DISCRETE and COMBINATORIAL MATHEMATICS

An Applied Introduction

Ralph P. Grimaldi

Fifth Edition

DISCRETE AND COMBINATORIAL MATHEMATICS

An Applied Introduction

FIFTH EDITION

RALPH P. GRIMALDI
Rose-Hulman Institute of Technology

PEARSON
Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

NOTATION

LOGIC	p, q	statements (or propositions)
	$\neg p$	the negation of (statement) p: *not p*
	$p \wedge q$	the conjunction of p, q: *p and q*
	$p \vee q$	the disjunction of p, q: *p or q*
	$p \rightarrow q$	the implication of q by p: *p implies q*
	$p \leftrightarrow q$	the biconditional of p and q: *p if and only if q*
	iff	if and only if
	$p \Rightarrow q$	logical implication: *p logically implies q*
	$p \Longleftrightarrow q$	logical equivalence: p is *logically equivalent* to q
	T_0	tautology
	F_0	contradiction
	$\forall x$	For *all* x (the universal quantifier)
	$\exists x$	For *some* x (the existential quantifier)
SET THEORY	$x \in A$	element x is a member of set A
	$x \notin A$	element x is not a member of set A
	$\mathcal{U}$	the universal set
	$A \subseteq B$, $B \supseteq A$	A is a subset of B
	$A \subset B$, $B \supset A$	A is a proper subset of B
	$A \nsubseteq B$	A is not a subset of B
	$A \not\subset B$	A is not a proper subset of B
	$\lvert A \rvert$	the cardinality, or size, of set A — that is, the number of elements in A
	$\emptyset = \{\ \}$	the empty, or null, set
	$\mathcal{P}(A)$	the power set of A — that is, the collection of all subsets of A
	$A \cap B$	the intersection of sets A, B: $\{x \mid x \in A \text{ and } x \in B\}$
	$A \cup B$	the union of sets A, B: $\{x \mid x \in A \text{ or } x \in B\}$
	$A \triangle B$	the symmetric difference of sets A, B: $\{x \mid x \in A \text{ or } x \in B, \text{ but } x \notin A \cap B\}$
	$\overline{A}$	the complement of set A: $\{x \mid x \in \mathcal{U} \text{ and } x \notin A\}$
	$A - B$	the (relative) complement of set B in set A: $\{x \mid x \in A \text{ and } x \notin B\}$
	$\bigcup_{i \in I} A_i$	$\{x \mid x \in A_i, \text{ for at least one } i \in I\}$, where I is an index set
	$\bigcap_{i \in I} A_i$	$\{x \mid x \in A_i, \text{ for every } i \in I\}$, where I is an index set
PROBABILITY	S	the sample space for an experiment $\mathcal{E}$
	$A \subseteq S$	A is an event
	$Pr(A)$	the probability of event A
	$Pr(A \mid B)$	the probability of A given B; conditional probability
	X	random variable
	$E(X)$	the expected value of X, a random variable
	$\text{Var}(X) = \sigma_X^2$	the variance of X, a random variable
	σ_X	the standard deviation of X, a random variable
NUMBERS	$a \mid b$	a divides b, for $a, b \in \mathbf{Z}$, $a \neq 0$
	$a \nmid b$	a does not divide b, for $a, b \in \mathbf{Z}$, $a \neq 0$
	$\gcd(a, b)$	the greatest common divisor of the integers a, b
	$\text{lcm}(a, b)$	the least common multiple of the integers a, b
	$\phi(n)$	Euler's phi function for $n \in \mathbf{Z}^+$
	$\lfloor x \rfloor$	the greatest integer less than or equal to the real number x: the greatest integer in x: the *floor* of x

NOTATION

	$\lceil x \rceil$	the smallest integer greater than or equal to the real number x: the *ceiling* of x
	$a \equiv b \pmod{n}$	a is congruent to b modulo n
RELATIONS	$A \times B$	the Cartesian, or cross, product of sets A, B: $\{(a, b) \mid a \in A, b \in B\}$
	$\mathcal{R} \subseteq A \times B$	$\mathcal{R}$ is a relation from A to B
	$a \mathcal{R} b$; $(a, b) \in \mathcal{R}$	a is related to b
	$a \not\mathcal{R} b$; $(a, b) \notin \mathcal{R}$	a is not related to b
	$\mathcal{R}^c$	the converse of relation $\mathcal{R}$: $(a, b) \in \mathcal{R}$ iff $(b, a) \in \mathcal{R}^c$
	$\mathcal{R} \circ \mathcal{S}$	the composite relation for $\mathcal{R} \subseteq A \times B$, $\mathcal{S} \subseteq B \times C$: $(a, c) \in \mathcal{R} \circ \mathcal{S}$ if $(a, b) \in \mathcal{R}$, $(b, c) \in \mathcal{S}$ for some $b \in B$
	$\text{lub}\{a, b\}$	the least upper bound of a and b
	$\text{glb}\{a, b\}$	the greatest lower bound of a and b
	$[a]$	the equivalence class of element a (relative to an equivalence relation $\mathcal{R}$ on a set A): $\{x \in A \mid x \mathcal{R} a\}$
FUNCTIONS	$f: A \to B$	f is a function from A to B
	$f(A_1)$	for $f: A \to B$ and $A_1 \subseteq A$, $f(A_1)$ is the image of A_1 under f — that is, $\{f(a) \mid a \in A_1\}$
	$f(A)$	for $f: A \to B$, $f(A)$ is the range of f
	$f: A \times A \to B$	f is a binary operation on A
	$f: A \times A \to B \ (\subseteq A)$	f is a closed binary operation on A
	$1_A: A \to A$	the identity function on A: $1_A(a) = a$ for each $a \in A$
	$f\|_{A_1}$	the restriction of $f: A \to B$ to $A_1 \subseteq A$
	$g \circ f$	the composite function for $f: A \to B$, $g: B \to C$: $(g \circ f)a = g(f(a))$, for $a \in A$
	f^{-1}	the inverse of function f
	$f^{-1}(B_1)$	the preimage of $B_1 \subseteq B$ for $f: A \to B$
	$f \in O(g)$	f is "big Oh" of g; f is of order g
THE ALGEBRA OF STRINGS	$\mathbf{\Sigma}$	a finite set of symbols called an alphabet
	λ	the empty string
	$\\|x\\|$	the length of string x
	$\mathbf{\Sigma}^n$	$\{x_1 x_2 \cdots x_n \mid x_i \in \mathbf{\Sigma}\}$, $n \in \mathbf{Z}^+$
	$\mathbf{\Sigma}^0$	$\{\lambda\}$
	$\mathbf{\Sigma}^+$	$\bigcup_{n \in \mathbf{Z}^+} \mathbf{\Sigma}^n$: the set of all strings of positive length
	$\mathbf{\Sigma}^*$	$\bigcup_{n \geq 0} \mathbf{\Sigma}^n$: the set of all finite strings
	$A \subseteq \mathbf{\Sigma}^*$	A is a language
	AB	the concatenation of languages $A, B \subseteq \mathbf{\Sigma}^*$: $\{ab \mid a \in A, b \in B\}$
	A^n	$\{a_1 a_2 \cdots a_n \mid a_i \in A \subseteq \mathbf{\Sigma}^*\}$, $n \in \mathbf{Z}^+$
	A^0	$\{\lambda\}$
	A^+	$\bigcup_{n \in \mathbf{Z}^+} A^n$
	A^*	$\bigcup_{n \geq 0} A^n$: the Kleene closure of language A
	$M = (S, \mathcal{S}, \mathcal{O}, \nu, \omega)$	a finite state machine M with internal states S, input alphabet $\mathcal{S}$, output alphabet $\mathcal{O}$, next state function $\nu: S \times \mathcal{S} \to S$ and output function $\omega: S \times \mathcal{S} \to \mathcal{O}$

5

Relations and Functions

In this chapter we extend the set theory of Chapter 3 to include the concepts of relation and function. Algebra, trigonometry, and calculus all involve functions. Here, however, we shall study functions from a set-theoretic approach that includes finite functions, and we shall introduce some new counting ideas in the study. Furthermore, we shall examine the concept of function complexity and its role in the study of the analysis of algorithms.

We take a path along which we shall find the answers to the following (closely related) six problems:

1) The Defense Department has seven different contracts that deal with a high-security project. Four companies can manufacture the distinct parts called for in each contract, and in order to maximize the security of the overall project, it is best to have all four companies working on some part. In how many ways can the contracts be awarded so that every company is involved?
2) How many seven-symbol quaternary (0, 1, 2, 3) sequences have at least one occurrence of each of the symbols 0, 1, 2, and 3?
3) An $m \times n$ *zero-one matrix* is a matrix A with m rows and n columns, such that in row i, for all $1 \leq i \leq m$, and column j, for all $1 \leq j \leq n$, the entry a_{ij} that appears is either 0 or 1. How many 7×4 zero-one matrices have exactly one 1 in each row and at least one 1 in each column? (The zero-one matrix is a data structure that arises in computer science. We shall learn more about it in later chapters.)
4) Seven (unrelated) people enter the lobby of a building which has four additional floors, and they all get on an elevator. What is the probability that the elevator must stop at every floor in order to let passengers off?
5) For positive integers m, n with $m < n$, prove that

$$\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m = 0.$$

6) For every positive integer n, verify that

$$n! = \sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^n.$$

Do you recognize the connection among the first four problems? The first three are the same problem in different settings. However, it is not obvious that the last two problems are related or that there is a connection between them and the first four. These identities, however, will be established using the same counting technique that we develop to solve the first four problems.

5.1 Cartesian Products and Relations

We start with an idea that was introduced earlier in Definition 3.11. However, we repeat the definition now in order to make the presentation here independent of this prior encounter.

Definition 5.1 For sets A, B the *Cartesian product*, or *cross product*, of A and B is denoted by $A \times B$ and equals $\{(a, b) | a \in A, b \in B\}$.

We say that the elements of $A \times B$ are *ordered pairs*. For $(a, b), (c, d) \in A \times B$, we have $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$.

If A, B are finite, it follows from the rule of product that $|A \times B| = |A| \cdot |B|$. Although we generally will not have $A \times B = B \times A$, we will have $|A \times B| = |B \times A|$.

Here $A \subseteq \mathcal{U}_1$ and $B \subseteq \mathcal{U}_2$, and we may find that the universes are different—that is, $\mathcal{U}_1 \neq \mathcal{U}_2$. Also, even if $A, B \subseteq \mathcal{U}$, it is not necessary that $A \times B \subseteq \mathcal{U}$, so unlike the cases for union and intersection, here $\mathcal{P}(\mathcal{U})$ is not necessarily closed under this binary operation.

We can extend the definition of the Cartesian product, or cross product, to more than two sets. Let $n \in \mathbf{Z}^+$, $n \geq 3$. For sets $A_1, A_2, \ldots, A_n$, the *(n-fold) product* of $A_1, A_2, \ldots, A_n$ is denoted by $A_1 \times A_2 \times \cdots \times A_n$ and equals $\{(a_1, a_2, \ldots, a_n) | a_i \in A_i, 1 \leq i \leq n\}$.† The elements of $A_1 \times A_2 \times \cdots \times A_n$ are called ordered *n-tuples*, although we generally use the term *triple* in place of 3-tuple. As with ordered pairs, if $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \in A_1 \times A_2 \times \cdots \times A_n$, then $(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n)$ if and only if $a_i = b_i$ for all $1 \leq i \leq n$.

EXAMPLE 5.1 Let $A = \{2, 3, 4\}$, $B = \{4, 5\}$. Then

a) $A \times B = \{(2, 4), (2, 5), (3, 4), (3, 5), (4, 4), (4, 5)\}$.

b) $B \times A = \{(4, 2), (4, 3), (4, 4), (5, 2), (5, 3), (5, 4)\}$.

c) $B^2 = B \times B = \{(4, 4), (4, 5), (5, 4), (5, 5)\}$.

d) $B^3 = B \times B \times B = \{(a, b, c) | a, b, c \in B\}$; for instance, $(4, 5, 5) \in B^3$.

EXAMPLE 5.2 The set $\mathbf{R} \times \mathbf{R} = \{(x, y) | x, y \in \mathbf{R}\}$ is recognized as the real plane of coordinate geometry and two-dimensional calculus. The subset $\mathbf{R}^+ \times \mathbf{R}^+$ is the interior of the first quadrant of this plane. Likewise $\mathbf{R}^3$ represents Euclidean three-space, where the three-dimensional interior of any sphere (of positive radius), two-dimensional planes, and one-dimensional lines are subsets of importance.

EXAMPLE 5.3 Once again let $A = \{2, 3, 4\}$ and $B = \{4, 5\}$, as in Example 5.1, and let $C = \{x, y\}$. The construction of the Cartesian product $A \times B$ can be represented pictorially with the aid of a *tree diagram*, as in part (a) of Fig. 5.1. This diagram proceeds from left to right. From

†When dealing with the Cartesian product of three or more sets, we must be careful about the lack of associativity. In the case of three sets, for example, there is a difference between any two of the sets $A_1 \times A_2 \times A_3$, $(A_1 \times A_2) \times A_3$, and $A_1 \times (A_2 \times A_3)$ because their respective elements are ordered triples (a_1, a_2, a_3), and the distinct ordered pairs $((a_1, a_2), a_3)$ and $(a_1, (a_2, a_3))$. Although such differences are important in certain instances, we shall not concentrate on them here and shall always use the nonparenthesized form $A_1 \times A_2 \times A_3$. This will also be our convention when dealing with the Cartesian product of four or more sets.

the left-most endpoint, three branches originate — one for each of the elements of A. Then from each point, labeled 2, 3, 4, two branches emanate — one for each of the elements 4, 5 of B. The six ordered pairs at the right endpoints constitute the elements (ordered pairs) of $A \times B$. Part (b) of the figure provides a tree diagram to demonstrate the construction of $B \times A$. Finally, the tree diagram in Fig. 5.1 (c) shows us how to envision the construction of $A \times B \times C$, and demonstrates that $|A \times B \times C| = 12 = 3 \times 2 \times 2 = |A||B||C|$.

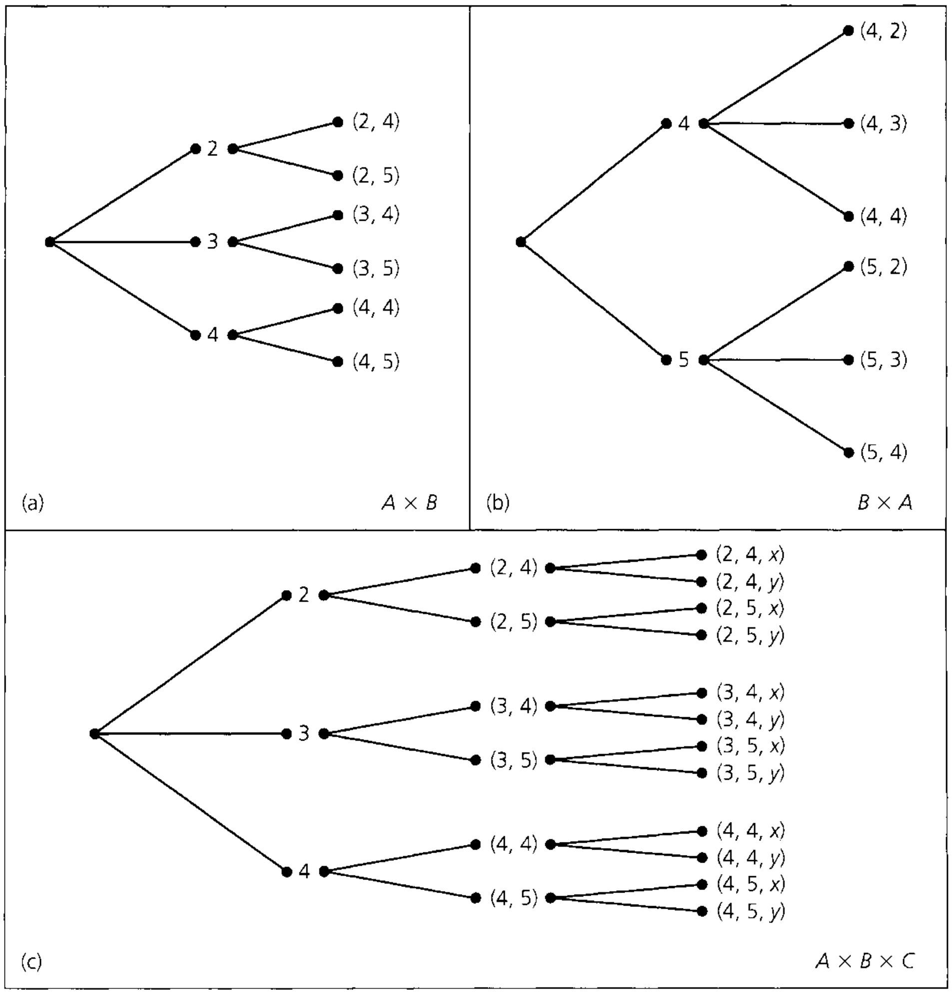


Figure 5.1

In addition to their tie-in with Cartesian products, tree diagrams also arise in other situations.

EXAMPLE 5.4

At the Wimbledon Tennis Championships, women play at most three sets in a match. The winner is the first to win two sets. If we let N and E denote the two players, the tree diagram in Fig. 5.2 indicates the six ways in which this match can be won. For example, the starred line segment (edge) indicates that player E won the first set. The double-starred edge indicates that player N has won the match by winning the first and third sets.

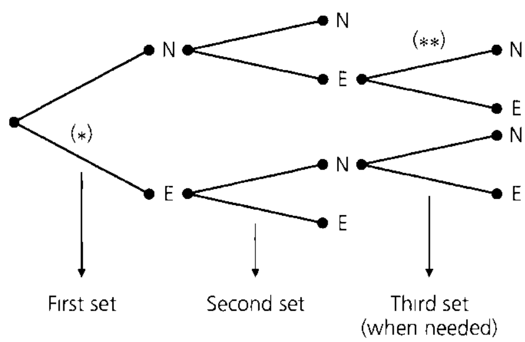


Figure 5.2

Tree diagrams are examples of a general structure called a *tree*. Trees and graphs are important structures that arise in computer science and optimization theory. These will be investigated in later chapters.

For the cross product of two sets, we find the subsets of this structure of great interest.

Definition 5.2 For sets A, B, any subset of $A \times B$ is called a (*binary*) *relation* from A to B. Any subset of $A \times A$ is called a (*binary*) *relation* on A.

Since we will primarily deal with binary relations, for us the word "relation" will mean binary relation, unless something otherwise is specified.

EXAMPLE 5.5

With A, B as in Example 5.1, the following are some of the relations from A to B.

a) $\emptyset$
b) $\{(2, 4)\}$
c) $\{(2, 4), (2, 5)\}$
d) $\{(2, 4), (3, 4), (4, 4)\}$
e) $\{(2, 4), (3, 4), (4, 5)\}$
f) $A \times B$

Since $|A \times B| = 6$, it follows from Definition 5.2 that there are 2^6 possible relations from A to B (for there are 2^6 possible subsets of $A \times B$).

For finite sets A, B with $|A| = m$ and $|B| = n$, there are 2^{mn} relations from A to B, including the empty relation as well as the relation $A \times B$ itself.

There are also $2^{nm} (= 2^{mn})$ relations from B to A, one of which is also $\emptyset$ and another of which is $B \times A$. The reason we get the same number of relations from B to A as we have from A to B is that any relation $\mathcal{R}_1$ from B to A can be obtained from a unique relation $\mathcal{R}_2$ from A to B by simply reversing the components of each ordered pair in $\mathcal{R}_2$ (and vice versa).

