Extra Examples ›

Solution: An integer is a lower bound of $\{3, 9, 12\}$ if 3, 9, and 12 are divisible by this integer. The only such integers are 1 and 3. Because $1 \mid 3$, 3 is the greatest lower bound of $\{3, 9, 12\}$. The only lower bound for the set $\{1, 2, 4, 5, 10\}$ with respect to $|$ is the element 1. Hence, 1 is the greatest lower bound for $\{1, 2, 4, 5, 10\}$.

An integer is an upper bound for $\{3, 9, 12\}$ if and only if it is divisible by 3, 9, and 12. The integers with this property are those divisible by the least common multiple of 3, 9, and 12, which is 36. Hence, 36 is the least upper bound of $\{3, 9, 12\}$. A positive integer is an upper bound for the set $\{1, 2, 4, 5, 10\}$ if and only if it is divisible by 1, 2, 4, 5, and 10. The integers with this property are those integers divisible by the least common multiple of these integers, which is 20. Hence, 20 is the least upper bound of $\{1, 2, 4, 5, 10\}$. ◀

9.6.5 Lattices

A partially ordered set in which every pair of elements has both a least upper bound and a greatest lower bound is called a **lattice**. Lattices have many special properties. Furthermore, lattices are used in many different applications such as models of information flow and play an important role in Boolean algebra.

EXAMPLE 21 Determine whether the posets represented by each of the Hasse diagrams in Figure 8 are lattices.

Solution: The posets represented by the Hasse diagrams in (a) and (c) are both lattices because in each poset every pair of elements has both a least upper bound and a greatest lower bound, as the reader should verify. On the other hand, the poset with the Hasse diagram shown in (b) is not a lattice, because the elements b and c have no least upper bound. To see this, note that each of the elements d, e, and f is an upper bound, but none of these three elements precedes the other two with respect to the ordering of this poset. ◀

EXAMPLE 22 Is the poset $(\mathbf{Z}^+, |)$ a lattice?

Solution: Let a and b be two positive integers. The least upper bound and greatest lower bound of these two integers are the least common multiple and the greatest common divisor of these integers, respectively, as the reader should verify. It follows that this poset is a lattice. ◀

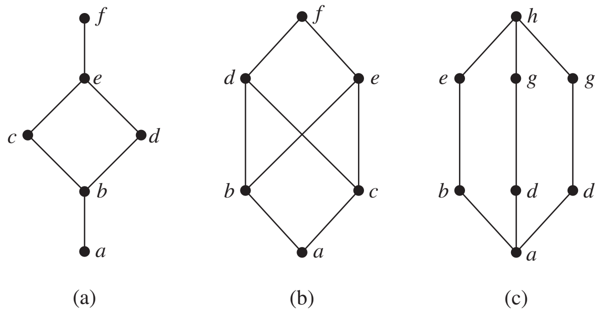


FIGURE 8 **Hasse diagrams of three posets.**

EXAMPLE 23 Determine whether the posets $(\{1, 2, 3, 4, 5\}, |)$ and $(\{1, 2, 4, 8, 16\}, |)$ are lattices.

Solution: Because 2 and 3 have no upper bounds in $(\{1, 2, 3, 4, 5\}, |)$, they certainly do not have a least upper bound. Hence, the first poset is not a lattice.

Every two elements of the second poset have both a least upper bound and a greatest lower bound. The least upper bound of two elements in this poset is the larger of the elements and the greatest lower bound of two elements is the smaller of the elements, as the reader should verify. Hence, this second poset is a lattice. ◀

EXAMPLE 24 Determine whether $(P(S), \subseteq)$ is a lattice where S is a set.

Solution: Let A and B be two subsets of S. The least upper bound and the greatest lower bound of A and B are $A \cup B$ and $A \cap B$, respectively, as the reader can show. Hence, $(P(S), \subseteq)$ is a lattice. ◀

EXAMPLE 25 **The Lattice Model of Information Flow** In many settings the flow of information from one person or computer program to another is restricted via security clearances. We can use a lattice model to represent different information flow policies. For example, one common information flow policy is the *multilevel security policy* used in government and military systems. Each piece of information is assigned to a security class, and each security class is represented by a pair (A, C) where A is an *authority level* and C is a *category*. People and computer programs are then allowed access to information from a specific restricted set of security classes.

Links ❯

There are billions of pages of classified U.S. government documents.

The typical authority levels used in the U.S. government are unclassified (0), confidential (1), secret (2), and top secret (3). (Information is said to be classified if it is confidential, secret, or top secret.) Categories used in security classes are the subsets of a set of all *compartments* relevant to a particular area of interest. Each compartment represents a particular subject area. For example, if the set of compartments is {*spies, moles, double agents*}, then there are eight different categories, one for each of the eight subsets of the set of compartments, such as {*spies, moles*}.

We can order security classes by specifying that $(A_1, C_1) \preccurlyeq (A_2, C_2)$ if and only if $A_1 \le A_2$ and $C_1 \subseteq C_2$. Information is permitted to flow from security class (A_1, C_1) into security class (A_2, C_2) if and only if $(A_1, C_1) \preccurlyeq (A_2, C_2)$. For example, information is permitted to flow from the security class (*secret,* {*spies, moles*}) into the security class (*top secret,* {*spies, moles, double agents*}), whereas information is not allowed to flow from the security class (*top secret,* {*spies, moles*}) into either of the security classes (*secret,* {*spies, moles, double agents*}) or (*top secret,* {*spies*}).

We leave it to the reader (see Exercise 48) to show that the set of all security classes with the ordering defined in this example forms a lattice. ◀

9.6.6 Topological Sorting

Links ❯

Suppose that a project is made up of 20 different tasks. Some tasks can be completed only after others have been finished. How can an order be found for these tasks? To model this problem we set up a partial order on the set of tasks so that $a \prec b$ if and only if a and b are tasks where b cannot be started until a has been completed. To produce a schedule for the project, we need to produce an order for all 20 tasks that is compatible with this partial order. We will show how this can be done.

We begin with a definition. A total ordering $\preccurlyeq$ is said to be **compatible** with the partial ordering R if $a \preccurlyeq b$ whenever aRb. Constructing a compatible total ordering from a partial ordering is called **topological sorting**.* We will need to use Lemma 1.

LEMMA 1 Every finite nonempty poset $(S, \preccurlyeq)$ has at least one minimal element.

Proof: Choose an element a_0 of S. If a_0 is not minimal, then there is an element a_1 with $a_1 \prec a_0$. If a_1 is not minimal, there is an element a_2 with $a_2 \prec a_1$. Continue this process, so that if a_n is not minimal, there is an element a_{n+1} with $a_{n+1} \prec a_n$. Because there are only a finite number of elements in the poset, this process must end with a minimal element a_n. ◁

The topological sorting algorithm we will describe works for any finite nonempty poset. To define a total ordering on the poset $(A, \preccurlyeq)$, first choose a minimal element a_1; such an element exists by Lemma 1. Next, note that $(A - \{a_1\}, \preccurlyeq)$ is also a poset, as the reader should verify. (Here by $\preccurlyeq$ we mean the restriction of the original relation $\preccurlyeq$ on A to $A - \{a_1\}$.) If it is nonempty, choose a minimal element a_2 of this poset. Then remove a_2 as well, and if there are additional elements left, choose a minimal element a_3 in $A - \{a_1, a_2\}$. Continue this process by choosing a_{k+1} to be a minimal element in $A - \{a_1, a_2, \dots, a_k\}$, as long as elements remain.

Because A is a finite set, this process must terminate. The end product is a sequence of elements $a_1, a_2, \dots, a_n$. The desired total ordering $\preccurlyeq_t$ is defined by

$$a_1 \prec_t a_2 \prec_t \cdots \prec_t a_n.$$

This total ordering is compatible with the original partial ordering. To see this, note that if $b \prec c$ in the original partial ordering, c is chosen as the minimal element at a phase of the algorithm where b has already been removed, for otherwise c would not be a minimal element. Pseudocode for this topological sorting algorithm is shown in Algorithm 1.

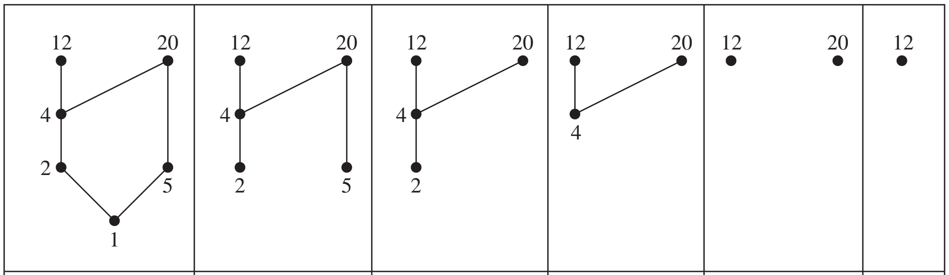
ALGORITHM 1 Topological Sorting.

```
procedure topological sort ((S, $\preccurlyeq$): finite poset)
k := 1
while $S \neq \emptyset$
    $a_k$ := a minimal element of S {such an element exists by Lemma 1}
    $S := S - \{a_k\}$
    $k := k + 1$
return $a_1, a_2, \dots, a_n$ {$a_1, a_2, \dots, a_n$ is a compatible total ordering of S}
```

EXAMPLE 26 Find a compatible total ordering for the poset $(\{1, 2, 4, 5, 12, 20\}, |)$.

Solution: The first step is to choose a minimal element. This must be 1, because it is the only minimal element. Next, select a minimal element of $(\{2, 4, 5, 12, 20\}, |)$. There are two minimal elements in this poset, namely, 2 and 5. We select 5. The remaining elements are $\{2, 4, 12, 20\}$. The only minimal element at this stage is 2. Next, 4 is chosen because it is the only minimal

*"Topological sorting" is terminology used by computer scientists; mathematicians use the terminology "linearization of a partial ordering" for the same thing. In mathematics, topology is the branch of geometry dealing with properties of geometric figures that hold for all figures that can be transformed into one another by continuous bijections. In computer science, a topology is any arrangement of objects that can be connected with edges.



Minimal element chosen 1	5	2	4	20	12

FIGURE 9 A topological sort of ({1, 2, 4, 5, 12, 20}, |).

element of ({4, 12, 20}, |). Because both 12 and 20 are minimal elements of ({12, 20}, |), either can be chosen next. We select 20, which leaves 12 as the last element left. This produces the total ordering

$$1 \prec 5 \prec 2 \prec 4 \prec 20 \prec 12.$$

The steps used by this sorting algorithm are displayed in Figure 9. ◀

Topological sorting has an application to the scheduling of projects.

EXAMPLE 27 A development project at a computer company requires the completion of seven tasks. Some of these tasks can be started only after other tasks are finished. A partial ordering on tasks is set up by considering task $X \prec$ task Y if task Y cannot be started until task X has been completed. The Hasse diagram for the seven tasks, with respect to this partial ordering, is shown in Figure 10. Find an order in which these tasks can be carried out to complete the project.

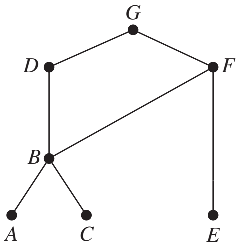
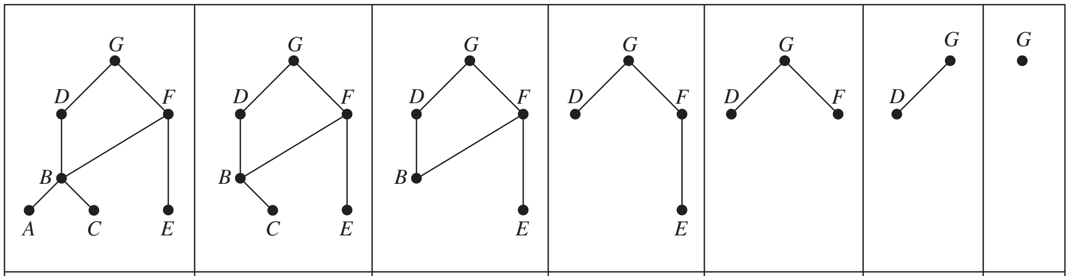


FIGURE 10 The Hasse diagram for seven tasks.

Solution: An ordering of the seven tasks can be obtained by performing a topological sort. The steps of a sort are illustrated in Figure 11. The result of this sort, $A \prec C \prec B \prec E \prec F \prec D \prec G$, gives one possible order for the tasks. ◀



Minimal element chosen *A*	*C*	*B*	*E*	*F*	*D*	*G*

FIGURE 11 A topological sort of the tasks.

Exercises

1. Which of these relations on $\{0, 1, 2, 3\}$ are partial orderings? Determine the properties of a partial ordering that the others lack.
   a) $\{(0, 0), (1, 1), (2, 2), (3, 3)\}$
   b) $\{(0, 0), (1, 1), (2, 0), (2, 2), (2, 3), (3, 2), (3, 3)\}$
   c) $\{(0, 0), (1, 1), (1, 2), (2, 2), (3, 3)\}$
   d) $\{(0, 0), (1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$
   e) $\{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 2), (3, 3)\}$

2. Which of these relations on $\{0, 1, 2, 3\}$ are partial orderings? Determine the properties of a partial ordering that the others lack.
   a) $\{(0, 0), (2, 2), (3, 3)\}$
   b) $\{(0, 0), (1, 1), (2, 0), (2, 2), (2, 3), (3, 3)\}$
   c) $\{(0, 0), (1, 1), (1, 2), (2, 2), (3, 1), (3, 3)\}$
   d) $\{(0, 0), (1, 1), (1, 2), (1, 3), (2, 0), (2, 2), (2, 3), (3, 0), (3, 3)\}$
   e) $\{(0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 2), (1, 3), (2, 0), (2, 2), (3, 3)\}$

3. Is (S, R) a poset if S is the set of all people in the world and $(a, b) \in R$, where a and b are people, if
   a) a is taller than b?
   b) a is not taller than b?
   c) $a = b$ or a is an ancestor of b?
   d) a and b have a common friend?

4. Is (S, R) a poset if S is the set of all people in the world and $(a, b) \in R$, where a and b are people, if
   a) a is no shorter than b?
   b) a weighs more than b?
   c) $a = b$ or a is a descendant of b?
   d) a and b do not have a common friend?

5. Which of these are posets?
   a) $(\mathbf{Z}, =)$ **b)** $(\mathbf{Z}, \neq)$ **c)** $(\mathbf{Z}, \geq)$ **d)** $(\mathbf{Z}, \nmid)$

6. Which of these are posets?
   a) $(\mathbf{R}, =)$ **b)** $(\mathbf{R}, <)$ **c)** $(\mathbf{R}, \leq)$ **d)** $(\mathbf{R}, \neq)$

7. Determine whether the relations represented by these zero–one matrices are partial orders.

   a) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

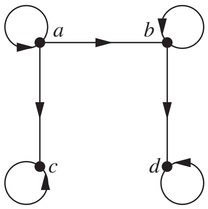
   c) $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$

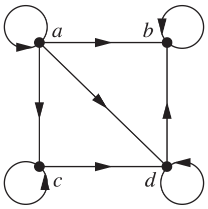
8. Determine whether the relations represented by these zero–one matrices are partial orders.

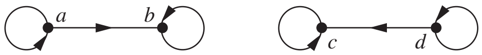
   a) $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

   c) $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$

In Exercises 9–11 determine whether the relation with the directed graph shown is a partial order.

9.


10.


11.


12. Let (S, R) be a poset. Show that (S, R^{-1}) is also a poset, where R^{-1} is the inverse of R. The poset (S, R^{-1}) is called the **dual** of (S, R).

13. Find the duals of these posets.
   a) $(\{0, 1, 2\}, \leq)$ **b)** $(\mathbf{Z}, \geq)$
   c) $(P(\mathbf{Z}), \supseteq)$ **d)** $(\mathbf{Z}^+, |)$

14. Which of these pairs of elements are comparable in the poset $(\mathbf{Z}^+, |)$?
   a) 5, 15 **b)** 6, 9 **c)** 8, 16 **d)** 7, 7

15. Find two incomparable elements in these posets.
   a) $(P(\{0, 1, 2\}), \subseteq)$ **b)** $(\{1, 2, 4, 6, 8\}, |)$

16. Let $S = \{1, 2, 3, 4\}$. With respect to the lexicographic order based on the usual less than elation,
   a) find all pairs in $S \times S$ less than (2, 3).
   b) find all pairs in $S \times S$ greater than (3, 1).
   c) draw the Hasse diagram of the poset $(S \times S, \preccurlyeq)$.

17. Find the lexicographic ordering of these n-tuples:
   a) (1, 1, 2), (1, 2, 1) **b)** (0, 1, 2, 3), (0, 1, 3, 2)
   c) (1, 0, 1, 0, 1), (0, 1, 1, 1, 0)

18. Find the lexicographic ordering of these strings of lowercase English letters:
   a) *quack, quick, quicksilver, quicksand, quacking*
   b) *open, opener, opera, operand, opened*
   c) *zoo, zero, zoom, zoology, zoological*

19. Find the lexicographic ordering of the bit strings 0, 01, 11, 001, 010, 011, 0001, and 0101 based on the ordering $0 < 1$.

20. Draw the Hasse diagram for the greater than or equal to relation on $\{0, 1, 2, 3, 4, 5\}$.

21. Draw the Hasse diagram for the less than or equal to relation on $\{0, 2, 5, 10, 11, 15\}$.

22. Draw the Hasse diagram for divisibility on the set

a) $\{1, 2, 3, 4, 5, 6\}$. **b)** $\{3, 5, 7, 11, 13, 16, 17\}$.
c) $\{2, 3, 5, 10, 11, 15, 25\}$. **d)** $\{1, 3, 9, 27, 81, 243\}$.

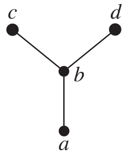
23. Draw the Hasse diagram for divisibility on the set

a) $\{1, 2, 3, 4, 5, 6, 7, 8\}$. **b)** $\{1, 2, 3, 5, 7, 11, 13\}$.
c) $\{1, 2, 3, 6, 12, 24, 36, 48\}$.
d) $\{1, 2, 4, 8, 16, 32, 64\}$.

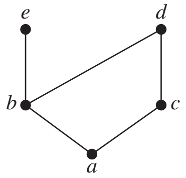
24. Draw the Hasse diagram for inclusion on the set $P(S)$, where $S = \{a, b, c, d\}$.

In Exercises 25–27 list all ordered pairs in the partial ordering with the accompanying Hasse diagram.

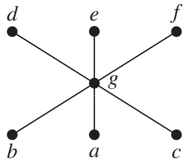
25.



26.



27.



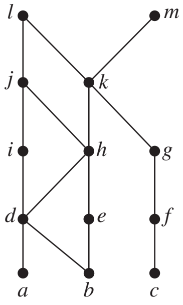
28. What is the covering relation of the partial ordering $\{(a, b) \mid a \text{ divides } b\}$ on $\{1, 2, 3, 4, 6, 12\}$?

29. What is the covering relation of the partial ordering $\{(A, B) \mid A \subseteq B\}$ on the power set of S, where $S = \{a, b, c\}$?

30. What is the covering relation of the partial ordering for the poset of security classes defined in Example 25?

31. Show that a finite poset can be reconstructed from its covering relation. [*Hint:* Show that the poset is the reflexive transitive closure of its covering relation.]

32. Answer these questions for the partial order represented by this Hasse diagram.



a) Find the maximal elements.
b) Find the minimal elements.
c) Is there a greatest element?
d) Is there a least element?
e) Find all upper bounds of $\{a, b, c\}$.
f) Find the least upper bound of $\{a, b, c\}$, if it exists.
g) Find all lower bounds of $\{f, g, h\}$.
h) Find the greatest lower bound of $\{f, g, h\}$, if it exists.

33. Answer these questions for the poset $(\{3, 5, 9, 15, 24, 45\}, |)$.

a) Find the maximal elements.
b) Find the minimal elements.
c) Is there a greatest element?
d) Is there a least element?
e) Find all upper bounds of $\{3, 5\}$.
f) Find the least upper bound of $\{3, 5\}$, if it exists.
g) Find all lower bounds of $\{15, 45\}$.
h) Find the greatest lower bound of $\{15, 45\}$, if it exists.

34. Answer these questions for the poset $(\{2, 4, 6, 9, 12, 18, 27, 36, 48, 60, 72\}, |)$.

a) Find the maximal elements.
b) Find the minimal elements.
c) Is there a greatest element?
d) Is there a least element?
e) Find all upper bounds of $\{2, 9\}$.
f) Find the least upper bound of $\{2, 9\}$, if it exists.
g) Find all lower bounds of $\{60, 72\}$.
h) Find the greatest lower bound of $\{60, 72\}$, if it exists.

35. Answer these questions for the poset $(\{\{1\}, \{2\}, \{4\}, \{1, 2\}, \{1, 4\}, \{2, 4\}, \{3, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}, \subseteq)$.

a) Find the maximal elements.
b) Find the minimal elements.
c) Is there a greatest element?
d) Is there a least element?
e) Find all upper bounds of $\{\{2\}, \{4\}\}$.
f) Find the least upper bound of $\{\{2\}, \{4\}\}$, if it exists.
g) Find all lower bounds of $\{\{1, 3, 4\}, \{2, 3, 4\}\}$.
h) Find the greatest lower bound of $\{\{1, 3, 4\}, \{2, 3, 4\}\}$, if it exists.

36. Give a poset that has

a) a minimal element but no maximal element.
b) a maximal element but no minimal element.
c) neither a maximal nor a minimal element.

37. Show that lexicographic order is a partial ordering on the Cartesian product of two posets.

38. Show that lexicographic order is a partial ordering on the set of strings from a poset.

39. Suppose that $(S, \preccurlyeq_1)$ and $(T, \preccurlyeq_2)$ are posets. Show that $(S \times T, \preccurlyeq)$ is a poset where $(s, t) \preccurlyeq (u, v)$ if and only if $s \preccurlyeq_1 u$ and $t \preccurlyeq_2 v$.

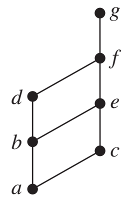
40. a) Show that there is exactly one greatest element of a poset, if such an element exists.
b) Show that there is exactly one least element of a poset, if such an element exists.

41. a) Show that there is exactly one maximal element in a poset with a greatest element.
b) Show that there is exactly one minimal element in a poset with a least element.

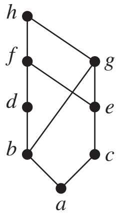
42. a) Show that the least upper bound of a set in a poset is unique if it exists.
b) Show that the greatest lower bound of a set in a poset is unique if it exists.

43. Determine whether the posets with these Hasse diagrams are lattices.

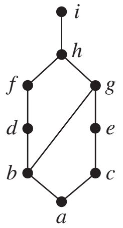
a)



b)



c)



44. Determine whether these posets are lattices.

a) $(\{1, 3, 6, 9, 12\}, |)$ **b)** $(\{1, 5, 25, 125\}, |)$

c) $(\mathbf{Z}, \geq)$

d) $(P(S), \supseteq)$, where $P(S)$ is the power set of a set S

45. Show that every nonempty finite subset of a lattice has a least upper bound and a greatest lower bound.

46. Show that if the poset (S, R) is a lattice then the dual poset (S, R^{-1}) is also a lattice.

47. In a company, the lattice model of information flow is used to control sensitive information with security classes represented by ordered pairs (A, C). Here A is an authority level, which may be nonproprietary (0), proprietary (1), restricted (2), or registered (3). A category C is a subset of the set of all projects {*Cheetah, Impala, Puma*}. (Names of animals are often used as code names for projects in companies.)

a) Is information permitted to flow from (*Proprietary*, {*Cheetah, Puma*}) into (*Restricted*, {*Puma*})?

b) Is information permitted to flow from (*Restricted*, {*Cheetah*}) into (*Registered*, {*Cheetah, Impala*})?

c) Into which classes is information from (*Proprietary*, {*Cheetah, Puma*}) permitted to flow?

d) From which classes is information permitted to flow into the security class (*Restricted*, {*Impala, Puma*})?

48. Show that the set S of security classes (A, C) is a lattice, where A is a positive integer representing an authority class and C is a subset of a finite set of compartments, with $(A_1, C_1) \preccurlyeq (A_2, C_2)$ if and only if $A_1 \leq A_2$ and $C_1 \subseteq C_2$. [*Hint:* First show that $(S, \preccurlyeq)$ is a poset and then show that the least upper bound and greatest lower bound of (A_1, C_1) and (A_2, C_2) are $(\max(A_1, A_2), C_1 \cup C_2)$ and $(\min(A_1, A_2), C_1 \cap C_2)$, respectively.]

***49.** Show that the set of all partitions of a set S with the relation $P_1 \preccurlyeq P_2$ if the partition P_1 is a refinement of the partition P_2 is a lattice. (See the preamble to Exercise 49 of Section 9.5.)

50. Show that every totally ordered set is a lattice.

51. Show that every finite lattice has a least element and a greatest element.

52. Give an example of an infinite lattice with

a) neither a least nor a greatest element.

b) a least but not a greatest element.

c) a greatest but not a least element.

d) both a least and a greatest element.

53. Verify that $(\mathbf{Z}^+ \times \mathbf{Z}^+, \preccurlyeq)$ is a well-ordered set, where $\preccurlyeq$ is lexicographic order, as claimed in Example 8.

54. Determine whether each of these posets is well-ordered.

a) $(S, \leq)$, where $S = \{10, 11, 12, \ldots\}$

b) $(\mathbf{Q} \cap [0, 1], \leq)$ (the set of rational numbers between 0 and 1 inclusive)

c) $(S, \leq)$, where S is the set of positive rational numbers with denominators not exceeding 3

d) $(\mathbf{Z}^-, \geq)$, where $\mathbf{Z}^-$ is the set of negative integers

A poset $(R, \preccurlyeq)$ is **well-founded** if there is no infinite decreasing sequence of elements in the poset, that is, elements $x_1, x_2, \ldots, x_n$ such that $\cdots \prec x_n \prec \cdots \prec x_2 \prec x_1$. A poset $(R, \preccurlyeq)$ is **dense** if for all $x \in S$ and $y \in S$ with $x \prec y$, there is an element $z \in R$ such that $x \prec z \prec y$.

55. Show that the poset $(\mathbf{Z}, \preccurlyeq)$, where $x \prec y$ if and only if $|x| < |y|$ is well-founded but is not a totally ordered set.

56. Show that a dense poset with at least two elements that are comparable is not well-founded.

57. Show that the poset of rational numbers with the usual less than or equal to relation, $(\mathbf{Q}, \leq)$, is a dense poset.

***58.** Show that the set of strings of lowercase English letters with lexicographic order is neither well-founded nor dense.

59. Show that a poset is well-ordered if and only if it is totally ordered and well-founded.

60. Show that a finite nonempty poset has a maximal element.

61. Find a compatible total order for the poset with the Hasse diagram shown in Exercise 32.

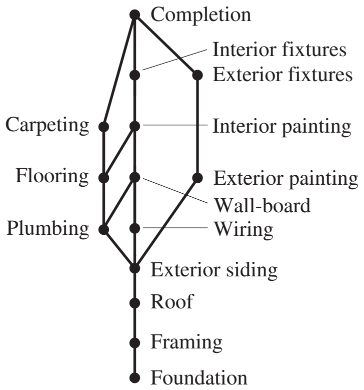
62. Find a compatible total order for the divisibility relation on the set $\{1, 2, 3, 6, 8, 12, 24, 36\}$.

63. Find all compatible total orderings for the poset $(\{1, 2, 4, 5, 12, 20\}, |\}$ from Example 26.

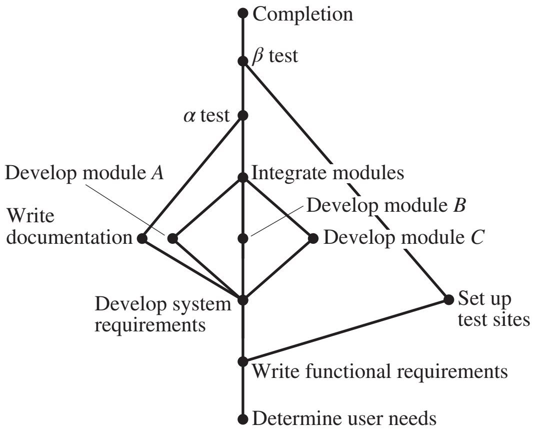
64. Find all compatible total orderings for the poset with the Hasse diagram in Exercise 27.

65. Find all possible orders for completing the tasks in the development project in Example 27.

66. Schedule the tasks needed to build a house, by specifying their order, if the Hasse diagram representing these tasks is as shown in the figure.