EXAMPLE 5.6

For $B = \{1, 2\}$, let $A = \mathcal{P}(B) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. The following is an example of a *relation* on A: $\mathcal{R} = \{(\emptyset, \emptyset), (\emptyset, \{1\}), (\emptyset, \{2\}), (\emptyset, \{1, 2\}), (\{1\}, \{1\}), (\{1\}, \{1, 2\}), (\{2\}, \{2\}), (\{2\}, \{1, 2\}), (\{1, 2\}, \{1, 2\})\}$. We can say that the relation $\mathcal{R}$ is the *subset relation* where $(C, D) \in \mathcal{R}$ if and only if $C, D \subseteq B$ and $C \subseteq D$.

EXAMPLE 5.7

With $A = \mathbf{Z}^+$, we may define a relation $\mathcal{R}$ on set A as $\{(x, y)|x \leq y\}$. This is the familiar "is less than or equal to" relation for the set of positive integers. It can be represented graphically as the set of points, with positive integer components, located on or above the line $y = x$ in the Euclidean plane, as partially shown in Fig. 5.3. Here we cannot list the entire relation as we did in Example 5.6, but we note, for example, that $(7, 7), (7, 11) \in \mathcal{R}$, but $(8, 2) \notin \mathcal{R}$. The fact that $(7, 11) \in \mathcal{R}$ can also be denoted by $7\ \mathcal{R}\ 11$; $(8, 2) \notin \mathcal{R}$ becomes $8\ \not{\mathcal{R}}\ 2$. Here $7\ \mathcal{R}\ 11$ and $8\ \not{\mathcal{R}}\ 2$ are examples of the *infix* notation for a relation.

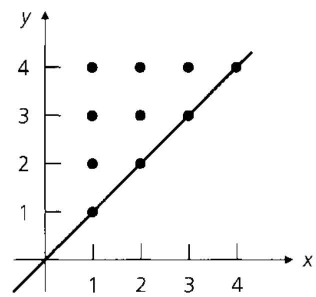


Figure 5.3

Our last example helps us to review the idea of a recursively defined set.

EXAMPLE 5.8

Let $\mathcal{R}$ be the subset of $\mathbf{N} \times \mathbf{N}$ where $\mathcal{R} = \{(m, n)|n = 7m\}$. Consequently, among the ordered pairs in $\mathcal{R}$ one finds (0, 0), (1, 7), (11, 77), and (15, 105). This relation $\mathcal{R}$ on $\mathbf{N}$ can also be given recursively by

1) $(0, 0) \in \mathcal{R}$; and

2) If $(s, t) \in \mathcal{R}$, then $(s + 1, t + 7) \in \mathcal{R}$.

We use the recursive definition to show that the ordered pair (3, 21) (from $\mathbf{N} \times \mathbf{N}$) is in $\mathcal{R}$. Our derivation is as follows: From part (1) of the recursive definition we start with $(0, 0) \in \mathcal{R}$. Then part (2) of the definition gives us

i) $(0, 0) \in \mathcal{R} \Rightarrow (0 + 1, 0 + 7) = (1, 7) \in \mathcal{R}$;

ii) $(1, 7) \in \mathcal{R} \Rightarrow (1 + 1, 7 + 7) = (2, 14) \in \mathcal{R}$; and

iii) $(2, 14) \in \mathcal{R} \Rightarrow (2 + 1, 14 + 7) = (3, 21) \in \mathcal{R}$.

We close this section with these final observations.

1) For any set A, $A \times \emptyset = \emptyset$. (If $A \times \emptyset \neq \emptyset$, let $(a, b) \in A \times \emptyset$. Then $a \in A$ and $b \in \emptyset$. Impossible!) Likewise, $\emptyset \times A = \emptyset$.

2) The Cartesian product and the binary operations of union and intersection are interrelated in the following theorem.

THEOREM 5.1

For any sets $A, B, C \subseteq \mathcal{U}$:

a) $A \times (B \cap C) = (A \times B) \cap (A \times C)$

b) $A \times (B \cup C) = (A \times B) \cup (A \times C)$

c) $(A \cap B) \times C = (A \times C) \cap (B \times C)$
d) $(A \cup B) \times C = (A \times C) \cup (B \times C)$

Proof: We prove part (a) and leave the other parts for the reader. We use the same concept of set equality (as in Definition 3.2 of Section 3.1) even though the elements here are ordered pairs. For all $a, b \in \mathcal{U}$, $(a, b) \in A \times (B \cap C) \Leftrightarrow a \in A$ and $b \in B \cap C \Leftrightarrow a \in A$ and $b \in B, C \Leftrightarrow a \in A, b \in B$ and $a \in A, b \in C \Leftrightarrow (a, b) \in A \times B$ and $(a, b) \in A \times C \Leftrightarrow (a, b) \in (A \times B) \cap (A \times C)$.

EXERCISES 5.1

1. If $A = \{1, 2, 3, 4\}$, $B = \{2, 5\}$, and $C = \{3, 4, 7\}$, determine $A \times B$; $B \times A$; $A \cup (B \times C)$; $(A \cup B) \times C$; $(A \times C) \cup (B \times C)$.

2. If $A = \{1, 2, 3\}$, and $B = \{2, 4, 5\}$, give examples of (a) three nonempty relations from A to B; (b) three nonempty relations on A.

3. For A, B as in Exercise 2, determine the following: (a) $|A \times B|$; (b) the number of relations from A to B; (c) the number of relations on A; (d) the number of relations from A to B that contain $(1, 2)$ and $(1, 5)$; (e) the number of relations from A to B that contain exactly five ordered pairs; and (f) the number of relations on A that contain at least seven elements.

4. For which sets A, B is it true that $A \times B = B \times A$?

5. Let A, B, C, D be nonempty sets.
   a) Prove that $A \times B \subseteq C \times D$ if and only if $A \subseteq C$ and $B \subseteq D$.
   b) What happens to the result in part (a) if any of the sets A, B, C, D is empty?

6. The men's final at Wimbledon is won by the first player to win three sets of the five-set match. Let C and M denote the players. Draw a tree diagram to show all the ways in which the match can be decided.

7. a) If $A = \{1, 2, 3, 4, 5\}$ and $B = \{w, x, y, z\}$, how many elements are there in $\mathcal{P}(A \times B)$?
   b) Generalize the result in part (a).

8. Logic chips are taken from a container, tested individually, and labeled defective or good. The testing process is continued until either two defective chips are found or five chips are tested in total. Using a tree diagram, exhibit a sample space for this process.

9. Complete the proof of Theorem 5.1.

10. A rumor is spread as follows. The originator calls two people. Each of these people phones three friends, each of whom in turn calls five associates. If no one receives more than one call, and no one calls the originator, how many people now know the rumor? How many phone calls were made?

11. For $A, B, C \subseteq \mathcal{U}$, prove that
$$A \times (B - C) = (A \times B) - (A \times C).$$

12. Let A, B be sets with $|B| = 3$. If there are 4096 relations from A to B, what is $|A|$?

13. Let $\mathcal{R} \subseteq \mathbf{N} \times \mathbf{N}$ where $(m, n) \in \mathcal{R}$ if (and only if) $n = 5m + 2$. (a) Give a recursive definition for $\mathcal{R}$. (b) Use the recursive definition from part (a) to show that $(4, 22) \in \mathcal{R}$.

14. a) Give a recursive definition for the relation $\mathcal{R} \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+$ where $(m, n) \in \mathcal{R}$ if (and only if) $m \geq n$.
    b) From the definition in part (a) verify that $(5, 2)$ and $(4, 4)$ are in $\mathcal{R}$.

5.2 Functions: Plain and One-to-One

In this section we concentrate on a special kind of relation called a *function*. One finds functions in many different settings throughout mathematics and computer science. As for general relations, they will reappear in Chapter 7, where we shall examine them much more thoroughly.

Definition 5.3 For nonempty sets A, B, a *function*, or *mapping*, *f from A to B*, denoted $f\colon A \to B$, is a relation from A to B in which every element of A appears exactly once as the first component of an ordered pair in the relation.

We often write $f(a) = b$ when (a, b) is an ordered pair in the function f. For $(a, b) \in f$, b is called *the image* of a under f, whereas a is a *preimage* of b. In addition, the definition suggests that f is a method for *associating* with each $a \in A$ the *unique* element $f(a) = b \in B$. Consequently, $(a, b), (a, c) \in f$ implies $b = c$.

EXAMPLE 5.9

For $A = \{1, 2, 3\}$ and $B = \{w, x, y, z\}$, $f = \{(1, w), (2, x), (3, x)\}$ is a function, and consequently a relation, from A to B. $\mathcal{R}_1 = \{(1, w), (2, x)\}$ and $\mathcal{R}_2 = \{(1, w), (2, w), (2, x), (3, z)\}$ are relations, but not functions, from A to B. (Why?)

Definition 5.4

For the function $f: A \to B$, A is called the *domain* of f and B the *codomain* of f. The subset of B consisting of those elements that appear as second components in the ordered pairs of f is called the *range* of f and is also denoted by $f(A)$ because it is the set of images (of the elements of A) under f.

In Example 5.9, the domain of $f = \{1, 2, 3\}$, the codomain of $f = \{w, x, y, z\}$, and the range of $f = f(A) = \{w, x\}$.

A pictorial representation of these ideas appears in Fig. 5.4. This diagram suggests that a may be regarded as an *input* that is *transformed* by f into the corresponding *output*, $f(a)$. In this context, a C++ compiler can be thought of as a function that transforms a source program (the input) into its corresponding object program (the output).

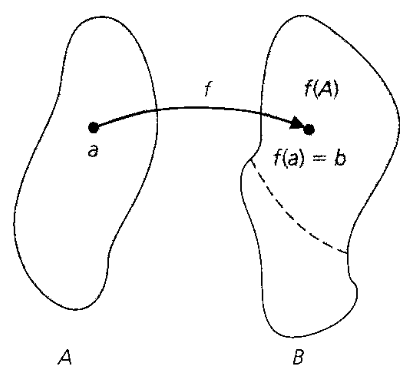


Figure 5.4

EXAMPLE 5.10

Many interesting functions arise in computer science.

a) A common function encountered is the *greatest integer function,* or *floor function.* This function $f: \mathbf{R} \to \mathbf{Z}$, is given by

$$f(x) = \lfloor x \rfloor = \text{the greatest integer less than or equal to } x.$$

Consequently, $f(x) = x$, if $x \in \mathbf{Z}$; and, when $x \in \mathbf{R} - \mathbf{Z}$, $f(x)$ is the integer to the immediate left of x on the real number line.

For this function we find that

1) $\lfloor 3.8 \rfloor = 3$, $\lfloor 3 \rfloor = 3$, $\lfloor -3.8 \rfloor = -4$, $\lfloor -3 \rfloor = -3$;

2) $\lfloor 7.1 + 8.2 \rfloor = \lfloor 15.3 \rfloor = 15 = 7 + 8 = \lfloor 7.1 \rfloor + \lfloor 8.2 \rfloor$; and

3) $\lfloor 7.7 + 8.4 \rfloor = \lfloor 16.1 \rfloor = 16 \neq 15 = 7 + 8 = \lfloor 7.7 \rfloor + \lfloor 8.4 \rfloor$.

b) A second function — one related to the floor function in part (a) — is the *ceiling function*. This function $g\colon \mathbf{R} \to \mathbf{Z}$ is defined by

$$g(x) = \lceil x \rceil = \text{the least integer greater than or equal to } x.$$

So $g(x) = x$ when $x \in \mathbf{Z}$, but when $x \in \mathbf{R} - \mathbf{Z}$, then $g(x)$ is the integer to the immediate right of x on the real number line. In dealing with the ceiling function one finds that

1) $\lceil 3 \rceil = 3$, $\lceil 3.01 \rceil = \lceil 3.7 \rceil = 4 = \lceil 4 \rceil$, $\lceil -3 \rceil = -3$, $\lceil -3.01 \rceil = \lceil -3.7 \rceil = -3$;

2) $\lceil 3.6 + 4.5 \rceil = \lceil 8.1 \rceil = 9 = 4 + 5 = \lceil 3.6 \rceil + \lceil 4.5 \rceil$; and

3) $\lceil 3.3 + 4.2 \rceil = \lceil 7.5 \rceil = 8 \neq 9 = 4 + 5 = \lceil 3.3 \rceil + \lceil 4.2 \rceil$.

c) The function trunc (for truncation) is another integer-valued function defined on $\mathbf{R}$. This function deletes the fractional part of a real number. For example, trunc(3.78) = 3, trunc(5) = 5, trunc(−7.22) = −7. Note that trunc(3.78) = $\lfloor 3.78 \rfloor = 3$ while trunc(−3.78) = $\lceil -3.78 \rceil = -3$.

d) In storing a matrix in a one-dimensional array, many computer languages use the *row major* implementation. Here, if $A = (a_{ij})_{m \times n}$ is an $m \times n$ matrix, the first row of A is stored in locations $1, 2, 3, \ldots, n$ of the array if we start with a_{11} in location 1. The entry a_{21} is then found in position $n + 1$, while entry a_{34} occupies position $2n + 4$ in the array. In order to determine the location of an entry a_{ij} from A, where $1 \leq i \leq m$, $1 \leq j \leq n$, one defines the *access function* f from the entries of A to the positions $1, 2, 3, \ldots, mn$ of the array. A formula for the access function here is $f(a_{ij}) = (i - 1)n + j$.

a_{11}	a_{12}	$\cdots$	a_{1n}	a_{21}	a_{22}	$\cdots$	a_{2n}	a_{31}	$\cdots$	a_{ij}	$\cdots$	a_{mn}
1	2	$\cdots$	n	$n+1$	$n+2$	$\cdots$	$2n$	$2n+1$	$\cdots$	$(i-1)n+j$	$\cdots$	$(m-1)n+n\ (=mn)$

EXAMPLE 5.11

We may use the floor and ceiling functions in parts (a) and (b), respectively, of Example 5.10 to restate some of the ideas we examined in Chapter 4.

a) When studying the division algorithm, we learned that for all $a, b \in \mathbf{Z}$, where $b > 0$, it was possible to find unique $q, r \in \mathbf{Z}$ with $a = qb + r$ and $0 \leq r < b$. Now we may add that $q = \lfloor \frac{a}{b} \rfloor$ and $r = a - \lfloor \frac{a}{b} \rfloor b$.

b) In Example 4.44 we found that the positive integer

$$29{,}338{,}848{,}000 = 2^8 3^5 5^3 7^3 11$$

has

$$60 = (5)(3)(2)(2)(1) = \left\lceil \frac{(8+1)}{2} \right\rceil \left\lceil \frac{(5+1)}{2} \right\rceil \left\lceil \frac{(3+1)}{2} \right\rceil \left\lceil \frac{(3+1)}{2} \right\rceil \left\lceil \frac{(1+1)}{2} \right\rceil$$

positive divisors that are perfect squares. In general, if $n \in \mathbf{Z}^+$ with $n > 1$, we know that we can write

$$n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$$

where $k \in \mathbf{Z}^+$, p_i is prime for all $1 \leq i \leq k$, $p_i \neq p_j$ for all $1 \leq i < j \leq k$, and $e_i \in \mathbf{Z}^+$ for all $1 \leq i \leq k$. This is due to the Fundamental Theorem of Arithmetic. Then if $r \in \mathbf{Z}^+$, we find that the number of positive divisors of n that are perfect rth powers is $\prod_{i=1}^{k} \left\lceil \frac{e_i + 1}{r} \right\rceil$. When $r = 1$ we get $\prod_{i=1}^{k} \lceil e_i + 1 \rceil = \prod_{i=1}^{k} (e_i + 1)$, which is the number of positive divisors of n.

EXAMPLE 5.12

In Sections 4.1 and 4.2 we were introduced to the concept of a sequence in conjunction with our study of recursive definitions. We should now realize that a sequence of real numbers $r_1, r_2, r_3, \ldots$ can be thought of as a function $f: \mathbf{Z}^+ \rightarrow \mathbf{R}$ where $f(n) = r_n$, for all $n \in \mathbf{Z}^+$. Likewise, an integer sequence $a_0, a_1, a_2, \ldots$ can be defined by means of a function $g: \mathbf{N} \rightarrow \mathbf{Z}$ where $g(n) = a_n$, for all $n \in \mathbf{N}$.

In Example 5.9 there are $2^{12} = 4096$ relations from A to B. We have examined one function among these relations, and now we wish to count the total number of functions from A to B.

For the general case, let A, B be nonempty sets with $|A| = m$, $|B| = n$. Consequently, if $A = \{a_1, a_2, a_3, \ldots, a_m\}$ and $B = \{b_1, b_2, b_3, \ldots, b_n\}$, then a typical function $f: A \rightarrow B$ can be described by $\{(a_1, x_1), (a_2, x_2), (a_3, x_3), \ldots, (a_m, x_m)\}$. We can select any of the n elements of B for x_1 and then do the same for x_2. (We can select any element of B for x_2 so that the same element of B may be selected for both x_1 and x_2.) We continue this selection process until one of the n elements of B is finally selected for x_m. In this way, using the rule of product, there are $n^m = |B|^{|A|}$ functions from A to B.

Therefore, for A, B in Example 5.9, there are $4^3 = |B|^{|A|} = 64$ functions from A to B, and $3^4 = |A|^{|B|} = 81$ functions from B to A. In general, we do not expect $|A|^{|B|}$ to equal $|B|^{|A|}$. Unlike the situation for relations, we cannot always obtain a function from B to A by simply interchanging the components in the ordered pairs of a function from A to B (or vice versa).

Now that we have the concept of a function as a special type of relation, we turn our attention to a special type of function.

Definition 5.5

A function $f: A \rightarrow B$ is called *one-to-one*, or *injective*, if each element of B appears at most once as the image of an element of A.

If $f: A \rightarrow B$ is one-to-one, with A, B finite, we must have $|A| \leq |B|$. For arbitrary sets A, B, $f: A \rightarrow B$ is one-to-one if and only if for all $a_1, a_2 \in A$, $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$.

EXAMPLE 5.13

Consider the function $f: \mathbf{R} \rightarrow \mathbf{R}$ where $f(x) = 3x + 7$ for all $x \in \mathbf{R}$. Then for all $x_1, x_2 \in \mathbf{R}$, we find that

$$f(x_1) = f(x_2) \Rightarrow 3x_1 + 7 = 3x_2 + 7 \Rightarrow 3x_1 = 3x_2 \Rightarrow x_1 = x_2,$$

so the given function f is one-to-one.

On the other hand, suppose that $g: \mathbf{R} \rightarrow \mathbf{R}$ is the function defined by $g(x) = x^4 - x$ for each real number x. Then

$$g(0) = (0)^4 - 0 = 0 \quad \text{and} \quad g(1) = (1)^4 - (1) = 1 - 1 = 0.$$

Consequently, g is *not* one-to-one, since $g(0) = g(1)$ but $0 \neq 1$ — that is, g is *not* one-to-one because there exist real numbers x_1, x_2 where $g(x_1) = g(x_2) \not\Rightarrow x_1 = x_2$.

EXAMPLE 5.14

Let $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4, 5\}$. The function

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

is a one-to-one function from A to B;

$$g = \{(1, 1), (2, 3), (3, 3)\}$$

is a function from A to B, but it fails to be one-to-one because $g(2) = g(3)$ but $2 \neq 3$.

For A, B in Example 5.14 there are 2^{15} relations from A to B and 5^3 of these are functions from A to B. The next question we want to answer is how many functions $f: A \to B$ are one-to-one. Again we argue for general finite sets.

With $A = \{a_1, a_2, a_3, \ldots, a_m\}$, $B = \{b_1, b_2, b_3, \ldots, b_n\}$, and $m \leq n$, a one-to-one function $f: A \to B$ has the form $\{(a_1, x_1), (a_2, x_2), (a_3, x_3), \ldots, (a_m, x_m)\}$, where there are n choices for x_1 (that is, any element of B), $n - 1$ choices for x_2 (that is, any element of B except the one chosen for x_1), $n - 2$ choices for x_3, and so on, finishing with $n - (m - 1) = n - m + 1$ choices for x_m. By the rule of product, the number of one-to-one functions from A to B is

$$n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!} = P(n, m) = P(|B|, |A|).$$

Consequently, for A, B in Example 5.14, there are $5 \cdot 4 \cdot 3 = 60$ one-to-one functions $f: A \to B$.

Definition 5.6

If $f: A \to B$ and $A_1 \subseteq A$, then

$$f(A_1) = \{b \in B | b = f(a), \text{ for some } a \in A_1\},$$

and $f(A_1)$ is called the *image of* A_1 *under* f.

EXAMPLE 5.15

For $A = \{1, 2, 3, 4, 5\}$ and $B = \{w, x, y, z\}$, let $f: A \to B$ be given by $f = \{(1, w), (2, x), (3, x), (4, y), (5, y)\}$. Then for $A_1 = \{1\}$, $A_2 = \{1, 2\}$, $A_3 = \{1, 2, 3\}$, $A_4 = \{2, 3\}$, and $A_5 = \{2, 3, 4, 5\}$, we find the following corresponding images under f:

$$f(A_1) = \{f(a)|a \in A_1\} = \{f(a)|a \in \{1\}\} = \{f(a)|a = 1\} = \{f(1)\} = \{w\};$$

$$f(A_2) = \{f(a)|a \in A_2\} = \{f(a)|a \in \{1, 2\}\} = \{f(a)|a = 1 \text{ or } 2\}$$
$$= \{f(1), f(2)\} = \{w, x\};$$

$f(A_3) = \{f(1), f(2), f(3)\} = \{w, x\}$, and $f(A_3) = f(A_2)$ because $f(2) = x = f(3)$;

$f(A_4) = \{x\}$; and $f(A_5) = \{x, y\}$.

EXAMPLE 5.16

a) Let $g: \mathbf{R} \to \mathbf{R}$ be given by $g(x) = x^2$. Then $g(\mathbf{R})$ = the range of $g = [0, +\infty)$. The image of $\mathbf{Z}$ under g is $g(\mathbf{Z}) = \{0, 1, 4, 9, 16, \ldots\}$, and for $A_1 = [-2, 1]$ we get $g(A_1) = [0, 4]$.

b) Let $h: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$ where $h(x, y) = 2x + 3y$. The domain of h is $\mathbf{Z} \times \mathbf{Z}$, not $\mathbf{Z}$, and the codomain is $\mathbf{Z}$. We find, for example, that $h(0, 0) = 2(0) + 3(0) = 0$ and $h(-3, 7) = 2(-3) + 3(7) = 15$. In addition, $h(2, -1) = 2(2) + 3(-1) = 1$, and for each $n \in \mathbf{Z}$, $h(2n, -n) = 2(2n) + 3(-n) = 4n - 3n = n$. Consequently, $h(\mathbf{Z} \times \mathbf{Z})$ = the range of $h = \mathbf{Z}$. For $A_1 = \{(0, n)|n \in \mathbf{Z}^+\} = \{0\} \times \mathbf{Z}^+ \subseteq \mathbf{Z} \times \mathbf{Z}$, the image of A_1 under h is $h(A_1) = \{3, 6, 9, \ldots\} = \{3n|n \in \mathbf{Z}^+\}$.

Our next result deals with the interplay between the images of subsets (of the domain) under a function f and the set operations of union and intersection.

THEOREM 5.2

Let $f: A \to B$, with $A_1, A_2 \subseteq A$. Then

a) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$; b) $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$;

c) $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$ when f is one-to-one.

Proof: We prove part (b) and leave the remaining parts for the reader.

For each $b \in B$, $b \in f(A_1 \cap A_2) \Rightarrow b = f(a)$, for some $a \in A_1 \cap A_2 \Rightarrow [b = f(a)$ for some $a \in A_1]$ and $[b = f(a)$ for some $a \in A_2] \Rightarrow b \in f(A_1)$ and $b \in f(A_2) \Rightarrow b \in f(A_1) \cap f(A_2)$, so $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$.

Definition 5.7

If $f: A \to B$ and $A_1 \subseteq A$, then $f|_{A_1}: A_1 \to B$ is called the *restriction of f to A_1* if $f|_{A_1}(a) = f(a)$ for all $a \in A_1$.

Definition 5.8

Let $A_1 \subseteq A$ and $f: A_1 \to B$. If $g: A \to B$ and $g(a) = f(a)$ for all $a \in A_1$, then we call g an *extension of f to A*.

EXAMPLE 5.17

For $A = \{1, 2, 3, 4, 5\}$, let $f: A \to \mathbf{R}$ be defined by $f = \{(1, 10), (2, 13), (3, 16), (4, 19), (5, 22)\}$. Let $g: \mathbf{Q} \to \mathbf{R}$ where $g(q) = 3q + 7$ for all $q \in \mathbf{Q}$. Finally, let $h: \mathbf{R} \to \mathbf{R}$ with $h(r) = 3r + 7$ for all $r \in \mathbf{R}$. Then

i) g is *an* extension of f (from A) to $\mathbf{Q}$;
ii) f is *the* restriction of g (from $\mathbf{Q}$) to A;
iii) h is *an* extension of f (from A) to $\mathbf{R}$;
iv) f is *the* restriction of h (from $\mathbf{R}$) to A;
v) h is *an* extension of g (from $\mathbf{Q}$) to $\mathbf{R}$; and
vi) g is *the* restriction of h (from $\mathbf{R}$) to $\mathbf{Q}$.

EXAMPLE 5.18

Let $A = \{w, x, y, z\}$, $B = \{1, 2, 3, 4, 5\}$, and $A_1 = \{w, y, z\}$. Let $f: A \to B$, $g: A_1 \to B$ be represented by the diagrams in Fig. 5.5. Then $g = f|_{A_1}$ and f is an extension of g from A_1 to A. We note that for the given function $g: A_1 \to B$, there are five ways to extend g from A_1 to A.

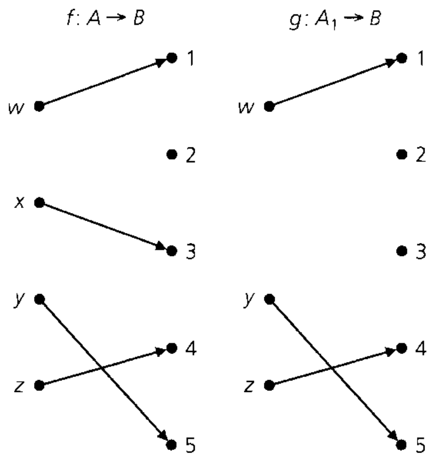


Figure 5.5

EXERCISES 5.2

1. Determine whether or not each of the following relations is a function. If a relation is a function, find its range.

a) $\{(x, y) | x, y \in \mathbf{Z}, y = x^2 + 7\}$, a relation from $\mathbf{Z}$ to $\mathbf{Z}$

b) $\{(x, y) | x, y \in \mathbf{R}, y^2 = x\}$, a relation from $\mathbf{R}$ to $\mathbf{R}$

c) $\{(x, y) | x, y \in \mathbf{R}, y = 3x + 1\}$, a relation from $\mathbf{R}$ to $\mathbf{R}$

d) $\{(x, y) | x, y \in \mathbf{Q}, x^2 + y^2 = 1\}$, a relation from $\mathbf{Q}$ to $\mathbf{Q}$

e) $\mathcal{R}$ is a relation from A to B where $|A| = 5$, $|B| = 6$, and $|\mathcal{R}| = 6$.

2. Does the formula $f(x) = 1/(x^2 - 2)$ define a function $f: \mathbf{R} \to \mathbf{R}$? A function $f: \mathbf{Z} \to \mathbf{R}$?

3. Let $A = \{1, 2, 3, 4\}$ and $B = \{x, y, z\}$. (a) List five functions from A to B. (b) How many functions $f: A \to B$ are there? (c) How many functions $f: A \to B$ are one-to-one? (d) How many functions $g: B \to A$ are there? (e) How many functions $g: B \to A$ are one-to-one? (f) How many functions $f: A \to B$ satisfy $f(1) = x$? (g) How many functions $f: A \to B$ satisfy $f(1) = f(2) = x$? (h) How many functions $f: A \to B$ satisfy $f(1) = x$ and $f(2) = y$?

4. If there are 2187 functions $f: A \to B$ and $|B| = 3$, what is $|A|$?

5. Let $A, B, C \subseteq \mathbf{R}^2$ where $A = \{(x, y) | y = 2x + 1\}$, $B = \{(x, y) | y = 3x\}$, and $C = \{(x, y) | x - y = 7\}$. Determine each of the following:

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

6. Let $A, B, C \subseteq \mathbf{Z}^2$ where $A = \{(x, y) | y = 2x + 1\}$, $B = \{(x, y) | y = 3x\}$, and $C = \{(x, y) | x - y = 7\}$.

a) Determine

i) $A \cap B$ **ii)** $B \cap C$
iii) $\overline{\overline{A} \cup \overline{C}}$ **iv)** $\overline{B} \cup \overline{C}$

b) How are the answers for (i)–(iv) affected if $A, B, C \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+$?

7. Determine each of the following:

a) $\lfloor 2.3 - 1.6 \rfloor$ **b)** $\lfloor 2.3 \rfloor - \lfloor 1.6 \rfloor$ **c)** $\lceil 3.4 \rceil \lfloor 6.2 \rfloor$
d) $\lfloor 3.4 \rfloor \lceil 6.2 \rceil$ **e)** $\lfloor 2\pi \rfloor$ **f)** $2\lceil \pi \rceil$

8. Determine whether each of the following statements is true or false. If the statement is false, provide a counterexample.

a) $\lfloor a \rfloor = \lceil a \rceil$ for all $a \in \mathbf{Z}$.

b) $\lfloor a \rfloor = \lceil a \rceil$ for all $a \in \mathbf{R}$.

c) $\lfloor a \rfloor = \lceil a \rceil - 1$ for all $a \in \mathbf{R} - \mathbf{Z}$.

d) $-\lceil a \rceil = \lceil -a \rceil$ for all $a \in \mathbf{R}$.

9. Find all real numbers x such that

a) $7\lfloor x \rfloor = \lfloor 7x \rfloor$ **b)** $\lfloor 7x \rfloor = 7$
c) $\lfloor x + 7 \rfloor = x + 7$ **d)** $\lfloor x + 7 \rfloor = \lfloor x \rfloor + 7$

10. Determine all $x \in \mathbf{R}$ such that $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = \lfloor 2x \rfloor$.

11. a) Find all real numbers x where $\lceil 3x \rceil = 3\lceil x \rceil$.

b) Let $n \in \mathbf{Z}^+$ where $n > 1$. Determine all $x \in \mathbf{R}$ such that $\lceil nx \rceil = n\lceil x \rceil$.

12. For $n, k \in \mathbf{Z}^+$, prove that $\lceil n/k \rceil = \lfloor (n-1)/k \rfloor + 1$.

13. a) Let $a \in \mathbf{R}^+$ where $a \geq 1$. Prove that (i) $\lfloor \lceil a \rceil / a \rfloor = 1$; and (ii) $\lceil \lfloor a \rfloor / a \rceil = 1$.

b) If $a \in \mathbf{R}^+$ and $0 < a < 1$, which result(s) in part (a) is (are) true?

14. Let $a_1, a_2, a_3, \ldots$ be the integer sequence defined recursively by

1) $a_1 = 1$; and

2) For all $n \in \mathbf{Z}^+$ where $n \geq 2$, $a_n = 2a_{\lfloor n/2 \rfloor}$.

a) Determine a_n for all $2 \leq n \leq 8$.

b) Prove that $a_n \leq n$ for all $n \in \mathbf{Z}^+$.

15. For each of the following functions, determine whether it is one-to-one and determine its range.

a) $f: \mathbf{Z} \to \mathbf{Z}$, $f(x) = 2x + 1$

b) $f: \mathbf{Q} \to \mathbf{Q}$, $f(x) = 2x + 1$

c) $f: \mathbf{Z} \to \mathbf{Z}$, $f(x) = x^3 - x$

d) $f: \mathbf{R} \to \mathbf{R}$, $f(x) = e^x$

e) $f: [-\pi/2, \pi/2] \to \mathbf{R}$, $f(x) = \sin x$

f) $f: [0, \pi] \to \mathbf{R}$, $f(x) = \sin x$

16. Let $f: \mathbf{R} \to \mathbf{R}$ where $f(x) = x^2$. Determine $f(A)$ for the following subsets A taken from the domain $\mathbf{R}$.

a) $A = \{2, 3\}$

b) $A = \{-3, -2, 2, 3\}$

c) $A = (-3, 3)$

d) $A = (-3, 2]$

e) $A = [-7, 2]$

f) $A = (-4, -3] \cup [5, 6]$

17. Let $A = \{1, 2, 3, 4, 5\}$, $B = \{w, x, y, z\}$, $A_1 = \{2, 3, 5\} \subseteq A$, and $g: A_1 \to B$. In how many ways can g be extended to a function $f: A \to B$?

18. Give an example of a function $f: A \to B$ and $A_1, A_2 \subseteq A$ for which $f(A_1 \cap A_2) \neq f(A_1) \cap f(A_2)$. [Thus the inclusion in Theorem 5.2(b) may be proper.]

19. Prove parts (a) and (c) of Theorem 5.2.

20. If $A = \{1, 2, 3, 4, 5\}$ and there are 6720 injective functions $f: A \to B$, what is $|B|$?

21. Let $f: A \to B$, where $A = X \cup Y$ with $X \cap Y = \emptyset$. If $f|_X$ and $f|_Y$ are one-to-one, does it follow that f is one-to-one?

22. For $n \in \mathbf{Z}^+$ define $X_n = \{1, 2, 3, \ldots, n\}$. Given $m, n \in \mathbf{Z}^+$, $f: X_m \to X_n$ is called *monotone increasing* if for all $i, j \in X_m$, $1 \leq i < j \leq m \Rightarrow f(i) \leq f(j)$. (a) How many monotone increasing functions are there with domain X_7 and codomain X_5? (b) Answer part (a) for the domain X_6 and codomain X_9. (c) Generalize the results in parts (a) and (b). (d) Determine the number of monotone increasing functions $f: X_{10} \to X_8$ where $f(4) = 4$. (e) How many monotone increasing functions $f: X_7 \to X_{12}$ satisfy $f(5) = 9$? (f) Generalize the results in parts (d) and (e).

23. Determine the access function $f(a_{ij})$, as described in Example 5.10(d), for a matrix $A = (a_{ij})_{m \times n}$, where (a) $m = 12$, $n = 12$; (b) $m = 7$, $n = 10$; (c) $m = 10$, $n = 7$.

24. For the access function developed in Example 5.10(d), the matrix $A = (a_{ij})_{m \times n}$ was stored in a one-dimensional array using the row major implementation. It is also possible to store this matrix using the column major implementation, where each entry a_{i1}, $1 \leq i \leq m$, in the first column of A is stored in locations $1, 2, 3, \ldots, m$, respectively, of the array, when a_{11} is stored in location 1. Then the entries a_{i2}, $1 \leq i \leq m$, of the second column of A are stored in locations $m + 1, m + 2, m + 3, \ldots, 2m$, respectively, of the array, and so on. Find a formula for the access function $g(a_{ij})$ under these conditions.

25. a) Let A be an $m \times n$ matrix that is to be stored (in a contiguous manner) in a one-dimensional array of r entries. Find a formula for the access function if a_{11} is to be stored in location k (≥ 1) of the array [as opposed to location 1 as in Example 5.10(d)] and we use (i) the row major implementation; (ii) the column major implementation.

b) State any conditions involving m, n, r, and k that must be satisfied in order for the results in part (a) to be valid.

26. The following exercise provides a combinatorial proof for a summation formula we have seen in four earlier results: (1) Exercise 22 in Section 1.4; (2) Example 4.4; (3) Exercise 3 in Section 4.1; and (4) Exercise 19 in Section 4.2.

Let $A = \{a, b, c\}$, $B = \{1, 2, 3, \ldots, n, n + 1\}$, and $S = \{f: A \to B \mid f(a) < f(c) \text{ and } f(b) < f(c)\}$.

a) If $S_1 = \{f: A \to B \mid f \in S \text{ and } f(c) = 2\}$, what is $|S_1|$?

b) If $S_2 = \{f: A \to B \mid f \in S \text{ and } f(c) = 3\}$, what is $|S_2|$?

c) For $1 \leq i \leq n$, let $S_i = \{f: A \to B \mid f \in S \text{ and } f(c) = i + 1\}$. What is $|S_i|$?

d) Let $T_1 = \{f: A \to B \mid f \in S \text{ and } f(a) = f(b)\}$. Explain why $|T_1| = \binom{n+1}{2}$.

e) Let $T_2 = \{f: A \to B \mid f \in S \text{ and } f(a) < f(b)\}$ and $T_3 = \{f: A \to B \mid f \in S \text{ and } f(a) > f(b)\}$. Explain why $|T_2| = |T_3| = \binom{n+1}{3}$.

f) What can we conclude about the sets

$$S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_n \text{ and } T_1 \cup T_2 \cup T_3?$$

g) Use the results from parts (c), (d), (e), and (f) to verify that

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

27. One version of *Ackermann's function A(m,n)* is defined recursively for $m, n \in \mathbf{N}$ by

$$A(0, n) = n + 1, n \geq 0;$$
$$A(m, 0) = A(m - 1, 1), m > 0; \text{ and}$$
$$A(m, n) = A(m - 1, A(m, n - 1)), m, n > 0.$$

[Such functions were defined in the 1920s by the German mathematician and logician Wilhelm Ackermann (1896–1962), who was a student of David Hilbert (1862–1943). These functions play an important role in computer science — in the theory of recursive functions and in the analysis of algorithms that involve the union of sets.]

a) Calculate $A(1, 3)$ and $A(2, 3)$.

b) Prove that $A(1, n) = n + 2$ for all $n \in \mathbf{N}$.

c) For all $n \in \mathbf{N}$ show that $A(2, n) = 3 + 2n$.

d) Verify that $A(3, n) = 2^{n+3} - 3$ for all $n \in \mathbf{N}$.

28. Given sets A, B, we define a *partial function* f with domain A and codomain B as a function from A' to B, where $\emptyset \neq A' \subseteq A$. [Here $f(x)$ is not defined for $x \in A - A'$.] For example, $f: \mathbf{R}^* \to \mathbf{R}$, where $f(x) = 1/x$, is a partial function on $\mathbf{R}$ since $f(0)$ is not defined. On the finite side, $\{(1, x), (2, x), (3, y)\}$ is a partial function for domain $A = \{1, 2, 3, 4, 5\}$ and codomain $B = \{w, x, y, z\}$. Furthermore, a computer program may be thought of as a partial function. The program's input is the input for the partial function and the program's output is the output of the function. Should the program fail to terminate, or terminate abnormally (perhaps, because of an attempt to divide by 0), then the partial function is considered to be undefined for that input. (a) For $A = \{1, 2, 3, 4, 5\}$, $B = \{w, x, y, z\}$, how many partial functions have domain A and codomain B? (b) Let A, B be sets where $|A| = m > 0$, $|B| = n > 0$. How many partial functions have domain A and codomain B?