67. Find an ordering of the tasks of a software project if the Hasse diagram for the tasks of the project is as shown.



Key Terms and Results

TERMS

binary relation from A to B: a subset of $A \times B$

relation on A: a binary relation from A to itself (that is, a subset of $A \times A$)

$S \circ R$: composite of R and S

R^{-1}: inverse relation of R

R^n: nth power of R

reflexive: a relation R on A is reflexive if $(a, a) \in R$ for all $a \in A$

symmetric: a relation R on A is symmetric if $(b, a) \in R$ whenever $(a, b) \in R$

antisymmetric: a relation R on A is antisymmetric if $a = b$ whenever $(a, b) \in R$ and $(b, a) \in R$

transitive: a relation R on A is transitive if $(a, b) \in R$ and $(b, c) \in R$ implies that $(a, c) \in R$

n-ary relation on $A_1, A_2, \ldots, A_n$: a subset of $A_1 \times A_2 \times \cdots \times A_n$

relational data model: a model for representing databases using n-ary relations

primary key: a domain of an n-ary relation such that an n-tuple is uniquely determined by its value for this domain

composite key: the Cartesian product of domains of an n-ary relation such that an n-tuple is uniquely determined by its values in these domains

selection operator: a function that selects the n-tuples in an n-ary relation that satisfy a specified condition

projection: a function that produces relations of smaller degree from an n-ary relation by deleting fields

join: a function that combines n-ary relations that agree on certain fields

itemset: a collection of items

count of an itemset: the number of transactions that are supersets of the itemset

frequent itemset: an itemset with frequency greater than or equal to the support threshold

support of an itemset: the frequency of transactions that contain the itemset

association rule: an implication of the form $I \rightarrow J$, where I and J are itemsets

support of the association rule $I \rightarrow J$: the fraction of transactions that contain both the itemsets I and J

confidence of an association rule: the conditional probability that J is a subset of a transaction given that I is

directed graph or digraph: a set of elements called vertices and ordered pairs of these elements, called edges

loop: an edge of the form (a, a)

closure of a relation R with respect to a property P: the relation S (if it exists) that contains R, has property **P**, and is contained within any relation that contains R and has property **P**

path in a digraph: a sequence of edges $(a, x_1), (x_1, x_2), \ldots, (x_{n-2}, x_{n-1}), (x_{n-1}, b)$ such that the terminal vertex of each edge is the initial vertex of the succeeding edge in the sequence

circuit (or cycle) in a digraph: a path that begins and ends at the same vertex

R^* (connectivity relation): the relation consisting of those ordered pairs (a, b) such that there is a path from a to b

equivalence relation: a reflexive, symmetric, and transitive relation

equivalent: if R is an equivalence relation, a is equivalent to b if aRb

$[a]_R$ (equivalence class of a with respect to R): the set of all elements of A that are equivalent to a

$[a]_m$ (congruence class modulo m): the set of integers congruent to a modulo m

partition of a set S: a collection of pairwise disjoint nonempty subsets that have S as their union

partial ordering: a relation that is reflexive, antisymmetric, and transitive

poset (S, R): a set S and a partial ordering R on this set

comparable: the elements a and b in the poset $(A, \preccurlyeq)$ are comparable if $a \preccurlyeq b$ or $b \preccurlyeq a$

incomparable: elements in a poset that are not comparable

total (or linear) ordering: a partial ordering for which every pair of elements are comparable

totally (or linearly) ordered set: a poset with a total (or linear) ordering

well-ordered set: a poset $(S, \preccurlyeq)$, where $\preccurlyeq$ is a total order and every nonempty subset of S has a least element

lexicographic order: a partial ordering of Cartesian products or strings

Hasse diagram: a graphical representation of a poset where loops and all edges resulting from the transitive property are not shown, and the direction of the edges is indicated by the position of the vertices

maximal element: an element of a poset that is not less than any other element of the poset

minimal element: an element of a poset that is not greater than any other element of the poset

greatest element: an element of a poset greater than all other elements in this set

least element: an element of a poset less than all other elements in this set

upper bound of a set: an element in a poset greater than all other elements in the set

lower bound of a set: an element in a poset less than all other elements in the set

least upper bound of a set: an upper bound of the set that is less than all other upper bounds

greatest lower bound of a set: a lower bound of the set that is greater than all other lower bounds

lattice: a partially ordered set in which every two elements have a greatest lower bound and a least upper bound

compatible total ordering for a partial ordering: a total ordering that contains the given partial ordering

topological sort: the construction of a total ordering compatible with a given partial ordering

RESULTS

The reflexive closure of a relation R on the set A equals $R \cup \Delta$, where $\Delta = \{(a, a) \mid a \in A\}$.

The symmetric closure of a relation R on the set A equals $R \cup R^{-1}$, where $R^{-1} = \{(b, a) \mid (a, b) \in R\}$.

The transitive closure of a relation equals the connectivity relation formed from this relation.

Warshall's algorithm for finding the transitive closure of a relation

Let R be an equivalence relation. Then the following three statements are equivalent: (1) $a\ R\ b$; (2) $[a]_R \cap [b]_R \neq \emptyset$; (3) $[a]_R = [b]_R$.

The equivalence classes of an equivalence relation on a set A form a partition of A. Conversely, an equivalence relation can be constructed from any partition so that the equivalence classes are the subsets in the partition.

The principle of well-ordered induction

The topological sorting algorithm

Review Questions

1. a) What is a relation on a set?
   b) How many relations are there on a set with n elements?
2. a) What is a reflexive relation?
   b) What is a symmetric relation?
   c) What is an antisymmetric relation?
   d) What is a transitive relation?
3. Give an example of a relation on the set $\{1, 2, 3, 4\}$ that is
   a) reflexive, symmetric, and not transitive.
   b) not reflexive, symmetric, and transitive.
   c) reflexive, antisymmetric, and not transitive.
   d) reflexive, symmetric, and transitive.
   e) reflexive, antisymmetric, and transitive.
4. a) How many reflexive relations are there on a set with n elements?
   b) How many symmetric relations are there on a set with n elements?
   c) How many antisymmetric relations are there on a set with n elements?
5. a) Explain how an n-ary relation can be used to represent information about students at a university.
   b) How can the 5-ary relation containing names of students, their addresses, telephone numbers, majors, and grade point averages be used to form a 3-ary relation containing the names of students, their majors, and their grade point averages?
   c) How can the 4-ary relation containing names of students, their addresses, telephone numbers, and majors and the 4-ary relation containing names of students, their student numbers, majors, and numbers of credit hours be combined into a single n-ary relation?
6. a) Explain how to use a zero–one matrix to represent a relation on a finite set.

b) Explain how to use the zero–one matrix representing a relation to determine whether the relation is reflexive, symmetric, and/or antisymmetric.

7. a) Explain how to use a directed graph to represent a relation on a finite set.
   b) Explain how to use the directed graph representing a relation to determine whether a relation is reflexive, symmetric, and/or antisymmetric.

8. a) Define the reflexive closure and the symmetric closure of a relation.
   b) How can you construct the reflexive closure of a relation?
   c) How can you construct the symmetric closure of a relation?
   d) Find the reflexive closure and the symmetric closure of the relation $\{(1, 2), (2, 3), (2, 4), (3, 1)\}$ on the set $\{1, 2, 3, 4\}$.

9. a) Define the transitive closure of a relation.
   b) Can the transitive closure of a relation be obtained by including all pairs (a, c) such that (a, b) and (b, c) belong to the relation?
   c) Describe two algorithms for finding the transitive closure of a relation.
   d) Find the transitive closure of the relation $\{(1,1), (1,3), (2,1), (2,3), (2,4), (3,2), (3,4), (4,1)\}$.

10. a) Define an equivalence relation.
    b) Which relations on the set $\{a, b, c, d\}$ are equivalence relations and contain (a, b) and (b, d)?

11. a) Show that congruence modulo m is an equivalence relation whenever m is a positive integer.
    b) Show that the relation $\{(a, b) \mid a \equiv \pm b \pmod{7}\}$ is an equivalence relation on the set of integers.

12. a) What are the equivalence classes of an equivalence relation?
    b) What are the equivalence classes of the "congruent modulo 5" relation?
    c) What are the equivalence classes of the equivalence relation in Question 11(b)?

13. Explain the relationship between equivalence relations on a set and partitions of this set.

14. a) Define a partial ordering.
    b) Show that the divisibility relation on the set of positive integers is a partial order.

15. Explain how partial orderings on the sets A_1 and A_2 can be used to define a partial ordering on the set $A_1 \times A_2$.

16. a) Explain how to construct the Hasse diagram of a partial order on a finite set.
    b) Draw the Hasse diagram of the divisibility relation on the set $\{2, 3, 5, 9, 12, 15, 18\}$.

17. a) Define a maximal element of a poset and the greatest element of a poset.
    b) Give an example of a poset that has three maximal elements.
    c) Give an example of a poset with a greatest element.

18. a) Define a lattice.
    b) Give an example of a poset with five elements that is a lattice and an example of a poset with five elements that is not a lattice.

19. a) Show that every finite subset of a lattice has a greatest lower bound and a least upper bound.
    b) Show that every lattice with a finite number of elements has a least element and a greatest element.

20. a) Define a well-ordered set.
    b) Describe an algorithm for producing a totally ordered set compatible with a given partially ordered set.
    c) Explain how the algorithm from (b) can be used to order the tasks in a project if tasks are done one at a time and each task can be done only after one or more of the other tasks have been completed.

Supplementary Exercises

1. Let S be the set of all strings of English letters. Determine whether these relations are reflexive, irreflexive, symmetric, antisymmetric, and/or transitive.
   a) $R_1 = \{(a, b) \mid a \text{ and } b \text{ have no letters in common}\}$
   b) $R_2 = \{(a, b) \mid a \text{ and } b \text{ are not the same length}\}$
   c) $R_3 = \{(a, b) \mid a \text{ is longer than } b\}$

2. Construct a relation on the set $\{a, b, c, d\}$ that is
   a) reflexive, symmetric, but not transitive.
   b) irreflexive, symmetric, and transitive.
   c) irreflexive, antisymmetric, and not transitive.
   d) reflexive, neither symmetric nor antisymmetric, and transitive.
   e) neither reflexive, irreflexive, symmetric, antisymmetric, nor transitive.

3. Show that the relation R on $\mathbf{Z} \times \mathbf{Z}$ defined by $(a, b)\, R\, (c, d)$ if and only if $a + d = b + c$ is an equivalence relation.

4. Show that a subset of an antisymmetric relation is also antisymmetric.

5. Let R be a reflexive relation on a set A. Show that $R \subseteq R^2$.

6. Suppose that R_1 and R_2 are reflexive relations on a set A. Show that $R_1 \oplus R_2$ is irreflexive.

7. Suppose that R_1 and R_2 are reflexive relations on a set A. Is $R_1 \cap R_2$ also reflexive? Is $R_1 \cup R_2$ also reflexive?

8. Suppose that R is a symmetric relation on a set A. Is $\overline{R}$ also symmetric?

9. Let R_1 and R_2 be symmetric relations. Is $R_1 \cap R_2$ also symmetric? Is $R_1 \cup R_2$ also symmetric?

10. A relation R is called **circular** if aRb and bRc imply that cRa. Show that R is reflexive and circular if and only if it is an equivalence relation.

11. Show that a primary key in an n-ary relation is a primary key in any projection of this relation that contains this key as one of its fields.

12. Is the primary key in an n-ary relation also a primary key in a larger relation obtained by taking the join of this relation with a second relation?

13. Show that the reflexive closure of the symmetric closure of a relation is the same as the symmetric closure of its reflexive closure.

14. Let R be the relation on the set of all mathematicians that contains the ordered pair (a, b) if and only if a and b have written a published mathematical paper together.

a) Describe the relation R^2.

b) Describe the relation R^*.

c) The **Erdős number** of a mathematician is 1 if this mathematician wrote a paper with the prolific Hungarian mathematician Paul Erdős, it is 2 if this mathematician did not write a joint paper with Erdős but wrote a joint paper with someone who wrote a joint paper with Erdős, and so on (except that the Erdős number of Erdős himself is 0). Give a definition of the Erdős number in terms of paths in R.

15. a) Give an example to show that the transitive closure of the symmetric closure of a relation is not necessarily the same as the symmetric closure of the transitive closure of this relation.

b) Show, however, that the transitive closure of the symmetric closure of a relation must contain the symmetric closure of the transitive closure of this relation.

16. a) Let S be the set of subroutines of a computer program. Define the relation R by $\mathbf{P}\,R\,\mathbf{Q}$ if subroutine **P** calls subroutine **Q** during its execution. Describe the transitive closure of R.

b) For which subroutines **P** does $(\mathbf{P}, \mathbf{P})$ belong to the transitive closure of R?

c) Describe the reflexive closure of the transitive closure of R.

17. Suppose that R and S are relations on a set A with $R \subseteq S$ such that the closures of R and S with respect to a property **P** both exist. Show that the closure of R with respect to **P** is a subset of the closure of S with respect to **P**.

18. Show that the symmetric closure of the union of two relations is the union of their symmetric closures.

* **19.** Devise an algorithm, based on the concept of interior vertices, that finds the length of the longest path between two vertices in a directed graph, or determines that there are arbitrarily long paths between these vertices.

20. Which of these are equivalence relations on the set of all people?

a) $\{(x, y) \mid x$ and y have the same sign of the zodiac$\}$

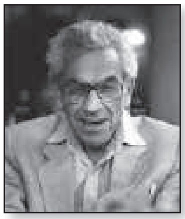
b) $\{(x, y) \mid x$ and y were born in the same year$\}$

c) $\{(x, y) \mid x$ and y have been in the same city$\}$

* **21.** How many different equivalence relations with exactly three different equivalence classes are there on a set with five elements?

22. Show that $\{(x, y) \mid x - y \in \mathbf{Q}\}$ is an equivalence relation on the set of real numbers, where **Q** denotes the set of rational numbers. What are $[1]$, $[\frac{1}{2}]$, and $[\pi]$?

Links



Courtesy of George Csicsery

PAUL ERDŐS (1913–1996) Paul Erdős, born in Budapest, Hungary, was the son of two high school mathematics teachers. He was a child prodigy; at age 3 he could multiply three-digit numbers in his head, and at 4 he discovered negative numbers on his own. Because his mother did not want to expose him to contagious diseases, he was mostly home-schooled. At 17 Erdős entered Eőtvős University, graduating four years later with a Ph.D. in mathematics. After graduating he spent four years at Manchester, England, on a postdoctoral fellowship. In 1938 he went to the United States because of the difficult political situation in Hungary, especially for Jews. He spent much of his time in the United States, except for 1954 to 1962, when he was banned as part of the paranoia of the McCarthy era. He also spent considerable time in Israel.

Erdős made many significant contributions to combinatorics and to number theory. One of the discoveries of which he was most proud is his elementary proof (in the sense that it does not use any complex analysis) of the prime number theorem, which provides an estimate for the number of primes not exceeding a fixed positive integer. He also participated in the modern development of the Ramsey theory.

Erdős traveled extensively throughout the world to work with other mathematicians, visiting conferences, universities, and research laboratories. He had no permanent home. He devoted himself almost entirely to mathematics, traveling from one mathematician to the next, proclaiming “My brain is open.” Erdős was the author or coauthor of more than 1500 papers and had more than 500 coauthors. Copies of his articles are kept by Ron Graham, a famous discrete mathematician with whom he collaborated extensively and who took care of many of his worldly needs.

Erdős offered rewards, ranging from $10 to $10,000, for the solution of problems that he found particularly interesting, with the size of the reward depending on the difficulty of the problem. He paid out close to $4000. Erdős had his own special language, using such terms as “epsilon” (child), “boss” (woman), “slave” (man), “captured” (married), “liberated” (divorced), “Supreme Fascist” (God), “Sam” (United States), and “Joe” (Soviet Union). Although he was curious about many things, he concentrated almost all his energy on mathematical research. He had no hobbies and no full-time job. He never married and apparently remained celibate. Erdős was extremely generous, donating much of the money he collected from prizes, awards, and stipends for scholarships and to worthwhile causes. He traveled extremely lightly and did not like having many material possessions.

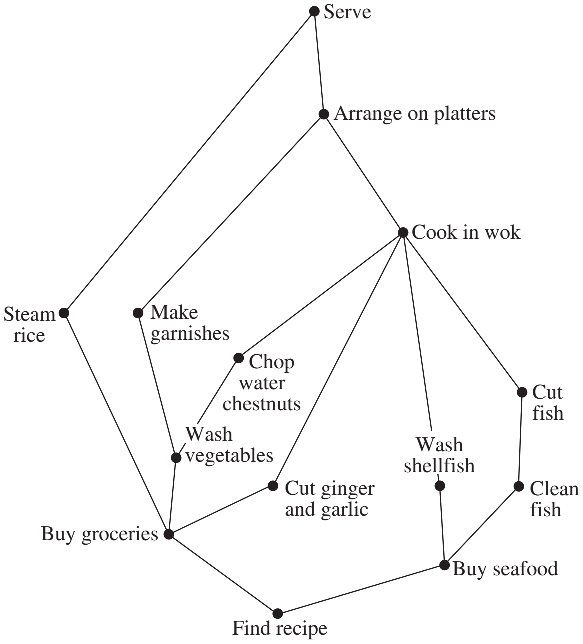
23. Suppose that $P_1 = \{A_1, A_2, \ldots, A_m\}$ and $P_2 = \{B_1, B_2, \ldots, B_n\}$ are both partitions of the set S. Show that the collection of nonempty subsets of the form $A_i \cap B_j$ is a partition of S that is a refinement of both P_1 and P_2 (see the preamble to Exercise 49 of Section 9.5).

***24.** Show that the transitive closure of the symmetric closure of the reflexive closure of a relation R is the smallest equivalence relation that contains R.

25. Let $\mathbf{R}(S)$ be the set of all relations on a set S. Define the relation $\preccurlyeq$ on $\mathbf{R}(S)$ by $R_1 \preccurlyeq R_2$ if $R_1 \subseteq R_2$, where R_1 and R_2 are relations on S. Show that $(\mathbf{R}(S), \preccurlyeq)$ is a poset.

26. Let $\mathbf{P}(S)$ be the set of all partitions of the set S. Define the relation $\preccurlyeq$ on $\mathbf{P}(S)$ by $P_1 \preccurlyeq P_2$ if P_1 is a refinement of P_2 (see Exercise 49 of Section 9.5). Show that $(\mathbf{P}(S), \preccurlyeq)$ is a poset.

27. Schedule the tasks needed to cook a Chinese meal by specifying their order, if the Hasse diagram representing these tasks is as shown here.



A subset of a poset such that every two elements of this subset are comparable is called a **chain**. A subset of a poset is called an **antichain** if every two elements of this subset are incomparable.

28. Find all chains in the posets with the Hasse diagrams shown in Exercises 25–27 in Section 9.6.

29. Find all antichains in the posets with the Hasse diagrams shown in Exercises 25–27 in Section 9.6.

30. Find an antichain with the greatest number of elements in the poset with the Hasse diagram of Exercise 32 in Section 9.6.

31. Show that every maximal chain in a finite poset $(S, \preccurlyeq)$ contains a minimal element of S. (A maximal chain is a chain that is not a subset of a larger chain.)

****32.** Show that every finite poset can be partitioned into k chains, where k is the largest number of elements in an antichain in this poset.

***33.** Show that in any group of $mn + 1$ people there is either a list of $m + 1$ people where a person in the list (except for the first person listed) is a descendant of the previous person on the list, or there are $n + 1$ people such that none of these people is a descendant of any of the other n people. [*Hint:* Use Exercise 32.]

Suppose that $(S, \preccurlyeq)$ is a well-founded partially ordered set. The *principle of well-founded induction* states that $P(x)$ is true for all $x \in S$ if $\forall x(\forall y(y \prec x \rightarrow P(y)) \rightarrow P(x))$.

34. Show that no separate basis case is needed for the principle of well-founded induction. That is, $P(u)$ is true for all minimal elements u in S if $\forall x(\forall y(y \prec x \rightarrow P(y)) \rightarrow P(x))$.

***35.** Show that the principle of well-founded induction is valid.

A relation R on a set A is a **quasi-ordering** on A if R is reflexive and transitive.

36. Let R be the relation on the set of all functions from $\mathbf{Z}^+$ to $\mathbf{Z}^+$ such that (f, g) belongs to R if and only if f is $O(g)$. Show that R is a quasi-ordering.

37. Let R be a quasi-ordering on a set A. Show that $R \cap R^{-1}$ is an equivalence relation.

***38.** Let R be a quasi-ordering and let S be the relation on the set of equivalence classes of $R \cap R^{-1}$ such that (C, D) belongs to S, where C and D are equivalence classes of R, if and only if there are elements c of C and d of D such that (c, d) belongs to R. Show that S is a partial ordering.

Let L be a lattice. Define the **meet** ($\wedge$) and **join** ($\vee$) operations by $x \wedge y = \text{glb}(x, y)$ and $x \vee y = \text{lub}(x, y)$.

39. Show that the following properties hold for all elements x, y, and z of a lattice L.

a) $x \wedge y = y \wedge x$ and $x \vee y = y \vee x$ **(commutative laws)**

b) $(x \wedge y) \wedge z = x \wedge (y \wedge z)$ and $(x \vee y) \vee z = x \vee (y \vee z)$ **(associative laws)**

c) $x \wedge (x \vee y) = x$ and $x \vee (x \wedge y) = x$ **(absorption laws)**

d) $x \wedge x = x$ and $x \vee x = x$ **(idempotent laws)**

40. Show that if x and y are elements of a lattice L, then $x \vee y = y$ if and only if $x \wedge y = x$.

A lattice L is **bounded** if it has both an **upper bound**, denoted by 1, such that $x \preccurlyeq 1$ for all $x \in L$ and a **lower bound**, denoted by 0, such that $0 \preccurlyeq x$ for all $x \in L$.

41. Show that if L is a bounded lattice with upper bound 1 and lower bound 0 then these properties hold for all elements $x \in L$.

a) $x \vee 1 = 1$ **b)** $x \wedge 1 = x$

c) $x \vee 0 = x$ **d)** $x \wedge 0 = 0$

42. Show that every finite lattice is bounded.

A lattice is called **distributive** if $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ for all x, y, and z in L.

***43.** Give an example of a lattice that is not distributive.

44. Show that the lattice $(P(S), \subseteq)$ where $P(S)$ is the power set of a finite set S is distributive.

45. Is the lattice $(\mathbf{Z}^+, |)$ distributive?

The **complement** of an element a of a bounded lattice L with upper bound 1 and lower bound 0 is an element b such that $a \vee b = 1$ and $a \wedge b = 0$. Such a lattice is **complemented** if every element of the lattice has a complement.

46. Give an example of a finite lattice where at least one element has more than one complement and at least one element has no complement.

47. Show that the lattice $(P(S), \subseteq)$ where $P(S)$ is the power set of a finite set S is complemented.

***48.** Show that if L is a finite distributive lattice, then an element of L has at most one complement.

The game of Chomp, introduced in Example 12 in Section 1.8, can be generalized for play on any finite partially ordered set $(S, \preceq)$ with a least element a. In this game, a move consists of selecting an element x in S and removing x and all elements larger than it from S. The loser is the player who is forced to select the least element a.

49. Show that the game of Chomp with cookies arranged in an $m \times n$ rectangular grid, described in Example 12 in Section 1.8, is the same as the game of Chomp on the poset $(S, |)$, where S is the set of all positive integers that divide $p^{m-1}q^{n-1}$, where p and q are distinct primes.

50. Show that if $(S, \preceq)$ has a greatest element b, then a winning strategy for Chomp on this poset exists. [*Hint:* Generalize the argument in Example 12 in Section 1.8.]

Computer Projects

Write programs with these input and output.

1. Given the matrix representing a relation on a finite set, determine whether the relation is reflexive and/or irreflexive.

2. Given the matrix representing a relation on a finite set, determine whether the relation is symmetric and/or antisymmetric.

3. Given the matrix representing a relation on a finite set, determine whether the relation is transitive.

4. Given a positive integer n, display all the relations on a set with n elements.

***5.** Given a positive integer n, determine the number of transitive relations on a set with n elements.

***6.** Given a positive integer n, determine the number of equivalence relations on a set with n elements.

***7.** Given a positive integer n, display all the equivalence relations on the set of the n smallest positive integers.

8. Given an n-ary relation, find the projection of this relation when specified fields are deleted.

9. Given an m-ary relation and an n-ary relation, and a set of common fields, find the join of these relations with respect to these common fields.

10. Given the matrix representing a relation on a finite set, find the matrix representing the reflexive closure of this relation.

11. Given the matrix representing a relation on a finite set, find the matrix representing the symmetric closure of this relation.

12. Given the matrix representing a relation on a finite set, find the matrix representing the transitive closure of this relation by computing the join of the Boolean powers of the matrix representing the relation.

13. Given the matrix representing a relation on a finite set, find the matrix representing the transitive closure of this relation using Warshall's algorithm.

14. Given the matrix representing a relation on a finite set, find the matrix representing the smallest equivalence relation containing this relation.

15. Given a partial ordering on a finite set, find a total ordering compatible with it using topological sorting.

Computations and Explorations

Use a computational program or programs you have written to do these exercises.

1. Display all the different relations on a set with four elements.

2. Display all the different reflexive and symmetric relations on a set with six elements.

3. Display all the reflexive and transitive relations on a set with five elements.

***4.** Determine how many transitive relations there are on a set with n elements for all positive integers n with $n \leq 7$.

5. Find the transitive closure of a relation of your choice on a set with at least 20 elements. Either use a relation that corresponds to direct links in a particular transportation or communications network or use a randomly generated relation.

6. Compute the number of different equivalence relations on a set with n elements for all positive integers n not exceeding 20.

7. Display all the equivalence relations on a set with seven elements.

***8.** Display all the partial orders on a set with five elements.

***9.** Display all the lattices on a set with five elements.

Writing Projects

Respond to these with essays using outside sources.

1. Discuss the concept of a fuzzy relation. How are fuzzy relations used?

2. Describe the basic principles of relational databases, going beyond what was covered in Section 9.2. How widely used are relational databases as compared with other types of databases?

3. Explain how the Apriori algorithm is used to find frequent itemsets and strong association rules.

4. Describe some applications of association rules in detail.

5. Look up the original papers by Warshall and by Roy (in French) in which they develop algorithms for finding transitive closures. Discuss their approaches. Why do you suppose that what we call Warshall's algorithm was discovered independently by more than one person?

6. Describe how equivalence classes can be used to define the rational numbers as classes of pairs of integers and how the basic arithmetic operations on rational numbers can be defined following this approach. (See Exercise 40 in Section 9.5.)

7. Explain how Helmut Hasse used what we now call Hasse diagrams.

8. Describe some of the mechanisms used to enforce information flow policies in computer operating systems.

9. Discuss the use of the Program Evaluation and Review Technique (PERT) to schedule the tasks of a large complicated project. How widely is PERT used?

10. Discuss the use of the Critical Path Method (CPM) to find the shortest time for the completion of a project. How widely is CPM used?

11. Discuss the concept of *duality* in a lattice. Explain how duality can be used to establish new results.

12. Explain what is meant by a *modular lattice*. Describe some of the properties of modular lattices and describe how modular lattices arise in the study of projective geometry.

CHAPTER 7

FUNCTIONS

Functions are ubiquitous in mathematics and computer science. That means you can hardly take two steps in these subjects without running into one. In this book we have previously discussed truth tables and input/output tables (which can be regarded as Boolean functions), sequences (which are functions defined on sets of integers), *mod* and *div* (which are functions defined on Cartesian products of integers), and floor and ceiling (which are functions from **R** to **Z**).

In this chapter we consider an additional wide variety of functions, focusing on those defined on discrete sets (such as finite sets or sets of integers). We then look at properties of functions such as one-to-one and onto, existence of inverse functions, and the interaction of composition of functions and the properties of one-to-one and onto. We end the chapter with the surprising result that there are different sizes of infinite sets and give an application to computability.

7.1 Functions Defined on General Sets

The theory that has had the greatest development in recent times is without any doubt the theory of functions. — Vito Volterra, 1888

As used in ordinary language, the word *function* indicates dependence of one varying quantity on another. If your teacher tells you that your grade in a course will be a function of your performance on the exams, you interpret this to mean that the teacher has some rule for translating exam scores into grades. To each collection of exam scores there corresponds a certain grade.

In Section 1.3 we defined a function as a certain type of relation. In this chapter we focus on the more dynamic way functions are used in mathematics. The following is a restatement of the definition of function that includes additional terminology associated with the concept.