5.3 Onto Functions: Stirling Numbers of the Second Kind

The results we develop in this section will provide the answers to the first five problems stated at the beginning of this chapter. We find that the *onto* function is the key to all of the answers.

Definition 5.9 A function $f: A \to B$ is called *onto*, or *surjective*, if $f(A) = B$ — that is, if for all $b \in B$ there is at least one $a \in A$ with $f(a) = b$.

EXAMPLE 5.19

The function $f: \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^3$ is an onto function. For here we find that if r is any real number in the codomain of f, then the real number $\sqrt[3]{r}$ is in the domain of f and $f(\sqrt[3]{r}) = (\sqrt[3]{r})^3 = r$. Hence the codomain of f = $\mathbf{R}$ = the range of f, and the function f is onto.

The function $g: \mathbf{R} \to \mathbf{R}$, where $g(x) = x^2$ for each real number x, is *not* an onto function. In this case no negative real number appears in the range of g. For example, for -9 to be in the range of g, we would have to be able to find a *real* number r with $g(r) = r^2 = -9$. Unfortunately, $r^2 = -9 \Rightarrow r = 3i$ or $r = -3i$, where $3i, -3i \in \mathbf{C}$, but $3i, -3i \notin \mathbf{R}$. So here the range of $g = g(\mathbf{R}) = [0, +\infty) \subset \mathbf{R}$, and the function g is *not* onto. Note, however, that the function $h: \mathbf{R} \to [0, +\infty)$ defined by $h(x) = x^2$ *is* an onto function.

EXAMPLE 5.20

Consider the function $f: \mathbf{Z} \to \mathbf{Z}$ where $f(x) = 3x + 1$ for each $x \in \mathbf{Z}$. Here the range of $f = \{\ldots, -8, -5, -2, 1, 4, 7, \ldots\} \subset \mathbf{Z}$, so f is *not* an onto function. If we examine the situation here a little more closely, we find that the integer 8, for example, is not in the range of f even though the equation

$$3x + 1 = 8$$

can be easily solved — giving us $x = 7/3$. But that is the problem, for the rational number $7/3$ is *not* an integer — so there is no x in the domain $\mathbf{Z}$ with $f(x) = 8$.

On the other hand, each of the functions

1) $g: \mathbf{Q} \to \mathbf{Q}$, where $g(x) = 3x + 1$ for $x \in \mathbf{Q}$; and
2) $h: \mathbf{R} \to \mathbf{R}$, where $h(x) = 3x + 1$ for $x \in \mathbf{R}$

is an onto function. Furthermore, $3x_1 + 1 = 3x_2 + 1 \Rightarrow 3x_1 = 3x_2 \Rightarrow x_1 = x_2$, regardless of whether x_1 and x_2 are integers, rational numbers, or real numbers. Consequently, all three of the functions f, g, and h are one-to-one.

EXAMPLE 5.21

If $A = \{1, 2, 3, 4\}$ and $B = \{x, y, z\}$, then

$$f_1 = \{(1, z), (2, y), (3, x), (4, y)\} \quad \text{and} \quad f_2 = \{(1, x), (2, x), (3, y), (4, z)\}$$

are both functions from A *onto* B. However, the function $g = \{(1, x), (2, x), (3, y), (4, y)\}$ is not onto, because $g(A) = \{x, y\} \subset B$.

If A, B are finite sets, then for an onto function $f: A \to B$ to possibly exist we must have $|A| \geq |B|$. Considering the development in the first two sections of this chapter, the reader undoubtedly feels it is time once again to use the rule of product and count the number of onto functions $f: A \to B$ where $|A| = m \geq n = |B|$. Unfortunately, the rule of product proves inadequate here. We shall obtain the needed result for some specific examples and then conjecture a general formula. In Chapter 8 we shall establish the conjecture using the Principle of Inclusion and Exclusion.

EXAMPLE 5.22

If $A = \{x, y, z\}$ and $B = \{1, 2\}$, then all functions $f: A \to B$ are onto except $f_1 = \{(x, 1), (y, 1), (z, 1)\}$, and $f_2 = \{(x, 2), (y, 2), (z, 2)\}$, the *constant* functions. So there are $|B|^{|A|} - 2 = 2^3 - 2 = 6$ onto functions from A to B.

In general, if $|A| = m \geq 2$ and $|B| = 2$, then there are $2^m - 2$ onto functions from A to B. (Does this formula tell us anything when $m = 1$?)

EXAMPLE 5.23

For $A = \{w, x, y, z\}$ and $B = \{1, 2, 3\}$, there are 3^4 functions from A to B. Considering subsets of B of size 2, there are 2^4 functions from A to $\{1, 2\}$, 2^4 functions from A to $\{2, 3\}$, and 2^4 functions from A to $\{1, 3\}$. So we have $3(2^4) = \binom{3}{2}2^4$ functions from A to B that are definitely not onto. However, before we acknowledge $3^4 - \binom{3}{2}2^4$ as the final answer, we must realize that not all of these $\binom{3}{2}2^4$ functions are distinct. For when we consider all the functions from A to $\{1, 2\}$, we are removing, among these, the function $\{(w, 2), (x, 2), (y, 2), (z, 2)\}$. Then, considering the functions from A to $\{2, 3\}$, we remove the same function: $\{(w, 2), (x, 2), (y, 2), (z, 2)\}$. Consequently, in the result $3^4 - \binom{3}{2}2^4$, we have twice removed each of the constant functions $f: A \to B$, where $f(A)$ is one of the sets $\{1\}$, $\{2\}$, or $\{3\}$. Adjusting our present result for this, we find that there are $3^4 - \binom{3}{2}2^4 + 3 = \binom{3}{3}3^4 - \binom{3}{2}2^4 + \binom{3}{1}1^4 = 36$ onto functions from A to B.

Keeping $B = \{1, 2, 3\}$, for any set A with $|A| = m \geq 3$, there are $\binom{3}{3}3^m - \binom{3}{2}2^m + \binom{3}{1}1^m$ functions from A onto B. (What result does this formula yield when $m = 1$? when $m = 2$?)

The last two examples suggest a pattern that we now state, without proof, as our general formula.

For finite sets A, B with $|A| = m$ and $|B| = n$, there are

$$\binom{n}{n}n^m - \binom{n}{n-1}(n-1)^m + \binom{n}{n-2}(n-2)^m - \cdots$$
$$+(-1)^{n-2}\binom{n}{2}2^m + (-1)^{n-1}\binom{n}{1}1^m = \sum_{k=0}^{n-1}(-1)^k\binom{n}{n-k}(n-k)^m$$
$$= \sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m$$

onto functions from A to B.

EXAMPLE 5.24

Let $A = \{1, 2, 3, 4, 5, 6, 7\}$ and $B = \{w, x, y, z\}$. Applying the general formula with $m = 7$ and $n = 4$, we find that there are

$$\binom{4}{4}4^7 - \binom{4}{3}3^7 + \binom{4}{2}2^7 - \binom{4}{1}1^7 = \sum_{k=0}^{3}(-1)^k\binom{4}{4-k}(4-k)^7$$
$$= \sum_{k=0}^{4}(-1)^k\binom{4}{4-k}(4-k)^7 = 8400 \text{ functions from } A \text{ onto } B.$$

The result in Example 5.24 is also the answer to the first three questions proposed at the start of this chapter. Once we remove the unnecessary vocabulary, we recognize that in all three cases we want to distribute seven different objects into four distinct containers with no container left empty. We can do this in terms of onto functions.

For Problem 4 we have a sample space $\mathscr{S}$ consisting of the $4^7 = 16{,}384$ ways in which seven people can each select one of the four floors. (Note that 4^7 is also the total number of functions $f: A \to B$ where $|A| = 7$, $|B| = 4$.) The event that we are concerned with contains 8400 of those selections, so the probability that the elevator must stop at every floor is $8400/16384 \doteq 0.5127$, slightly more than half of the time.

Finally, for Problem 5, since $\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m$ is the number of onto functions $f: A \to B$ for $|A| = m$, $|B| = n$, for the case where $m < n$ there are no such functions and the summation is 0.

Problem 6 will be addressed in Section 5.6.

Before going on to anything new, however, we consider one more problem.

EXAMPLE 5.25

At the CH Company, Joan, the supervisor, has a secretary, Teresa, and three other administrative assistants. If seven accounts must be processed, in how many ways can Joan assign the accounts so that each assistant works on at least one account and Teresa's work includes the most expensive account?

First and foremost, the answer is not 8400 as in Example 5.24. Here we must consider two disjoint subcases and then apply the rule of sum.

a) If Teresa, the secretary, works only on the most expensive account, then the other six accounts can be distributed among the three administrative assistants in $\sum_{k=0}^{3}(-1)^k\binom{3}{3-k}(3-k)^6 = 540$ ways. (540 = the number of onto functions $f: A \to B$ with $|A| = 6$, $|B| = 3$.)

b) If Teresa does more than just the most expensive account, the assignments can be made in $\sum_{k=0}^{4}(-1)^k\binom{4}{4-k}(4-k)^6 = 1560$ ways. (1560 = the number of onto functions $g: C \to D$ with $|C| = 6$, $|D| = 4$.)

Consequently, the assignments can be given under the prescribed conditions in $540 + 1560 = 2100$ ways. [We mentioned earlier that the answer would not be 8400, but it is $(1/4)(8400) = (1/|B|)(8400)$, where 8400 is the number of onto functions $f: A \to B$, with $|A| = 7$ and $|B| = 4$. This is no coincidence, as we shall learn when we discuss Theorem 5.3.]

We now continue our discussion with the distribution of distinct objects into containers with none left empty, but now the containers become identical.

EXAMPLE 5.26

If $A = \{a, b, c, d\}$ and $B = \{1, 2, 3\}$, then there are 36 onto functions from A to B or, equivalently, 36 ways to distribute four distinct objects into three distinguishable containers, with no container empty (and no regard for the location of objects in a given container). Among these 36 distributions we find the following collection of six (one of six such possible collections of six):

1) $\{a, b\}_1 \quad \{c\}_2 \quad \{d\}_3$

2) $\{a, b\}_1 \quad \{d\}_2 \quad \{c\}_3$

3) $\{c\}_1 \quad \{a, b\}_2 \quad \{d\}_3$

4) $\{c\}_1 \quad \{d\}_2 \quad \{a, b\}_3$

5) $\{d\}_1 \quad \{a, b\}_2 \quad \{c\}_3$

6) $\{d\}_1 \quad \{c\}_2 \quad \{a, b\}_3,$

where, for example, the notation $\{c\}_2$ means that c is in the second container. Now if we no longer distinguish the containers, these $6 = 3!$ distributions become identical, so there are $36/(3!) = 6$ ways to distribute the distinct objects a, b, c, d among three identical containers, leaving no container empty.

For $m \geq n$ there are $\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m$ ways to distribute m distinct objects into n numbered (but otherwise identical) containers with no container left empty. Removing the numbers on the containers, so that they are now identical in appearance, we find that one distribution into these n (nonempty) identical containers corresponds with $n!$ such distributions into the numbered containers. So the number of ways in which it is possible to distribute the m distinct objects into n identical containers, with no container left empty, is

$$\frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m.$$

This will be denoted by $S(m, n)$ and is called a *Stirling number of the second kind.*

We note that for $|A| = m \geq n = |B|$, there are $n! \cdot S(m, n)$ onto functions from A to B.

Table 5.1 lists some Stirling numbers of the second kind.

EXAMPLE 5.27

For $m \geq n$, $\sum_{i=1}^{n} S(m, i)$ is the number of possible ways to distribute m distinct objects into n identical containers with empty containers allowed. From the fourth row of Table 5.1

Table 5.1

	$S(m, n)$							
$m \backslash n$	1	2	3	4	5	6	7	8
1	1							
2	1	1						
3	1	3	1					
4	1	7	6	1				
5	1	15	25	10	1			
6	1	31	90	65	15	1		
7	1	63	301	350	140	21	1	
8	1	127	966	1701	1050	266	28	1

we see that there are $1 + 7 + 6 = 14$ ways to distribute the objects a, b, c, d among three identical containers, with some container(s) possibly empty.

We continue now with the derivation of an identity involving Stirling numbers of the second kind. The proof is combinatorial in nature.

THEOREM 5.3 Let m, n be positive integers with $1 < n \leq m$. Then

$$S(m + 1, n) = S(m, n - 1) + nS(m, n).$$

Proof: Let $A = \{a_1, a_2, \ldots, a_m, a_{m+1}\}$. Then $S(m + 1, n)$ counts the number of ways in which the objects of A can be distributed among n identical containers, with no container left empty.

There are $S(m, n - 1)$ ways of distributing $a_1, a_2, \ldots, a_m$ among $n - 1$ identical containers, with none left empty. Then, placing a_{m+1} in the remaining empty container results in $S(m, n - 1)$ of the distributions counted in $S(m + 1, n)$—namely, those distributions where a_{m+1} is in a container by itself. Alternatively, distributing $a_1, a_2, \ldots, a_m$ among the n identical containers with none left empty, we have $S(m, n)$ distributions. Now, however, for each of these $S(m, n)$ distributions the n containers become distinguished by their contents. Selecting one of the n distinct containers for a_{m+1}, we have $nS(m, n)$ distributions of the total $S(m + 1, n)$—namely, those where a_{m+1} is in the same container as another object from A. The result then follows by the rule of sum.

To illustrate Theorem 5.3 consider the triangle shown in Table 5.1. Here the largest number corresponds with $S(m + 1, n)$, for $m = 7$ and $n = 3$, and we see that $S(7 + 1, 3) = 966 = 63 + 3(301) = S(7, 2) + 3S(7, 3)$. The identity in Theorem 5.3 can be used to extend Table 5.1 if necessary.

If we multiply the result in Theorem 5.3 by $(n - 1)!$ we have

$$\left(\frac{1}{n}\right) [n!S(m + 1, n)] = [(n - 1)!S(m, n - 1)] + [n!S(m, n)] .$$

This new form of the equation tells us something about numbers of onto functions. If $A = \{a_1, a_2, \ldots, a_m, a_{m+1}\}$ and $B = \{b_1, b_2, \ldots, b_{n-1}, b_n\}$ with $m \geq n - 1$, then

$$\left(\frac{1}{n}\right)\text{(The number of onto functions } h: A \to B)$$
$$= \text{(The number of onto functions } f: A - \{a_{m+1}\} \to B - \{b_n\})$$
$$+ \text{(The number of onto functions } g: A - \{a_{m+1}\} \to B).$$

Thus the relationship at the end of Example 5.25 is not just a coincidence.

We close this section with an application that deals with a counting problem in which the Stirling numbers of the second kind are used in conjunction with the Fundamental Theorem of Arithmetic.

EXAMPLE 5.28

Consider the positive integer $30{,}030 = 2 \times 3 \times 5 \times 7 \times 11 \times 13$. Among the unordered factorizations of this number one finds

i) $30 \times 1001 = (2 \times 3 \times 5)(7 \times 11 \times 13)$
ii) $110 \times 273 = (2 \times 5 \times 11)(3 \times 7 \times 13)$
iii) $2310 \times 13 = (2 \times 3 \times 5 \times 7 \times 11)(13)$
iv) $14 \times 33 \times 65 = (2 \times 7)(3 \times 11)(5 \times 13)$
v) $22 \times 35 \times 39 = (2 \times 11)(5 \times 7)(3 \times 13)$

The results given in (i), (ii), and (iii) demonstrate three of the ways to distribute the six distinct objects 2, 3, 5, 7, 11, 13 into two identical containers with no container left empty. So these first three examples are three of the $S(6, 2) = 31$ unordered two-factor factorizations of 30,030—that is, there are $S(6, 2)$ ways to factor 30,030 as mn where $m, n \in \mathbf{Z}^+$ for $1 < m, n < 30{,}030$ and where order is not relevant. Likewise, the results in (iv) and (v) are two of the $S(6, 3) = 90$ unordered ways to factor 30,030 into three integer factors, each greater than 1. If we want at least two factors (greater than 1) in each of these unordered factorizations, then we find that there are $\sum_{i=2}^{6} S(6, i) = 202$ such factorizations. If we want to include the *one-factor* factorization 30,030—where we distribute the six distinct objects 2, 3, 5, 7, 11, 13 into one (identical) container—then we have 203 such factorizations in total.

EXERCISES 5.3

1. Give an example of finite sets A and B with $|A|, |B| \geq 4$ and a function $f: A \to B$ such that (a) f is neither one-to-one nor onto; (b) f is one-to-one but not onto; (c) f is onto but not one-to-one; (d) f is onto and one-to-one.

2. For each of the following functions $f: \mathbf{Z} \to \mathbf{Z}$, determine whether the function is one-to-one and whether it is onto. If the function is not onto, determine the range $f(\mathbf{Z})$.

a) $f(x) = x + 7$ **b)** $f(x) = 2x - 3$
c) $f(x) = -x + 5$ **d)** $f(x) = x^2$
e) $f(x) = x^2 + x$ **f)** $f(x) = x^3$

3. For each of the following functions $g: \mathbf{R} \to \mathbf{R}$, determine whether the function is one-to-one and whether it is onto. If the function is not onto, determine the range $g(\mathbf{R})$.

a) $g(x) = x + 7$ **b)** $g(x) = 2x - 3$
c) $g(x) = -x + 5$ **d)** $g(x) = x^2$
e) $g(x) = x^2 + x$ **f)** $g(x) = x^3$

4. Let $A = \{1, 2, 3, 4\}$ and $B = \{1, 2, 3, 4, 5, 6\}$. (a) How many functions are there from A to B? How many of these are one-to-one? How many are onto? (b) How many functions are there from B to A? How many of these are onto? How many are one-to-one?

5. Verify that $\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^m = 0$ for $n = 5$ and $m = 2, 3, 4$.

6. **a)** Verify that $5^7 = \sum_{i=1}^{5} \binom{5}{i}(i!)S(7, i)$.

b) Provide a combinatorial argument to prove that for all $m, n \in \mathbf{Z}^+$,

$$m^n = \sum_{i=1}^{m} \binom{m}{i}(i!)S(n, i).$$

7. **a)** Let $A = \{1, 2, 3, 4, 5, 6, 7\}$ and $B = \{v, w, x, y, z\}$. Determine the number of functions $f: A \to B$ where (i) $f(A) = \{v, x\}$; (ii) $|f(A)| = 2$; (iii) $f(A) = \{w, x, y\}$; (iv) $|f(A)| = 3$; (v) $f(A) = \{v, x, y, z\}$; and (vi) $|f(A)| = 4$.

b) Let A, B be sets with $|A| = m \geq n = |B|$. If $k \in \mathbf{Z}^+$ with $1 \leq k \leq n$, how many functions $f: A \to B$ are such that $|f(A)| = k$?

8. A chemist who has five assistants is engaged in a research project that calls for nine compounds that must be synthesized. In how many ways can the chemist assign these syntheses to the five assistants so that each is working on at least one synthesis?

9. Use the fact that every polynomial equation having real-number coefficients and odd degree has a real root in order to show that the function $f: \mathbf{R} \to \mathbf{R}$, defined by $f(x) = x^5 - 2x^2 + x$, is an onto function. Is f one-to-one?

10. Suppose we have seven different colored balls and four containers numbered I, II, III, and IV. (a) In how many ways can we distribute the balls so that no container is left empty? (b) In this collection of seven colored balls, one of them is blue. In how many ways can we distribute the balls so that no container is empty and the blue ball is in container II? (c) If we remove the numbers from the containers so that we can no longer distinguish them, in how many ways can we distribute the seven colored balls among the four identical containers, with some container(s) possibly empty?

11. Determine the next two rows ($m = 9, 10$) of Table 5.1 for the Stirling numbers $S(m, n)$, where $1 \leq n \leq m$.

12. **a)** In how many ways can 31,100,905 be factored into three factors, each greater than 1, if the order of the factors is not relevant?

b) Answer part (a), assuming the order of the three factors is relevant.

c) In how many ways can one factor 31,100,905 into two or more factors where each factor is greater than 1 and no regard is paid to the order of the factors?

d) Answer part (c), assuming the order of the factors is to be taken into consideration.

13. **a)** How many two-factor unordered factorizations, where each factor is greater than 1, are there for 156,009?

b) In how many ways can 156,009 be factored into two or more factors, each greater than 1, with no regard to the order of the factors?

c) Let $p_1, p_2, p_3, \ldots, p_n$ be n distinct primes. In how many ways can one factor the product $\prod_{i=1}^{n} p_i$ into two or more factors, each greater than 1, where the order of the factors is not relevant?

14. Write a computer program (or develop an algorithm) to compute the Stirling numbers $S(m, n)$ when $1 \leq m \leq 12$ and $1 \leq n \leq m$.

15. A lock has n buttons labeled $1, 2, \ldots, n$. To open this lock we press each of the n buttons exactly once. If no two or more buttons may be pressed simultaneously, then there are $n!$ ways to do this. However, if one may press two or more buttons simultaneously, then there are more than $n!$ ways to press all of the buttons. For instance, if $n = 3$ there are six ways to press the buttons one at a time. But if one may also press two or more buttons simultaneously, then we find 13 cases — namely,

(1) 1, 2, 3	(2) 1, 3, 2	(3) 2, 1, 3
(4) 2, 3, 1	(5) 3, 1, 2	(6) 3, 2, 1
(7) {1, 2}, 3	(8) 3, {1, 2}	(9) {1, 3}, 2
(10) 2, {1, 3}	(11) {2, 3}, 1	(12) 1, {2, 3}
(13) {1, 2, 3}.		

[Here, for example, case (12) indicates that one presses button 1 first and then buttons 2, 3 (together) second.] (a) How many ways are there to press the buttons when $n = 4$? $n = 5$? How many for n in general? (b) Suppose a lock has 15 buttons. To open this lock one must press 12 different buttons (one at a time, or simultaneously in sets of two or more). In how many ways can this be done?

16. At St. Xavier High School ten candidates $C_1, C_2, \ldots, C_{10}$, run for senior class president.

a) How many outcomes are possible where (i) there are no ties (that is, no two, or more, candidates receive the same number of votes? (ii) ties are permitted? [Here we may have an outcome such as $\{C_2, C_3, C_7\}, \{C_1, C_4, C_9, C_{10}\}, \{C_5\}, \{C_6, C_8\}$, where C_2, C_3, C_7 tie for first place, C_1, C_4, C_9, C_{10} tie for fourth place, C_5 is in eighth place, and C_6, C_8 are tied for ninth place.] (iii) three candidates tie for first place (and other ties are permitted)?

b) How many of the outcomes in section (iii) of part (a) have C_3 as one of the first-place candidates?

c) How many outcomes have C_3 in first place (alone, or tied with others)?

17. For $m, n, r \in \mathbf{Z}^+$ with $m \geq rn$, let $S_r(m, n)$ denote the number of ways to distribute m distinct objects among n identical containers where each container receives at least r of the objects. Verify that

$$S_r(m+1, n) = nS_r(m, n) + \binom{m}{r-1}S_r(m+1-r, n-1).$$

18. We use $s(m, n)$ to denote the number of ways to seat m people at n circular tables with at least one person at each table. The arrangements at any one table are not distinguished if one can be rotated into another (as in Example 1.16). The ordering of the tables is *not* taken into account. For instance, the arrangements in parts (a), (b), (c) of Fig. 5.6 are considered the same; those in parts (a), (d), (e) are distinct (in pairs).

The numbers $s(m, n)$ are referred to as the *Stirling numbers of the first kind.*

a) If $n > m$, what is $s(m, n)$?

b) For $m \geq 1$, what are $s(m, m)$ and $s(m, 1)$?

c) Determine $s(m, m-1)$ for $m \geq 2$.

d) Show that for $m \geq 3$,

$$s(m, m-2) = \left(\frac{1}{24}\right) m(m-1)(m-2)(3m-1).$$

19. As in the previous exercise, $s(m, n)$ denotes a Stirling number of the first kind.

a) For $m \geq n > 1$ prove that

$$s(m, n) = (m-1)s(m-1, n) + s(m-1, n-1).$$

b) Verify that for $m \geq 2$,

$$s(m, 2) = (m-1)! \sum_{i=1}^{m-1} \frac{1}{i}.$$

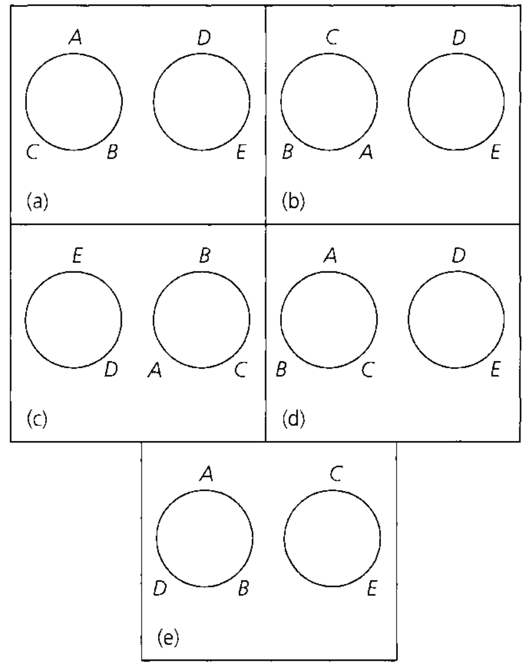


Figure 5.6

5.4 Special Functions

In Section 2 of Chapter 3 we mentioned that addition is a closed binary operation on the set $\mathbf{Z}^+$, whereas $\cap$ is a closed binary operation on $\mathcal{P}(\mathcal{U})$ for any given universe $\mathcal{U}$. We also noted in that section that "taking the minus" of an integer is a unary operation on $\mathbf{Z}$. Now it is time to make these notions of (closed) binary and unary operations more precise in terms of functions.

Definition 5.10 For any nonempty sets A, B, any function $f\colon A \times A \to B$ is called a *binary operation* on A. If $B \subseteq A$, then the binary operation is said to be *closed* (*on* A). (When $B \subseteq A$ we may also say that *A is closed under f*.)

Definition 5.11 A function $g\colon A \to A$ is called a *unary*, or *monary*, operation on A.

EXAMPLE 5.29

a) The function $f\colon \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$, defined by $f(a, b) = a - b$, is a closed binary operation on $\mathbf{Z}$.

b) If $g\colon \mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}$ is the function where $g(a, b) = a - b$, then g is a binary operation on $\mathbf{Z}^+$, but it is *not* closed. For example, we find that $3, 7 \in \mathbf{Z}^+$, but $g(3, 7) = 3 - 7 = -4 \notin \mathbf{Z}^+$.

c) The function $h\colon \mathbf{R}^+ \to \mathbf{R}^+$ defined by $h(a) = 1/a$ is a unary operation on $\mathbf{R}^+$.

EXAMPLE 5.30

Let $\mathcal{U}$ be a universe, and let $A, B \subseteq \mathcal{U}$. (a) If $f: \mathcal{P}(\mathcal{U}) \times \mathcal{P}(\mathcal{U}) \to \mathcal{P}(\mathcal{U})$ is defined by $f(A, B) = A \cup B$, then f is a closed binary operation on $\mathcal{P}(\mathcal{U})$. (b) The function $g: \mathcal{P}(\mathcal{U}) \to \mathcal{P}(\mathcal{U})$ defined by $g(A) = \overline{A}$ is a unary operation on $\mathcal{P}(\mathcal{U})$.

Definition 5.12

Let $f: A \times A \to B$; that is, f is a binary operation on A.

a) f is said to be *commutative* if $f(a, b) = f(b, a)$ for all $(a, b) \in A \times A$.

b) When $B \subseteq A$ (that is, when f is closed), f is said to be *associative* if for all $a, b, c \in A$, $f(f(a, b), c) = f(a, f(b, c))$.

EXAMPLE 5.31

The binary operation of Example 5.30 is commutative and associative, whereas the binary operation in part (a) of Example 5.29 is neither.

EXAMPLE 5.32

a) Define the closed binary operation $f: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$ by $f(a, b) = a + b - 3ab$. Since both the addition and the multiplication of integers are commutative binary operations, it follows that

$$f(a, b) = a + b - 3ab = b + a - 3ba = f(b, a),$$

so f is commutative.

To determine whether f is associative, consider $a, b, c \in \mathbf{Z}$. Then

$$f(a, b) = a + b - 3ab \quad \text{and} \quad \begin{aligned} f(f(a, b), c) &= f(a, b) + c - 3f(a, b)c \\ &= (a + b - 3ab) + c - 3(a + b - 3ab)c \\ &= a + b + c - 3ab - 3ac - 3bc + 9abc, \end{aligned}$$

whereas

$$f(b, c) = b + c - 3bc \quad \text{and} \quad \begin{aligned} f(a, f(b, c)) &= a + f(b, c) - 3af(b, c) \\ &= a + (b + c - 3bc) - 3a(b + c - 3bc) \\ &= a + b + c - 3ab - 3ac - 3bc + 9abc. \end{aligned}$$

Since $f(f(a, b), c) = f(a, f(b, c))$ for all $a, b, c \in \mathbf{Z}$, the closed binary operation f is associative as well as commutative.

b) Consider the closed binary operation $h: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$, where $h(a, b) = a|b|$. Then $h(3, -2) = 3|-2| = 3(2) = 6$, but $h(-2, 3) = -2|3| = -6$. Consequently, h is *not* commutative. However, with regard to the associative property, if $a, b, c \in \mathbf{Z}$, we find that

$$h(h(a, b), c) = h(a, b)|c| = a|b||c| \quad \text{and}$$
$$h(a, h(b, c)) = a|h(b, c)| = a|b|c|| = a|b||c|,$$

so the closed binary operation h is associative.

EXAMPLE 5.33

If $A = \{a, b, c, d\}$, then $|A \times A| = 16$. Consequently, there are 4^{16} functions $f: A \times A \to A$; that is, there are 4^{16} closed binary operations on A.

To determine the number of commutative closed binary operations g on A, we realize that there are four choices for each of the assignments $g(a, a)$, $g(b, b)$, $g(c, c)$, and $g(d, d)$.

We are then left with the $4^2 - 4 = 16 - 4 = 12$ other ordered pairs (in $A \times A$) of the form (x, y), $x \neq y$. These 12 ordered pairs must be considered in sets of two in order to insure commutativity. For example, we need $g(a, b) = g(b, a)$ and may select any one of the four elements of A for $g(a, b)$. But then this choice must also be assigned to $g(b, a)$. Therefore, since there are four choices for each of these $12/2 = 6$ sets of two ordered pairs, we find that the number of commutative closed binary operations g on A is $4^4 \cdot 4^6 = 4^{10}$.

Definition 5.13 Let $f: A \times A \to B$ be a binary operation on A. An element $x \in A$ is called an *identity* (or *identity element*) for f if $f(a, x) = f(x, a) = a$, for all $a \in A$.

EXAMPLE 5.34

a) Consider the (closed) binary operation $f: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$, where $f(a, b) = a + b$. Here the integer 0 is an identity since $f(a, 0) = a + 0 = 0 + a = f(0, a) = a$, for each integer a.

b) We find that there is no identity for the function in part (a) of Example 5.29. For if f had an identity x, then for *any* $a \in \mathbf{Z}$, $f(a, x) = a \Rightarrow a - x = a \Rightarrow x = 0$. But then $f(x, a) = f(0, a) = 0 - a \neq a$, unless $a = 0$.

c) Let $A = \{1, 2, 3, 4, 5, 6, 7\}$, and let $g: A \times A \to A$ be the (closed) binary operation defined by $g(a, b) = \min\{a, b\}$—that is, the minimum (or smaller) of a, b. This binary operation is commutative and associative, and for any $a \in A$ we have $g(a, 7) = \min\{a, 7\} = a = \min\{7, a\} = g(7, a)$. So 7 is an identity element for g.

In parts (a) and (c) of Example 5.34 we examined two (closed) binary operations, each of which has an identity. Part (b) of that example showed that such an operation need not have an identity element. Could a binary operation have more than one identity? We find that the answer is no when we consider the following theorem.

THEOREM 5.4 Let $f: A \times A \to B$ be a binary operation. If f has an identity, then that identity is unique.

Proof: If f has more than one identity, let $x_1, x_2 \in A$ with

$$f(a, x_1) = a = f(x_1, a), \quad \text{for all } a \in A, \qquad \text{and}$$
$$f(a, x_2) = a = f(x_2, a), \quad \text{for all } a \in A.$$

Consider x_1 as an element of A and x_2 as an identity. Then $f(x_1, x_2) = x_1$. Now reverse the roles of x_1 and x_2—that is, consider x_2 as an element of A and x_1 as an identity. We find that $f(x_1, x_2) = x_2$. Consequently, $x_1 = x_2$, and f has at most one identity.

Now that we have settled the issue of the uniqueness of the identity element, let us see how this type of element enters into one more enumeration problem.

EXAMPLE 5.35

If $A = \{x, a, b, c, d\}$, how many closed binary operations on A have x as the identity?

Let $f: A \times A \to A$ with $f(x, y) = y = f(y, x)$ for all $y \in A$. Then we may represent f by a table as in Table 5.2. Here the nine values, where x is the first component—as in (x, c), or the second component—as in (d, x), are determined by the fact that x is the identity element. Each of the 16 remaining (vacant) entries in Table 5.2 can be filled with any one of the five elements in A.

Table 5.2

f	x	a	b	c	d
x	x	a	b	c	d
a	a	–	–	–	–
b	b	–	–	–	–
c	c	–	–	–	–
d	d	–	–	–	–

Hence there are 5^{16} closed binary operations on A where x is the identity. Of these $5^{10} = 5^4 \cdot 5^{(4^2-4)/2}$ are commutative. We also realize that there are 5^{16} closed binary operations on A where b is the identity. So there are $5^{17} = \binom{5}{1}5^{16} = \binom{5}{1}5^{5^2-[2(5)-1]} = \binom{5}{1}5^{(5-1)^2}$ closed binary operations on A that have an identity, and of these $5^{11} = \binom{5}{1}5^{10} = \binom{5}{1}5^4 5^{(4^2-4)/2}$ are commutative.

Having seen several examples of functions (in Examples 5.16(b), 5.29, 5.30, 5.32, 5.33, 5.34, and 5.35) where the domain is a cross product of sets, we now investigate functions where the domain is a subset of a cross product.

Definition 5.14 For sets A and B, if $D \subseteq A \times B$, then $\pi_A: D \to A$, defined by $\pi_A(a, b) = a$, is called the *projection* on the first coordinate. The function $\pi_B: D \to B$, defined by $\pi_B(a, b) = b$, is called the *projection* on the second coordinate.

We note that if $D = A \times B$ then π_A and π_B are both onto.

EXAMPLE 5.36

If $A = \{w, x, y\}$ and $B = \{1, 2, 3, 4\}$, let $D = \{(x, 1), (x, 2), (x, 3), (y, 1), (y, 4)\}$. Then the projection $\pi_A: D \to A$ satisfies $\pi_A(x, 1) = \pi_A(x, 2) = \pi_A(x, 3) = x$, and $\pi_A(y, 1) = \pi_A(y, 4) = y$. Since $\pi_A(D) = \{x, y\} \subset A$, this function is *not* onto.

For $\pi_B: D \to B$ we find that $\pi_B(x, 1) = \pi_B(y, 1) = 1$, $\pi_B(x, 2) = 2$, $\pi_B(x, 3) = 3$, and $\pi_B(y, 4) = 4$, so $\pi_B(D) = B$ and this projection is an onto function.

EXAMPLE 5.37

Let $A = B = \mathbf{R}$ and consider the set $D \subseteq A \times B$ where $D = \{(x, y) | y = x^2\}$. Then D represents the subset of the Euclidean plane that contains the points on the parabola $y = x^2$.

Among the infinite number of points in D we find the point (3, 9). Here $\pi_A(3, 9) = 3$, the x-coordinate of (3, 9), whereas $\pi_B(3, 9) = 9$, the y-coordinate of the point.

For this example, $\pi_A(D) = \mathbf{R} = A$, so π_A is onto. (The projection π_A is also one-to-one.) However, $\pi_B(D) = [0, +\infty) \subset \mathbf{R}$, so π_B is *not* onto. [Nor is it one-to-one — for example, $\pi_B(2, 4) = 4 = \pi_B(-2, 4)$.]

We now extend the notion of projection as follows. Let $A_1, A_2, \ldots, A_n$ be sets, and $\{i_1, i_2, \ldots, i_m\} \subseteq \{1, 2, \ldots, n\}$ with $i_1 < i_2 < \cdots < i_m$ and $m \leq n$. If $D \subseteq A_1 \times A_2 \times \cdots \times A_n = \times_{i=1}^{n} A_i$, then the function $\pi: D \to A_{i_1} \times A_{i_2} \times \cdots \times A_{i_m}$ defined by $\pi(a_1, a_2, \ldots, a_n) = (a_{i_1}, a_{i_2}, \ldots, a_{i_m})$ is the projection of D on the i_1th, i_2th, $\ldots$, i_mth coordinates. The elements of D are called (ordered) *n-tuples*; an element in $\pi(D)$ is an (ordered) *m-tuple*.

These projections arise in a natural way in the study of *relational data bases*, a standard technique for organizing and describing large quantities of data by modern large-scale computing systems. In situations like credit card transactions, not only must existing data be organized but new data must be inserted, as when credit cards are processed for new cardholders. When bills on existing accounts are paid, or when new purchases are made on these accounts, data must be updated. Another example arises when records are searched for special considerations, as when a college admissions office searches educational records seeking, for its mailing lists, high school students who have demonstrated certain levels of mathematical achievement.

The following example demonstrates the use of projections in a method for organizing and describing data on a somewhat smaller scale.

EXAMPLE 5.38

At a certain university the following sets are related for purposes of registration:

A_1 = the set of course numbers for courses offered in mathematics.

A_2 = the set of course titles offered in mathematics.

A_3 = the set of mathematics faculty.

A_4 = the set of letters of the alphabet.

Consider the *table*, or relation,[†] $D \subseteq A_1 \times A_2 \times A_3 \times A_4$ given in Table 5.3.

Table 5.3

Course Number	Course Title	Professor	Section Letter
MA 111	Calculus I	P. Z. Chinn	A
MA 111	Calculus I	V. Larney	B
MA 112	Calculus II	J. Kinney	A
MA 112	Calculus II	A. Schmidt	B
MA 112	Calculus II	R. Mines	C
MA 113	Calculus III	J. Kinney	A

The sets A_1, A_2, A_3, A_4 are called the *domains of the relational data base*, and *table D* is said to have *degree* 4. Each element of D is often called a *list*.

The projection of D on $A_1 \times A_3 \times A_4$ is shown in Table 5.4. Table 5.5 shows the results for the projection of D on $A_1 \times A_2$.

Table 5.4

Course Number	Professor	Section Letter
MA 111	P. Z. Chinn	A
MA 111	V. Larney	B
MA 112	J. Kinney	A
MA 112	A. Schmidt	B
MA 112	R. Mines	C
MA 113	J. Kinney	A

Table 5.5

Course Number	Course Title
MA 111	Calculus I
MA 112	Calculus II
MA 113	Calculus III

[†]Here the relation D is *not* binary. In fact, D is a *quaternary* relation.