• Definition

A **function f from a set X to a set Y**, denoted $f \colon X \to Y$, is a relation from X, the **domain**, to Y, the **co-domain**, that satisfies two properties: (1) every element in X is related to some element in Y, and (2) no element in X is related to more than one element in Y. Thus, given any element x in X, there is a unique element in Y that is related to x by f. If we call this element y, then we say that "f sends x to y" or "f maps x to y" and write $x \xrightarrow{f} y$ or $f \colon x \to y$. The unique element to which f sends x is denoted

$\boldsymbol{f(x)}$	and is called	$\boldsymbol{f}$ **of** $\boldsymbol{x}$, or
		the output of $\boldsymbol{f}$ for the input $\boldsymbol{x}$, or
		the value of $\boldsymbol{f}$ at $\boldsymbol{x}$, or
		the image of $\boldsymbol{x}$ under $\boldsymbol{f}$.

The set of all values of f taken together is called the *range of f* or the *image of X under f*. Symbolically,

$$\textbf{range of } \boldsymbol{f} = \textbf{image of } \boldsymbol{X} \textbf{ under } \boldsymbol{f} = \{y \in Y \mid y = f(x), \text{ for some } x \text{ in } X\}.$$

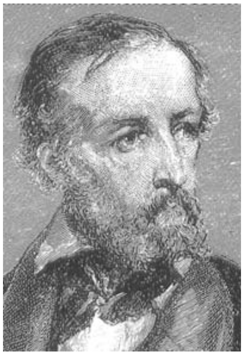
Given an element y in Y, there may exist elements in X with y as their image. If $f(x) = y$, then x is called **a preimage of $\boldsymbol{y}$** or **an inverse image of $\boldsymbol{y}$**. The set of all inverse images of y is called *the inverse image of y*. Symbolically,

$$\textbf{the inverse image of } \boldsymbol{y} = \{x \in X \mid f(x) = y\}.$$

Caution! Use $f(x)$ to refer to the value of the function f at x. Generally avoid using $f(x)$ to refer to the function f itself.

In some mathematical contexts, the notation $f(x)$ is used to refer both to the value of f at x and to the function f itself. Because using the notation this way can lead to confusion, we avoid it whenever possible. In this book, unless explicitly stated otherwise, the symbol $f(x)$ always refers to the value of the function f at x and not to the function f itself.

The concept of function was developed over a period of centuries. A definition similar to that given above was first formulated for sets of numbers by the German mathematician Lejeune Dirichlet (DEER-ish-lay) in 1837.



Stock Montage

Johann Peter Gustav Lejeune Dirichlet (1805–1859)

Arrow Diagrams

Recall from Section 1.3 that if X and Y are finite sets, you can define a function f from X to Y by drawing an arrow diagram. You make a list of elements in X and a list of elements in Y, and draw an arrow from each element in X to the corresponding element in Y, as shown in Figure 7.1.1.

This arrow diagram does define a function because

1. Every element of X has an arrow coming out of it.
2. No element of X has two arrows coming out of it that point to two different elements of Y.

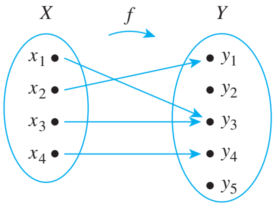
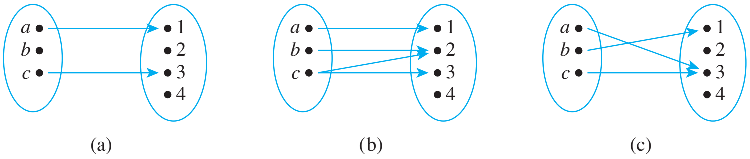


Figure 7.1.1

Example 7.1.1 Functions and Nonfunctions

Which of the arrow diagrams in Figure 7.1.2 define functions from $X = \{a, b, c\}$ to $Y = \{1, 2, 3, 4\}$?



(a) (b) (c)

Figure 7.1.1

Solution Only (c) defines a function. In (a) there is an element of X, namely b, that is not sent to any element of Y; that is, there is no arrow coming out of b. And in (b) the element c is not sent to a *unique* element of Y; that is, there are two arrows coming out of c, one pointing to 2 and the other to 3. ■

Example 7.1.2 A Function Defined by an Arrow Diagram

Let $X = \{a, b, c\}$ and $Y = \{1, 2, 3, 4\}$. Define a function f from X to Y by the arrow diagram in Figure 7.1.3.

a. Write the domain and co-domain of f.

b. Find $f(a)$, $f(b)$, and $f(c)$.

c. What is the range of f?

d. Is c an inverse image of 2? Is b an inverse image of 3?

e. Find the inverse images of 2, 4, and 1.

f. Represent f as a set of ordered pairs.

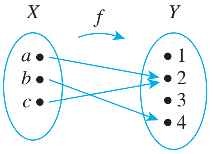


Figure 7.1.1

Solution

a. domain of $f = \{a, b, c\}$, co-domain of $f = \{1, 2, 3, 4\}$

b. $f(a) = 2$, $f(b) = 4$, $f(c) = 2$

c. range of $f = \{2, 4\}$

d. Yes, No

e. inverse image of $2 = \{a, c\}$
inverse image of $4 = \{b\}$
inverse image of $1 = \emptyset$ (*since no arrows point to* 1)

f. $\{(a, 2), (b, 4), (c, 2)\}$ ■

In Example 7.1.2 there are no arrows pointing to the 1 or the 3. This illustrates the fact that although each element of the domain of a function must have an arrow pointing out from it, there can be elements of the co-domain to which no arrows point. Note also that there are two arrows pointing to the 2—one coming from a and the other from c.

In Section 1.3 we gave a test for determining whether two functions with the same domain and co-domain are equal, saying that the test results from the definition of a function as a binary relation. We formalize this justification in Theorem 7.1.1.

Theorem 7.1.1 A Test for Function Equality

If $F: X \to Y$ and $G: X \to Y$ are functions, then $F = G$ if, and only if, $F(x) = G(x)$ for all $x \in X$.

Proof:

Note So $(x, y) \in F \Leftrightarrow y = F(x)$ and $(x, y) \in G \Leftrightarrow y = G(x)$.

Suppose $F: X \to Y$ and $G: X \to Y$ are functions, that is, F and G are binary relations from X to Y that satisfy the two additional function properties. Then F and G are subsets of $X \times Y$, and for (x, y) to be in F means that y is the unique element related to x by F, which we denote as $F(x)$. Similarly, for (x, y) to be in G means that y is the unique element related to x by G, which we denote as $G(x)$.

Now suppose that $F(x) = G(x)$ for all $x \in X$. Then if x is any element of X,

$$(x, y) \in F \Leftrightarrow y = F(x) \Leftrightarrow y = G(x) \Leftrightarrow (x, y) \in G \qquad \text{because } F(x) = G(x)$$

So F and G consist of exactly the same elements and hence $F = G$.

Conversely, if $F = G$, then for all $x \in X$,

$$y = F(x) \Leftrightarrow (x, y) \in F \Leftrightarrow (x, y) \in G \Leftrightarrow y = G(x) \qquad \text{because } F \text{ and } G \text{ consist of exactly the same elements}$$

Thus, since both $F(x)$ and $G(x)$ equal y, we have that

$$F(x) = G(x).$$

Example 7.1.3 Equality of Functions

a. Let $J_3 = \{0, 1, 2\}$, and define functions f and g from J_3 to J_3 as follows: For all x in J_3,

$$f(x) = (x^2 + x + 1) \text{ } mod \text{ } 3 \quad \text{and} \quad g(x) = (x+2)^2 \text{ } mod \text{ } 3.$$

Does $f = g$?

b. Let $F: \mathbf{R} \to \mathbf{R}$ and $G: \mathbf{R} \to \mathbf{R}$ be functions. Define new functions $F + G: \mathbf{R} \to \mathbf{R}$ and $G + F: \mathbf{R} \to \mathbf{R}$ as follows: For all $x \in \mathbf{R}$,

$$(F + G)(x) = F(x) + G(x) \quad \text{and} \quad (G + F)(x) = G(x) + F(x).$$

Does $F + G = G + F$?

Solution

a. Yes, the table of values shows that $f(x) = g(x)$ for all x in J_3.

x	x^2+x+1	$f(x) = (x^2+x+1)$ *mod* 3	$(x+2)^2$	$g(x) = (x+2)^2$ *mod* 3
0	1	1 *mod* 3 = 1	4	4 *mod* 3 = 1
1	3	3 *mod* 3 = 0	9	9 *mod* 3 = 0
2	7	7 *mod* 3 = 1	16	16 *mod* 3 = 1

b. Again the answer is yes. For all real numbers x,

$$\begin{aligned}(F + G)(x) &= F(x) + G(x) && \text{by definition of } F + G \\ &= G(x) + F(x) && \text{by the commutative law for addition of real numbers} \\ &= (G + F)(x) && \text{by definition of } G + F\end{aligned}$$

Hence $F + G = G + F$. ■

Examples of Functions

The following examples illustrate some of the wide variety of different types of functions.

Example 7.1.4 The Identity Function on a Set

Given a set X, define a function I_X from X to X by

$$I_X(x) = x \quad \text{for all } x \text{ in } X.$$

The function I_X is called the **identity function on X** because it sends each element of X to the element that is identical to it. Thus the identity function can be pictured as a machine that sends each piece of input directly to the output chute without changing it in any way.

Let X be any set and suppose that a_{ij}^k and $\phi(z)$ are elements of X. Find $I_X\left(a_{ij}^k\right)$ and $I_X(\phi(z))$.

Solution Whatever is input to the identity function comes out unchanged, so $I_X\left(a_{ij}^k\right) = a_{ij}^k$ and $I_X(\phi(z)) = \phi(z)$. ■

Example 7.1.5 Sequences

The formal definition of sequences specifies that an infinite sequence is a function defined on the set of integers that are greater than or equal to a particular integer. For example, the sequence denoted

$$1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \frac{1}{5}, \ldots, \frac{(-1)^n}{n+1}, \ldots$$

can be thought of as the function f from the nonnegative integers to the real numbers that associates $0 \to 1$, $1 \to -\frac{1}{2}$, $2 \to \frac{1}{3}$, $3 \to -\frac{1}{4}$, $4 \to \frac{1}{5}$, and, in general, $n \to \frac{(-1)^n}{n+1}$. In other words, $f\colon \mathbf{Z}^{nonneg} \to \mathbf{R}$ is the function defined as follows:

$$\text{Send each integer } n \geq 0 \text{ to } f(n) = \frac{(-1)^n}{n+1}.$$

In fact, there are many functions that can be used to define a given sequence. For instance, express the sequence above as a function from the set of *positive* integers to the set of real numbers.

Solution Define $g\colon \mathbf{Z}^+ \to \mathbf{R}$ by $g(n) = \frac{(-1)^{n+1}}{n}$, for each $n \in \mathbf{Z}^+$. Then $g(1) = 1$, $g(2) = -\frac{1}{2}$, $g(3) = \frac{1}{3}$, and in general

$$g(n+1) = \frac{(-1)^{n+2}}{n+1} = \frac{(-1)^n}{n+1} = f(n).$$ ■

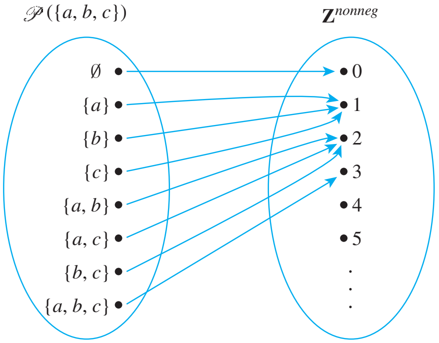
Example 7.1.6 A Function Defined on a Power Set

Recall from Section 6.1 that $\mathscr{P}(A)$ denotes the set of all subsets of the set A. Define a function $F\colon \mathscr{P}(\{a, b, c\}) \to \mathbf{Z}^{nonneg}$ as follows: For each $X \in \mathscr{P}(\{a, b, c\})$,

$$F(X) = \text{the number of elements in } X.$$

Draw an arrow diagram for F.

Solution



■

Example 7.1.7 Functions Defined on a Cartesian Product

Note It is customary to omit one set of parentheses when referring to functions defined on Cartesian products. For example, we write $M(a, b)$ rather than $M((a, b))$.

Define functions $M: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ and $R: \mathbf{R} \times \mathbf{R} \to \mathbf{R} \times \mathbf{R}$ as follows: For all ordered pairs (a, b) of integers,

$$M(a, b) = ab \quad \text{and} \quad R(a, b) = (-a, b).$$

Then M is the multiplication function that sends each pair of real numbers to the product of the two, and R is the reflection function that sends each point in the plane that corresponds to a pair of real numbers to the mirror image of the point across the vertical axis. Find the following:

a. $M(-1, -1)$ b. $M\left(\frac{1}{2}, \frac{1}{2}\right)$ c. $M(\sqrt{2}, \sqrt{2})$

d. $R(2, 5)$ e. $R(-2, 5)$ f. $R(3, -4)$

Solution

a. $(-1)(-1) = 1$ b. $(1/2)(1/2) = 1/4$ c. $\sqrt{2} \cdot \sqrt{2} = 2$

d. $(-2, 5)$ e. $(-(-2), 5) = (2, 5)$ f. $(-3, -4)$ ■

Note It is not obvious, but it is true, that for any positive real number x there is a unique real number y such that $b^y = x$. Most calculus books contain a discussion of this result.

• Definition Logarithms and Logarithmic Functions

Let b be a positive real number with $b \neq 1$. For each positive real number x, the **logarithm with base b of x,** written $\log_b x$, is the exponent to which b must be raised to obtain x. Symbolically,

$$\log_b x = y \quad \Leftrightarrow \quad b^y = x.$$

The **logarithmic function with base b** is the function from $\mathbf{R}^+$ to $\mathbf{R}$ that takes each positive real number x to $\log_b x$.

Example 7.1.8 The Logarithmic Function with Base b

Find the following:

a. $\log_3 9$ b. $\log_2\left(\frac{1}{2}\right)$ c. $\log_{10}(1)$ d. $\log_2(2^m)$ (m is any real number)

e. $2^{\log_2 m}$ $(m > 0)$

Solution

a. $\log_3 9 = 2$ because $3^2 = 9$. b. $\log_2 \left(\frac{1}{2}\right) = -1$ because $2^{-1} = \frac{1}{2}$.

c. $\log_{10}(1) = 0$ because $10^0 = 1$.

d. $\log_2(2^m) = m$ because the exponent to which 2 must be raised to obtain 2^m is m.

e. $2^{\log_2 m} = m$ because $\log_2 m$ is the exponent to which 2 must be raised to obtain m. ■

Recall from Section 5.9 that if S is a nonempty, finite set of characters, then a **string over *S*** is a finite sequence of elements of S. The number of characters in a string is called the **length** of the string. The **null string over *S*** is the "string" with no characters. It is usually denoted ϵ and is said to have length 0.

Example 7.1.9 Encoding and Decoding Functions

Digital messages consist of finite sequences of 0's and 1's. When they are communicated across a transmission channel, they are frequently coded in special ways to reduce the possibility that they will be garbled by interfering noise in the transmission lines. For example, suppose a message consists of a sequence of 0's and 1's. A simple way to encode the message is to write each bit three times. Thus the message

$$00101111$$

would be encoded as

$$000000111000111111111111.$$

The receiver of the message decodes it by replacing each section of three identical bits by the one bit to which all three are equal.

Let A be the set of all strings of 0's and 1's, and let T be the set of all strings of 0's and 1's that consist of consecutive triples of identical bits. The encoding and decoding processes described above are actually functions from A to T and from T to A. The encoding function E is the function from A to T defined as follows: For each string $s \in A$,

$$\begin{aligned} E(s) = {} & \text{the string obtained from } s \text{ by replacing each} \\ & \text{bit of } s \text{ by the same bit written three times.} \end{aligned}$$

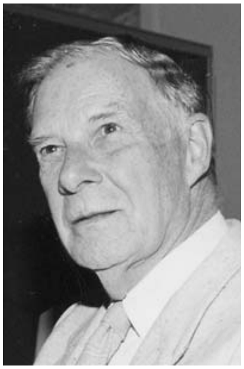
The decoding function D is defined as follows: For each string $t \in T$,

$$\begin{aligned} D(t) = {} & \text{the string obtained from } t \text{ by replacing each consecutive} \\ & \text{triple of three identical bits of } t \text{ by a single copy of that bit.} \end{aligned}$$

The advantage of this particular coding scheme is that it makes it possible to do a certain amount of error correction when interference in the transmission channels has introduced errors into the stream of bits. If the receiver of the coded message observes that one of the sections of three consecutive bits that should be identical does not consist of identical bits, then one bit differs from the other two. In this case, if errors are rare, it is likely that the single bit that is different is the one in error, and this bit is changed to agree with the other two before decoding. ■

Example 7.1.10 The Hamming Distance Function

The Hamming distance function, named after the computer scientist Richard W. Hamming, is very important in coding theory. It gives a measure of the "difference" between two strings of 0's and 1's that have the same length. Let S_n be the set of all strings of 0's



Courtesy of U.S. Naval Academy

Richard Hamming (1915–1998)

and 1's of length n. Define a function $H\colon S_n \times S_n \to \mathbf{Z}^{nonneg}$ as follows: For each pair of strings $(s, t) \in S_n \times S_n$,

$$H(s, t) = \text{the number of positions in which } s \text{ and } t \text{ have different values.}$$

Thus, letting $n = 5$, $$H(11111, 00000) = 5$$

because 11111 and 00000 differ in all five positions, whereas

$$H(11000, 00000) = 2$$

because 11000 and 00000 differ only in the first two positions.

a. Find $H(00101, 01110)$. b. Find $H(10001, 01111)$.

Solution

a. 3 b. 4 ■

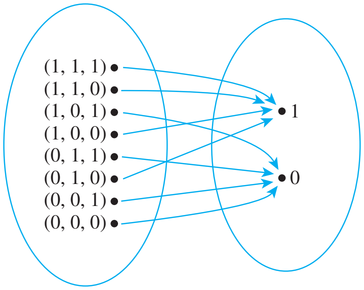
Boolean Functions

In Section 2.4 we showed how to find input/output tables for certain digital logic circuits. Any such input/output table defines a function in the following way: The elements in the input column can be regarded as ordered tuples of 0's and 1's; the set of all such ordered tuples is the domain of the function. The elements in the output column are all either 0 or 1; thus {0, 1} is taken to be the co-domain of the function. The relationship is that which sends each input element to the output element in the same row. Thus, for instance, the input/output table of Figure 7.1.4(a) defines the function with the arrow diagram shown in Figure 7.1.4(b).

More generally, the input/output table corresponding to a circuit with n input wires has n input columns. Such a table defines a function from the set of all n-tuples of 0's and 1's to the set {0, 1}.

Input			Output
P	***Q***	***R***	***S***
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	1
0	0	1	0
0	0	0	0

(a)



(b)

Figure 7.1.2 **Two Representations of a Boolean Function**

• Definition

An **(n-place) Boolean function** f is a function whose domain is the set of all ordered n-tuples of 0's and 1's and whose co-domain is the set {0, 1}. More formally, the domain of a Boolean function can be described as the Cartesian product of n copies of the set {0, 1}, which is denoted $\{0, 1\}^n$. Thus $f\colon \{0, 1\}^n \to \{0, 1\}$.

Example 7.1.11 A Boolean Function

Consider the three-place Boolean function defined from the set of all 3-tuples of 0's and 1's to {0, 1} as follows: For each triple (x_1, x_2, x_3) of 0's and 1's,

$$f(x_1, x_2, x_3) = (x_1 + x_2 + x_3)\ mod\ 2.$$

Describe f using an input/output table.

Solution

$$f(1, 1, 1) = (1 + 1 + 1)\ mod\ 2 = 3\ mod\ 2 = 1$$
$$f(1, 1, 0) = (1 + 1 + 0)\ mod\ 2 = 2\ mod\ 2 = 0$$

The rest of the values of f can be calculated similarly to obtain the following table.

Input			Output
x_1	x_2	x_3	$(x_1 + x_2 + x_3)\ mod\ 2$
1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	1
0	0	1	1
0	0	0	0

■

Checking Whether a Function Is Well Defined

It can sometimes happen that what appears to be a function defined by a rule is not really a function at all. To give an example, suppose we wrote, "Define a function $f: \mathbf{R} \to \mathbf{R}$ by specifying that for all real numbers x,

$$f(x) \text{ is the real number } y \text{ such that } x^2 + y^2 = 1.$$

There are two distinct reasons why this description does not define a function. For almost all values of x, either (1) there is no y that satisfies the given equation or (2) there are two different values of y that satisfy the equation. For instance, when $x = 2$, there is no real number y such that $2^2 + y^2 = 1$, and when $x = 0$, both $y = -1$ and $y = 1$ satisfy the equation $0^2 + y^2 = 1$. In general, we say that a "function" is **not well defined** if it fails to satisfy at least one of the requirements for being a function.

Example 7.1.12 A Function That Is Not Well Defined

Recall that **Q** represents the set of all rational numbers. Suppose you read that a function $f: \mathbf{Q} \to \mathbf{Z}$ is to be defined by the formula

$$f\left(\frac{m}{n}\right) = m \quad \text{for all integers } m \text{ and } n \text{ with } n \neq 0.$$

That is, the integer associated by f to the number $\frac{m}{n}$ is m. Is f well defined? Why?

Solution The function f is not well defined. The reason is that fractions have more than one representation as quotients of integers. For instance, $\frac{1}{2} = \frac{3}{6}$. Now if f were a function, then the definition of a function would imply that $f\left(\frac{1}{2}\right) = \left(\frac{3}{6}\right)$ since $\frac{1}{2} = \frac{3}{6}$. But applying the formula for f, you find that

$$f\left(\frac{1}{2}\right) = 1 \quad \text{and} \quad f\left(\frac{3}{6}\right) = 3,$$

and so

$$f\left(\frac{1}{2}\right) \neq f\left(\frac{3}{6}\right).$$

This contradiction shows that f is not well defined and, therefore, is not a function. ■

Note that the phrase *well-defined function* is actually redundant; for a function to be well defined really means that it is worthy of being called a function.

Functions Acting on Sets

Given a function from a set X to a set Y, you can consider the set of images in Y of all the elements in a subset of X and the set of inverse images in X of all the elements in a subset of Y.

Note For $y \in Y$, $f^{-1}(y) = f^{-1}(\{y\})$.

• Definition

If $f: X \to Y$ is a function and $A \subseteq X$ and $C \subseteq Y$, then

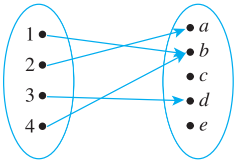
$$f(A) = \{y \in Y \mid y = f(x) \text{ for some } x \text{ in } A\}$$

and
$$f^{-1}(C) = \{x \in X \mid f(x) \in C\}.$$

$f(A)$ is called the **image of A**, and $f^{-1}(C)$ is called the **inverse image of C**.

Example 7.1.13 The Action of a Function on Subsets of a Set

Let $X = \{1, 2, 3, 4\}$ and $Y = \{a, b, c, d, e\}$, and define $F\colon\ X \to Y$ by the following arrow diagram:



Let $A = \{1, 4\}$, $C = \{a, b\}$, and $D = \{c, e\}$. Find $F(A)$, $F(X)$, $F^{-1}(C)$, and $F^{-1}(D)$.

Solution

$F(A) = \{b\} \quad F(X) = \{a, b, d\} \quad F^{-1}(C) = \{1, 2, 4\} \quad F^{-1}(D) = \emptyset$ ■

Example 7.1.14 Interaction of a Function with Union

Let X and Y be sets, let F be a function from X to Y, and let A and B be any subsets of X. Prove that $F(A \cup B) \subseteq F(A) \cup F(B)$.

Solution

The fact that X, Y, F, A, and B were formally introduced prior to the word "Prove" allows you to regard their existence and relationships as part of your background knowledge. Thus to prove that $F(A \cup B) \subseteq F(A) \cup F(B)$, you only need show that if y is any element in $F(A \cup B)$, then y is an element of $F(A) \cup F(B)$.

Proof:

Suppose $y \in F(A \cup B)$. *[We must show that $y \in F(A) \cup F(B)$.]* By definition of function, $y = F(x)$ for some $x \in A \cup B$. By definition of union, $x \in A$ or $x \in B$.

Case 1, $x \in A$: In this case, $y = F(x)$ for some x in A. Hence $y \in F(A)$, and so by definition of union, $y \in F(A) \cup F(B)$.

Case 2, $x \in B$: In this case, $y = F(x)$ for some x in B. Hence $y \in F(B)$, and so by definition of union, $y \in F(A) \cup F(B)$.

Thus in either case $y \in F(A) \cup F(B)$ *[as was to be shown].* ■

Exercise 38 asks you to prove the opposite containment from the one in example 7.1.14. Taken together, the example and the solution to the exercise establish the full equality that $F(A \cup B) = F(A) \cup F(B)$.

Test Yourself

Answers to Test Yourself questions are located at the end of each section.

1. Given a function f from a set X to a set Y, $f(x)$ is _____.

2. Given a function f from a set X to a set Y, if $f(x) = y$, then y is called _____ or _____ or _____.

3. Given a function f from a set X to a set Y, the range of f (or the image of X under f) is _____.

4. Given a function f from a set X to a set Y, if $f(x) = y$, then x is called _____ or _____.

5. Given a function f from a set X to a set Y, if $y \in Y$, then $f^{-1}(y) =$ _____ and is called _____.

6. Given functions f and g from a set X to a set Y, $f = g$ if, and only if, _____.

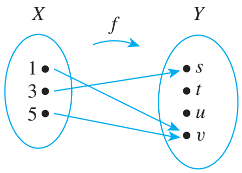
7. Given positive real numbers x and b with $b \neq 1$, $\log_b x =$ _____.

8. Given a function f from a set X to a set Y and a subset A of X, $f(A) =$ _____.

9. Given a function f from a set X to a set Y and a subset C of Y, $f^{-1}(C) =$ _____.

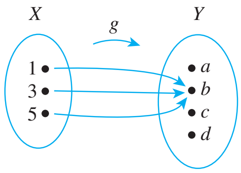
Exercise Set 7.1*

1. Let $X = \{1, 3, 5\}$ and $Y = \{s, t, u, v\}$. Define $f\colon X \to Y$ by the following arrow diagram.



a. Write the domain of f and the co-domain of f.
b. Find $f(1)$, $f(3)$, and $f(5)$.
c. What is the range of f?
d. Is 3 an inverse image of s? Is 1 an inverse image of u?
e. What is the inverse image of s? of u? of v?
f. Represent f as a set of ordered pairs.

2. Let $X = \{1, 3, 5\}$ and $Y = \{a, b, c, d\}$. Define $g\colon X \to Y$ by the following arrow diagram.



a. Write the domain of g and the co-domain of g.
b. Find $g(1)$, $g(3)$, and $g(5)$.
c. What is the range of g?
d. Is 3 an inverse image of a? Is 1 an inverse image of b?
e. What is the inverse image of b? of c?
f. Represent g as a set of ordered pairs.

*For exercises with blue numbers or letters, solutions are given in Appendix B. The symbol ***H*** indicates that only a hint or a partial solution is given. The symbol ✶ signals that an exercise is more challenging than usual.

3. Indicate whether the statements in parts (a)–(d) are true or false. Justify your answers.
   a. If two elements in the domain of a function are equal, then their images in the co-domain are equal.
   b. If two elements in the co-domain of a function are equal, then their preimages in the domain are also equal.
   c. A function can have the same output for more than one input.
   d. A function can have the same input for more than one output.

4. **a.** Find all functions from $X = \{a, b\}$ to $Y = \{u, v\}$.
   b. Find all functions from $X = \{a, b, c\}$ to $Y = \{u\}$.
   c. Find all functions from $X = \{a, b, c\}$ to $Y = \{u, v\}$.