Tables 5.4 and 5.5 are another way of representing the same data that appear in Table 5.3. Given Tables 5.4 and 5.5, one can recapture Table 5.3.

The theory of relational data bases is concerned with representing data in different ways and with the operations, such as projections, needed for such representations. The computer implementation of such techniques is also considered. More on this topic is mentioned in the exercises and chapter references.

EXERCISES 5.4

1. For $A = \{a, b, c\}$, let $f: A \times A \to A$ be the closed binary operation given in Table 5.6. Give an example to show that f is *not* associative.

Table 5.6

f	a	b	c
a	b	a	c
b	a	c	b
c	c	b	a

2. Let $f: \mathbf{R} \times \mathbf{R} \to \mathbf{Z}$ be the closed binary operation defined by $f(a, b) = \lceil a + b \rceil$. (a) Is f commutative? (b) Is f associative? (c) Does f have an identity element?

3. Each of the following functions $f: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z}$ is a closed binary operation on $\mathbf{Z}$. Determine in each case whether f is commutative and/or associative.

   a) $f(x, y) = x + y - xy$

   b) $f(x, y) = \max\{x, y\}$, the maximum (or larger) of x, y

   c) $f(x, y) = x^y$

   d) $f(x, y) = x + y - 3$

4. Which of the closed binary operations in Exercise 3 have an identity?

5. Let $|A| = 5$. (a) What is $|A \times A|$? (b) How many functions $f: A \times A \to A$ are there? (c) How many closed binary operations are there on A? (d) How many of these closed binary operations are commutative?

6. Let $A = \{x, a, b, c, d\}$.

   a) How many closed binary operations f on A satisfy $f(a, b) = c$?

   b) How many of the functions f in part (a) have x as an identity?

   c) How many of the functions f in part (a) have an identity?

   d) How many of the functions f in part (c) are commutative?

7. Let $f: \mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ be the closed binary operation defined by $f(a, b) = \gcd(a, b)$. (a) Is f commutative? (b) Is f associative? (c) Does f have an identity element?

8. Let $A = \{2, 4, 8, 16, 32\}$, and consider the closed binary operation $f: A \times A \to A$ where $f(a, b) = \gcd(a, b)$. Does f have an identity element?

9. For distinct primes p, q let $A = \{p^m q^n | 0 \le m \le 31, 0 \le n \le 37\}$. (a) What is $|A|$? (b) If $f: A \times A \to A$ is the closed binary operation defined by $f(a, b) = \gcd(a, b)$, does f have an identity element?

10. State a result that generalizes the ideas presented in the previous two exercises.

11. For $\emptyset \neq A \subseteq \mathbf{Z}^+$, let $f, g: A \times A \to A$ be the closed binary operations defined by $f(a, b) = \min\{a, b\}$ and $g(a, b) = \max\{a, b\}$. Does f have an identity element? Does g?

12. Let $A = B = \mathbf{R}$. Determine $\pi_A(D)$ and $\pi_B(D)$ for each of the following sets $D \subseteq A \times B$.

    a) $D = \{(x, y) | x = y^2\}$

    b) $D = \{(x, y) | y = \sin x\}$

    c) $D = \{(x, y) | x^2 + y^2 = 1\}$

13. Let A_i, $1 \le i \le 5$, be the domains for a table $D \subseteq A_1 \times A_2 \times A_3 \times A_4 \times A_5$, where $A_1 = \{$U, V, W, X, Y, Z$\}$ (used as code names for different cereals in a test), and $A_2 = A_3 = A_4 = A_5 = \mathbf{Z}^+$. The table D is given as Table 5.7.

    a) What is the degree of the table?

    b) Find the projection of D on $A_3 \times A_4 \times A_5$.

    c) A domain of a table is called a *primary key* for the table if its value uniquely identifies each list of D. Determine the primary key(s) for this table.

14. Let A_i, $1 \le i \le 5$, be the domains for a table $D \subseteq A_1 \times A_2 \times A_3 \times A_4 \times A_5$, where $A_1 = \{1, 2\}$ (used to identify the daily vitamin capsule produced by two pharmaceutical companies), $A_2 = \{$A, D, E$\}$, and $A_3 = A_4 = A_5 = \mathbf{Z}^+$. The table D is given as Table 5.8.

    a) What is the degree of the table?

    b) What is the projection of D on $A_1 \times A_2$? on $A_3 \times A_4 \times A_5$?

    c) This table has no primary key. (See Exercise 13.) We can, however, define a *composite primary key* as the cross product of a *minimal* number of domains of the table, whose components, taken collectively, uniquely identify each list of D. Determine some composite primary keys for this table.

Table 5.7

Code Name of Cereal	Grams of Sugar per 1-oz Serving	% of RDA[a] of Vitamin A per 1-oz Serving	% of RDA of Vitamin C per 1-oz Serving	% of RDA of Protein per 1-oz Serving
U	1	25	25	6
V	7	25	2	4
W	12	25	2	4
X	0	60	40	20
Y	3	25	40	10
Z	2	25	40	10

[a]RDA = recommended daily allowance

Table 5.8

Vitamin Capsule	Vitamin Present in Capsule	Amount of Vitamin in Capsule in IU[a]	Dosage: Capsules / Day	No. of Capsules per Bottle
1	A	10,000	1	100
1	D	400	1	100
1	E	30	1	100
2	A	4,000	1	250
2	D	400	1	250
2	E	15	1	250

[a]IU = international units

5.5 The Pigeonhole Principle

A change of pace is in order as we introduce an interesting distribution principle. This principle may seem to have nothing in common with what we have been doing so far, but it will prove to be helpful nonetheless.

In mathematics one sometimes finds that an almost obvious idea, when applied in a rather subtle manner, is the key needed to solve a troublesome problem. On the list of such obvious ideas many would undoubtedly place the following rule, known as the *pigeonhole principle*.

The Pigeonhole Principle: **If m pigeons occupy n pigeonholes and $m > n$, then at least one pigeonhole has two or more pigeons roosting in it.**

One situation for 6 ($= m$) pigeons and 4 ($= n$) pigeonholes (actually birdhouses) is shown in Fig. 5.7. The general result readily follows by the method of proof by contradiction. If the result is not true, then each pigeonhole has at most one pigeon roosting in it—for a total of at most n ($< m$) pigeons. (Somewhere we have lost at least $m - n$ pigeons!)

But now what can pigeons roosting in pigeonholes have to do with mathematics—discrete, combinatorial, or otherwise? Actually, this principle can be applied in various problems in which we seek to establish whether a certain situation can actually occur. We

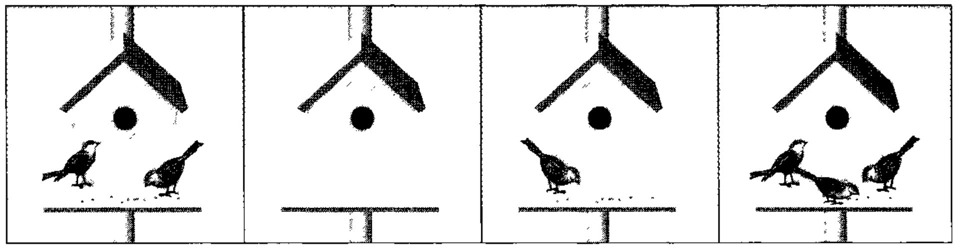


Figure 5.7

illustrate this principle in the following examples and shall find it useful in Section 5.6 and at other points in the text.

EXAMPLE 5.39

An office employs 13 file clerks, so at least two of them must have birthdays during the same month. Here we have 13 pigeons (the file clerks) and 12 pigeonholes (the months of the year).

Here is a second rather immediate application of our principle.

EXAMPLE 5.40

Larry returns from the laundromat with 12 pairs of socks (each pair a different color) in a laundry bag. Drawing the socks from the bag randomly, he'll have to draw at most 13 of them to get a matched pair.

From this point on, application of the pigeonhole principle may be more subtle.

EXAMPLE 5.41

Wilma operates a computer with a magnetic tape drive. One day she is given a tape that contains 500,000 "words" of four or fewer lowercase letters. (Consecutive words on the tape are separated by a blank character.) Can it be that the 500,000 words are all distinct?

From the rules of sum and product, the total number of different possible words, using four or fewer letters, is

$$26^4 + 26^3 + 26^2 + 26 = 475{,}254.$$

With these 475,254 words as the pigeonholes, and the 500,000 words on the tape as the pigeons, it follows that at least one word is repeated on the tape.

EXAMPLE 5.42

Let $S \subset \mathbf{Z}^+$, where $|S| = 37$. Then S contains two elements that have the same remainder upon division by 36.

Here the pigeons are the 37 positive integers in S. We know from the division algorithm (of Theorem 4.5) that when any positive integer n is divided by 36, there exists a unique quotient q and unique remainder r, where

$$n = 36q + r, \qquad 0 \leq r < 36.$$

The 36 possible values of r constitute the pigeonholes, and the result is now established by the pigeonhole principle.

EXAMPLE 5.43

Prove that if 101 integers are selected from the set $S = \{1, 2, 3, \ldots, 200\}$, then there are two integers such that one divides the other.

For each $x \in S$, we may write $x = 2^k y$, with $k \geq 0$, and $\gcd(2, y) = 1$. (This result follows from the Fundamental Theorem of Arithmetic.) Then y must be odd, so $y \in T = \{1, 3, 5, \ldots, 199\}$, where $|T| = 100$. Since 101 integers are selected from S, by the pigeonhole principle there are two distinct integers of the form $a = 2^m y$, $b = 2^n y$ for some (the same) $y \in T$. If $m < n$, then $a|b$; otherwise, we have $m > n$ and then $b|a$.

EXAMPLE 5.44

Any subset of size 6 from the set $S = \{1, 2, 3, \ldots, 9\}$ must contain two elements whose sum is 10.

Here the pigeons constitute a six-element subset of $\{1, 2, 3, \ldots, 9\}$, and the pigeonholes are the subsets $\{1, 9\}$, $\{2, 8\}$, $\{3, 7\}$, $\{4, 6\}$, $\{5\}$. When the six pigeons go to their respective pigeonholes, they must fill at least one of the two-element subsets whose members sum to 10.

EXAMPLE 5.45

Triangle ACE is equilateral with $AC = 1$. If five points are selected from the interior of the triangle, there are at least two whose distance apart is less than $1/2$.

For the triangle in Fig. 5.8, the four smaller triangles are congruent equilateral triangles and $AB = 1/2$. We break up the interior of triangle ACE into the following four regions, which are mutually disjoint in pairs:

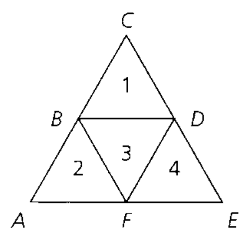


Figure 5.8

R_1: the interior of triangle BCD together with the points on the segment BD, excluding B and D.

R_2: the interior of triangle ABF.

R_3: the interior of triangle BDF together with the points on the segments BF and DF, excluding B, D, and F.

R_4: the interior of triangle FDE.

Now we apply the pigeonhole principle. Five points in the interior of triangle ACE must be such that at least two of them are in one of the four regions R_i, $1 \leq i \leq 4$, where any two points are separated by a distance less than $1/2$.

EXAMPLE 5.46

Let S be a set of six positive integers whose maximum is at most 14. Show that the sums of the elements in all the nonempty subsets of S cannot all be distinct.

For each nonempty subset A of S, the sum of the elements in A, denoted s_A, satisfies $1 \leq s_A \leq 9 + 10 + \cdots + 14 = 69$, and there are $2^6 - 1 = 63$ nonempty subsets of S. We should like to draw the conclusion from the pigeonhole principle by letting the possible sums, from 1 to 69, be the pigeonholes, with the 63 nonempty subsets of S as the pigeons, but then we have too few pigeons.

So instead of considering all nonempty subsets of S, we cut back to those nonempty subsets A of S where $|A| \leq 5$. Then for each such subset A it follows that $1 \leq s_A \leq 10 + 11 + \cdots + 14 = 60$. There are 62 nonempty subsets A of S with $|A| \leq 5$ —namely, all the subsets of S except for $\emptyset$ and the set S itself. With 62 pigeons (the nonempty subsets A of S where $|A| \leq 5$) and 60 pigeonholes (the possible sums s_A), it follows by the pigeonhole principle that the elements of at least two of these 62 subsets must yield the same sum.

EXAMPLE 5.47

Let $m \in \mathbf{Z}^+$ with m odd. Prove that there exists a positive integer n such that m divides $2^n - 1$.

Consider the $m + 1$ positive integers $2^1 - 1, 2^2 - 1, 2^3 - 1, \ldots, 2^m - 1, 2^{m+1} - 1$. By the pigeonhole principle and the division algorithm there exist $s, t \in \mathbf{Z}^+$ with $1 \leq s < t \leq m + 1$, where $2^s - 1$ and $2^t - 1$ have the same remainder upon division by m. Hence $2^s - 1 = q_1 m + r$ and $2^t - 1 = q_2 m + r$, for $q_1, q_2 \in \mathbf{N}$, and $(2^t - 1) - (2^s - 1) = (q_2 m + r) - (q_1 m + r)$, so $2^t - 2^s = (q_2 - q_1)m$. But $2^t - 2^s = 2^s(2^{t-s} - 1)$; and since m is odd, we have $\gcd(2^s, m) = 1$. Hence $m | (2^{t-s} - 1)$, and the result follows with $n = t - s$.

EXAMPLE 5.48

While on a four-week vacation, Herbert will play at least one set of tennis each day, but he won't play more than 40 sets total during this time. Prove that no matter how he distributes his sets during the four weeks, there is a span of consecutive days during which he will play exactly 15 sets.

For $1 \leq i \leq 28$, let x_i be the total number of sets Herbert will play from the start of the vacation to the end of the ith day. Then $1 \leq x_1 < x_2 < \cdots < x_{28} \leq 40$, and $x_1 + 15 < \cdots < x_{28} + 15 \leq 55$. We now have the 28 distinct numbers $x_1, x_2, \ldots, x_{28}$ and the 28 distinct numbers $x_1 + 15, x_2 + 15, \ldots, x_{28} + 15$. These 56 numbers can take on only 55 different values, so at least two of them must be equal, and we conclude that there exist $1 \leq j < i \leq 28$ with $x_i = x_j + 15$. Hence, from the start of day $j + 1$ to the end of day i, Herbert will play exactly 15 sets of tennis.

Our last example for this section deals with a classic result that was first discovered in 1935 by Paul Erdös and George Szekeres.

EXAMPLE 5.49

Let us start by considering two particular examples:

1) Note how the sequence 6, 5, 8, 3, 7 (of length 5) contains the decreasing subsequence 6, 5, 3 (of length 3).
2) Now note how the sequence 11, 8, 7, 1, 9, 6, 5, 10, 3, 12 (of length 10) contains the increasing subsequence 8, 9, 10, 12 (of length 4).

These two instances demonstrate the general result: For each $n \in \mathbf{Z}^+$, a sequence of $n^2 + 1$ distinct real numbers contains a decreasing or increasing subsequence of length $n + 1$.

To verify this claim let $a_1, a_2, \ldots, a_{n^2+1}$ be a sequence of $n^2 + 1$ distinct real numbers. For $1 \leq k \leq n^2 + 1$, let

x_k = the maximum length of a decreasing subsequence that ends with a_k, and

y_k = the maximum length of an increasing subsequence that ends with a_k.

For instance, our second particular example would provide

k	1	2	3	4	5	6	7	8	9	10
a_k	11	8	7	1	9	6	5	10	3	12
x_k	1	2	3	4	2	4	5	2	6	1
y_k	1	1	1	1	2	2	2	3	2	4

If, in general, there is no decreasing or increasing subsequence of length $n + 1$, then $1 \leq x_k \leq n$ and $1 \leq y_k \leq n$ for all $1 \leq k \leq n^2 + 1$. Consequently, there are at most n^2 *distinct* ordered pairs (x_k, y_k). But we have $n^2 + 1$ ordered pairs (x_k, y_k), since $1 \leq k \leq n^2 + 1$. So the pigeonhole principle implies that there are two identical ordered pairs (x_i, y_i), (x_j, y_j), where $i \neq j$ — say $i < j$. Now the real numbers $a_1, a_2, \ldots, a_{n^2+1}$ are distinct, so if $a_i < a_j$ then $y_i < y_j$, while if $a_j < a_i$ then $x_j > x_i$. In either case we no longer have $(x_i, y_i) = (x_j, y_j)$. This contradiction tells us that $x_k = n + 1$ or $y_k = n + 1$ for some $n + 1 \leq k \leq n^2 + 1$; the result then follows.

For an interesting application of this result, consider $n^2 + 1$ sumo wrestlers facing forward and standing shoulder to shoulder. (Here no two wrestlers have the same weight.) We can select $n + 1$ of these wrestlers to take one step forward so that, as they are scanned from left to right, their successive weights either decrease or increase.

EXERCISES 5.5

1. In Example 5.40, what plays the roles of the pigeons and of the pigeonholes?

2. Show that if eight people are in a room, at least two of them have birthdays that occur on the same day of the week.

3. An auditorium has a seating capacity of 800. How many seats must be occupied to guarantee that at least two people seated in the auditorium have the same first and last initials?

4. Let $S = \{3, 7, 11, 15, 19, \ldots, 95, 99, 103\}$. How many elements must we select from S to insure that there will be at least two whose sum is 110?

5. a) Prove that if 151 integers are selected from $\{1, 2, 3, \ldots, 300\}$, then the selection must include two integers x, y where $x|y$ or $y|x$.

b) Write a statement that generalizes the results of part (a) and Example 5.43.

6. Prove that if we select 101 integers from the set $S = \{1, 2, 3, \ldots, 200\}$, there exist m, n in the selection where $\gcd(m, n) = 1$.

7. a) Show that if any 14 integers are selected from the set $S = \{1, 2, 3, \ldots, 25\}$, there are at least two whose sum is 26.

b) Write a statement that generalizes the results of part (a) and Example 5.44.

8. a) If $S \subseteq \mathbf{Z}^+$ and $|S| \geq 3$, prove that there exist distinct $x, y \in S$ where $x + y$ is even.

b) Let $S \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+$. Find the minimal value of $|S|$ that guarantees the existence of distinct ordered pairs $(x_1, x_2), (y_1, y_2) \in S$ such that $x_1 + y_1$ and $x_2 + y_2$ are both even.

c) Extending the ideas in parts (a) and (b), consider $S \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+ \times \mathbf{Z}^+$. What size must $|S|$ be to guarantee the existence of distinct ordered triples $(x_1, x_2, x_3), (y_1, y_2, y_3) \in S$ where $x_1 + y_1$, $x_2 + y_2$, and $x_3 + y_3$ are all even?

d) Generalize the results of parts (a), (b), and (c).

e) A point $P(x, y)$ in the Cartesian plane is called a *lattice point* if $x, y \in \mathbf{Z}$. Given distinct lattice points $P_1(x_1, y_1), P_2(x_2, y_2), \ldots, P_n(x_n, y_n)$, determine the smallest value of n that guarantees the existence of $P_i(x_i, y_i)$, $P_j(x_j, y_j)$, $1 \leq i < j \leq n$, such that the midpoint of the line segment connecting $P_i(x_i, y_i)$ and $P_j(x_j, y_j)$ is also a lattice point.

9. a) If 11 integers are selected from $\{1, 2, 3, \ldots, 100\}$, prove that there are at least two, say x and y, such that $0 < |\sqrt{x} - \sqrt{y}| < 1$.

b) Write a statement that generalizes the result of part (a).