5. Let $I_{\mathbf{Z}}$ be the identity function defined on the set of all integers, and suppose that $e, b_i^{jk}, K(t)$, and u_{kj} all represent integers. Find
   a. $I_{\mathbf{Z}}(e)$ **b.** $I_{\mathbf{Z}}\left(b_i^{jk}\right)$ c. $I_{\mathbf{Z}}(K(t))$ d. $I_{\mathbf{Z}}(u_{kj})$

6. Find functions defined on the set of nonnegative integers that define the sequences whose first six terms are given below.
   a. $1, -\frac{1}{3}, \frac{1}{5}, -\frac{1}{7}, \frac{1}{9}, -\frac{1}{11}$ b. $0, -2, 4, -6, 8, -10$

7. Let $A = \{1, 2, 3, 4, 5\}$ and define a function $F\colon \mathscr{P}(A) \to \mathbf{Z}$ as follows: For all sets X in $\mathscr{P}(A)$,
$$F(X) = \begin{cases} 0 & \text{if } X \text{ has an even number of elements} \\ 1 & \text{if } X \text{ has an odd number of elements.} \end{cases}$$
   Find the following:
   a. $F(\{1, 3, 4\})$ b. $F(\emptyset)$
   c. $F(\{2, 3\})$ d. $F(\{2, 3, 4, 5\})$

8. Let $J_5 = \{0, 1, 2, 3, 4\}$, and define a function $F\colon J_5 \to J_5$ as follows: For each $x \in J_5$, $F(x) = (x^3 + 2x + 4)\ mod\ 5$. Find the following:
   a. $F(0)$ **b.** $F(1)$ c. $F(2)$ d. $F(3)$ e. $F(4)$

9. Define a function $S\colon \mathbf{Z}^+ \to \mathbf{Z}^+$ as follows: For each positive integer n,
$$S(n) = \text{the sum of the positive divisors of } n.$$
   Find the following:
   a. $S(1)$ **b.** $S(15)$ **c.** $S(17)$
   d. $S(5)$ e. $S(18)$ f. $S(21)$

10. Let D be the set of all finite subsets of positive integers. Define a function $T\colon \mathbf{Z}^+ \to D$ as follows: For each positive integer n, $T(n) =$ the set of positive divisors of n.
   Find the following:
   a. $T(1)$ **b.** $T(15)$ **c.** $T(17)$
   d. $T(5)$ e. $T(18)$ f. $T(21)$

11. Define $F\colon \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z} \times \mathbf{Z}$ as follows: For all ordered pairs (a, b) of integers, $F(a, b) = (2a + 1, 3b - 2)$.
   Find the following:
   a. $F(4, 4)$ **b.** $F(2, 1)$ c. $F(3, 2)$ d. $F(1, 5)$

12. Define $G\colon J_5 \times J_5 \to J_5 \times J_5$ as follows: For all $(a, b) \in J_5 \times J_5$,
$$G(a, b) = ((2a + 1)\ mod\ 5, (3b - 2)\ mod\ 5).$$
   Find the following:
   a. $G(4, 4)$ **b.** $G(2, 1)$ c. $G(3, 2)$ d. $G(1, 5)$

13. Let $J_5 = \{0, 1, 2, 3, 4\}$, and define functions $f\colon J_5 \to J_5$ and $g\colon J_5 \to J_5$ as follows: For each $x \in J_5$,
$$f(x) = (x + 4)^2\ mod\ 5 \quad\text{and}\quad g(x) = (x^2 + 3x + 1)\ mod\ 5.$$
   Is $f = g$? Explain.

14. Let $J_5 = \{0, 1, 2, 3, 4\}$, and define functions $h\colon J_5 \to J_5$ and $k\colon J_5 \to J_5$ as follows: For each $x \in J_5$,
$$h(x) = (x + 3)^3\ mod\ 5 \quad\text{and}\quad k(x) = (x^3 + 4x^2 + 2x + 2)\ mod\ 5.$$
   Is $h = k$? Explain.

15. Let F and G be functions from the set of all real numbers to itself. Define the product functions $F \cdot G\colon \mathbf{R} \to \mathbf{R}$ and $G \cdot F\colon \mathbf{R} \to \mathbf{R}$ as follows: For all $x \in \mathbf{R}$,
$$(F \cdot G)(x) = F(x) \cdot G(x)$$
$$(G \cdot F)(x) = G(x) \cdot F(x)$$
   Does $F \cdot G = G \cdot F$? Explain.

16. Let F and G be functions from the set of all real numbers to itself. Define new functions $F - G\colon \mathbf{R} \to \mathbf{R}$ and $G - F\colon \mathbf{R} \to \mathbf{R}$ as follows: For all $x \in \mathbf{R}$,
$$(F - G)(x) = F(x) - G(x)$$
$$(G - F)(x) = G(x) - F(x)$$
   Does $F - G = G - F$? Explain.

17. Use the definition of logarithm to fill in the blanks below.
   a. $\log_2 8 = 3$ because ______.
   b. $\log_5\left(\frac{1}{25}\right) = 2$ because ______.
   c. $\log_4 4 = 1$ because ______.
   d. $\log_3(3^n) = n$ because ______.
   e. $\log_4 1 = 0$ because ______.

18. Find exact values for each of the following quantities. Do not use a calculator.
   a. $\log_3 81$ b. $\log_2 1024$ **c.** $\log_3\left(\frac{1}{27}\right)$ d. $\log_2 1$
   e. $\log_{10}\left(\frac{1}{10}\right)$ f. $\log_3 3$ g. $\log_2(2^k)$

19. Use the definition of logarithm to prove that for any positive real number b with $b \neq 1$, $\log_b b = 1$.

20. Use the definition of logarithm to prove that for any positive real number b with $b \neq 1$, $\log_b 1 = 0$.

21. If b is any positive real number with $b \neq 1$ and x is any real number, b^{-x} is defined as follows: $b^{-x} = \frac{1}{b^x}$. Use this definition and the definition of logarithm to prove that $\log_b\left(\frac{1}{u}\right) = -\log_b(u)$ for all positive real numbers u and b, with $b \neq 1$.

H 22. Use the unique factorization for the integers theorem (Section 4.3) and the definition of logarithm to prove that $\log_3(7)$ is irrational.

23. If b and y are positive real numbers such that $\log_b y = 3$, what is $\log_{1/b}(y)$? Why?

24. If b and y are positive real numbers such that $\log_b y = 2$, what is $\log_{b^2}(y)$? Why?

25. Let $A = \{2, 3, 5\}$ and $B = \{x, y\}$. Let p_1 and p_2 be the **projections of $A \times B$ onto the first and second coordinates.** That is, for each pair $(a, b) \in A \times B$, $p_1(a, b) = a$ and $p_2(a, b) = b$.

a. Find $p_1(2, y)$ and $p_1(5, x)$. What is the range of p_1?
b. Find $p_2(2, y)$ and $p_2(5, x)$. What is the range of p_2?

26. Observe that *mod* and *div* can be defined as functions from $\mathbf{Z}^{nonneg} \times \mathbf{Z}^+$ to $\mathbf{Z}$. For each ordered pair (n, d) consisting of a nonnegative integer n and a positive integer d, let

$$mod(n, d) = n \ mod \ d \text{ (the nonnegative remainder obtained when } n \text{ is divided by } d\text{).}$$

$$div(n, d) = n \ div \ d \text{ (the integer quotient obtained when } n \text{ is divided by } d\text{).}$$

Find each of the following:

a. *mod* (67, 10) and *div* (67, 10)
b. *mod* (59, 8) and *div* (59, 8)
c. *mod* (30, 5) and *div* (30, 5)

27. Let S be the set of all strings of a's and b's.

a. Define $f: S \to Z$ as follows: For each string s in S

$$f(s) \begin{cases} \text{the number of } b\text{'s to the left of the left-most } a \text{ in } s \\ 0 \quad \text{if } s \text{ contains no } a\text{'s.} \end{cases}$$

Find $f(aba)$, $f(bbab)$ and $f(b)$. What is the range of f?

b. Define $g: S \to S$ as follows: For each string s in S,

$$g(s) = \text{the string obtained by writing the characters of } s \text{ in reverse order.}$$

Find $g(aba)$, $g(bbab)$, and $g(b)$. What is the range of g?

28. Consider the coding and decoding functions E and D defined in Example 7.1.9.

a. Find $E(0110)$ and $D(111111000111)$.
b. Find $E(1010)$ and $D(000000111111)$.

29. Consider the Hamming distance function defined in Example 7.1.10.

a. Find $H(10101, 00011)$
b. Find $H(00110, 10111)$.

30. Draw arrow diagrams for the Boolean functions defined by the following input/output tables.

a.

Input		**Output**
P	***Q***	***R***
1	1	0
1	0	1
0	1	0
0	0	1

b.

Input			**Output**
P	***Q***	***R***	***S***
1	1	1	1
1	1	0	0
1	0	1	1
1	0	0	1
0	1	1	0
0	1	0	0
0	0	1	0
0	0	0	1

31. Fill in the following table to show the values of all possible two-place Boolean functions.

Input	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
1 1																
1 0																
0 1																
0 0																

32. Consider the three-place Boolean function f defined by the following rule: For each triple (x_1, x_2, x_3) of 0's and 1's,

$$f(x_1, x_2, x_3) = (4x_1 + 3x_2 + 2x_3) \ mod \ 2.$$

a. Find $f(1, 1, 1)$ and $f(0, 0, 1)$.
b. Describe f using an input/output table.

33. Student A tries to define a function $g: \mathbf{Q} \to \mathbf{Z}$ by the rule

$$g\left(\frac{m}{n}\right) = m - n, \text{ for all integers } m \text{ and } n \text{ with } n \neq 0.$$

Student B claims that g is not well defined. Justify student B's claim.

34. Student C tries to define a function $h: \mathbf{Q} \to \mathbf{Q}$ by the rule

$$h\left(\frac{m}{n}\right) = \frac{m^2}{n}, \text{ for all integers } m \text{ and } n \text{ with } n \neq 0.$$

Student D claims that h is not well defined. Justify student D's claim.

35. Let $J_5 = \{0, 1, 2, 3, 4\}$. Then $J_5 - \{0\} = \{1, 2, 3, 4\}$. Student A tries to define a function $R: J_5 - \{0\} \to J_5 - \{0\}$ as follows: For each $x \in J_5 - \{0\}$,

$$R(x) \text{ is the number } y \text{ so that } (xy)\ mod\ 5 = 1.$$

Student B claims that R is not well defined. Who is right: student A or student B? Justify your answer.

36. Let $J_4 = \{0, 1, 2, 3\}$. Then $J_4 - \{0\} = \{1, 2, 3\}$. Student C tries to define a function $S: J_4 - \{0\} \to J_4 - \{0\}$ as follows: For each $x \in J_4 - \{0\}$,

$$S(x) \text{ is the number } y \text{ so that } (xy)\ mod\ 4 = 1.$$

Student F claims that S is not well defined. Who is right: student C or student D? Justify your answer.

37. On certain computers the integer data type goes from $-2,147,483,648$ through $2,147,483,647$. Let S be the set of all integers from $-2,147,483,648$ through $2,147,483,647$. Try to define a function $f: S \to S$ by the rule $f(n) = n^2$ for each n in S. Is f well defined? Why?

38. Let $X = \{a, b, c\}$ and $Y = \{r, s, t, u, v, w\}$. Define $f: X \to Y$ as follows: $f(a) = v$, $f(b) = v$, and $f(c) = t$.
a. Draw an arrow diagram for f.
b. Let $A = \{a, b\}$, $C = \{t\}$, $D = \{u, v\}$, and $E = \{r, s\}$. Find $f(A)$, $f(X)$, $f^{-1}(C)$, $f^{-1}(D)$, $f^{-1}(E)$, and $f^{-1}(Y)$.

39. Let $X = \{1, 2, 3, 4\}$ and $Y = \{a, b, c, d, e\}$. Define $g: X \to Y$ as follows: $g(1) = a$, $g(2) = a$, $g(3) = a$, and $g(4) = d$.
a. Draw an arrow diagram for g.
b. Let $A = \{2, 3\}$, $C = \{a\}$, and $D = \{b, c\}$. Find $g(A)$, $g(X)$, $g^{-1}(C)$, $g^{-1}(D)$, and $g^{-1}(Y)$.

H **40.** Let X and Y be sets, let A and B be any subsets of X, and let F be a function from X to Y. Fill in the blanks in the following proof that $F(A) \cup F(B) \subseteq F(A \cup B)$.

Proof: Let y be any element in $F(A) \cup F(B)$. *[We must show that y is in $F(A \cup B)$.]* By definition of union, $\underline{(a)}$.

Case 1, $y \in F(A)$: In this case, by definition of $F(A)$, $y = F(x)$ for $\underline{(b)}$ $x \in A$. Since $A \subseteq A \cup B$, it follows from the definition of union that $x \in \underline{(c)}$. Hence, $y = F(x)$ for some $x \in A \cup B$, and thus, by definition of $F(A \cup B)$, $y \in \underline{(d)}$.

Case 2, $y \in F(B)$: In this case, by definition of $F(B)$, $\underline{(e)}$ $x \in B$. Since $B \subseteq A \cup B$ it follows from the definition of union that $\underline{(f)}$.

Therefore, regardless of whether $y \in F(A)$ or $y \in F(B)$, we have that $y \in F(A \cup B)$ *[as was to be shown]*.

In 41–49 let X and Y be sets, let A and B be any subsets of X, and let C and D be any subsets of Y. Determine which of the properties are true for all functions F from X to Y and which are false for at least one function F from X to Y. Justify your answers.

41. If $A \subseteq B$ then $F(A) \subseteq F(B)$.

42. $F(A \cap B) \subseteq F(A) \cap F(B)$

43. $F(A) \cap F(B) \subseteq F(A \cap B)$

44. For all subsets A and B of X, $F(A - B) = F(A) - F(B)$.

45. For all subsets C and D of Y, if $C \subseteq D$, then

$$F^{-1}(C) \subseteq F^{-1}(D).$$

H **46.** For all subsets C and D of Y,

$$F^{-1}(C \cup D) = F^{-1}(C) \cup F^{-1}(D).$$

47. For all subsets C and D of Y,

$$F^{-1}(C \cap D) = F^{-1}(C) \cap F^{-1}(D).$$

48. For all subsets C and D of Y,

$$F^{-1}(C - D) = F^{-1}(C) - F^{-1}(D).$$

49. $F(F^{-1}(C)) \subseteq C$

50. Given a set S and a subset A, the **characteristic function of A,** denoted χ_A, is the function defined from S to $\mathbf{Z}$ with the property that for all $u \in S$,

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

Show that each of the following holds for all subsets A and B of S and all $u \in S$.
a. $\chi_{A \cap B}(u) = \chi_A(u) \cdot \chi_B(u)$
b. $\chi_{A \cup B}(u) = \chi_A(u) + \chi_B(u) - \chi_A(u) \cdot \chi_B(u)$

Each of exercises 51–53 refers to the Euler phi function, denoted ϕ, which is defined as follows: For each integer $n \geq 1$, $\phi(n)$ is the number of positive integers less than or equal to n that have no common factors with n except ±1. For example, $\phi(10) = 4$ because there are four positive integers less than or equal to 10 that have no common factors with 10 except ±1; namely, 1, 3, 7, and 9.

51. Find each of the following:
a. $\phi(15)$ **b.** $\phi(2)$ **c.** $\phi(5)$
d. $\phi(12)$ e. $\phi(11)$ f. $\phi(1)$

✶ **52.** Prove that if p is a prime number and n is an integer with $n \geq 1$, then $\phi(p^n) = p^n - p^{n-1}$.

H **53.** Prove that there are infinitely many integers n for which $\phi(n)$ is a perfect square.

Answers for Test Yourself

1. the unique output element in Y that is related to x by f 2. the value of f at x; the image of x under f; the output of f for the input x 3. the set of all y in Y such that $f(x) = y$ 4. an inverse image of y under f; a preimage of y 5. $\{x \in X \mid f(x) = y\}$; the inverse image of y 6. $f(x) = g(x)$ for all $x \in X$ 7. the exponent to which b must be raised to obtain x (*Or*: the real number y such that $x = b^y$) 8. $\{y \in Y \mid y = f(x)$ for some $x \in A\}$ (*Or*: $\{f(x) \mid x \in A\}$) 9. $\{x \in X \mid f(x) \in C\}$

7.2 One-to-One and Onto, Inverse Functions

Don't accept a statement just because it is printed. — Anna Pell Wheeler, 1883–1966

In this section we discuss two important properties that functions may satisfy: the property of being *one-to-one* and the property of being *onto*. Functions that satisfy both properties are called *one-to-one correspondences* or *one-to-one onto functions*. When a function is a one-to-one correspondence, the elements of its domain and co-domain match up perfectly, and we can define an *inverse function* from the co-domain to the domain that "undoes" the action of the function.

One-to-One Functions

In Section 7.1 we noted that a function may send several elements of its domain to the same element of its co-domain. In terms of arrow diagrams, this means that two or more arrows that start in the domain can point to the same element in the co-domain. On the other hand, if no two arrows that start in the domain point to the same element of the co-domain then the function is called *one-to-one* or *injective*. For a one-to-one function, each element of the range is the image of at most one element of the domain.

• **Definition**

Let F be a function from a set X to a set Y. F is **one-to-one** (or **injective**) if, and only if, for all elements x_1 and x_2 in X,

$$\text{if } F(x_1) = F(x_2), \text{ then } x_1 = x_2,$$

or, equivalently,

$$\text{if } x_1 \neq x_2, \text{ then } F(x_1) \neq F(x_2).$$

Symbolically,

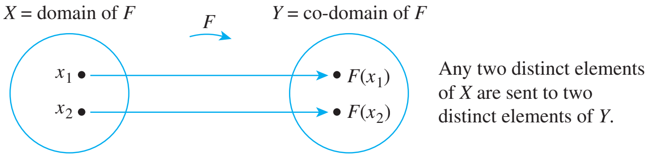
$$F\colon X \to Y \text{ is one-to-one} \quad \Leftrightarrow \quad \forall x_1, x_2 \in X, \text{if } F(x_1) = F(x_2) \text{ then } x_1 = x_2.$$

To obtain a precise statement of what it means for a function *not* to be one-to-one, take the negation of one of the equivalent versions of the definition above. Thus:

A function $F\colon X \to Y$ is *not* one-to-one $\Leftrightarrow$ $\exists$ elements x_1 and x_2 in X with $F(x_1) = F(x_2)$ and $x_1 \neq x_2$.

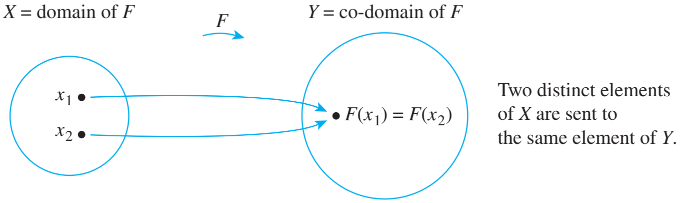
That is, if elements x_1 and x_2 can be found that have the same function value but are not equal, then F is not one-to-one.

In terms of arrow diagrams, a one-to-one function can be thought of as a function that separates points. That is, it takes distinct points of the domain to distinct points of the co-domain. A function that is not one-to-one fails to separate points. That is, at least two points of the domain are taken to the same point of the co-domain. This is illustrated in Figure 7.2.1 on the next page.



X = domain of F
F
Y = co-domain of F
x_1
x_2
$F(x_1)$
$F(x_2)$
Any two distinct elements of X are sent to two distinct elements of Y.

Figure 7.2.1(a) **A One-to-One Function Separates Points**



X = domain of F
F
Y = co-domain of F
x_1
x_2
$F(x_1) = F(x_2)$
Two distinct elements of X are sent to the same element of Y.

Figure 7.2.1(b) **A Function That Is Not One-to-One Collapses Points Together**

Example 7.2.1 Identifying One-to-One Functions Defined on Finite Sets

a. Do either of the arrow diagrams in Figure 7.2.2 define one-to-one functions?

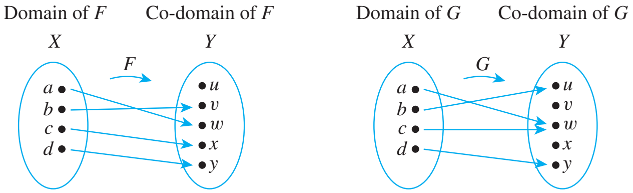


Figure 7.2.2

b. Let $X = \{1, 2, 3\}$ and $Y = \{a, b, c, d\}$. Define $H: X \to Y$ as follows: $H(1) = c$, $H(2) = a$, and $H(3) = d$. Define $K: X \to Y$ as follows: $K(1) = d$, $K(2) = b$, and $K(3) = d$. Is either H or K one-to-one?

Solution

a. F is one-to-one but G is not. F is one-to-one because no two different elements of X are sent by F to the same element of Y. G is not one-to-one because the elements a and c are both sent by G to the same element of Y: $G(a) = G(c) = w$ but $a \neq c$.

b. H is one-to-one but K is not. H is one-to-one because each of the three elements of the domain of H is sent by H to a different element of the co-domain: $H(1) \neq H(2)$, $H(1) \neq H(3)$, and $H(2) \neq H(3)$. K, however, is not one-to-one because $K(1) = K(3) = d$ but $1 \neq 3$. ■

Consider the problem of writing a computer algorithm to check whether a function F is one-to-one. If F is defined on a finite set and there is an independent algorithm to compute values of F, then an algorithm to check whether F is one-to-one can be written as follows: Represent the domain of F as a one-dimensional array $a[1], a[2], \ldots, a[n]$ and use a nested loop to examine all possible pairs $(a[i], a[j])$, where $i < j$. If there is a pair $(a[i], a[j])$ for which $F(a[i]) = F(a[j])$ and $a[i] \neq a[j]$, then F is not one-to-one. If, however, all pairs have been examined without finding such a pair, then F is one-to-one. You are asked to write such an algorithm in exercise 57 at the end of this section.

One-to-One Functions on Infinite Sets

Now suppose f is a function defined on an infinite set X. By definition, f is one-to-one if, and only if, the following universal statement is true:

$$\forall x_1, x_2 \in X, \text{if } f(x_1) = f(x_2) \text{ then } x_1 = x_2.$$

Thus, to prove f is one-to-one, you will generally use the method of direct proof:

suppose x_1 and x_2 are elements of X such that $f(x_1) = f(x_2)$

and **show** that $x_1 = x_2$.

To show that f is *not* one-to-one, you will ordinarily

find elements x_1 and x_2 in X so that $f(x_1) = f(x_2)$ but $x_1 \neq x_2$.

Example 7.2.2 Proving or Disproving That Functions Are One-to-One

Define $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{Z} \to \mathbf{Z}$ by the rules

$$f(x) = 4x - 1 \quad \text{for all} \quad x \in \mathbf{R}$$

and

$$g(n) = n^2 \quad \text{for all} \quad n \in \mathbf{Z}.$$

a. Is f one-to-one? Prove or give a counterexample.

b. Is g one-to-one? Prove or give a counterexample.

Solution It is usually best to start by taking a positive approach to answering questions like these. Try to prove the given functions are one-to-one and see whether you run into difficulty. If you finish without running into any problems, then you have a proof. If you do encounter a problem, then analyzing the problem may lead you to discover a counterexample.

a. The function $f: \mathbf{R} \to \mathbf{R}$ is defined by the rule

$$\boxed{f(x) = 4x - 1 \quad \text{for all real numbers } x.}$$

To prove that f is one-to-one, you need to prove that

$$\forall \text{ real numbers } x_1 \text{ and } x_2, \text{if } f(x_1) = f(x_2) \text{ then } x_1 = x_2.$$

Substituting the definition of f into the outline of a direct proof, you

suppose x_1 and x_2 are any real numbers such that $4x_1 - 1 = 4x_2 - 1$,

and **show** that $x_1 = x_2$.

Can you reach what is to be shown from the supposition? Of course. Just add 1 to both sides of the equation in the supposition and then divide both sides by 4.

This discussion is summarized in the following formal answer.

Answer to (a):

If the function $f: \mathbf{R} \to \mathbf{R}$ is defined by the rule $f(x) = 4x - 1$, for all real numbers x, then f is one-to-one.

Proof:

Suppose x_1 and x_2 are real numbers such that $f(x_1) = f(x_2)$. *[We must show that* $x_1 = x_2$.*]* By definition of f,

$$4x_1 - 1 = 4x_2 - 1.$$

Adding 1 to both sides gives

$$4x_1 = 4x_2,$$

and dividing both sides by 4 gives

$$x_1 = x_2,$$

which is what was to be shown.

b. The function $g: \mathbf{Z} \to \mathbf{Z}$ is defined by the rule

$$\boxed{g(n) = n^2 \quad \text{for all integers } n.}$$

As above, you start as though you were going to prove that g is one-to-one. Substituting the definition of g into the outline of a direct proof, you

suppose n_1 and n_2 are integers such that $n_1^2 = n_2^2$,

and **try to show** that $n_1 = n_2$.

Can you reach what is to be shown from the supposition? No! It is quite possible for two numbers to have the same squares and yet be different. For example, $2^2 = (-2)^2$ but $2 \neq -2$.

Thus, in trying to prove that g is one-to-one, you run into difficulty. But analyzing this difficulty leads to the discovery of a counterexample, which shows that g is not one-to-one.

This discussion is summarized as follows:

Answer to (b):

If the function $g: \mathbf{Z} \to \mathbf{Z}$ is defined by the rule $g(n) = n^2$, for all $n \in \mathbf{Z}$, then g is not one-to-one.

Counterexample:

Let $n_1 = 2$ and $n_2 = -2$. Then by definition of g,

$$g(n_1) = g(2) = 2^2 = 4 \quad \text{and also}$$
$$g(n_2) = g(-2) = (-2)^2 = 4.$$

Hence $$g(n_1) = g(n_2) \quad \text{but} \quad n_1 \neq n_2,$$

and so g is not one-to-one.

■

Application: Hash Functions

Imagine a set of student records, each of which includes the student's social security number, and suppose the records are to be stored in a table in which a record can be located if the social security number is known. One way to do this would be to place the record with social security number n into position n of the table. However, since social security numbers have nine digits, this method would require a table with 999,999,999 positions. The problem is that creating such a table for a small set of records would be very wasteful of computer memory space. **Hash functions** are functions defined from larger to smaller sets of integers, frequently using the *mod* function, which provide part of the solution to this problem. We illustrate how to define and use a *hash* function with a very simple example.

Example 7.2.3 A Hash Function

Suppose there are no more than seven student records. Define a function $Hash$ from the set of all social security numbers (ignoring hyphens) to the set {0, 1, 2, 3, 4, 5, 6} as follows:

$$Hash(n) = n \; mod \; 7 \quad \text{for all social security numbers } n.$$

To use your calculator to find $n \; mod \; 7$, use the formula $n \; mod \; 7 = n - 7 \cdot (n \; div \; 7)$. (See Section 4.4.) In other words, divide n by 7, multiply the integer part of the result by 7, and subtract that number from n. For instance, since $328343419/7 = 46906202.71\ldots$,

$$Hash(328\text{-}34\text{-}3419) = 328343419 - (7 \cdot 46906202) = 5.$$

Table 7.2.1

0	356-63-3102
1	
2	513-40-8716
3	223-79-9061
4	
5	328-34-3419
6	

As a first approximation to solving the problem of storing the records, try to place the record with social security number n in position $Hash(n)$. For instance, if the social security numbers are 328-34-3419, 356-63-3102, 223-79-9061, and 513-40-8716, the positions of the records are as shown in Table 7.2.1.

The problem with this approach is that *Hash* may not be one-to one; *Hash* might assign the same position in the table to records with different social security numbers. Such an assignment is called a **collision.** When collisions occur, various **collision resolution methods** are used. One of the simplest is the following: If, when the record with social security number n is to be placed, position $Hash(n)$ is already occupied, start from that position and search downward to place the record in the first empty position that occurs, going back up to the beginning of the table if necessary. To locate a record in the table from its social security number, n, you compute $Hash(n)$ and search downward from that position to find the record with social security number n. If there are not too many collisions, this is a very efficient way to store and locate records.

Suppose the social security number for another record to be stored is 908-37-1011. Find the position in Table 7.2.1 into which this record would be placed.

Solution When you compute *Hash* you find that *Hash*(908-37-1011) = 2, which is already occupied by the record with social security number 513-40-8716. Searching downward from position 2, you find that position 3 is also occupied but position 4 is free.

$$\begin{array}{ccccccc} 908\text{-}37\text{-}1011 & \xrightarrow{Hash} & 2 & \rightarrow & 3 & \rightarrow & 4 \\ & & \uparrow & & \uparrow & & \uparrow \\ & & \text{occupied} & & \text{occupied} & & \text{free} \end{array}$$

Therefore, you place the record with social security number n into position 4. ■

Onto Functions

It was noted in Section 7.1 that there may be an element of the co-domain of a function that is not the image of any element in the domain. On the other hand, *every* element of a function's co-domain may be the image of some element of its domain. Such a function is called *onto* or *surjective*. When a function is onto, its range is equal to its co-domain.

• Definition

Let F be a function from a set X to a set Y. F is **onto** (or **surjective**) if, and only if, given any element y in Y, it is possible to find an element x in X with the property that $y = F(x)$.

Symbolically:

$$F\colon X \to Y \text{ is onto} \quad \Leftrightarrow \quad \forall y \in Y, \exists x \in X \text{ such that } F(x) = y.$$

To obtain a precise statement of what it means for a function *not* to be onto, take the negation of the definition of onto:

$$\boxed{F\colon X \to Y \text{ is } \textit{not} \text{ onto} \quad \Leftrightarrow \quad \exists y \text{ in } Y \text{ such that } \forall x \in X,\ F(x) \neq y.}$$

That is, there is some element in Y that is *not* the image of *any* element in X.

In terms of arrow diagrams, a function is onto if each element of the co-domain has an arrow pointing to it from some element of the domain. A function is not onto if at least one element in its co-domain does not have an arrow pointing to it. This is illustrated in Figure 7.2.3.

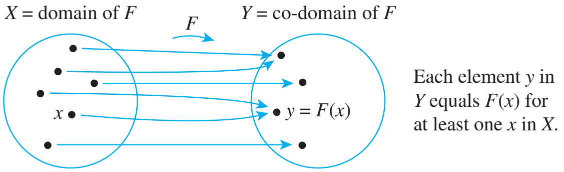


Figure 7.2.3(a) **A Function That Is Onto**

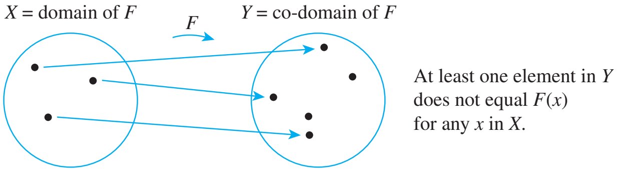

X = domain of F
F
Y = co-domain of F
At least one element in Y does not equal F(x) for any x in X.

Figure 7.2.3(b) **A Function That Is Not Onto**

Example 7.2.4 Identifying Onto Functions Defined on Finite Sets

a. Do either of the arrow diagrams in Figure 7.2.4 define onto functions?

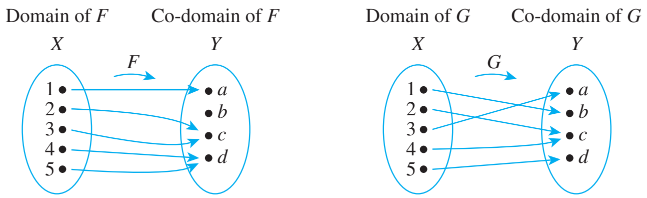


Figure 7.2.4

b. Let $X = \{1, 2, 3, 4\}$ and $Y = \{a, b, c\}$. Define $H: X \to Y$ as follows: $H(1) = c$, $H(2) = a$, $H(3) = c$, $H(4) = b$. Define $K: X \to Y$ as follows: $K(1) = c$, $K(2) = b$, $K(3) = b$, and $K(4) = c$. Is either H or K onto?

Solution

a. F is not onto because $b \neq F(x)$ for any x in X. G is onto because each element of Y equals $G(x)$ for some x in X: $a = G(3)$, $b = G(1)$, $c = G(2) = G(4)$, and $d = G(5)$.

b. H is onto but K is not. H is onto because each of the three elements of the co-domain of H is the image of some element of the domain of H: $a = H(2)$, $b = H(4)$, and $c = H(1) = H(3)$. K, however, is not onto because $a \neq K(x)$ for any x in $\{1, 2, 3, 4\}$. ■

It is possible to write a computer algorithm to check whether a function F is onto, provided F is defined from a finite set X to a finite set Y and there is an independent algorithm to compute values of F. Represent X and Y as one-dimensional arrays $a[1], a[2], \ldots, a[n]$ and $b[1], b[2], \ldots, b[m]$, respectively, and use a nested loop to pick each element y of Y in turn and search through the elements of X to find an x such that y is the image of x. If any search is unsuccessful, then F is not onto. If each such search is successful, then F is onto. You are asked to write such an algorithm in exercise 58 at the end of this section.

Onto Functions on Infinite Sets

Now suppose F is a function from a set X to a set Y, and suppose Y is infinite. By definition, F is onto if, and only if, the following universal statement is true:

$$\forall y \in Y, \exists x \in X \text{ such that } F(x) = y.$$

Thus to prove F is onto, you will ordinarily use the method of generalizing from the generic particular:

suppose that y is any element of Y

and **show** that there is an element X of X with $F(x) = y$.

To prove F is *not* onto, you will usually

find an element y of Y such that $y \neq F(x)$ for *any* x in X.

Example 7.2.5 Proving or Disproving That Functions Are Onto

Define $f: \mathbf{R} \to \mathbf{R}$ and $h: \mathbf{Z} \to \mathbf{Z}$ by the rules

$$f(x) = 4x - 1 \quad \text{for all } x \in \mathbf{R}$$

and

$$h(n) = 4n - 1 \quad \text{for all } n \in \mathbf{Z}.$$

a. Is f onto? Prove or give a counterexample.

b. Is h onto? Prove or give a counterexample.

Solution

a. The best approach is to start trying to prove that f is onto and be alert for difficulties that might indicate that it is not. Now $f: \mathbf{R} \to \mathbf{R}$ is the function defined by the rule

$$\boxed{f(x) = 4x - 1 \quad \text{for all real numbers } x.}$$

To prove that f is onto, you must prove

$$\forall y \in Y, \exists x \in X \text{ such that } f(x) = y.$$

Substituting the definition of f into the outline of a proof by the method of generalizing from the generic particular, you

suppose y is a real number

and **show** that there exists a real number x such that $y = 4x - 1$.

Caution! This scratch work only proves what x has to be *if* it exists. The scratch work does not prove that x exists.

Scratch Work: **If** such a real number x exists, then

$$\begin{aligned} 4x - 1 &= y \\ 4x &= y + 1 && \text{by adding 1 to both sides} \\ x &= \frac{y+1}{4} && \text{by dividing both sides by 4.} \end{aligned}$$

Thus *if* such a number x exists, it must equal $(y + 1)/4$. Does such a number exist? Yes. To show this, let $x = (y + 1)/4$, and then made sure that (1) x is a real number and that (2) f really does send x to y. The following formal answer summarizes this process.

Answer to (a):

If $f: \mathbf{R} \to \mathbf{R}$ is the function defined by the rule $f(x) = 4x - 1$ for all real numbers x, then f is onto.

Proof:

Let $y \in \mathbf{R}$. *[We must show that $\exists x$ in* $\mathbf{R}$ *such that $f(x) = y$.]* Let $x = (y + 1)/4$. Then x is a real number since sums and quotients (other than by 0) of real numbers are real numbers. It follows that

$$\begin{aligned} f(x) &= f\left(\frac{y+1}{4}\right) && \text{by substitution} \\ &= 4 \cdot \left(\frac{y+1}{4}\right) - 1 && \text{by definition of } f \\ &= (y + 1) - 1 = y && \text{by basic algebra.} \end{aligned}$$

[This is what was to be shown.]

b. The function $h: \mathbf{Z} \to \mathbf{Z}$ is defined by the rule

$$\boxed{h(n) = 4n - 1 \quad \text{for all integers } n.}$$

To prove that h is onto, it would be necessary to prove that

$$\forall \text{ integers } m, \exists \text{ an integer } n \text{ such that } h(n) = m.$$

Substituting the definition of h into the outline of a proof by the method of generalizing from the generic particular, you

suppose m is any integer

and **try to show** that there is an integer n with $4n - 1 = m$.

Can you reach what is to be shown from the supposition? No! If $4n - 1 = m$, then

$$n = \frac{m+1}{4} \quad \text{by adding 1 and dividing by 4.}$$

But n must be an integer. And when, for example, $m = 0$, then

$$n = \frac{0+1}{4} = \frac{1}{4},$$

which is *not* an integer.

Thus, in trying to prove that h is onto, you run into difficulty, and this difficulty reveals a counterexample that shows h is not onto.

This discussion is summarized in the following formal answer.

Answer to (b):

If the function $h: \mathbf{Z} \to \mathbf{Z}$ is defined by the rule $h(n) = 4n - 1$ for all integers n, then h is not onto.

Counterexample:

The co-domain of h is $\mathbf{Z}$ and $0 \in \mathbf{Z}$. But $h(n) \neq 0$ for any integer n. For if $h(n) = 0$, then

$$4n - 1 = 0 \quad \text{by definition of } h$$

which implies that

$$4n = 1 \quad \text{by adding 1 to both sides}$$

and so

$$n = \frac{1}{4} \quad \text{by dividing both sides by 4.}$$

But $1/4$ is not an integer. Hence there is no integer n for which $f(n) = 0$, and thus f is not onto.

■

Relations between Exponential and Logarithmic Functions

Note That the quantity b^x is a real number for any real number x follows from the least-upper-bound property of the real number system. (See Appendix A.)

For positive numbers $b \neq 1$, the **exponential function with base b,** denoted $\exp_b$, is the function from $\mathbf{R}$ to $\mathbf{R}^+$ defined as follows: For all real numbers x,

$$\exp_b(x) = b^x$$

where $b^0 = 1$ and $b^{-x} = 1/b^x$.

When working with the exponential function, it is useful to recall the laws of exponents from elementary algebra.

Laws of Exponents

If b and c are any positive real numbers and u and v are any real numbers, the following laws of exponents hold true:

$$b^u b^v = b^{u+v} \qquad 7.2.1$$

$$(b^u)^v = b^{uv} \qquad 7.2.2$$

$$\frac{b^u}{b^v} = b^{u-v} \qquad 7.2.3$$

$$(bc)^u = b^u c^u \qquad 7.2.4$$

In Section 7.1 the logarithmic function with base b was defined for any positive number b $\neq 1$ to be the function from $\mathbf{R}^+$ to $\mathbf{R}$ with the property that for each positive real number x,

$$\log_b(x) = \text{ the exponent to which } b \text{ must be raised to obtain } x.$$

Or, equivalently, for each positive real number x and real number y,

$$\log_b x = y \quad \Leftrightarrow \quad b^y = x.$$

It can be shown using calculus that both the exponential and logarithmic functions are one-to-one and onto. Therefore, by definition of one-to-one, the following properties hold true:

For any positive real number b with $b \neq 1$,

$$\text{if } b^u = b^v \text{ then } u = v \quad \text{for all real numbers } u \text{ and } v, \qquad 7.2.5$$

and

$$\text{if } \log_b u = \log_b v \text{ then } u = v \quad \text{for all positive real numbers } u \text{ and } v. \qquad 7.2.6$$

These properties are used to derive many additional facts about exponents and logarithms. In particular we have the following properties of logarithms.

Theorem 7.2.1 Properties of Logarithms

For any positive real numbers b, c and x with $b \neq 1$ and $c \neq 1$:

a. $\log_b(xy) = \log_b x + \log_b y$

b. $\log_b\left(\dfrac{x}{y}\right) = \log_b x - \log_b y$

c. $\log_b(x^a) = a \log_b x$

d. $\log_c x = \dfrac{\log_b x}{\log_b c}$

Theorem 7.2.1(d) is proved in the next example. You are asked to prove the remainder of the theorem in exercises 33–35 at the end of this section.

Example 7.2.6 Using the One-to-Oneness of the Exponential Function

Use the definition of logarithm, the laws of exponents, and the one-to-oneness of the exponential function (property 7.2.5) to prove part (d) of Theorem 7.2.1: For any positive real numbers b, c, and x, with $b \neq 1$ and $c \neq 1$,

$$\boxed{\log_c x = \frac{\log_b x}{\log_b c}.}$$

Solution Suppose positive real numbers b, c, and x are given. Let

$$(1)\ \ u = \log_b c \qquad (2)\ \ v = \log_c x \qquad (3)\ \ w = \log_b x.$$

Then, by definition of logarithm,

$$(1')\ \ c = b^u \qquad (2')\ \ x = c^v \qquad (3')\ \ x = b^w.$$

Substituting (1′) into (2′) and using one of the laws of exponents gives

$$x = c^v = (b^u)^v = b^{uv} \qquad \text{by 7.2.2}$$

But by (3), $x = b^w$ also. Hence

$$b^{uv} = b^w,$$

and so by the one-to-oneness of the exponential function (property 7.2.5),

$$uv = w.$$

Substituting from (1), (2), and (3) gives that

$$(\log_b c)(\log_c x) = \log_b x.$$

And dividing both sides by $\log_b c$ (which is nonzero because $c \neq 1$) results in

$$\log_c x = \frac{\log_b x}{\log_b c}. \qquad ■$$

Example 7.2.7 Computing Logarithms with Base 2 on a Calculator

In computer science it is often necessary to compute logarithms with base 2. Most calculators do not have keys to compute logarithms with base 2 but do have keys to compute logarithms with base 10 (called **common logarithms** and often denoted simply log) and logarithms with base e (called **natural logarithms** and usually denoted ln). Suppose your calculator shows that $\ln 5 \cong 1.609437912$ and $\ln 2 \cong 0.6931471806$. Use Theorem 7.2.1(d) to find an approximate value for $\log_2 5$.

Solution By Theorem 7.2.1(d),

$$\log_2 5 = \frac{\ln 5}{\ln 2} \cong \frac{1.609437912}{0.6931471806} \cong 2.321928095. \qquad ■$$

One-to-One Correspondences

Consider a function $F: X \to Y$ that is both one-to-one and onto. Given any element x in X, there is a unique corresponding element $y = F(x)$ in Y (since F is a function). Also given any element y in Y, there is an element x in X such that $F(x) = y$ (since F is onto) and there is only one such x (since F is one-to-one). Thus, a function that is one-to-one and onto sets up a pairing between the elements of X and the elements of Y that matches each element of X with exactly one element of Y and each element of Y with exactly one element of X. Such a pairing is called a *one-to-one correspondence* or *bijection* and is illustrated by the arrow diagram in Figure 7.2.5. One-to-one correspondences are often used as aids to counting. The pairing of Figure 7.2.5, for example, shows that there are five elements in the set X.

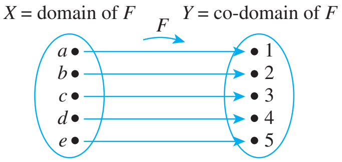


Figure 7.2.5 **An Arrow Diagram for a One-to-One Correspondence**

• Definition

A **one-to-one correspondence** (or **bijection**) from a set X to a set Y is a function $F: X \to Y$ that is both one-to-one and onto.

Example 7.2.8 A Function from a Power Set to a Set of Strings

Let $\mathscr{P}(\{a, b\})$ be the set of all subsets of $\{a, b\}$ and let S be the set of all strings of length 2 made up of 0's and 1's. Then $\mathscr{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$ and $S = \{00, 01, 10, 11\}$. Define a function h from $\mathscr{P}(\{a, b\})$ to S as follows: Given any subset A of $\{a, b\}$, a is either in A or not in A, and b is either in A or not in A. If a is in A, write a 1 in the first position of the string $h(A)$. If a is not in A, write a 0 in the first position of the string $h(A)$. Similarly, if b is in A, write a 1 in the second position of the string $h(A)$. If b is not in A, write a 0 in the second position of the string $h(A)$. This definition is summarized in the following table.

h

Subset of $\{a, b\}$	Status of a	Status of b	String in S
$\emptyset$	not in	not in	00
$\{a\}$	in	not in	10
$\{b\}$	not in	in	01
$\{a, b\}$	in	in	11

Is h a one-to-one correspondence?

Solution The arrow diagram shown in Figure 7.2.6 shows clearly that h is a one-to-one correspondence. It is onto because each element of S has an arrow pointing to it. It is one-to-one because each element of S has no more than one arrow pointing to it.

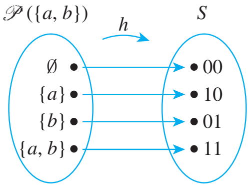


Figure 7.2.6 ■

Example 7.2.9 A String-Reversing Function

Let T be the set of all finite strings of x's and y's. Define $g: T \to T$ by the rule: For all strings $s \in T$,

$$g(s) = \text{the string obtained by writing the characters of } s \text{ in reverse order.}$$

Is g a one-to-one correspondence from T to itself?

Solution The answer is yes. To show that g is a one-to-one correspondence, it is necessary to show that g is one-to-one and onto.

To see that g is one-to-one, suppose that for some strings s_1 and s_2 in T, $g(s_1) = g(s_2)$. *[We must show that $s_1 = s_2$.]* Now to say that $g(s_1) = g(s_2)$ is the same as saying that the string obtained by writing the characters of s_1 in reverse order equals the string obtained by writing the characters of s_2 in reverse order. But if s_1 and s_2 are equal when written in reverse order, then they must be equal to start with. In other words, $s_1 = s_2$ *[as was to be shown].*

To show that g is onto, suppose t is a string in T. *[We must find a string s in T such that $g(s) = t$.]* Let $s = g(t)$. By definition of g, $s = g(t)$ is the string in T obtained by writing the characters of t in reverse order. But when the order of the characters of a string is reversed once and then reversed again, the original string is recovered. Thus

$$\begin{aligned} g(s) = g(g(t)) &= \text{the string obtained by writing the characters} \\ &\quad \text{of } t \text{ in reverse order and then writing those} \\ &\quad \text{characters in reverse order again} \\ &= t. \end{aligned}$$

This is what was to be shown. ■

Example 7.2.10 A Function of Two Variables

Define a function $F: \mathbf{R} \times \mathbf{R} \to \mathbf{R} \times \mathbf{R}$ as follows: For all $(x, y) \in \mathbf{R} \times \mathbf{R}$,

$$F(x, y) = (x + y, x - y).$$

Is F a one-to-one correspondence from $\mathbf{R} \times \mathbf{R}$ to itself?

Solution The answer is yes. To show that F is a one-to-one correspondence, you need to show both that F is one-to-one and that F is onto.

Proof that F is one-to-one: Suppose that (x_1, y_1) and (x_2, y_2) are any ordered pairs in $\mathbf{R} \times \mathbf{R}$ such that

$$F(x_1, y_1) = F(x_2, y_2).$$

[We must show that $(x_1, y_1) = (x_2, y_2)$.] By definition of F,

$$(x_1 + y_1, x_1 - y_1) = (x_2 + y_2, x_2 - y_2).$$

For two ordered pairs to be equal, both the first and second components must be equal. Thus x_1, y_1, x_2, and y_2 satisfy the following system of equations:

$$x_1 + y_1 = x_2 + y_2 \quad (1)$$

$$x_1 - y_1 = x_2 - y_2 \quad (2)$$

Adding equations (1) and (2) gives that

$$2x_1 = 2x_2, \quad \text{and so} \quad x_1 = x_2.$$

Substituting $x_1 = x_2$ into equation (1) yields

$$x_1 + y_1 = x_1 + y_2, \quad \text{and so} \quad y_1 = y_2.$$

Thus, by definition of equality of ordered pairs, $(x_1, y_1) = (x_2, y_2)$ *[as was to be shown].*

Caution! This scratch work only shows what (r, s) has to be *if* it exists. The scratch work does not prove that (r, s) exists.

Scratch Work for the Proof that F is onto: To prove that F is onto, you suppose you have any ordered pair in the co-domain $\mathbf{R} \times \mathbf{R}$, say (u, v), and then you show that there is an ordered pair in the domain that is sent to (u, v) by F. To do this, you suppose temporarily that you have found such an ordered pair, say (r, s). Then

$$F(r, s) = (u, v) \qquad \text{because you are supposing that } F \text{ sends} (r, s) \text{ to } (u, v),$$

and

$$F(r, s) = (r + s, r - s) \qquad \text{by definition of } F.$$

Equating the right-hand sides gives

$$(r + s, r - s) = (u, v).$$

By definition of equality of ordered pairs this means that

$$r + s = u \qquad (1)$$

$$r - s = v \qquad (2)$$

Adding equations (1) and (2) gives

$$2r = u + v, \quad \text{and so} \quad r = \frac{u+v}{2}.$$

Subtracting equation (2) from equation (1) yields

$$2s = u - v, \quad \text{and so} \quad s = \frac{u-v}{2}.$$

Thus, ***if*** F sends (r, s) to (u, v), then $r = (u + v)/2$ and $s = (u - v)/2$. To turn this scratch work into a proof, you need to make sure that (1) $\left(\frac{u+v}{2}, \frac{u-v}{2}\right)$ is in the domain of F, and (2) that F really does send $\left(\frac{u+v}{2}, \frac{u-v}{2}\right)$ to (u, v).

Proof that F is onto: Suppose (u, v) is any ordered pair in the co-domain of F. *[We will show that there is an ordered pair in the domain of F that is sent to (u, v) by F.]* Let

$$r = \frac{u+v}{2} \quad \text{and} \quad s = \frac{u-v}{2}.$$

Then (r, s) is an ordered pair of real numbers and so is in the domain of F. In addition:

$$\begin{aligned}
F(r, s) &= F\left(\frac{u+v}{2}, \frac{u-v}{2}\right) && \text{by definition of } F \\
&= \left(\frac{u+v}{2} + \frac{u-v}{2}, \frac{u+v}{2} - \frac{u-v}{2}\right) && \text{by substitution} \\
&= \left(\frac{u+v+u-v}{2}, \frac{u+v-u+v}{2}\right) \\
&= \left(\frac{2u}{2}, \frac{2v}{2}\right) \\
&= (u, v) && \text{by algebra.}
\end{aligned}$$

[This is what was to be shown.] ■

Inverse Functions

If F is a one-to-one correspondence from a set X to a set Y, then there is a function from Y to X that "undoes" the action of F; that is, it sends each element of Y back to the element of X that it came from. This function is called the *inverse function* for F.

Theorem 7.2.2

Suppose $F: X \to Y$ is a one-to-one correspondence; that is, suppose F is one-to-one and onto. Then there is a function $\boldsymbol{F^{-1}: Y \to X}$ that is defined as follows:

Given any element y in Y,

$$\boldsymbol{F^{-1}(y)} = \text{that unique element } x \text{ in } X \text{ such that } F(x) \text{ equals } y.$$

In other words,

$$F^{-1}(y) = x \quad \Leftrightarrow \quad y = F(x).$$

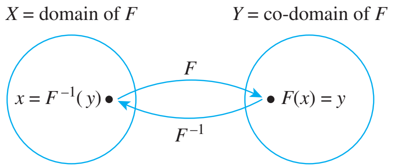
The proof of Theorem 7.2.2 follows immediately from the definition of one-to-one and onto. Given an element y in Y, there is an element x in X with $F(x) = y$ because F is onto; x is unique because F is one-to-one.

• Definition

The function F^{-1} of Theorem 7.2.2 is called the **inverse function** for F.

Note that according to this definition, the logarithmic function with base $b > 0$ is the inverse of the exponential function with base b.

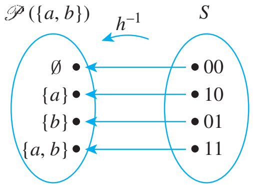
The diagram that follows illustrates the fact that an inverse function sends each element back to where it came from.



Example 7.2.11 Finding an Inverse Function for a Function Given by an Arrow Diagram

Define the inverse function for the one-to-one correspondence h given in Example 7.2.8.

Solution The arrow diagram for h^{-1} is obtained by tracing the h-arrows back from S to $\mathscr{P}(\{a, b\})$ as shown below.



$h^{-1}(00) = \emptyset \quad h^{-1}(10) = \{a\}$

$h^{-1}(01) = \{b\} \quad h^{-1}(11) = \{a, b\}$

■

Example 7.2.12 Finding an Inverse Function for a Function Given in Words

Define the inverse function for the one-to-one correspondence g given in Example 7.2.9.

Solution The function $g: T \to T$ is defined by the rule

For all strings t in T,

$$g(t) = \text{the string obtained by writing the characters of } t \text{ in reverse order.}$$

Now if the characters of t are written in reverse order and then written in reverse order again, the original string is recovered. Thus given any string t in T,

$$\begin{aligned} g^{-1}(t) &= \text{the unique string that, when written} \\ &\quad\ \text{in reverse order, equals } t \\ &= \text{the string obtained by writing the} \\ &\quad\ \text{characters of } t \text{ in reverse order} \\ &= g(t). \end{aligned}$$

Hence $g^{-1}\colon T \to T$ is the same as g, or, in other words, $g^{-1} = g$. ■

Example 7.2.13 Finding an Inverse Function for a Function Given by a Formula

The function $f\colon \mathbf{R} \to \mathbf{R}$ defined by the formula

$$f(x) = 4x - 1 \quad \text{for all real numbers } x$$

was shown to be one-to-one in Example 7.2.2 and onto in Example 7.2.5. Find its inverse function.

Solution For any *[particular but arbitrarily chosen]* y in $\mathbf{R}$, by definition of f^{-1},

$$f^{-1}(y) = \text{that unique real number } x \text{ such that } f(x) = y.$$

But

$$\begin{aligned} & f(x) = y \\ \Leftrightarrow\quad & 4x - 1 = y && \text{by definition of } f \\ \Leftrightarrow\quad & x = \frac{y+1}{4} && \text{by algebra.} \end{aligned}$$

Hence $f^{-1}(y) = \dfrac{y+1}{4}$. ■

The following theorem follows easily from the definitions.

Theorem 7.2.3

If X and Y are sets and $F\colon X \to Y$ is one-to-one and onto, then $F^{-1}\colon Y \to X$ is also one-to-one and onto.

Proof:

F^{-1} is one-to-one: Suppose y_1 and y_2 are elements of Y such that $F^{-1}(y_1) = F^{-1}(y_2)$. *[We must show that $y_1 = y_2$.]* Let $x = F^{-1}(y_1) = F^{-1}(y_2)$. Then $x \in X$, and by definition of F^{-1},

$$F(x) = y_1 \quad \text{since } x = F^{-1}(y_1)$$

and

$$F(x) = y_2 \quad \text{since } x = F^{-1}(y_2).$$

Consequently, $y_1 = y_2$ since each is equal to $F(x)$. This is what was to be shown.

F^{-1} is onto: Suppose $x \in X$. *[We must show that there exists an element y in Y such that $F^{-1}(y) = x$.]* Let $y = F(x)$. Then $y \in Y$, and by definition of F^{-1}, $F^{-1}(y) = x$. This is what was to be shown.

Example 7.2.14 Finding an Inverse Function for a Function of Two Variables

Define the inverse function $F^{-1}: \mathbf{R} \times \mathbf{R} \to \mathbf{R} \times \mathbf{R}$ for the one-to-one correspondence given in Example 7.2.10.

Solution

The solution to Example 7.2.10 shows that $F\left(\frac{u+v}{2}, \frac{u-v}{2}\right) = (u, v)$. Because F is one-to-one, this means that

$\left(\frac{u+v}{2}, \frac{u-v}{2}\right)$ is the unique ordered pair in the domain of F that is sent to (u, v) by F.

Thus, F^{-1} is defined as follows: For all $(u, v) \in \mathbf{R} \times \mathbf{R}$,

$$F^{-1}(u, v) = \left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$ ■

Test Yourself

1. If F is a function from a set X to a set Y, then F is one-to-one if, and only if, _____.
2. If F is a function from a set X to a set Y, then F is not one-to-one if, and only if, _____.
3. If F is a function from a set X to a set Y, then F is onto if, and only if, _____.
4. If F is a function from a set X to a set Y, then F is not onto if, and only if, _____.
5. The following two statements are _____:

$$\forall\, u, v \in U,\ \text{if } H(u) = H(v) \text{ then } u = v.$$
$$\forall\, u, v \in U,\ \text{if } u \neq v \text{ then } H(u) \neq H(v).$$

6. Given a function $F: X \to Y$ and an infinite set X, to prove that F is one-to-one, you suppose that _____ and then you show that _____.
7. Given a function $F: X \to Y$ and an infinite set X, to prove that F is onto, you suppose that _____ and then you show that _____.
8. Given a function $F: X \to Y$, to prove that F is not one-to-one, you _____.
9. Given a function $F: X \to Y$, to prove that F is not onto, you _____.
10. A one-to-one correspondence from a set X to a set Y is a _____ that is _____.
11. If F is a one-to-one correspondence from a set X to a set Y and y is in Y, then $F^{-1}(y)$ is _____.

Exercise Set 7.2

1. The definition of one-to-one is stated in two ways:

$$\forall x_1, x_2 \in X, \text{if } F(x_1) = F(x_2) \text{ then } x_1 = x_2$$

and $\quad \forall x_1, x_2 \in X, \text{if } x_1 \neq x_2 \text{ then } F(x_1) \neq F(x_2).$

Why are these two statements logically equivalent?

2. Fill in each blank with the word *most* or *least*.

a. A function F is one-to-one if, and only if, each element in the co-domain of F is the image of at ______ one element in the domain of F.

b. A function F is onto if, and only if, each element in the co-domain of F is the image of at ______ one element in the domain of F.

***H* 3.** When asked to state the definition of one-to-one, a student replies, "A function f is one-to-one if, and only if, every element of X is sent by f to exactly one element of Y." Give a counterexample to show that the student's reply is incorrect.

***H* 4.** Let $f: X \to Y$ be a function. True or false? A sufficient condition for f to be one-to-one is that for all elements y in Y, there is at most one x in X with $f(x) = y$.