10. Let triangle ABC be equilateral, with $AB = 1$. Show that if we select 10 points in the interior of this triangle, there must be at least two whose distance apart is less than $1/3$.

11. Let $ABCD$ be a square with $AB = 1$. Show that if we select five points in the interior of this square, there are at least two whose distance apart is less than $1/\sqrt{2}$.

12. Let $A \subseteq \{1, 2, 3, \ldots, 25\}$ where $|A| = 9$. For any subset B of A let s_B denote the sum of the elements in B. Prove that there are distinct subsets C, D of A such that $|C| = |D| = 5$ and $s_C = s_D$.

13. Let S be a set of five positive integers the maximum of which is at most 9. Prove that the sums of the elements in all the nonempty subsets of S cannot all be distinct.

14. During the first six weeks of his senior year in college, Brace sends out at least one resumé each day but no more than 60 resumés in total. Show that there is a period of consecutive days during which he sends out exactly 23 resumés.

15. Let $S \subset \mathbf{Z}^+$ with $|S| = 7$. For $\emptyset \neq A \subseteq S$, let s_A denote the sum of the elements in A. If m is the maximum element in S, find the possible values of m so that there will exist distinct subsets B, C of S with $s_B = s_C$.

16. Let $k \in \mathbf{Z}^+$. Prove that there exists a positive integer n such that $k|n$ and the only digits in n are 0's and 3's.

17. **a)** Find a sequence of four distinct real numbers with no decreasing or increasing subsequence of length 3.

b) Find a sequence of nine distinct real numbers with no decreasing or increasing subsequence of length 4.

c) Generalize the results in parts (a) and (b).

d) What do the preceding parts of this exercise tell us about Example 5.49?

18. The 50 members of Nardine's aerobics class line up to get their equipment. Assuming that no two of these people have the same height, show that eight of them (as the line is equipped from first to last) have successive heights that either decrease or increase.

19. For $k, n \in \mathbf{Z}^+$, prove that if $kn + 1$ pigeons occupy n pigeonholes, then at least one pigeonhole has $k + 1$ or more pigeons roosting in it.

20. How many times must we roll a single die in order to get the same score (a) at least twice? (b) at least three times? (c) at least n times, for $n \geq 4$?

21. **a)** Let $S \subset \mathbf{Z}^+$. What is the smallest value for $|S|$ that guarantees the existence of two elements x, $y \in S$ where x and y have the same remainder upon division by 1000?

b) What is the smallest value of n such that whenever $S \subseteq \mathbf{Z}^+$ and $|S| = n$, then there exist three elements x, y, $z \in S$ where all three have the same remainder upon division by 1000?

c) Write a statement that generalizes the results of parts (a) and (b) and Example 5.42.

22. For $m, n \in \mathbf{Z}^+$, prove that if m pigeons occupy n pigeonholes, then at least one pigeonhole has $\lfloor (m-1)/n \rfloor + 1$ or more pigeons roosting in it.

23. Let $p_1, p_2, \ldots, p_n \in \mathbf{Z}^+$. Prove that if $p_1 + p_2 + \cdots + p_n - n + 1$ pigeons occupy n pigeonholes, then either the first pigeonhole has p_1 or more pigeons roosting in it, or the second pigeonhole has p_2 or more pigeons roosting in it, ..., or the nth pigeonhole has p_n or more pigeons roosting in it.

24. Given 8 Perl books, 17 Visual BASIC† books, 6 Java books, 12 SQL books, and 20 C++ books, how many of these books must we select to insure that we have 10 books dealing with the same computer language?

5.6 Function Composition and Inverse Functions

When computing with the elements of $\mathbf{Z}$, we find that the (closed binary) operation of addition provides a method for combining two integers, say a and b, into a third integer, namely $a + b$. Furthermore, for each integer c there is a second integer d where $c + d = d + c = 0$, and we call d the additive *inverse* of c. (It is also true that c is the additive *inverse* of d.)

Turning to the elements of $\mathbf{R}$ and the (closed binary) operation of multiplication, we have a method for combining any $r, s \in \mathbf{R}$ into their product rs. And here, for each $t \in \mathbf{R}$, if $t \neq 0$, then there is a real number u such that $ut = tu = 1$. The real number u is called the multiplicative *inverse* of t. (The real number t is also the multiplicative *inverse* of u.)

In this section we first study a method for combining two functions into a single function. Then we develop the concept of the inverse (of a function) for functions with certain properties. To accomplish these objectives, we need the following preliminary ideas.

†Visual BASIC is a trademark of the Microsoft Corporation.

Having examined functions that are one-to-one and those that are onto, we turn now to functions with both of these properties.

Definition 5.15 If $f: A \to B$, then f is said to be *bijective*, or to be a *one-to-one correspondence*, if f is both one-to-one and onto.

EXAMPLE 5.50 If $A = \{1, 2, 3, 4\}$ and $B = \{w, x, y, z\}$, then $f = \{(1, w), (2, x), (3, y), (4, z)\}$ is a one-to-one correspondence from A (on)to B, and $g = \{(w, 1), (x, 2), (y, 3), (z, 4)\}$ is a one-to-one correspondence from B (on)to A.

It should be pointed out that whenever the term *correspondence* was used in Chapter 1 and in Examples 3.11 and 4.12, the adjective *one-to-one* was implied though never stated.

For any nonempty set A there is always a very simple but important one-to-one correspondence, as seen in the following definition.

Definition 5.16 The function $1_A: A \to A$, defined by $1_A(a) = a$ for all $a \in A$, is called the *identity function* for A.

Definition 5.17 If $f, g: A \to B$, we say that f and g are *equal* and write $f = g$, if $f(a) = g(a)$ for all $a \in A$.

A common pitfall in dealing with the equality of functions occurs when f and g are functions with a common domain A and $f(a) = g(a)$ for all $a \in A$. It may *not* be the case that $f = g$. The pitfall results from not paying attention to the codomains of the functions.

EXAMPLE 5.51 Let $f: \mathbf{Z} \to \mathbf{Z}$, $g: \mathbf{Z} \to \mathbf{Q}$ where $f(x) = x = g(x)$, for all $x \in \mathbf{Z}$. Then f, g share the common domain $\mathbf{Z}$, have the same range $\mathbf{Z}$, and act the same on every element of $\mathbf{Z}$. Yet $f \neq g$! Here f is a one-to-one correspondence, whereas g is one-to-one but not onto; so the codomains do make a difference.

EXAMPLE 5.52 Consider the functions $f, g: \mathbf{R} \to \mathbf{Z}$ defined as follows:

$$f(x) = \begin{cases} x, & \text{if } x \in \mathbf{Z} \\ \lfloor x \rfloor + 1, & \text{if } x \in \mathbf{R} - \mathbf{Z} \end{cases} \qquad g(x) = \lceil x \rceil, \text{ for all } x \in \mathbf{R}$$

If $x \in \mathbf{Z}$, then $f(x) = x = \lceil x \rceil = g(x)$.

For $x \in \mathbf{R} - \mathbf{Z}$, write $x = n + r$ where $n \in \mathbf{Z}$ and $0 < r < 1$. (For example, if $x = 2.3$, we write $2.3 = 2 + 0.3$, with $n = 2$ and $r = 0.3$; for $x = -7.3$ we have $-7.3 = -8 + 0.7$, with $n = -8$ and $r = 0.7$.) Then

$$f(x) = \lfloor x \rfloor + 1 = n + 1 = \lceil x \rceil = g(x).$$

Consequently, even though the functions f, g are defined by *different* formulas, we realize that they are the *same* function — because they have the same domain and codomain and $f(x) = g(x)$ for all x in the domain $\mathbf{R}$.

Now that we have dispensed with the necessary preliminaries, it is time to examine an operation for combining two appropriate functions.

Definition 5.18 If $f: A \to B$ and $g: B \to C$, we define the *composite function*, which is denoted $g \circ f: A \to C$, by $(g \circ f)(a) = g(f(a))$, for each $a \in A$.

EXAMPLE 5.53

Let $A = \{1, 2, 3, 4\}$, $B = \{a, b, c\}$, and $C = \{w, x, y, z\}$ with $f: A \to B$ and $g: B \to C$ given by $f = \{(1, a), (2, a), (3, b), (4, c)\}$ and $g = \{(a, x), (b, y), (c, z)\}$. For each element of A we find:

$$(g \circ f)(1) = g(f(1)) = g(a) = x \qquad (g \circ f)(3) = g(f(3)) = g(b) = y$$
$$(g \circ f)(2) = g(f(2)) = g(a) = x \qquad (g \circ f)(4) = g(f(4)) = g(c) = z$$

So

$$g \circ f = \{(1, x), (2, x), (3, y), (4, z)\}.$$

Note: The composition $f \circ g$ is *not* defined.

EXAMPLE 5.54

Let $f: \mathbf{R} \to \mathbf{R}$, $g: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$, $g(x) = x + 5$. Then

$$(g \circ f)(x) = g(f(x)) = g(x^2) = x^2 + 5,$$

whereas

$$(f \circ g)(x) = f(g(x)) = f(x + 5) = (x + 5)^2 = x^2 + 10x + 25.$$

Here $g \circ f: \mathbf{R} \to \mathbf{R}$ and $f \circ g: \mathbf{R} \to \mathbf{R}$, but $(g \circ f)(1) = 6 \neq 36 = (f \circ g)(1)$, so even though both composites $f \circ g$ and $g \circ f$ can be formed, we do not have $f \circ g = g \circ f$. Consequently, the composition of functions is not, in general, a commutative operation.

The definition and examples for composite functions required that the codomain of f = domain of g. If range of $f \subseteq$ domain of g, this will actually be enough to yield the composite function $g \circ f: A \to C$. Also, for any $f: A \to B$, we observe that $f \circ 1_A = f = 1_B \circ f$.

An important recurring idea in mathematics is the investigation of whether combining two entities with a common property yields a result with this property. For example, if A and B are finite sets, then $A \cap B$ and $A \cup B$ are also finite. However, for infinite sets A and B, we have $A \cup B$ infinite but $A \cap B$ could be finite.

For the composition of functions we have the following result.

THEOREM 5.5 Let $f: A \to B$ and $g: B \to C$.

a) If f and g are one-to-one, then $g \circ f$ is one-to-one.

b) If f and g are onto, then $g \circ f$ is onto.

Proof:

a) To prove that $g \circ f: A \to C$ is one-to-one, let $a_1, a_2 \in A$ with $(g \circ f)(a_1) = (g \circ f)(a_2)$. Then $(g \circ f)(a_1) = (g \circ f)(a_2) \Rightarrow g(f(a_1)) = g(f(a_2)) \Rightarrow f(a_1) = f(a_2)$, because g is one-to-one. Also, $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$, because f is one-to-one. Consequently, $g \circ f$ is one-to-one.

b) For $g \circ f: A \to C$, let $z \in C$. Since g is onto, there exists $y \in B$ with $g(y) = z$. With f onto and $y \in B$, there exists $x \in A$ with $f(x) = y$. Hence $z = g(y) = g(f(x)) = (g \circ f)(x)$, so the range of $g \circ f = C =$ the codomain of $g \circ f$, and $g \circ f$ is onto.

Although function composition is not commutative, if $f: A \to B$, $g: B \to C$, and $h: C \to D$, what can we say about the functions $(h \circ g) \circ f$ and $h \circ (g \circ f)$? Specifically, is $(h \circ g) \circ f = h \circ (g \circ f)$? That is, is function composition associative?

Before considering the general result, let us first investigate a particular example.

EXAMPLE 5.55

Let $f, g, h: \mathbf{R} \to \mathbf{R}$, where $f(x) = x^2$, $g(x) = x + 5$, and $h(x) = \sqrt{x^2 + 2}$.

Then $((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = (h \circ g)(x^2) = h(g(x^2)) = h(x^2 + 5) = \sqrt{(x^2 + 5)^2 + 2} = \sqrt{x^4 + 10x^2 + 27}$.

On the other hand, we see that $(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = h(g(x^2)) = h(x^2 + 5) = \sqrt{(x^2 + 5)^2 + 2} = \sqrt{x^4 + 10x^2 + 27}$, as above.

So in this particular example, $(h \circ g) \circ f$ and $h \circ (g \circ f)$ are two functions with the same domain and codomain, and for all $x \in \mathbf{R}$, $((h \circ g) \circ f)(x) = \sqrt{x^4 + 10x^2 + 27} = (h \circ (g \circ f))(x)$. Consequently, $(h \circ g) \circ f = h \circ (g \circ f)$.

We now find that the result in Example 5.55 is true in general.

THEOREM 5.6

If $f: A \to B$, $g: B \to C$, and $h: C \to D$, then $(h \circ g) \circ f = h \circ (g \circ f)$.

Proof: Since the two functions have the same domain, A, and codomain, D, the result will follow by showing that for every $x \in A$, $((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$. (See the diagram shown in Fig. 5.9.)

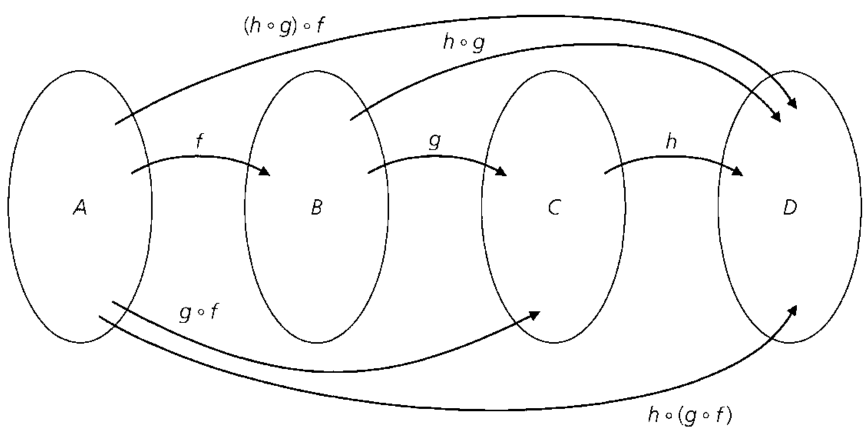


Figure 5.9

Using the definition of the composite function we know that for each $x \in A$ it takes two steps to determine $(g \circ f)(x)$. First we find $f(x)$, the image of x under f. This is an element of B. Then we apply the function g to the element $f(x)$ to determine $g(f(x))$, the image of $f(x)$ under g. This results in an element of C. At this point we apply the function h to the element $g(f(x))$ to determine $h(g(f(x))) = h((g \circ f)(x)) = (h \circ (g \circ f))(x)$. This result is an element of D. Similarly, starting once again with x in A, we have $f(x)$ in B, and now we apply the composite function $h \circ g$ to $f(x)$. This gives us $((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)))$.

Since $((h \circ g) \circ f)(x) = h(g(f(x))) = (h \circ (g \circ f))(x)$, for each x in A, it now follows that

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

Consequently, the composition of functions is an associative operation.

By virtue of the associative property for function composition, we can write $h \circ g \circ f$, $(h \circ g) \circ f$ or $h \circ (g \circ f)$ without any problem of ambiguity. In addition, this property enables us to define powers of functions, where appropriate.

Definition 5.19 If $f: A \to A$ we define $f^1 = f$, and for $n \in \mathbf{Z}^+$, $f^{n+1} = f \circ (f^n)$.

This definition is another example wherein the result is defined *recursively*. With $f^{n+1} = f \circ (f^n)$, we see the dependence of f^{n+1} on a previous power, namely, f^n.

EXAMPLE 5.56 With $A = \{1, 2, 3, 4\}$ and $f: A \to A$ defined by $f = \{(1, 2), (2, 2), (3, 1), (4, 3)\}$, we have $f^2 = f \circ f = \{(1, 2), (2, 2), (3, 2), (4, 1)\}$ and $f^3 = f \circ f^2 = f \circ f \circ f = \{(1, 2), (2, 2), (3, 2), (4, 2)\}$. (What are f^4, f^5?)

We now come to the last new idea for this section: the existence of the invertible function and some of its properties.

Definition 5.20 For sets A, B, if $\mathcal{R}$ is a relation from A to B, then the *converse* of $\mathcal{R}$, denoted $\mathcal{R}^c$, is the relation from B to A defined by $\mathcal{R}^c = \{(b, a) | (a, b) \in \mathcal{R}\}$.

To get $\mathcal{R}^c$ from $\mathcal{R}$, we simply interchange the components of each ordered pair in $\mathcal{R}$. So if $A = \{1, 2, 3, 4\}$, $B = \{w, x, y\}$, and $\mathcal{R} = \{(1, w), (2, w), (3, x)\}$, then $\mathcal{R}^c = \{(w, 1), (w, 2), (x, 3)\}$, a relation from B to A.

Since a function is a relation we can also form the converse of a function. For the same preceding sets A, B, let $f: A \to B$ where $f = \{(1, w), (2, x), (3, y), (4, x)\}$. Then $f^c = \{(w, 1), (x, 2), (y, 3), (x, 4)\}$, a relation, but not a function, from B to A. We wish to investigate when the converse of a function yields a function, but before getting too abstract let us consider the following example.

EXAMPLE 5.57 For $A = \{1, 2, 3\}$ and $B = \{w, x, y\}$, let $f: A \to B$ be given by $f = \{(1, w), (2, x), (3, y)\}$. Then $f^c = \{(w, 1), (x, 2), (y, 3)\}$ is a function from B to A, and we observe that $f^c \circ f = 1_A$ and $f \circ f^c = 1_B$.

This finite example leads us to the following definition.

Definition 5.21 If $f: A \to B$, then f is said to be *invertible* if there is a function $g: B \to A$ such that $g \circ f = 1_A$ and $f \circ g = 1_B$.

Note that the function g in Definition 5.21 is also invertible.

EXAMPLE 5.58

Let $f, g: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 2x + 5$, $g(x) = (1/2)(x - 5)$. Then $(g \circ f)(x) = g(f(x)) = g(2x + 5) = (1/2)\,[(2x + 5) - 5] = x$, and $(f \circ g)(x) = f(g(x)) = f((1/2)(x - 5)) = 2\,[(1/2)(x - 5)] + 5 = x$, so $f \circ g = 1_{\mathbf{R}}$ and $g \circ f = 1_{\mathbf{R}}$. Consequently, f and g are both invertible functions.

Having seen some examples of invertible functions, we now wish to show that the function g of Definition 5.21 is unique. Then we shall find the means to identify an invertible function.

THEOREM 5.7

If a function $f: A \to B$ is invertible and a function $g: B \to A$ satisfies $g \circ f = 1_A$ and $f \circ g = 1_B$, then this function g is unique.

Proof: If g is not unique, then there is another function $h: B \to A$ with $h \circ f = 1_A$ and $f \circ h = 1_B$. Consequently, $h = h \circ 1_B = h \circ (f \circ g) = (h \circ f) \circ g = 1_A \circ g = g$.

As a result of this theorem we shall call the function g *the inverse* of f and shall adopt the notation $g = f^{-1}$. Theorem 5.7 also implies that $f^{-1} = f^c$.

We also see that whenever f is an invertible function, so is the function f^{-1}, and $(f^{-1})^{-1} = f$, again by the uniqueness in Theorem 5.7. But we still do not know what conditions on f insure that f is invertible.

Before stating our next theorem we note that the invertible functions of Examples 5.57 and 5.58 are all bijective. Consequently, these examples provide some motivation for the following result.

THEOREM 5.8

A function $f: A \to B$ is invertible if and only if it is one-to-one and onto.