***H* 5.** All but two of the following statements are correct ways to express the fact that a function f is onto. Find the two that are incorrect.

a. f is onto $\Leftrightarrow$ every element in its co-domain is the image of some element in its domain.

b. f is onto $\Leftrightarrow$ every element in its domain has a corresponding image in its co-domain.

c. f is onto $\Leftrightarrow \forall y \in Y, \exists x \in X$ such that $f(x) = y$.

d. f is onto $\Leftrightarrow \forall x \in X, \exists y \in Y$ such that $f(x) = y$.

e. f is onto $\Leftrightarrow$ the range of f is the same as the co-domain of f.

6. Let $X = \{1, 5, 9\}$ and $Y = \{3, 4, 7\}$.

a. Define $f: X \to Y$ by specifying that

$$f(1) = 4, \quad f(5) = 7, \quad f(9) = 4.$$

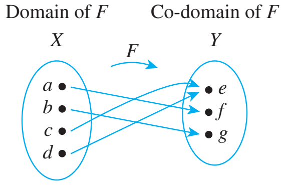
Is f one-to-one? Is f onto? Explain your answers.

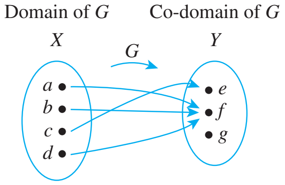
b. Define $g: X \to Y$ by specifying that

$$g(1) = 7, \quad g(5) = 3, \quad g(9) = 4.$$

Is g one-to-one? Is g onto? Explain your answers.

7. Let $X = \{a, b, c, d\}$ and $Y = \{e, f, g\}$. Define functions F and G by the arrow diagrams below.

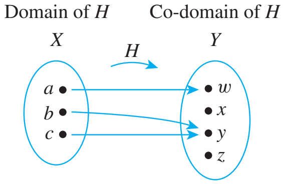


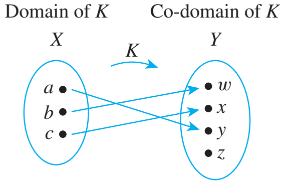


a. Is F one-to-one? Why or why not? Is it onto? Why or why not?

b. Is G one-to-one? Why or why not? Is it onto? Why or why not?

8. Let $X = \{a, b, c\}$ and $Y = \{w, x, y, z\}$. Define functions H and K by the arrow diagrams below.





a. Is H one-to-one? Why or why not? Is it onto? Why or why not?

b. Is K one-to-one? Why or why not? Is it onto? Why or why not?

9. Let $X = \{1, 2, 3\}$, $Y = \{1, 2, 3, 4\}$, and $Z = \{1, 2\}$.

a. Define a function $f: X \to Y$ that is one-to-one but not onto.

b. Define a function $g: X \to Z$ that is onto but not one-to-one.

c. Define a function $h: X \to X$ that is neither one-to-one nor onto.

d. Define a function $k: X \to X$ that is one-to-one and onto but is not the identity function on X.

10. a. Define $f: \mathbf{Z} \to \mathbf{Z}$ by the rule $f(n) = 2n$, for all integers n.

(i) Is f one-to-one? Prove or give a counterexample.

(ii) Is f onto? Prove or give a counterexample.

b. Let $2\mathbf{Z}$ denote the set of all even integers. That is, $2\mathbf{Z} = \{n \in \mathbf{Z} \mid n = 2k, \text{ for some integer } k\}$. Define $h: \mathbf{Z} \to 2\mathbf{Z}$ by the rule $h(n) = 2n$, for all integers n. Is h onto? Prove or give a counterexample.

H 11. a. Define $g: \mathbf{Z} \to \mathbf{Z}$ by the rule $g(n) = 4n - 5$, for all integers n.

(i) Is g one-to-one? Prove or give a counterexample.

(ii) Is g onto? Prove or give a counterexample.

b. Define $G: \mathbf{R} \to \mathbf{R}$ by the rule $G(x) = 4x - 5$ for all real numbers x. Is G onto? Prove or give a counterexample.

12. a. Define $F: \mathbf{Z} \to \mathbf{Z}$ by the rule $F(n) = 2 - 3n$, for all integers n.

(i) *Is* F one-to-one? Prove or give a counterexample.

(ii) Is F onto? Prove or give a counterexample.

b. Define $G: \mathbf{R} \to \mathbf{R}$ by the rule $G(x) = 2 - 3x$ for all real numbers x. Is G onto? Prove or give a counterexample.

13. a. Define $H: \mathbf{R} \to \mathbf{R}$ by the rule $H(x) = x^2$, for all real numbers x.

(i) Is H one-to-one? Prove or give a counterexample.

(ii) Is H onto? Prove or give a counterexample.

b. Define $K: \mathbf{R}^{nonneg} \to \mathbf{R}^{nonneg}$ by the rule $K(x) = x^2$, for all nonnegative real numbers x. Is K onto? Prove or give a counterexample.

14. Explain the mistake in the following "proof."

Theorem: The function $f: \mathbf{Z} \to \mathbf{Z}$ defined by the formula $f(n) = 4n + 3$, for all integers n, is one-to-one.

"**Proof:** Suppose any integer n is given. Then by definition of f, there is only one possible value for $f(n)$, namely, $4n + 3$. Hence f is one-to-one."

In each of 15–18 a function f is defined on a set of real numbers. Determine whether or not f is one-to-one and justify your answer.

15. $f(x) = \dfrac{x+1}{x}$, for all real numbers $x \neq 0$

16. $f(x) = \dfrac{x}{x^2+1}$, for all real numbers x

17. $f(x) = \dfrac{3x-1}{x}$, for all real numbers $x \neq 0$

18. $f(x) = \dfrac{x+1}{x-1}$, for all real numbers $x \neq 1$

19. Referring to Example 7.2.3, assume that records with the following social security numbers are to be placed in sequence into Table 7.2.1. Find the position into which each record is placed.

a. 417-30-2072 b. 364-98-1703 c. 283-09-0787

20. Define Floor: $\mathbf{R} \to \mathbf{Z}$ by the formula Floor$(x) = \lfloor x \rfloor$, for all real numbers x.
a. Is Floor one-to-one? Prove or give a counterexample.
b. Is Floor onto? Prove or give a counterexample.

21. Let S be the set of all strings of 0's and 1's, and define l: $S \to \mathbf{Z}^{nonneg}$ by

$$l(s) = \text{ the length of } s, \quad \text{for all strings } s \text{ in } S.$$

a. Is l one-to-one? Prove or give a counterexample.
b. Is l onto? Prove or give a counterexample.

22. Let S be the set of all strings of 0's and 1's, and define D: $S \to \mathbf{Z}$ as follows: For all $s \in S$,

$$D(s) = \text{the number of 1's in } s \text{ minus the number of 0's in } s.$$

a. Is D one-to-one? Prove or give a counterexample.
b. Is D onto? Prove or give a counterexample.

23. Define F: $\mathscr{P}(\{a, b, c\}) \to \mathbf{Z}$ as follows: For all A in $\mathscr{P}(\{a, b, c\})$,

$$F(A) = \text{the number of elements in } A.$$

a. Is F one-to-one? Prove or give a counterexample.
b. Is F onto? Prove or give a counterexample.

24. Let S be the set of all strings of a's and b's, and define N: $S \to \mathbf{Z}$ by

$$N(s) = \text{the number of } a\text{'s in } s, \quad \text{for all } s \in S.$$

a. Is N one-to-one? Prove or give a counterexample.
b. Is N onto? Prove or give a counterexample.

25. Let S be the set of all strings in a's and b's, and define C: $S \to S$ by

$$C(s) = as, \quad \text{for all } s \in S.$$

(C is called **concatenation** by a on the left.)
a. Is C one-to-one? Prove or give a counterexample.
b. Is C onto? Prove or give a counterexample.

26. Define S: $\mathbf{Z}^+ - \mathbf{Z}^+$ by the rule: For all integers n, $S(n) =$ the sum of the positive divisors of n.
a. Is S one-to-one? Prove or give a counterexample.
b. Is S onto? Prove or give a counterexample.

***H* 27.** Let D be the set of all finite subsets of positive integers, and define T: $\mathbf{Z}^+ \to D$ by the rule: For all integers n, $T(n) =$ the set of all of the positive divisors of n.
a. Is T one-to-one? Prove or give a counterexample.
b. Is T onto? Prove or give a counterexample.

28. Define G: $\mathbf{R} \times \mathbf{R} \to \mathbf{R} \times \mathbf{R}$ as follows: $G(x, y) = (2y, -x)$ for all $(x, y) \in \mathbf{R} \times \mathbf{R}$.
a. Is G one-to-one? Prove or give a counterexample.
b. Is G onto? Prove or give a counterexample.

29. Define H: $\mathbf{R} \times \mathbf{R} \to \mathbf{R} \times \mathbf{R}$ as follows: $H(x, y) = (x + 1, 2 - y)$ for all $(x, y) \in \mathbf{R} \times \mathbf{R}$.
a. Is H one-to-one? Prove or give a counterexample.
b. Is H onto? Prove or give a counterexample.

30. Define J: $\mathbf{Q} \times \mathbf{Q} \to \mathbf{R}$ by the rule $J(r, s) = r + \sqrt{2}s$ for all $(r, s) \in \mathbf{Q} \times \mathbf{Q}$.
a. Is J one-to-one? Prove or give a counterexample.
b. Is J onto? Prove or give a counterexample.

✶ 31. Define F: $\mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ and G: $\mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ as follows: For all $(n, m) \in \mathbf{Z}^+ \times \mathbf{Z}^+$,

$$F(n, m) = 3^n 5^m \quad \text{and} \quad G(n, m) = 3^n 6^m.$$

***H* a.** Is F one-to-one? Prove or give a counterexample.
b. Is G one-to-one? Prove or give a counterexample.

32. **a.** Is $\log_8 27 = \log_2 3$? Why or why not?
b. Is $\log_{16} 9 = \log_4 3$? Why or why not?

The properties of logarithm established in 33–35 are used in Sections 11.4 and 11.5.

33. Prove that for all positive real numbers b, x, and y with $b \neq 1$,

$$\log_b \left(\frac{x}{y}\right) = \log_b x - \log_b y.$$

34. Prove that for all positive real numbers b, x, and y with $b \neq 1$,

$$\log_b(xy) = \log_b x + \log_b y.$$

***H* 35.** Prove that for all real numbers a, b, and x with b and x positive and $b \neq 1$,

$$\log_b(x^a) = a \log_b x.$$

Exercises 36 and 37 use the following definition: If f: $\mathbf{R} \to \mathbf{R}$ and g: $\mathbf{R} \to \mathbf{R}$ are functions, then the function $(f + g)$: $\mathbf{R} \to \mathbf{R}$ is defined by the formula $(f + g)(x) = f(x) + g(x)$ for all real numbers x.

36. If f: $\mathbf{R} \to \mathbf{R}$ and g: $\mathbf{R} \to \mathbf{R}$ are both one-to-one, is $f + g$ also one-to-one? Justify your answer.

37. If f: $\mathbf{R} \to \mathbf{R}$ and g: $\mathbf{R} \to \mathbf{R}$ are both onto, is $f + g$ also onto? Justify your answer.

Exercises 38 and 39 use the following definition: If f: $\mathbf{R} \to \mathbf{R}$ is a function and c is a nonzero real number, the function $(c \cdot f)$: $\mathbf{R} \to \mathbf{R}$ is defined by the formula $(c \cdot f)(x) = c \cdot f(x)$ for all real numbers x.

38. Let f: $\mathbf{R} \to \mathbf{R}$ be a function and c a nonzero real number. If f is one-to-one, is $c \cdot f$ also one-to-one? Justify your answer.

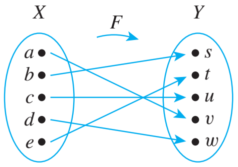
39. Let f: $\mathbf{R} \to \mathbf{R}$ be a function and c a nonzero real number. If f is onto, is $c \cdot f$ also onto? Justify your answer.

***H* 40.** Suppose F: $X \to Y$ is one-to-one.
a. Prove that for all subsets $A \subseteq X$, $F^{-1}(F(A)) = A$.
b. Prove that for all subsets A_1 and A_2 in X, $F(A_1 \cap A_2) = F(A_1) \cap F(A_2)$.

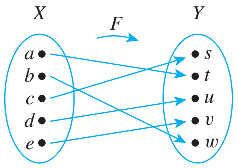
41. Suppose $F: X \to Y$ is onto. Prove that for all subsets $B \subseteq Y$, $F(F^{-1}(B)) = B$.

Let $X = \{a, b, c, d, e\}$ and $Y = \{s, t, u, v, w\}$. In each of 42 and 43 a one-to-one correspondence $F: X \to Y$ is defined by an arrow diagram. In each case draw an arrow diagram for F^{-1}.

42.



43.



In 44–55 indicate which of the functions in the referenced exercise are one-to-one correspondences. For each function that is a one-to-one correspondence, find the inverse function.

44. Exercise 10a

45. Exercise 10b

46. Exercise 11a

47. Exercise 11b

48. Exercise 12a

49. Exercise 12b

50. Exercise 21

51. Exercise 22

52. Exercise 15 with the co-domain taken to be the set of all real numbers not equal to 1.

H **53.** Exercise 16 with the co-domain taken to be the set of all real numbers.

54. Exercise 17 with the co-domain taken to be the set of all real numbers not equal to 3.

55. Exercise 18 with the co-domain taken to be the set of all real numbers not equal to 1.

56. In Example 7.2.8 a one-to-one correspondence was defined from the power set of $\{a, b\}$ to the set of all strings of 0's and 1's that have length 2. Thus the elements of these two sets can be matched up exactly, and so the two sets have the same number of elements.

   a. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set with n elements. Use Example 7.2.8 as a model to define a one-to-one correspondence from $\mathscr{P}(X)$, the set of all subsets of X, to the set of all strings of 0's and 1's that have length n.

   b. Use the one-to-one correspondence of part (a) to deduce that a set with n elements has 2^n subsets. (This provides an alternative proof of Theorem 6.3.1.)

H **57.** Write a computer algorithm to check whether a function from one finite set to another is one-to-one. Assume the existence of an independent algorithm to compute values of the function.

H **58.** Write a computer algorithm to check whether a function from one finite set to another is onto. Assume the existence of an independent algorithm to compute values of the function.

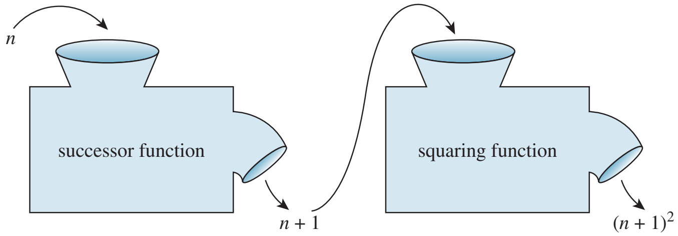
Answers for Test Yourself

1. for all x_1 and x_2 in X, if $F(x_1) = F(x_2)$ then $x_1 = x_2$ 2. there exist elements x_1 and x_2 in X such that $F(x_1) = F(x_2)$ and $x_1 \neq x_2$ 3. for all y in Y, there exists at least one element x in X such that $f(x) = y$ 4. there exists an element y in Y such that for all elements x in X, $f(x) \neq y$ 5. logically equivalent ways of expressing what it means for a function H to be one-to-one (The second is the contrapositive of the first.) 6. x_1 and x_2 are any *[particular but arbitrarily chosen]* elements in X with the property that $F(x_1) = F(x_2)$; $x_1 = x_2$ 7. y is any *[particular but arbitrarily chosen]* element in Y; there exists at least one element x in X such that $F(x) = y$ 8. show that there are concrete elements x_1 and x_2 in X with the property that $F(x_1) = F(x_2)$ and $x_1 \neq x_2$ 9. show that there is a concrete element y in Y with the property that $F(x) \neq y$ for any element x in X 10. function from X to Y; both one-to-one and onto 11. the unique element x in X such that $F(x) = y$ (in other words, $F^{-1}(y)$ is the unique preimage of y in X)

7.3 Composition of Functions

It is no paradox to say that in our most theoretical moods we may be nearest to our most practical applications. — Alfred North Whitehead

Consider two functions, the successor function and the squaring function, defined from **Z** (the set of integers) to **Z**, and imagine that each is represented by a machine. If the two machines are hooked up so that the output from the successor function is used as input to the squaring function, then they work together to operate as one larger machine. In this larger machine, an integer n is first increased by 1 to obtain $n + 1$; then the quantity $n + 1$ is squared to obtain $(n + 1)^2$. This is illustrated in the following drawing.



Combining functions in this way is called *composing* them; the resulting function is called the *composition* of the two functions. Note that the composition can be formed only if the output of the first function is acceptable input to the second function. That is, the range of the first function must be contained in the domain of the second function.

Note We put the f first when we say "the composition of f and g" because an element x is acted upon first by f and then by g.

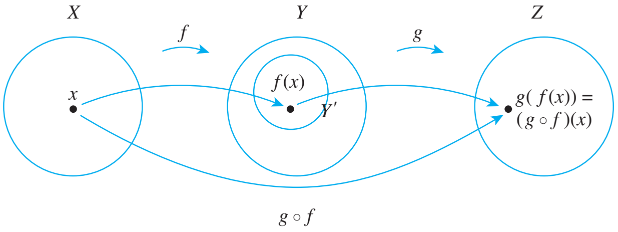
• Definition

Let $f\colon X \to Y'$ and $g\colon Y \to Z$ be functions with the property that the range of f is a subset of the domain of g. Define a new function $g \circ f\colon X \to Z$ as follows:

$$(g \circ f)(x) = g(f(x)) \quad \text{for all } x \in X,$$

where $g \circ f$ is read "g circle f" and $g(f(x))$ is read "g of f of x." The function $g \circ f$ is called the **composition of f and g.**

This definition is shown schematically below.



Example 7.3.1 Composition of Functions Defined by Formulas

Let $f\colon \mathbf{Z} \to \mathbf{Z}$ be the successor function and let $g\colon \mathbf{Z} \to \mathbf{Z}$ be the squaring function. Then $f(n) = n + 1$ for all $n \in \mathbf{Z}$ and $g(n) = n^2$ for all $n \in \mathbf{Z}$.

a. Find the compositions $g \circ f$ and $f \circ g$.

b. Is $g \circ f = f \circ g$? Explain.

Caution! Be careful not to confuse $g \circ f$ and $g(f(x))$: $g \circ f$ is the name of the function whereas $g(f(x))$ is the value of the function at x.

Solution

a. The functions $g \circ f$ and $f \circ g$ are defined as follows:

$$(g \circ f)(n) = g(f(n)) = g(n + 1) = (n + 1)^2 \quad \text{for all } n \in \mathbf{Z},$$

and

$$(f \circ g)(n) = f(g(n)) = f(n^2) = n^2 + 1 \quad \text{for all } n \in \mathbf{Z}.$$

b. Two functions from one set to another are equal if, and only if, they always take the same values. In this case,

$$(g \circ f)(1) = (1+1)^2 = 4, \text{ whereas } (f \circ g)(1) = 1^2 + 1 = 2.$$

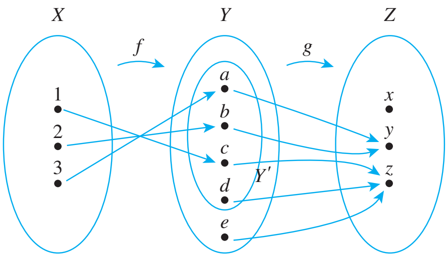
Thus the two functions $g \circ f$ and $f \circ g$ are not equal:

$$g \circ f \neq f \circ g.$$ ■

Example 7.3.1 illustrates the important fact that composition of functions is not a commutative operation: *For general functions F and G, $F \circ G$ need not necessarily equal $G \circ F$* (although the two *may* be equal).

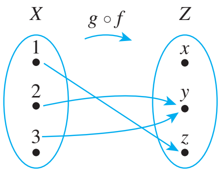
Example 7.3.2 Composition of Functions Defined on Finite Sets

Let $X = \{1, 2, 3\}$, $Y' = \{a, b, c, d\}$, $Y = \{a, b, c, d, e\}$, and $Z = \{x, y, z\}$. Define functions $f: X \to Y'$ and $g: Y \to Z$ by the arrow diagrams below.



Draw the arrow diagram for $g \circ f$. What is the range of $g \circ f$?

Solution To find the arrow diagram for $g \circ f$, just trace the arrows all the way across from X to Z through Y. The result is shown below.



$$(g \circ f)(1) = g(f(1)) = g(c) = z$$
$$(g \circ f)(2) = g(f(2)) = g(b) = y$$
$$(g \circ f)(3) = g(f(3)) = g(a) = y$$

The range of $g \circ f$ is $\{y, z\}$. ■

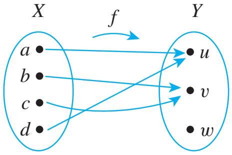
Recall that the identity function on a set X, I_X, is the function from X to X defined by the formula

$$I_X(x) = x \quad \text{for all } x \in X.$$

That is, the identity function on X sends each element of X to itself. What happens when an identity function is composed with another function?

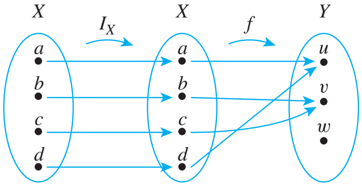
Example 7.3.3 Composition with the Identity Function

Let $X = \{a, b, c, d\}$ and $Y = \{u, v, w\}$, and suppose $f: X \to Y$ is given by the arrow diagram shown on the next page.



Find $f \circ I_X$ and $I_Y \circ f$.

Solution The values of $f \circ I_X$ are obtained by tracing through the arrow diagram shown below.



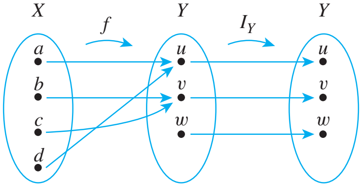
$$(f \circ I_X)(a) = f(I_X(a)) = f(a) = u$$
$$(f \circ I_X)(b) = f(I_X(b)) = f(b) = v$$
$$(f \circ I_X)(c) = f(I_X(c)) = f(c) = v$$
$$(f \circ I_X)(d) = f(I_X(d)) = f(d) = u$$

Note that for all elements x in X,

$$(f \circ I_X)(x) = f(x).$$

By definition of equality of functions, this means that $f \circ I_X = f$.

Similarly, the equality $I_Y \circ f = f$ can be verified by tracing through the arrow diagram below for each x in X and noting that in each case, $(I_Y \circ f)(x) = f(x)$.



■

More generally, the composition of any function with an identity function equals the function.

Theorem 7.3.1 Composition with an Identity Function

If f is a function from a set X to a set Y, and I_X is the identity function on X, and I_Y is the identity function on Y, then

(a) $f \circ I_X = f$ and (b) $I_Y \circ f = f$.

Proof:

Part (a): Suppose f is a function from a set X to a set Y and I_X is the identity function on X. Then, for all x in X,

$$(f \circ I_X)(x) = f(I_X(x)) = f(x).$$

Hence, by definition of equality of functions, $f \circ I_X = f$, as was to be shown.

Part (b): This is exercise 13 at the end of this section.

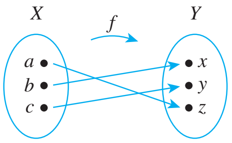
Now let f be a function from a set X to a set Y, and suppose f has an inverse function f^{-1}. Recall that f^{-1} is the function from Y to X with the property that

$$f^{-1}(y) = x \quad \Leftrightarrow \quad f(x) = y.$$

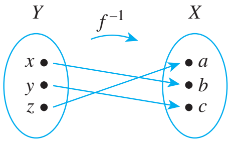
What happens when f is composed with f^{-1}? Or when f^{-1} is composed with f?

Example 7.3.4 Composing a Function with Its Inverse

Let $X = \{a, b, c\}$ and $Y = \{x, y, z\}$. Define $f\colon X \to Y$ by the following arrow diagram.



Then f is one-to-one and onto. Thus f^{-1} exists and is found by tracing the arrows backwards, as shown below.



Now $f^{-1} \circ f$ is found by following the arrows from X to Y by f and back to X by f^{-1}. If you do this, you will see that

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

and

$$(f^{-1} \circ f)(c) = f^{-1}(f(c)) = f^{-1}(y) = c.$$

Thus the composition of f and f^{-1} sends each element to itself. So by definition of the identity function,

$$f^{-1} \circ f = I_X.$$

In a similar way, you can see that

$$f \circ f^{-1} = I_Y.$$

■

More generally, the composition of any function with its inverse (if it has one) is an identity function. Intuitively, the function sends an element in its domain to an element in its co-domain and the inverse function sends it back again, so the composition of the two sends each element to itself. This reasoning is formalized in Theorem 7.3.2.

Theorem 7.3.2 Composition of a Function with Its Inverse

If $f: X \to Y$ is a one-to-one and onto function with inverse function $f^{-1}: Y \to X$, then

$$\text{(a)}\ \ f^{-1} \circ f = I_X \quad \text{and} \quad \text{(b)}\ \ f \circ f^{-1} = I_Y.$$

Proof:

Part (a): Suppose $f: X \to Y$ is a one-to-one and onto function with inverse function $f^{-1}: Y \to X$. *[To show that $f^{-1} \circ f = I_X$, we must show that for all $x \in X$, $(f^{-1} \circ f)(x) = x$.]* Let x be any element in X. Then

$$(f^{-1} \circ f)(x) = f^{-1}(f(x))$$

by definition of composition of functions. Now the inverse function f^{-1} satisfies the condition

$$f^{-1}(b) = a \quad \Leftrightarrow \quad f(a) = b \quad \text{for all } a \in X \text{ and } b \in Y. \tag{7.3.1}$$

Let

$$x' = f^{-1}(f(x)). \tag{7.3.2}$$

Apply property (7.3.1) with x' playing the role of a and $f(x)$ playing the role of b. Then

$$f(x') = f(x).$$

But since f is one-to-one, this implies that $x' = x$. Substituting x for x' in equation (7.3.2) gives

$$x = f^{-1}(f(x)).$$

Then by definition of composition of functions,

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

as was to be shown.

Part (b): This is exercise 14 at the end of this section.

Composition of One-to-One Functions

The composition of functions interacts in interesting ways with the properties of being one-to-one and onto. What happens, for instance, when two one-to-one functions are composed? Must their composition be one-to-one? For example, let $X = \{a, b, c\}$, $Y = \{w, x, y, z\}$, and $Z = \{1, 2, 3, 4, 5\}$, and define one-to-one functions $f: X \to Y$ and $g: Y \to Z$ as shown in the arrow diagrams of Figure 7.3.1.

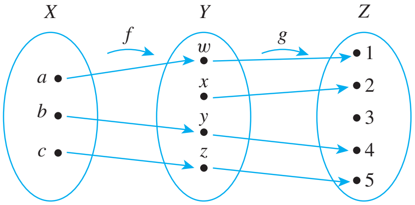


Figure 7.3.1

Then $g \circ f$ is the function with the arrow diagram shown in Figure 7.3.2.

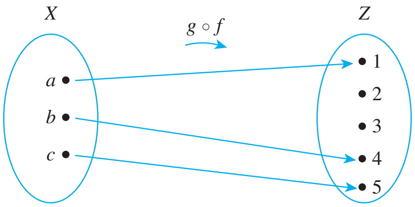


Figure 7.3.2

From the diagram it is clear that for these particular functions, the composition is one-to-one. This result is no accident. It turns out that the compositions of two one-to-one functions is always one-to-one.

Theorem 7.3.3

If $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are both one-to-one functions, then $g \circ f$ is one-to-one.

By the method of direct proof, the proof of Theorem 7.3.3 has the following starting point and conclusion to be shown.

Starting Point: Suppose f is a one-to-one function from X to Y and g is a one-to-one function from Y to Z.

To Show: $g \circ f$ is a one-to-one function from X to Z.

The conclusion to be shown says that a certain function is one-to-one. How do you show that? The crucial step is to realize that if you substitute $g \circ f$ into the definition of one-to-one, you see that

$$\boxed{g \circ f \text{ is one-to-one} \quad \Leftrightarrow \quad \forall x_1, x_2 \in X, \text{if } (g \circ f)(x_1) = (g \circ f)(x_2) \text{ then } x_1 = x_2.}$$

By the method of direct proof, then, to show $g \circ f$ is one-to-one, you

suppose x_1 and x_2 are elements of X such that $(g \circ f)(x_1) = (g \circ f)(x_2)$,

and you

show that $x_1 = x_2$.

Now the heart of the proof begins. To show that $x_1 = x_2$, you work forward from the supposition that $(g \circ f)(x_1) = (g \circ f)(x_2)$, using the fact that f and g are both one-to-one. By definition of composition,

$$(g \circ f)(x_1) = g(f(x_1)) \quad \text{and} \quad (g \circ f)(x_2) = g(f(x_2)).$$

Since the left-hand sides of the equations are equal, so are the right-hand sides. Thus

$$g(f(x_1)) = g(f(x_2)).$$

Now just stare at the above equation for a moment. It says that

$$g(\text{something}) = g(\text{something else}).$$

Because g is a one-to-one function, any time g of one thing equals g of another thing, those two things are equal. Hence

$$f(x_1) = f(x_2).$$

But f is also a one-to-one function. Any time f of one thing equals f of another thing, those two things are equal. Therefore,

$$x_1 = x_2.$$

This is what was to be shown!

This discussion is summarized in the following formal proof.

Proof of Theorem 7.3.3:

Suppose $f\colon X \to Y$ and $g\colon Y \to Z$ are both one-to-one functions. *[We must show that $g \circ f$ is one-to-one.]* Suppose x_1 and x_2 are elements of X such that

$$(g \circ f)(x_1) = (g \circ f)(x_2).$$

[We must show that $x_1 = x_2$.] By definition of composition of functions,

$$g(f(x_1)) = g(f(x_2)).$$

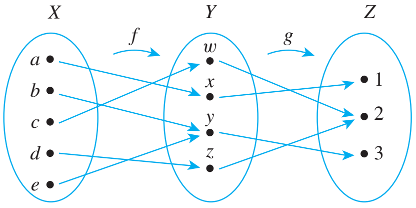
Since g is one-to-one, $\quad f(x_1) = f(x_2).$

And since f is one-to-one, $\quad x_1 = x_2.$

[This is what was to be shown.] Hence $g \circ f$ is one-to-one.

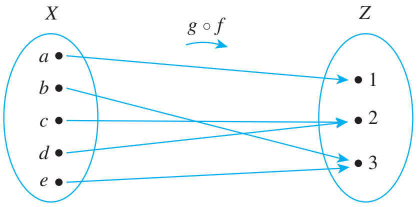
Composition of Onto Functions

Now consider what happens when two onto functions are composed. For example, let $X = \{a, b, c, d, e\}$, $Y = \{w, x, y, z\}$, and $Z = \{1, 2, 3\}$. Define onto functions f and g by the following arrow diagrams.



Then $g \circ f$ is the function with the arrow diagram shown below.

It is clear from the diagram that $g \circ f$ is onto.



It turns out that the composition of any two onto functions (that can be composed) is onto.

Theorem 7.3.4

If $f: X \to Y$ and $g: Y \to Z$ are both onto functions, then $g \circ f$ is onto.

A direct proof of Theorem 7.3.4 has the following starting point and conclusion to be shown:

Starting Point: Suppose f is an onto function from X to Y, and g is an onto function from Y to Z.

To Show: $g \circ f$ is an onto function from X to Z.

The conclusion to be shown says that a certain function is onto. How do you show that? The crucial step is to realize that if you substitute $g \circ f$ into the definition of onto, you see that

$g \circ f: X \to$ Z is onto $\Leftrightarrow$ given any element z of Z, it is possible to find an element x of X such that $(g \circ f)(x) = z$.

Caution! To show that a function is onto, you *must* start with on arbitrary element of the co-domain and deduce that it is the image of some element in the domain.

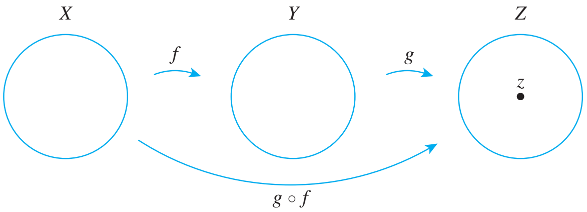
Since this statement is universal, to prove it you

suppose z is a *[particular but arbitrarily chosen]* element of Z

and **show** that there is an element x in X such that $(g \circ f)(x) = z$.

Hence you must start the proof by supposing you are given a particular but arbitrarily chosen element in Z. Let us call it z. Your job is to find an element x in X such that $(g \circ f)(x) = z$.

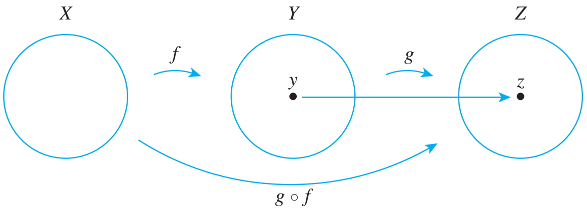
To find x, reason from the supposition that z is in Z, using the fact that both g and f are onto. Imagine arrow diagrams for the functions f and g.



You have a particular element z in Z, and you need to find an element x in X such that when x is sent over to Z by $g \circ f$, its image will be z. Since g is onto, z is at the tip of some arrow coming from Y. That is, there is an element y in Y such that

$$g(y) = z. \qquad 7.3.3$$

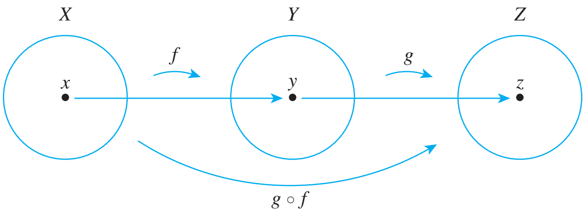
This means that the arrow diagrams can be drawn as follows:



But f also is onto, so every element in Y is at the tip of an arrow coming from X. In particular, y is at the tip of some arrow. That is, there is an element x in X such that

$$f(x) = y. \tag{7.3.4}$$

The diagram, therefore, can be drawn as shown below.



Now just substitute equation (7.3.4) into equation (7.3.3) to obtain

$$g(f(x)) = z.$$

But by definition of $g \circ f$,

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

Hence

$$(g \circ f)(x) = z.$$

Thus x is an element of X that is sent by $g \circ f$ to z, and so x is the element you were supposed to find.

This discussion is summarized in the following formal proof.

Proof of Theorem 7.3.4:

Suppose $f: X \to Y$ and $g: Y \to Z$ are both onto functions. *[We must show that $g \circ f$ is onto.]* Let z be a *[particular but arbitrarily chosen]* element of Z. *[We must show the existence of an element x in X such that $(g \circ f)(x) = z$.]* Since g is onto, there is an element y in Y such that $g(y) = z$. And since f is onto, there is an element x in X such that $f(x) = y$. Hence there exists an element x in X such that

$$(g \circ f)(x) = g(f(x)) = g(y) = z$$

[as was to be shown]. It follows that $g \circ f$ is onto.

Example 7.3.5 An Incorrect "Proof" That a Function Is Onto

To prove that a composition of onto functions is onto, a student wrote,

> "Suppose $f: X \to Y$ and $g: Y \to Z$ are both onto. Then
>
> $\forall y \in Y, \exists x \in X$ such that $f(x) = y$ (*)
>
> and
>
> $\forall z \in Z, \exists y \in Y$ such that $f(y) = z$.
>
> So
>
> $(g \circ f)(x) = g(f(x)) = g(y) = z,$
>
> and thus $g \circ f$ is onto."

Explain the mistakes in this "proof."

Solution To show that $g \circ f$ is onto, you must be able to meet the following challenge: If someone gives you an element z in Z (over which you have no control), you must be able to explain how to find an element x in X such that $(g \circ f)(x) = z$. Thus a proof that $g \circ f$ is onto must start with the assumption that you have been given a particular but arbitrarily chosen element of Z. This proof does not do that.

Moreover, note that statement (*) simply asserts that f is onto. An informal version of (*) is the following: Given any element in the co-domain of f, there is an element in the domain of f that is sent by f to the given element. Use of the symbols x and y to denote these elements is arbitrary. Any other two symbols could equally well have been used. Thus, if we replace the x and y in (*) by u and v, we obtain a logically equivalent statement, and the "proof" becomes the following:

> "Suppose $f\colon X \to Y$ and $g\colon Y \to Z$ are both onto. Then
>
> $$\forall v \in Y, \exists u \in X \text{ such that } f(u) = v$$
>
> and
>
> $$\forall z \in Z, \exists y \in Y \text{ such that } f(y) = z.$$
>
> So (??!)
>
> $$(g \circ f)(x) = g(f(x)) = g(y) = z,$$
>
> and thus $g \circ f$ is onto."

From this logically equivalent version of the "proof," you can see that the statements leading up to the word *So* do not provide a rationale for the statement that follows it. The original reason for writing *So* was based on a misinterpretation of the meaning of the notation. ■

Test Yourself

1. If f is a function from X to Y', g is a function from Y to Z, and $Y' \subseteq Y$, then $g \circ f$ is a function from _____ to _____, and $(g \circ f)(x) =$ _____ for all x in X.

2. If f is a function from X to Y and I_x and I_y are the identity functions from X to X and Y to Y, respectively, then $f \circ I_x =$ _____ and $I_y \circ f =$ _____.

3. If f is a one-to-one correspondence from X to Y, then $f^{-1} \circ f =$ _____ and $f \circ f^{-1} =$ _____.

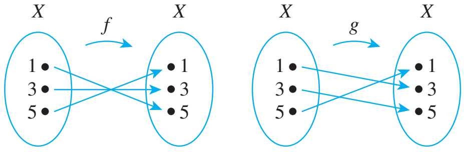
4. If f is a one-to-one function from X to Y and g is a one-to-one function from Y to Z, you prove that $g \circ f$ is one-to-one by supposing that _____ and then showing that _____.

5. If f is an onto function from X to Y and g is an onto function from Y to Z, you prove that $g \circ f$ is onto by supposing that _____ and then showing that _____.

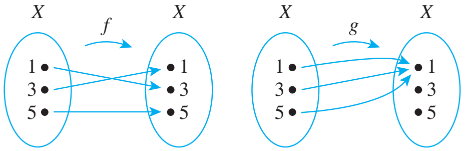
Exercise Set 7.3

In each of 1 and 2, functions f and g are defined by arrow diagrams. Find $G \circ F$ and $f \circ g$ and determine whether $G \circ F$ equals $f \circ g$.

1.



2.



In 3 and 4, functions F and G are defined by formulas. Find $G \circ F$ and $F \circ G$ and determine whether $G \circ F$ equals $F \circ G$.

3. $F(x) = x^3$ and $G(x) = x - 1$, for all real numbers x.

4. $F(x) = x^5$ and $G(x) = x^{1/5}$ for all real numbers x.

5. Define $f\colon \mathbf{R} \to \mathbf{R}$ by the rule $f(x) = -x$ for all real numbers x. Find $(f \circ f)(x)$.

6. Define $F\colon \mathbf{Z} \to \mathbf{Z}$ and $G\colon \mathbf{Z} \to \mathbf{Z}$ by the rules $F(a) = 7a$ and $G(a) = a \bmod 5$ for all integers a. Find $(G \circ F)(0)$, $(G \circ F)(1)$, $(G \circ F)(2)$, $(G \circ F)(3)$, and $(G \circ F)(4)$.

7. Define $H\colon \mathbf{Z} \to \mathbf{Z}$ and $K\colon \mathbf{Z} \to \mathbf{Z}$ by the rules $H(a) = 6a$ and $K(a) = a \mathit{mod}\ 4$ for all integers a. Find $(K \circ H)(0)$, $(K \circ H)(1)$, $(K \circ H)(2)$, and $(K \circ H)(3)$.

8. Define $L\colon \mathbf{Z} \to \mathbf{Z}$ and $M\colon \mathbf{Z} \to \mathbf{Z}$ by the rules $L(a) = a^2$ and $M(a) = a \mathit{mod}\ 5$ for all integers a.
   a. Find $(L \circ M)(12)$, $(M \circ L)(12)$, $(L \circ M)(9)$, and $(M \circ L)(9)$.
   b. Is $L \circ M = M \circ L$?

The functions of each pair in 9–11 are inverse to each other. For each pair, check that both compositions give the identity function.

9. $F\colon \mathbf{R} \to \mathbf{R}$ and $F^{-1}\colon \mathbf{R} \to \mathbf{R}$ are defined by

$$F(x) = 3x + 2 \quad \text{and} \quad F^{-1}(y) = \frac{y-2}{3},$$

for all $y \in \mathbf{R}$.

10. $G\colon \mathbf{R}^+ \to \mathbf{R}^+$ and $G^{-1}\colon \mathbf{R}^+ \to \mathbf{R}^+$ are defined by

$$G(x) = x^2 \quad \text{and} \quad G^{-1}(x) = \sqrt{x},$$

for all $x \in \mathbf{R}^+$.

11. H and H^{-1} are both defined from $\mathbf{R} - \{1\}$ to $\mathbf{R} - \{1\}$ by the formula

$$H(x) = H^{-1}(x) = \frac{x+1}{x-1}, \quad \text{for all } x \in \mathbf{R} - \{1\}.$$

12. Explain how it follows from the definition of logarithm that
   a. $\log_b(b^x) = x$, for all real numbers x.
   b. $b^{\log_b x} = x$, for all positive real numbers x.

H 13. Prove Theorem 7.3.1(b): If f is any function from a set X to a set Y, then $I_Y \circ f = f$, where I_Y is the identity function on Y.

14. Prove Theorem 7.3.2(b): If $f\colon X \to Y$ is a one-to-one and onto function with inverse function $f^{-1}\colon Y \to X$, then $f \circ f^{-1} = I_Y$, where I_Y is the identity function on Y.

15. Suppose Y and Z are sets and $g\colon Y \to Z$ is a one-to-one function. This means that if g takes the same value on any two elements of Y, then those elements are equal. Thus, for example, if a and b are elements of Y and $g(a) = g(b)$, then it can be inferred that $a = b$. What can be inferred in the following situations?
   a. s_k and s_m are elements of Y and $g(s_k) = g(s_m)$.
   b. $z/2$ and $t/2$ are elements of Y and $g(z/2) = g(t/2)$.
   c. $f(x_1)$ and $f(x_2)$ are elements of Y and $g(f(x_1)) = g(f(x_2))$.

16. If $f\colon X \to Y$ and $g\colon Y \to Z$ are functions and $g \circ f$ is one-to-one, must g be one-to-one? Prove or give a counterexample.

17. If $f\colon X \to Y$ and $g\colon Y \to Z$ are functions and $g \circ f$ is onto, must f be onto? Prove or give a counterexample.

H 18. If $f\colon X \to Y$ and $g\colon Y \to Z$ are functions and $g \circ f$ is one-to-one, must f be one-to-one? Prove or give a counterexample.

H 19. If $f\colon X \to Y$ and $g\colon Y \to Z$ are functions and $g \circ f$ is onto, must g be onto? Prove or give a counterexample.

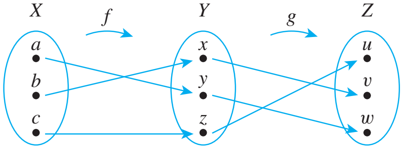
20. Let $f\colon W \to X$, $g\colon X \to Y$, and $h\colon Y \to Z$ be functions. Must $h \circ (g \circ f) = (h \circ g) \circ f$? Prove or give a counterexample.

21. True or False? Given any set X and given any functions $f\colon X \to X$, $g\colon X \to X$, and $h\colon X \to X$, if h is one-to-one and $h \circ f = h \circ g$, then $f = g$. Justify your answer.

22. True or False? Given any set X and given any functions $f\colon X \to X$, $g\colon X \to X$, and $h\colon X \to X$, if h is one-to-one and $f \circ h = g \circ h$, then $f = g$. Justify your answer.

In 23 and 24 find $g \circ f$, $(g \circ f)^{-1}$, g^{-1}, f^{-1}, and $f^{-1} \circ g^{-1}$, and state how $(g \circ f)^{-1}$ and $f^{-1} \circ g^{-1}$ are related.

23. Let $X = \{a, c, b\}$, $Y = \{x, y, z\}$, and $Z = \{u, v, w\}$. Define $f\colon X \to Y$ and $g\colon Y \to Z$ by the arrow diagrams below.



24. Define $f\colon \mathbf{R} \to \mathbf{R}$ and $g\colon \mathbf{R} \to \mathbf{R}$ by the formulas

$$f(x) = x + 3 \quad \text{and} \quad g(x) = -x \quad \text{for all } x \in \mathbf{R}.$$

25. Prove or give a counterexample: If $f\colon X \to Y$ and $g\colon Y \to X$ are functions such that $g \circ f = I_X$ and $f \circ g = I_Y$, then f and g are both one-to-one and onto and $g = f^{-1}$.

H 26. Suppose $f\colon X \to Y$ and $g\colon Y \to Z$ are both one-to-one and onto. Prove that $(g \circ f)^{-1}$ exists and that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

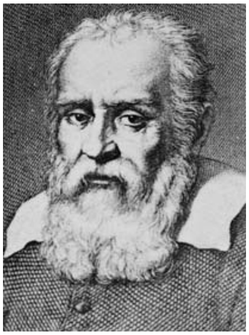
27. Let $f\colon X \to Y$ and $g\colon Y \to Z$. Is the following property true or false? For all subsets C in Z, $(g \circ f)^{-1}(C) = (f^{-1}(g^{-1}(C))$. Justify your answer.

Answers for Test Yourself

1. X; Z; $g(f(x))$ 2. f; f 3. I_X; I_Y 4. x_1 and x_2 are any *[particular but arbitrarily chosen]* elements in X with the property that $(g \circ f)(x_1) = (g \circ f)(x_2)$; $x_1 = x_2$ 5. z is any *[particular but arbitrarily chosen]* element in Z; there exists at least one element x in X such that $(g \circ f)(x) = z$

7.4 Cardinality with Applications to Computability

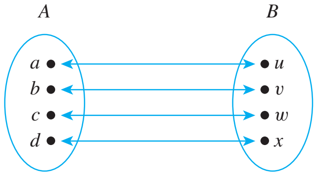
There are as many squares as there are numbers because they are just as numerous as their roots. —Galileo Galilei, 1632



iStockphoto.com/Steven Wynn

Galileo Galilei (1564–1642)

Historically, the term *cardinal number* was introduced to describe the size of a set ("This set has *eight* elements") as distinguished from an *ordinal number* that refers to the order of an element in a sequence ("This is the *eighth* element in the row"). The definition of cardinal number derives from the primitive technique of representing numbers by fingers or tally marks. Small children, when asked how old they are, will often answer by holding up a certain number of fingers, each finger being paired with a year of their life. As was discussed in Section 7.2, a pairing of the elements of two sets is called a one-to-one correspondence. We say that two finite sets whose elements can be paired by a one-to-one correspondence have the *same size.* This is illustrated by the following diagram.



The elements of set A can be put into one-to-one correspondence with the elements of B.

Now a **finite set** is one that has no elements at all or that can be put into one-to-one correspondence with a set of the form $\{1, 2, \ldots, n\}$ for some positive integer n. By contrast, an **infinite set** is a nonempty set that cannot be put into one-to-one correspondence with $\{1, 2, \ldots, n\}$ for any positive integer n. Suppose that, as suggested by the quote from Galileo at the beginning of this section, we extend the concept of size to infinite sets by saying that one infinite set has the same size as another if, and only if, the first set can be put into one-to-one correspondence with the second. What consequences follow from such a definition? Do all infinite sets have the same size, or are some infinite sets larger than others? These are the questions we address in this section. The answers are sometimes surprising and have the interesting consequence that there are functions defined on the set of integers whose values cannot be computed on a computer.

• Definition

Let A and B be any sets. ***A* has the same cardinality as *B*** if, and only if, there is a one-to-one correspondence from A to B. In other words, A has the same cardinality as B if, and only if, there is a function f from A to B that is one-to-one and onto.

The following theorem gives some basic properties of cardinality, most of which follow from statements proved earlier about one-to-one and onto functions.

Theorem 7.4.1 Properties of Cardinality

For all sets A, B, and C:

a. **Reflexive property of cardinality:** A has the same cardinality as A.

b. **Symmetric property of cardinality:** If A has the same cardinality as B, then B has the same cardinality as A.

c. **Transitive property of cardinality:** If A has the same cardinality as B and B has the same cardinality as C, then A has the same cardinality as C.

Proof:

Part (a), Reflexivity: Suppose A is any set. *[To show that A has the same cardinality as A, we must show there is a one-to-one correspondence from A to A.]* Consider the identity function I_A from A to A. This function is one-to-one because if x_1 and x_2 are any elements in A with $I_A(x_1) = I_A(x_2)$, then, by definition of I_A, $x_1 = x_2$. The identity function is also onto because if y is any element of A, then $y = I_A(y)$ by definition of I_A. Hence I_A is a one-to-one correspondence from A to A. *[So there exists a one-to-one correspondence from A to A, as was to be shown.]*

Part (b), Symmetry: Suppose A and B are any sets and A has the same cardinality as B. *[We must show that B has the same cardinality as A.]* Since A has the same cardinality as B, there is a function f from A to B that is one-to-one and onto. But then, by Theorems 7.2.2 and 7.2.3, there is a function f^{-1} from B to A that is also one-to-one and onto. Hence B has the same cardinality as A *[as was to be shown].*

Part (c), Transitivity: Suppose A, B, and C are any sets and A has the same cardinality as B and B has the same cardinality as C. *[We must show that A has the same cardinality as C.]* Since A has the same cardinality as B, there is a function f from A to B that is one-to-one and onto, and since B has the same cardinality as C, there is a function g from B to C that is one-to-one and onto. But then, by Theorems 7.3.3 and 7.3.4, $g \circ f$ is a function from A to C that is one-to-one and onto. Hence A has the same cardinality as C *[as was to be shown].*

Note that Theorem 7.4.1(b) makes it possible to say simply that two sets have the same cardinality instead of always having to say that one set has the same cardinality as another. That is, the following definition can be made.

• Definition

A and B **have the same cardinality** if, and only if, A has the same cardinality as B or B has the same cardinality as A.

The following example illustrates a very important property of infinite sets—namely, that an infinite set can have the same cardinality as a proper subset of itself. This property is sometimes taken as the definition of infinite set. The example shows that even though it may seem reasonable to say that there are twice as many integers as there are even integers, the elements of the two sets can be matched up exactly, and so, according to the definition, the two sets have the same cardinality.

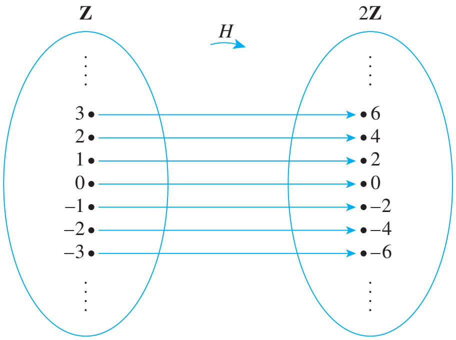
Example 7.4.1 An Infinite Set and a Proper Subset Can Have the Same Cardinality

Let $2\mathbf{Z}$ be the set of all even integers. Prove that $2\mathbf{Z}$ and $\mathbf{Z}$ have the same cardinality.

Solution Consider the function H from $\mathbf{Z}$ to $2\mathbf{Z}$ defined as follows:

$$H(n) = 2n \quad \text{for all } n \in \mathbf{Z}.$$

A (partial) arrow diagram for H is shown below.



To show that H is one-to-one, suppose $H(n_1) = H(n_2)$ for some integers n_1 and n_2. Then $2n_1 = 2n_2$ by definition of H, and dividing both sides by 2 gives $n_1 = n_2$. Hence h is one-to-one.

To show that H is onto, suppose m is any element of 2**Z**. Then m is an even integer, and so $m = 2k$ for some integer k. It follows that $H(k) = 2k = m$. Thus there exists k in **Z** with $H(k) = m$, and hence H is onto.

Note So there are "as many" even integers as there are integers!

Therefore, by definition of cardinality, **Z** and 2**Z** have the same cardinality. ■

In Section 9.4 we will show that a function from one finite set to another set of the same size is one-to-one if, and only if, it is onto. This result does not hold for infinite sets. Although it is true that for two infinite sets to have the same cardinality there must exist a function from one to the other that is both one-to-one and onto, it is also always the case that:

> If A and B are infinite sets with the same cardinality, then there exist functions from A to B that are one-to-one but not onto and functions from A to B that are onto but not one-to-one.

For instance, since the function H in Example 7.4.1 is one-to-one and onto, **Z** and 2**Z** have the same cardinality. But the "inclusion function" I from 2**Z** to **Z**, given by $I(n) = n$ for all even integers n, is one-to-one but not onto. And the function J from **Z** to 2**Z** defined by $J(n) = 2\lfloor n/2 \rfloor$, for all integers n, is onto but not one-to-one. (See exercise 6 at the end of this section.)

Countable Sets

The set $\mathbf{Z}^+$ of counting numbers $\{1, 2, 3, 4, \ldots\}$ is, in a sense, the most basic of all infinite sets. A set A having the same cardinality as this set is called *countably infinite*. The reason is that the one-to-one correspondence between the two sets can be used to "count" the elements of A: If F is a one-to-one and onto function from $\mathbf{Z}^+$ to A, then $F(1)$ can be designated as the first element of A, $F(2)$ as the second element of A, $F(3)$ as the third element of A, and so forth. This is illustrated graphically in Figure 7.4.1 on the next page. Because F is one-to-one, no element is ever counted twice, and because it is onto, every element of A is counted eventually.

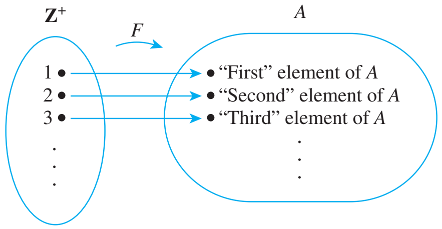


Figure 7.4.1 "Counting" a Countably Infinite Set

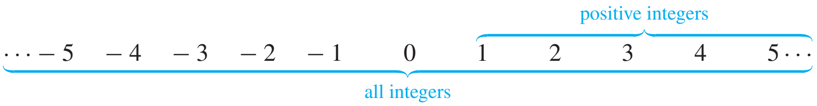
• Definition

A set is called **countably infinite** if, and only if, it has the same cardinality as the set of positive integers $\mathbf{Z}^+$. A set is called **countable** if, and only if, it is finite or countably infinite. A set that is not countable is called **uncountable.**

Example 7.4.2 Countability of Z, the Set of All Integers

Show that the set $\mathbf{Z}$ of all integers is countable.

Solution The set $\mathbf{Z}$ of all integers is certainly not finite, so if it is countable, it must be because it is countably infinite. To show that $\mathbf{Z}$ is countably infinite, find a function from the positive integers $\mathbf{Z}^+$ to $\mathbf{Z}$ that is one-to-one and onto. Looked at in one light, this contradicts common sense; judging from the diagram below, there appear to be more than twice as many integers as there are positive integers.



But you were alerted that results in this section might be surprising. Try to think of a way to "count" the set of all integers anyway.

The trick is to start in the middle and work outward systematically. Let the first integer be 0, the second 1, the third −1, the fourth 2, the fifth −2, and so forth as shown in Figure 7.4.2, starting at 0 and swinging outward in back-and-forth arcs from positive to negative integers and back again, picking up one additional integer at each swing.

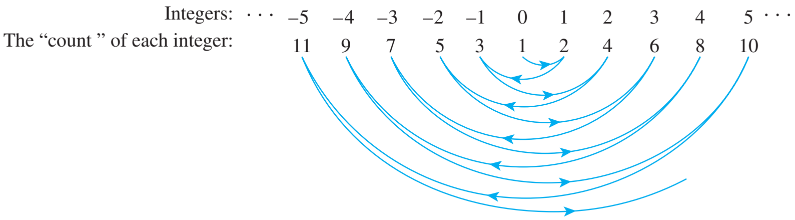


Figure 7.4.2 "Counting" the Set of All Integers

It is clear from the diagram that no integer is counted twice (so the function is one-to-one) and every integer is counted eventually (so the function is onto). Consequently, this diagram defines a function from $\mathbf{Z}^+$ to $\mathbf{Z}$ that is one-to-one and onto. Even though in one sense there seem to be more integers than positive integers, the elements of the two sets can be paired up one for one. It follows by definition of cardinality that $\mathbf{Z}^+$ has the same cardinality as $\mathbf{Z}$. Thus $\mathbf{Z}$ is countably infinite and hence countable.

The diagrammatic description of the previous function is acceptable as given. You can check, however, that the function can also be described by the explicit formula

$$F(n) = \begin{cases} \dfrac{n}{2} & \text{if } n \text{ is an even positive integer} \\ -\dfrac{n-1}{2} & \text{if } n \text{ is an odd positive integer.} \end{cases}$$ ■

Example 7.4.3 Countability of 2Z, the Set of All Even Integers

Show that the set $2\mathbf{Z}$ of all even integers is countable.

Solution Example 7.4.2 showed that $\mathbf{Z}^+$ has the same cardinality as $\mathbf{Z}$, and Example 7.4.1 showed that $\mathbf{Z}$ has the same cardinality as $2\mathbf{Z}$. Thus, by the transitive property of cardinality, $\mathbf{Z}^+$ has the same cardinality as $2\mathbf{Z}$. It follows by definition of countably infinite that $2\mathbf{Z}$ is countably infinite and hence countable. ■

The Search for Larger Infinities: The Cantor Diagonalization Process

Every infinite set we have discussed so far has been countably infinite. Do any larger infinities exist? Are there uncountable sets? Here is one candidate.

Imagine the number line as shown below.

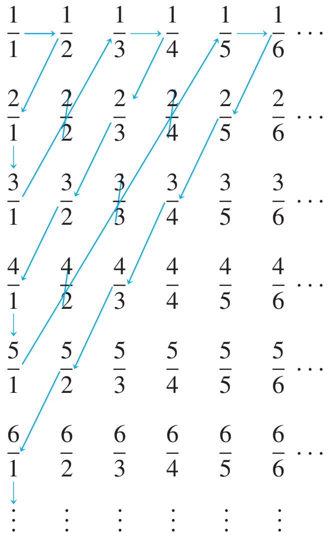
As noted in Section 1.2, the integers are spread along the number line at discrete intervals. The rational numbers, on the other hand, are *dense:* Between any two rational numbers, no matter how close, lies another rational number (the average of the two numbers, for instance; see exercise 17). This suggests the conjecture that the infinity of the set of rational numbers is larger than the infinity of the set of integers.

Amazingly, this conjecture is false. Despite the fact that the rational numbers are crowded onto the number line whereas the integers are quite separated, the set of all rational numbers can be put into one-to-one correspondence with the set of integers. The next example gives part of a proof of this fact. It shows that the set of all positive rational numbers can be put into one-to-one correspondence with the set of all positive integers. In exercise 16 at the end of this section you are asked to use this result, together with a technique similar to that of Example 7.4.2, to show that the set of *all* rational numbers is countable.

Example 7.4.4 The Set of All Positive Rational Numbers Is Countable

Show that the set $\mathbf{Q}^+$ of all positive rational numbers is countable.

Solution Display the elements of the set $\mathbf{Q}^+$ of positive rational numbers in a grid as shown in Figure 7.4.3 on the next page.



$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\cdots$
$\frac{2}{1}$	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	$\frac{2}{5}$	$\frac{2}{6}$	$\cdots$
$\frac{3}{1}$	$\frac{3}{2}$	$\frac{3}{3}$	$\frac{3}{4}$	$\frac{3}{5}$	$\frac{3}{6}$	$\cdots$
$\frac{4}{1}$	$\frac{4}{2}$	$\frac{4}{3}$	$\frac{4}{4}$	$\frac{4}{5}$	$\frac{4}{6}$	$\cdots$
$\frac{5}{1}$	$\frac{5}{2}$	$\frac{5}{3}$	$\frac{5}{4}$	$\frac{5}{5}$	$\frac{5}{6}$	$\cdots$
$\frac{6}{1}$	$\frac{6}{2}$	$\frac{6}{3}$	$\frac{6}{4}$	$\frac{6}{5}$	$\frac{6}{6}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Figure 7.4.3

Define a function F from $\mathbf{Z}^+$ to $\mathbf{Q}^+$ by starting to count at $\frac{1}{1}$ and following the arrows as indicated, skipping over any number that has already been counted.

To be specific: Set $F(1) = \frac{1}{1}$, $F(2) = \frac{1}{2}$, $F(3) = \frac{2}{1}$ and $F(4) = \frac{3}{1}$. Then skip $\frac{2}{2}$ since $\frac{2}{2} = \frac{1}{1}$, which was counted first. After that, set $F(5) = \frac{1}{3}$, $F(6) = \frac{1}{4}$, $F(7) = \frac{2}{3}$, $F(8) = \frac{3}{2}$, $F(9) = \frac{4}{1}$, and $F(10) = \frac{5}{1}$. Then skip $\frac{4}{2}$, $\frac{3}{3}$, and $\frac{2}{4}$ (since $\frac{4}{2} = \frac{2}{1}$, $\frac{3}{3} = \frac{1}{1}$, and $\frac{2}{4} = \frac{1}{2}$) and set $F(11) = \frac{1}{5}$. Continue in this way, defining $F(n)$ for each positive integer n.

Note that every positive rational number appears somewhere in the grid, and the counting procedure is set up so that every point in the grid is reached eventually. Thus the function F is onto. Also, skipping numbers that have already been counted ensures that no number is counted twice. Thus F is one-to-one. Consequently, F is a function from $\mathbf{Z}^+$ to $\mathbf{Q}^+$ that is one-to-one and onto, and so $\mathbf{Q}^+$ is countably infinite and hence countable. ■

al-Kashi
(1380–1429)

Bettmann/CORBIS

Simon Stevin
(1548–1620)

In 1874 the German mathematician Georg Cantor achieved success in the search for a larger infinity by showing that the set of all real numbers is uncountable. His method of proof was somewhat complicated, however. We give a proof of the uncountability of the set of all real numbers between 0 and 1 using a simpler technique introduced by Cantor in 1891 and now called the **Cantor diagonalization process.** Over the intervening years, this technique and variations on it have been used to establish a number of important results in logic and the theory of computation.

Before stating and proving Cantor's theorem, we note that every real number, which is a measure of location on a number line, can be represented by a decimal expansion of the form

$$a_0.a_1a_2a_3\ldots,$$

where a_0 is an integer (positive, negative, or zero) and for each $i \geq 1$, a_i is an integer from 0 through 9.

This way of thinking about numbers was developed over several centuries by mathematicians in the Chinese, Hindu, and Islamic worlds, culminating in the work of Ghiyāth al-Dīn Jamshīd al-Kashi in 1427. In Europe it was first clearly formulated and successfully promoted by the Flemish mathematician Simon Stevin in 1585. We illustrate the concept with an example.

Consider the point P in Figure 7.4.4. Figure 7.4.4(a) shows P located between 1 and 2. When the interval from 1 to 2 is divided into ten equal subintervals (see Figure 7.4.4(b)) P is seen to lie between 1.6 and 1.7. If the interval from 1.6 to 1.7 is itself divided into ten equal subintervals (see Figure 7.4.4(c)), the P is seen to lie between 1.62 and 1.63 but closer to 1.62 than to 1.63. So the first three digits of the decimal expansion for P are 1.62.

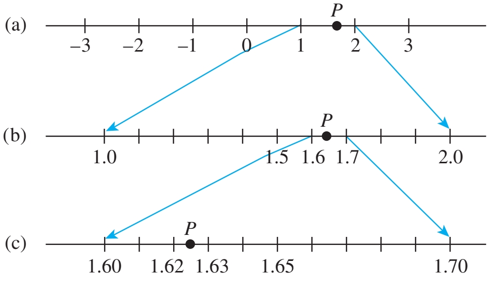


Figure 7.4.4

Assuming that any interval of real numbers, no matter how small, can be divided into ten equal subintervals, the process of obtaining additional digits in the decimal expansion for P can, in theory, be repeated indefinitely. If at any stage P is seen to be a subdivision point, then all further digits in the expansion may be taken to be 0. If not, then the process gives an expansion with an infinite number of digits.

The resulting decimal representation for P is unique except for numbers that end in infinitely repeating 9's or infinitely repeating 0's. For example (see exercise 25 at the end of this section),

$$0.199999\ldots = 0.200000\ldots.$$

Let us agree to express any such decimal in the form that ends in all 0's so that we will have a unique representation for every real number.

Theorem 7.4.2 (Cantor)

The set of all real numbers between 0 and 1 is uncountable.

Proof (by contradiction):

Suppose the set of all real numbers between 0 and 1 is countable. Then the decimal representations of these numbers can be written in a list as follows:

$$\begin{array}{c} 0.a_{11}a_{12}a_{13}\cdots a_{1n}\cdots \\ 0.a_{21}a_{22}a_{23}\cdots a_{2n}\cdots \\ 0.a_{31}a_{32}a_{33}\cdots a_{3n}\cdots \\ \vdots \\ 0.a_{n1}a_{n2}a_{n3}\cdots a_{nn}\cdots \\ \vdots \end{array}$$

[We will derive a contradiction by showing that there is a number between 0 *and* 1 *that does not appear on this list.]*

For each pair of positive integers i and j, the jth decimal digit of the ith number on the list is a_{ij}. In particular, the first decimal digit of the first number on the list is

a_{11}, the second decimal digit of the second number on the list is a_{22}, and so forth. As an example, suppose the list of real numbers between 0 and 1 starts out as follows:

$$\begin{array}{lcccccccl}
0. & \textcircled{2} & 0 & 1 & 4 & 8 & 8 & 0 & 2\ldots \\
0. & 1 & \textcircled{1} & 6 & 6 & 6 & 0 & 2 & 1\ldots \\
0. & 0 & 3 & \textcircled{3} & 5 & 3 & 3 & 2 & 0\ldots \\
0. & 9 & 6 & 7 & \textcircled{7} & 6 & 8 & 0 & 9\ldots \\
0. & 0 & 0 & 0 & 3 & \textcircled{1} & 0 & 0 & 2\ldots \\
 & & & & \vdots & & & &
\end{array}$$

The diagonal elements are circled: a_{11} is 2, a_{22} is 1, a_{33} is 3, a_{44} is 7, a_{55} is 1, and so forth.

Construct a new decimal number $d = 0.d_1d_2d_3 \cdots d_n \cdots$ as follows:

$$d_n = \begin{cases} 1 & \text{if } a_{nn} \neq 1 \\ 2 & \text{if } a_{nn} = 1 \end{cases}.$$

In the previous example,

$$\begin{aligned}
&d_1 \text{ is 1 because } a_{11} = 2 \neq 1, \\
&d_2 \text{ is 2 because } a_{22} = 1, \\
&d_3 \text{ is 1 because } a_{33} = 3 \neq 1, \\
&d_4 \text{ is 1 because } a_{44} = 7 \neq 1, \\
&d_5 \text{ is 2 because } a_{55} = 1,
\end{aligned}$$

and so forth. Hence d would equal $0.12112\ldots$.

The crucial observation is that for *each integer n, d differs in the nth decimal position from the nth number on the list*. But this implies that d is not on the list! In other words, d is a real number between 0 and 1 that is not on the list of *all* real numbers between 0 and 1. This contradiction shows the falseness of the supposition that the set of all numbers between 0 and 1 is countable. Hence the set of all real numbers between 0 and 1 is uncountable.

Along with demonstrating the existence of an uncountable set, Cantor developed a whole arithmetic theory of infinite sets of various sizes. One of the most basic theorems of the theory states that any subset of a countable set is countable.

Theorem 7.4.3

Any subset of any countable set is countable.

Proof:

Let A be a particular but arbitrarily chosen countable set and let B be any subset of A. *[We must show that B is countable.]* Either B is finite or it is infinite. If B is finite, then B is countable by definition of countable, and we are done. So suppose B is infinite. Since A is countable, the distinct elements of A can be represented as a sequence

$$a_1, a_2, a_3, \ldots.$$

Define a function $g: \mathbf{Z}^+ \to B$ inductively as follows:

continued on page 436

1. Search sequentially through elements of $a_1, a_2, a_3, \ldots$ until an element of B is found. *[This must happen eventually since $B \subseteq A$ and $B \neq \emptyset$.]* Call that element $g(1)$.
2. For each integer $k \geq 2$, suppose $g(k-1)$ has been defined. Then $g(k-1) = a_i$ for some a_i in $\{a_1, a_2, a_3, \ldots\}$. Starting with a_{i+1}, search sequentially through $a_{i+1}, a_{i+2}, a_{i+3}, \ldots$ trying to find an element of B. One must be found eventually because B is infinite, and $\{g(1), g(2), \ldots, g(k-1)\}$ is a finite set. When an element of B is found, define it to be $g(k)$.

Note If $g(k-1) = a_i$, then $g(k)$ could also be defined by applying the well-ordering principle for the integers to the set $\{n \in \mathbf{Z} \mid n > i$ and $a_i \in B\}$.

By (1) and (2) above, the function g is defined for each positive integer.

Since the elements of $a_1, a_2, a_3, \ldots$ are all distinct, g is one-to-one. Furthermore, the searches for elements of B are sequential: Each picks up where the previous one left off. Thus every element of A is reached during some search. But all the elements of B are located somewhere in the sequence $a_1, a_2, a_3, \ldots$, and so every element of B is eventually found and made the image of some integer. Hence g is onto. These remarks show that g is a one-to-one correspondence from $\mathbf{Z}^+$ to B. So B is countably infinite and thus countable.

An immediate consequence of Theorem 7.4.3 is the following corollary.

Corollary 7.4.4

Any set with an uncountable subset is uncountable.

Proof:

Consider the following equivalent phrasing of Theorem 7.4.3: For all sets S and for all subsets A of S, if S is countable, then A is countable. The contrapositive of this statement is logically equivalent to it and states: For all sets S and for all subsets A of S, if A is uncountable then S is uncountable. But this is an equivalent phrasing for the corollary. So the corrollary is proved.

Corollary 7.4.4 implies that the set of all real numbers is uncountable because the subset of numbers between 0 and 1 is uncountable. In fact, as Example 7.4.5 shows, the set of all real numbers has the same cardinality as the set of all real numbers between 0 and 1! This fact is further explored in exercises 13 and 14 at the end of this section.

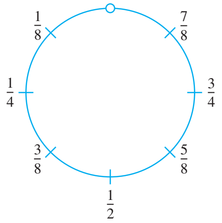
Example 7.4.5 The Cardinality of the Set of All Real Numbers

Show that the set of all real numbers has the same cardinality as the set of real numbers between 0 and 1.

Solution Let S be the open interval of real numbers between 0 and 1:

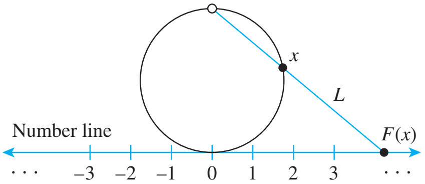
$$S = \{x \in \mathbf{R} \mid 0 < x < 1\}.$$

Imagine picking up S and bending it into a circle as shown below. Since S does not include either endpoint 0 or 1, the top-most point of the circle is omitted from the drawing.



Define a function $F: S \to \mathbf{R}$ as follows:

Draw a number line and place the interval, S, somewhat enlarged and bent into a circle, tangent to the line above the point 0. This is shown below.



For each point x on the circle representing S, draw a straight line L through the top-most point of the circle and x. Let $F(x)$ be the point of intersection of L and the number line. ($F(x)$ is called the *projection* of x onto the number line.)

It is clear from the geometry of the situation that distinct points on the circle go to distinct points on the number line, so F is one-to-one. In addition, given any point y on the number line, a line can be drawn through y and the top-most point of the circle. This line must intersect the circle at some point x, and, by definition, $y = F(x)$. Thus F is onto. Hence F is a one-to-one correspondence from S to **R,** and so S and **R** have the same cardinality. ■

You know that every positive integer is a real number, so putting Example 7.4.5 together with Cantor's theorem (Theorem 7.4.2) shows that the infinity of the set of all real numbers is "greater" than the infinity of the set of all positive integers. In exercise 35, you are asked to show that any set and its power set have different cardinalities. Because there is a one-to-one function from any set to its power set (the function that takes each element a to the singleton set $\{a\}$), this implies that the cardinality of any set is "less than" the cardinality of its power set. As a result, you can create an infinite sequence of larger and larger infinities! For example, you could begin with **Z**, the set of all integers, and take $\mathbf{Z}$, $\mathscr{P}(\mathbf{Z})$, $\mathscr{P}(\mathscr{P}(\mathbf{Z}))$, $\mathscr{P}(\mathscr{P}(\mathscr{P}(\mathbf{Z})))$, and so forth.

Application: Cardinality and Computability

Knowledge of the countability and uncountability of certain sets can be used to answer a question of computability. We begin by showing that a certain set is countable.

Example 7.4.6 Countability of the Set of Computer Programs in a Computer Language

Show that the set of all computer programs in a given computer language is countable.

Solution This result is a consequence of the fact that any computer program in any language can be regarded as a finite string of symbols in the (finite) alphabet of the language.

Given any computer language, let P be the set of all computer programs in the language. Either P is finite or P is infinite. If P is finite, then P is countable and we are done. If P is infinite, set up a binary code to translate the symbols of the alphabet of the language into strings of 0's and 1's. (For instance, either the seven-bit American Standard Code for Information Interchange, known as ASCII, or the eight-bit Extended Binary-Coded Decimal Interchange Code, known as EBCDIC, might be used.)

For each program in P, use the code to translate all the symbols in the program into 0's and 1's. Order these strings by length, putting shorter before longer, and order all strings of a given length by regarding each string as a binary number and writing the numbers in ascending order.

Define a function $F\colon \mathbf{Z}^+ \to P$ by specifying that

$$F(n) = \text{the } n\text{th program in the list} \qquad \text{for each } n \in \mathbf{Z}^+.$$

By construction, F is one-to-one and onto, and so P is countably infinite and hence countable. As a simple example, suppose the following are all the programs in P that translate into bit strings of length less than or equal to 5:

$$10111,\ 11,\ 0010,\ 1011,\ 01,\ 00100,\ 1010,\ 00010.$$

Ordering these by length gives

length 2: 11, 01

length 4: 0010, 1011, 1010

length 5: 10111, 00100, 00010

And ordering those of each given length by the size of the binary number they represent gives

$$\begin{aligned}
01 \quad &= F(1)\\
11 \quad &= F(2)\\
0010 \ &= F(3)\\
1010 \ &= F(4)\\
1011 \ &= F(5)\\
00010 &= F(6)\\
00100 &= F(7)\\
10111 &= F(8)
\end{aligned}$$

Note that when viewed purely as numbers, ignoring leading zeros, $0010 = 00010$. This shows the necessity of first ordering the strings by length before arranging them in ascending numeric order. ■

The final example of this section shows that a certain set is uncountable and hence that there must exist a noncomputable function.

Example 7.4.7 The Cardinality of a Set of Functions and Computability

a. Let T be the set of all functions from the positive integers to the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Show that T is uncountable.

b. Derive the consequence that there are noncomputable functions. Specifically, show that for any computer language there must be a function F from $\mathbf{Z}^+$ to $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ with the property that no computer program can be written in the language to take arbitrary values as input and output the corresponding function values.

Solution

a. Let S be the set of all real numbers between 0 and 1. As noted before, any number in S can be represented in the form

$$0.a_1a_2a_3 \ldots a_n \ldots,$$

where each a_i is an integer from 0 to 9. This representation is unique if decimals that end in all 9's are omitted.

Define a function F from S to a subset of T (the set of all functions from $\mathbf{Z}^+$ to $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$) as follows:

$$F(0.a_1a_2a_3\ldots a_n\ldots) = \text{the function that sends each positive integer } n \text{ to } a_n.$$

Choose the co-domain of F to be exactly that subset of T that makes F onto. That is, define the co-domain of F to equal the image of F. Note that F is one-to-one because if $F(x_1) = F(x_2)$, then each decimal digit of x_1 equals the corresponding decimal digit of x_2, and so $x_1 = x_2$. Thus F is a one-to-one correspondence from S to a subset of T. But S is uncountable by Theorem 7.4.2. Hence T has an uncountable subset, and so, by Corollary 7.4.4, T is uncountable.

b. Part (a) shows that the set T of all functions from $\mathbf{Z}^+$ to $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is uncountable. But Example 7.4.6 shows that given any computer language, the set of all programs in that language is countable. Consequently, in any computer language there are not enough programs to compute values of every function in T. There must exist functions that are not computable! ■

Test Yourself

1. A set is finite if, and only if, _____.

2. To prove that a set A has the same cardinality as a set B you must _____.

3. The reflexive property of cardinality says that given any set A, _____.

4. The symmetric property of cardinality says that given any sets A and B, _____.

5. The transitive property of cardinality says that given any sets A, B, and C, _____.

6. A set is called countably infinite if, and only if, _____.

7. A set is called countable if, and only if, _____.

8. In each of the following, fill in the blank with the word *countable* or the word *uncountable*.

   (a) The set of all integers is _____.

   (b) The set of all rational numbers is _____.

   (c) The set of all real numbers between 0 and 1 is _____.

   (d) The set of all real numbers is _____.

9. The Cantor diagonalization process is used to prove that _____.

Exercise Set 7.4

1. When asked what it means to say that set A has the same cardinality as set B, a student replies, "A and B are one-to-one and onto." What *should* the student have replied? Why?

2. Show that "there are as many squares as there are numbers" by exhibiting a one-to-one correspondence from the positive integers, $\mathbf{Z}^+$, to the set S of all squares of positive integers:

$$S = \{n \in \mathbf{Z}^+ \mid n = k^2, \text{ for some positive integer } k\}.$$

3. Let $3\mathbf{Z} = \{n \in \mathbf{Z} \mid n = 3k, \text{ for some integer } k\}$. Prove that $\mathbf{Z}$ and $3\mathbf{Z}$ have the same cardinality.

4. Let $\mathbf{O}$ be the set of all odd integers. Prove that $\mathbf{O}$ has the same cardinality as $2\mathbf{Z}$, the set of all even integers.

5. Let $25\mathbf{Z}$ be the set of all integers that are multiples of 25. Prove that $25\mathbf{Z}$ has the same cardinality as $2\mathbf{Z}$, the set of all even integers.

H **6.** Use the functions I and J defined in the paragraph following Example 7.4.1 to show that even though there is a one-to-one correspondence, H, from $2\mathbf{Z}$ to $\mathbf{Z}$, there is also a function from $2\mathbf{Z}$ to $\mathbf{Z}$ that is one-to-one but not onto and a function from $\mathbf{Z}$ to $2\mathbf{Z}$ that is onto but not one-to-one. In other words, show that I is one-to-one but not onto, and show that J is onto but not one-to-one.

7. a. Check that the formula for F given at the end of Example 7.4.2 produces the correct values for $n = 1, 2, 3$, and 4.

   b. Use the floor function to write a formula for F as a single algebraic expression for all positive integers n.

8. Use the result of exercise 3 to prove that $3\mathbf{Z}$ is countable.

9. Show that the set of all nonnegative integers is countable by exhibiting a one-to-one correspondence between $\mathbf{Z}^+$ and $\mathbf{Z}^{nonneg}$.

In 10–14, S denotes the set of real numbers strictly between 0 and 1. That is, $S = \{x \in \mathbf{R} \mid 0 < x < 1\}$.

10. Let $U = \{x \in \mathbf{R} \mid 0 < x < 2\}$. Prove that S and U have the same cardinality.

H **11.** Let $V = \{x \in \mathbf{R} \mid 2 < x < 5\}$. Prove that S and V have the same cardinality.

12. Let a and b be real numbers with $a < b$, and suppose that $W = \{x \in \mathbf{R} \mid a < x < b\}$. Prove that S and W have the same cardinality.

13. Draw the graph of the function f defined by the following formula:

For all real numbers x with $0 < x < 1$,

$$f(x) = \tan\left(\pi x - \frac{\pi}{2}\right).$$

Use the graph to explain why S and $\mathbf{R}$ have the same cardinality.

★ 14. Define a function g from the set of real numbers to S by the following formula:

For all real numbers x,

$$g(x) = \frac{1}{2} \cdot \left(\frac{x}{1 + |x|}\right) + \frac{1}{2}.$$

Prove that g is a one-to-one correspondence. (It is possible to prove this statement either with calculus or without it.) What conclusion can you draw from this fact?

15. Show that the set of all bit strings (strings of 0's and 1's) is countable.

16. Show that $\mathbf{Q}$, the set of all rational numbers, is countable.

17. Show that the set $\mathbf{Q}$ of all rational numbers is dense along the number line by showing that given any two rational numbers r_1 and r_2 with $r_1 < r_2$, there exists a rational number x such that $r_1 < x < r_2$.

H **18.** Must the average of two irrational numbers always be irrational? Prove or give a counterexample.

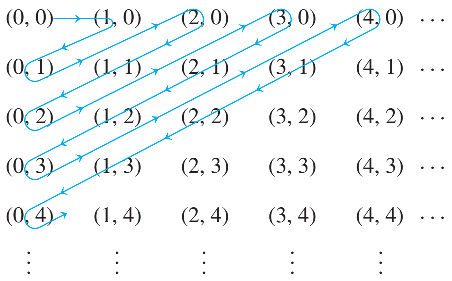
H★ **19.** Show that the set of all irrational numbers is dense along the number line by showing that given any two real numbers, there is an irrational number in between.

20. Give two examples of functions from $\mathbf{Z}$ to $\mathbf{Z}$ that are one-to-one but not onto.

21. Give two examples of functions from $\mathbf{Z}$ to $\mathbf{Z}$ that are onto but not one-to-one.

H **22.** Define a function $g\colon \mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ by the formula $g(m, n) = 2^m 3^n$ for all $(m, n) \in \mathbf{Z}^+ \times \mathbf{Z}^+$. Show that g is one-to-one and use this result to prove that $\mathbf{Z}^+ \times \mathbf{Z}^+$ is countable.

23. **a.** Explain how to use the following diagram to show that $\mathbf{Z}^{nonneg} \times \mathbf{Z}^{nonneg}$ and $\mathbf{Z}^{nonneg}$ have the same cardinality.



(0, 0)	(1, 0)	(2, 0)	(3, 0)	(4, 0)	...
(0, 1)	(1, 1)	(2, 1)	(3, 1)	(4, 1)	...
(0, 2)	(1, 2)	(2, 2)	(3, 2)	(4, 2)	...
(0, 3)	(1, 3)	(2, 3)	(3, 3)	(4, 3)	...
(0, 4)	(1, 4)	(2, 4)	(3, 4)	(4, 4)	...
⋮	⋮	⋮	⋮	⋮	

H★ **b.** Define a function $H\colon \mathbf{Z}^{nonneg} \times \mathbf{Z}^{nonneg} \to \mathbf{Z}^{nonneg}$ by the formula

$$H(m, n) = n + \frac{(m+n)(m+n+1)}{2}$$

for all nonnegative integers m and n. Interpret the action of H geometrically using the diagram of part (a).

★ 24. Prove that the function H defined analytically in exercise 23b is a one-to-one correspondence.

H **25.** Prove that $0.1999\ldots = 0.2$.

26. Prove that any infinite set contains a countably infinite subset.

27. If A is any countably infinite set, B is any set, and $g\colon A \to B$ is onto, then B is countable.

28. Prove that a disjoint union of any finite set and any countably infinite set is countably infinite.

H **29.** Prove that a union of any two countably infinite sets is countably infinite.

H **30.** Use the result of exercise 29 to prove that the set of all irrational numbers is uncountable.

H **31.** Use the results of exercises 28 and 29 to prove that a union of any two countable sets is countable.

H **32.** Prove that $\mathbf{Z} \times \mathbf{Z}$, the Cartesian product of the set of integers with itself, is countably infinite.

33. Use the results of exercises 27, 31, and 32 to prove the following: If R is the set of all solutions to all equations of the form $x^2 + bx + c = 0$, where b and c are integers, then R is countable.

H **34.** Let $\mathscr{P}(S)$ be the set of all subsets of set S, and let T be the set of all functions from S to $\{0, 1\}$. Show that $\mathscr{P}(S)$ and T have the same cardinality.

H **35.** Let S be a set and let $\mathscr{P}(S)$ be the set of all subsets of S. Show that S is "smaller than" $\mathscr{P}(S)$ in the sense that there is a one-to-one function from S to $\mathscr{P}(S)$ but there is no onto function from $\mathscr{P}(S)$ to S.

✶36. The Schroeder–Bernstein theorem states the following: If A and B are any sets with the property that there is a one-to-one function from A to B and a one-to-one function from B to A, then A and B have the same cardinality. Use this theorem to prove that there are as many functions from $\mathbf{Z}^+$ to $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ as there are functions from $\mathbf{Z}^+$ to $\{0, 1\}$.

H **37.** Prove that if A and B are any countably infinite sets, then $A \times B$ is countably infinite.

✶38. Suppose $A_1, A_2, A_3, \ldots$ is an infinite sequence of countable sets. Recall that

$$\bigcup_{i=1}^{\infty} A_i = \{x \mid x \in A_i \text{ for some positive integer } i\}.$$

Prove that $\bigcup_{i=1}^{\infty} A_i$ is countable. (In other words, prove that a countably infinite union of countable sets is countable.)

Answers for Test Yourself

1. it is the empty set or there is a one-to-one correspondence from $\{1, 2, \ldots, n\}$ to it, where n is a positive integer 2. show that there exists a function from A to B that is one-to-one and onto (*Or*: show that there exists a one-to-one correspondence from A to B) 3. A has the same cardinality as A. 4. if A has the same cardinality as B, then B has the same cardinality as A 5. if A has the same cardinality as B and B has the same cardinality as C, then A has the same cardinality as C 6. it has the same cardinality as the set of all positive integers 7. it is finite or countably infinite 8. countable; countable; uncountable; uncountable 9. the set of all real numbers between 0 and 1 is uncountable