Proof: Assuming that $f: A \to B$ is invertible, we have a unique function $g: B \to A$ with $g \circ f = 1_A$, $f \circ g = 1_B$. If $a_1, a_2 \in A$ with $f(a_1) = f(a_2)$, then $g(f(a_1)) = g(f(a_2))$, or $(g \circ f)(a_1) = (g \circ f)(a_2)$. With $g \circ f = 1_A$ it follows that $a_1 = a_2$, so f is one-to-one. For the onto property, let $b \in B$. Then $g(b) \in A$, so we can talk about $f(g(b))$. Since $f \circ g = 1_B$, we have $b = 1_B(b) = (f \circ g)(b) = f(g(b))$, so f is onto.

Conversely, suppose $f: A \to B$ is bijective. Since f is onto, for each $b \in B$ there is an $a \in A$ with $f(a) = b$. Consequently, we define the function $g: B \to A$ by $g(b) = a$, where $f(a) = b$. This definition yields a unique function. The only problem that could arise is if $g(b) = a_1 \neq a_2 = g(b)$ because $f(a_1) = b = f(a_2)$. However, this situation cannot arise because f is one-to-one. Our definition of g is such that $g \circ f = 1_A$ and $f \circ g = 1_B$, so we find that f is invertible, with $g = f^{-1}$.

EXAMPLE 5.59

From Theorem 5.8 it follows that the function $f_1: \mathbf{R} \to \mathbf{R}$ defined by $f_1(x) = x^2$ is not invertible (it is neither one-to-one nor onto), but $f_2: [0, +\infty) \to [0, +\infty)$ defined by $f_2(x) = x^2$ is invertible with $f_2^{-1}(x) = \sqrt{x}$.

The next result combines the ideas of function composition and inverse functions. The proof is left to the reader.

THEOREM 5.9 If $f: A \to B$, $g: B \to C$ are invertible functions, then $g \circ f: A \to C$ is invertible and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Having seen some examples of functions and their inverses, one might wonder whether there is an algebraic method to determine the inverse of an invertible function. If the function is finite, we simply interchange the components of the given ordered pairs. But what if the function is defined by a formula, as in Example 5.59? Fortunately, the algebraic manipulations prove to be little more than a careful analysis of "interchanging the components of the ordered pairs." This is demonstrated in the following examples.

EXAMPLE 5.60

For $m, b \in \mathbf{R}$, $m \neq 0$, the function $f: \mathbf{R} \to \mathbf{R}$ defined by $f = \{(x, y) | y = mx + b\}$ is an invertible function, because it is one-to-one and onto.

To get f^{-1} we note that

$$\begin{aligned} f^{-1} &= \{(x, y) | y = mx + b\}^c = \{(y, x) | y = mx + b\} \\ &= \underbrace{\{(x, y) | x = my + b\}}_{} = \{(x, y) | y = (1/m)(x - b)\}. \end{aligned}$$

This is where we rename the variables (replacing x by y and y by x) in order to change the components of the ordered pairs of f.

So $f: \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = mx + b$, and $f^{-1}: \mathbf{R} \to \mathbf{R}$ is defined by $f^{-1}(x) = (1/m)(x - b)$.

EXAMPLE 5.61

Let $f: \mathbf{R} \to \mathbf{R}^+$ be defined by $f(x) = e^x$, where $e \doteq 2.7183$, the base for the natural logarithm. From the graph in Fig. 5.10 we see that f is one-to-one and onto, so $f^{-1}: \mathbf{R}^+ \to \mathbf{R}$ does exist and $f^{-1} = \{(x, y) | y = e^x\}^c = \{(x, y) | x = e^y\} = \{(x, y) | y = \ln x\}$, so $f^{-1}(x) = \ln x$.

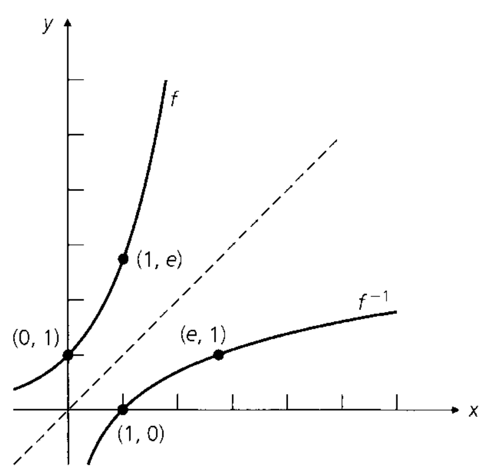


Figure 5.10

We should note that what happens in Fig. 5.10 happens in general. That is, the graphs of f and f^{-1} are symmetric about the line $y = x$. For example, the line segment connecting the points $(1, e)$ and $(e, 1)$ would be bisected by the line $y = x$. This is true for any corresponding pair of points $(x, f(x))$ and $(f(x), f^{-1}(f(x)))$.

This example also yields the following formulas:

$$x = 1_{\mathbf{R}}(x) = (f^{-1} \circ f)(x) = \ln(e^x), \quad \text{for all } x \in \mathbf{R}.$$
$$x = 1_{\mathbf{R}^+}(x) = (f \circ f^{-1})(x) = e^{\ln x}, \quad \text{for all } x > 0.$$

Even when a function $f: A \to B$ is not invertible, we find use for the symbol f^{-1} in the following sense.

Definition 5.22 If $f: A \to B$ and $B_1 \subseteq B$, then $f^{-1}(B_1) = \{x \in A | f(x) \in B_1\}$. The set $f^{-1}(B_1)$ is called the *preimage of* B_1 *under* f.

Be careful! We are now using the symbol f^{-1} in two different ways. Although we have the concept of a preimage for any function, not every function has an inverse function. Consequently, we cannot assume the existence of an inverse for a function f just because we find the symbol f^{-1} being used. A little caution is needed here.

EXAMPLE 5.62 Let $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{6, 7, 8, 9, 10\}$. If $f: A \to B$ with $f = \{(1, 7), (2, 7), (3, 8), (4, 6), (5, 9), (6, 9)\}$, then the following results are obtained.

a) For $B_1 = \{6, 8\} \subseteq B$, we have $f^{-1}(B_1) = \{3, 4\}$, since $f(3) = 8$ and $f(4) = 6$, and for any $a \in A$, $f(a) \notin B_1$ unless $a = 3$ or $a = 4$. Here we also note that $|f^{-1}(B_1)| = 2 = |B_1|$.

b) In the case of $B_2 = \{7, 8\} \subseteq B$, since $f(1) = f(2) = 7$ and $f(3) = 8$, we find that the preimage of B_2 under f is $\{1, 2, 3\}$. And here $|f^{-1}(B_2)| = 3 > 2 = |B_2|$.

c) Now consider the subset $B_3 = \{8, 9\}$ of B. For this case it follows that $f^{-1}(B_3) = \{3, 5, 6\}$ because $f(3) = 8$ and $f(5) = f(6) = 9$. Also we find that $|f^{-1}(B_3)| = 3 > 2 = |B_3|$.

d) If $B_4 = \{8, 9, 10\} \subseteq B$, then with $f(3) = 8$ and $f(5) = f(6) = 9$, we have $f^{-1}(B_4) = \{3, 5, 6\}$. So $f^{-1}(B_4) = f^{-1}(B_3)$ even though $B_4 \supset B_3$. This result follows because there is no element a in the domain A where $f(a) = 10$—that is, $f^{-1}(\{10\}) = \emptyset$.

e) Finally, when $B_5 = \{8, 10\}$ we find that $f^{-1}(B_5) = \{3\}$ since $f(3) = 8$ and, as in part (d), $f^{-1}(\{10\}) = \emptyset$. In this case $|f^{-1}(B_5)| = 1 < 2 = |B_5|$.

Whenever $f: A \to B$, then for each $b \in B$ we shall write $f^{-1}(b)$ instead of $f^{-1}(\{b\})$. For the function in Example 5.62, we find that

$$f^{-1}(6) = \{4\} \quad f^{-1}(7) = \{1, 2\} \quad f^{-1}(8) = \{3\} \quad f^{-1}(9) = \{5, 6\} \quad f^{-1}(10) = \emptyset.$$

EXAMPLE 5.63 Let $f: \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) = \begin{cases} 3x - 5, & x > 0 \\ -3x + 1, & x \le 0. \end{cases}$$

a) Determine $f(0)$, $f(1)$, $f(-1)$, $f(5/3)$, and $f(-5/3)$.

b) Find $f^{-1}(0)$, $f^{-1}(1)$, $f^{-1}(-1)$, $f^{-1}(3)$, $f^{-1}(-3)$, and $f^{-1}(-6)$.

c) What are $f^{-1}([-5, 5])$ and $f^{-1}([-6, 5])$?

a) $f(0) = -3(0) + 1 = 1$ $\quad f(5/3) = 3(5/3) - 5 = 0$
$f(1) = 3(1) - 5 = -2$ $\quad f(-5/3) = -3(-5/3) + 1 = 6$
$f(-1) = -3(-1) + 1 = 4$

b) $f^{-1}(0) = \{x \in \mathbf{R} | f(x) \in \{0\}\} = \{x \in \mathbf{R} | f(x) = 0\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } 3x - 5 = 0\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } -3x + 1 = 0\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } x = 5/3\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } x = 1/3\}$
$= \{5/3\} \cup \emptyset = \{5/3\}$

[Note how the horizontal line $y = 0$—that is, the x-axis—intersects the graph in Fig. 5.11 only at the point $(5/3, 0)$.]

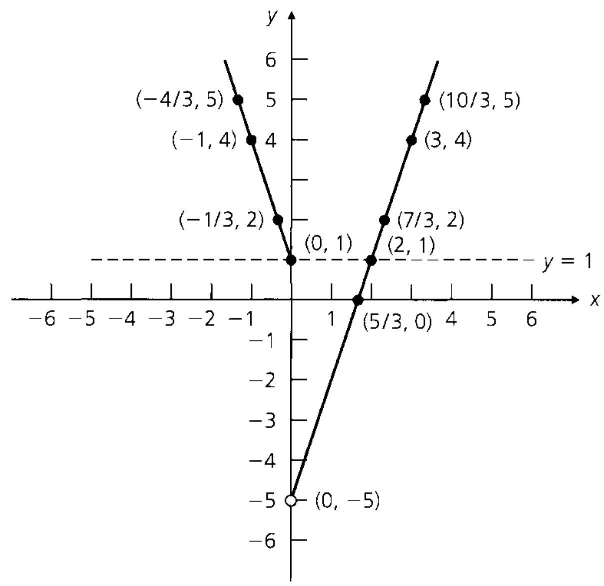


Figure 5.11

$f^{-1}(1) = \{x \in \mathbf{R} | f(x) \in \{1\}\} = \{x \in \mathbf{R} | f(x) = 1\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } 3x - 5 = 1\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } -3x + 1 = 1\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } x = 2\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } x = 0\}$
$= \{2\} \cup \{0\} = \{0, 2\}$

[Here we note how the dashed line $y = 1$ intersects the graph in Fig. 5.11 at the points $(0, 1)$ and $(2, 1)$.]

$f^{-1}(-1) = \{x \in \mathbf{R} | x > 0 \text{ and } 3x - 5 = -1\} \cup \{x \in \mathbf{R} | \; x \leq 0 \text{ and } -3x + 1 = -1\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } x = 4/3\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } x = 2/3\}$
$= \{4/3\} \cup \emptyset = \{4/3\}$

$$f^{-1}(3) = \{-2/3, 8/3\} \qquad f^{-1}(-3) = \{2/3\}$$

$f^{-1}(-6) = \{x \in \mathbf{R} | x > 0 \text{ and } 3x - 5 = -6\} \cup \{x \in \mathbf{R} | \; x \leq 0 \text{ and } -3x + 1 = -6\}$
$= \{x \in \mathbf{R} | x > 0 \text{ and } x = -1/3\} \cup \{x \in \mathbf{R} | x \leq 0 \text{ and } x = 7/3\}$
$= \emptyset \cup \emptyset = \emptyset$

c) $f^{-1}([-5, 5]) = \{x | f(x) \in [-5, 5]\} = \{x | -5 \leq f(x) \leq 5\}$.

(Case 1) $x > 0$: $\quad -5 \leq 3x - 5 \leq 5$
$0 \leq 3x \leq 10$
$0 \leq x \leq 10/3$—so we use $0 < x \leq 10/3$.

(Case 2) $x \le 0$: $-5 \le -3x + 1 \le 5$

$-6 \le -3x \le 4$

$2 \ge x \ge -4/3$—here we use $-4/3 \le x \le 0$.

Hence $f^{-1}([-5, 5]) = \{x \mid -4/3 \le x \le 0 \text{ or } 0 < x \le 10/3\} = [-4/3, 10/3]$. Since there are no points (x, y) on the graph (in Fig. 5.11) where $y \le -5$, it follows from our prior calculations that $f^{-1}([-6, 5]) = f^{-1}([-5, 5]) = [-4/3, 10/3]$.

EXAMPLE 5.64

a) Let $f: \mathbf{Z} \to \mathbf{R}$ be defined by $f(x) = x^2 + 5$. Table 5.9 lists $f^{-1}(B)$ for various subsets B of the codomain $\mathbf{R}$.

b) If $g: \mathbf{R} \to \mathbf{R}$ is defined by $g(x) = x^2 + 5$, the results in Table 5.10 show how a change in domain (from $\mathbf{Z}$ to $\mathbf{R}$) affects the preimages (in Table 5.9).

Table 5.9

B	$f^{-1}(B)$
$\{6\}$	$\{-1, 1\}$
$[6, 7]$	$\{-1, 1\}$
$[6, 10]$	$\{-2, -1, 1, 2\}$
$[-4, 5)$	$\emptyset$
$[-4, 5]$	$\{0\}$
$[5, +\infty)$	$\mathbf{Z}$

Table 5.10

B	$g^{-1}(B)$
$\{6\}$	$\{-1, 1\}$
$[6, 7]$	$[-\sqrt{2}, -1] \cup [1, \sqrt{2}]$
$[6, 10]$	$[-\sqrt{5}, -1] \cup [1, \sqrt{5}]$
$[-4, 5)$	$\emptyset$
$[-4, 5]$	$\{0\}$
$[5, +\infty)$	$\mathbf{R}$

The concept of a preimage appears in conjunction with the set operations of intersection, union, and complementation in our next result. The reader should note the difference between part (a) of this theorem and part (b) of Theorem 5.2.

THEOREM 5.10

If $f: A \to B$ and $B_1, B_2 \subseteq B$, then (a) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$; (b) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$; and (c) $f^{-1}(\overline{B_1}) = \overline{f^{-1}(B_1)}$.

Proof: We prove part (b) and leave parts (a) and (c) for the reader.

For $a \in A$, $a \in f^{-1}(B_1 \cup B_2) \Longleftrightarrow f(a) \in B_1 \cup B_2 \Longleftrightarrow f(a) \in B_1$ or $f(a) \in B_2 \Longleftrightarrow a \in f^{-1}(B_1)$ or $a \in f^{-1}(B_2) \Longleftrightarrow a \in f^{-1}(B_1) \cup f^{-1}(B_2)$.

Using the notation of the preimage, we see that a function $f: A \to B$ is one-to-one if and only if $|f^{-1}(b)| \le 1$ for each $b \in B$.

Discrete mathematics is primarily concerned with finite sets, and the last result of this section demonstrates how the property of finiteness can yield results that fail to be true in general. In addition, it provides an application of the pigeonhole principle.

THEOREM 5.11

Let $f: A \to B$ for finite sets A and B, where $|A| = |B|$. Then the following statements are equivalent: (a) f is one-to-one; (b) f is onto; and (c) f is invertible.

Proof: We have already shown in Theorem 5.8 that (c) $\Rightarrow$ (a) and (b), and that together (a), (b) $\Rightarrow$ (c). Consequently, this theorem will follow when we show that for these conditions on A, B, (a) $\Longleftrightarrow$ (b). Assuming (b), if f is not one-to-one, then there are elements $a_1, a_2 \in A$, with $a_1 \neq a_2$, but $f(a_1) = f(a_2)$. Then $|A| > |f(A)| = |B|$, contradicting $|A| = |B|$. Conversely, if f is not onto, then $|f(A)| < |B|$. With $|A| = |B|$ we have $|A| > |f(A)|$, and it follows from the pigeonhole principle that f is not one-to-one.

Using Theorem 5.11 we now verify the combinatorial identity introduced in Problem 6 at the start of this chapter. For if $n \in \mathbf{Z}^+$ and $|A| = |B| = n$, there are $n!$ one-to-one functions from A to B and $\sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^n$ onto functions from A to B. The equality $n! = \sum_{k=0}^{n}(-1)^k\binom{n}{n-k}(n-k)^n$ is then the numerical equivalent of parts (a) and (b) of Theorem 5.11. [This is also the reason why the diagonal elements $S(n, n)$, $1 \leq n \leq 8$, shown in Table 5.1 all equal 1.]

EXERCISES 5.6

1. a) For $A = \{1, 2, 3, 4, \ldots, 7\}$, how many bijective functions $f\colon A \to A$ satisfy $f(1) \neq 1$?

b) Answer part (a) where $A = \{x | x \in \mathbf{Z}^+, 1 \leq x \leq n\}$, for some fixed $n \in \mathbf{Z}^+$.

2. a) For $A = (-2, 7] \subseteq \mathbf{R}$ define the functions $f, g\colon A \to \mathbf{R}$ by

$$f(x) = 2x - 4 \quad \text{and} \quad g(x) = \frac{2x^2 - 8}{x + 2}.$$

Verify that $f = g$.

b) Is the result in part (a) affected if we change A to $[-7, 2)$?

3. Let $f, g\colon \mathbf{R} \to \mathbf{R}$, where $g(x) = 1 - x + x^2$ and $f(x) = ax + b$. If $(g \circ f)(x) = 9x^2 - 9x + 3$, determine a, b.

4. Let $g\colon \mathbf{N} \to \mathbf{N}$ be defined by $g(n) = 2n$. If $A = \{1, 2, 3, 4\}$ and $f\colon A \to \mathbf{N}$ is given by $f = \{(1, 2), (2, 3), (3, 5), (4, 7)\}$, find $g \circ f$.

5. If $\mathscr{U}$ is a given universe with (fixed) $S, T \subseteq \mathscr{U}$, define $g\colon \mathscr{P}(\mathscr{U}) \to \mathscr{P}(\mathscr{U})$ by $g(A) = T \cap (S \cup A)$ for $A \subseteq \mathscr{U}$. Prove that $g^2 = g$.

6. Let $f, g\colon \mathbf{R} \to \mathbf{R}$ where $f(x) = ax + b$ and $g(x) = cx + d$ for all $x \in \mathbf{R}$, with a, b, c, d real constants. What relationship(s) must be satisfied by a, b, c, d if $(f \circ g)(x) = (g \circ f)(x)$ for all $x \in \mathbf{R}$?

7. Let $f, g, h\colon \mathbf{Z} \to \mathbf{Z}$ be defined by $f(x) = x - 1$, $g(x) = 3x$,

$$h(x) = \begin{cases} 0, & x \text{ even} \\ 1, & x \text{ odd.} \end{cases}$$

Determine (a) $f \circ g$, $g \circ f$, $g \circ h$, $h \circ g$, $f \circ (g \circ h)$, $(f \circ g) \circ h$; (b) f^2, f^3, g^2, g^3, h^2, h^3, h^{500}.

8. Let $f\colon A \to B$, $g\colon B \to C$. Prove that (a) if $g \circ f\colon A \to C$ is onto, then g is onto; and (b) if $g \circ f\colon A \to C$ is one-to-one, then f is one-to-one.

9. a) Find the inverse of the function $f\colon \mathbf{R} \to \mathbf{R}^+$ defined by $f(x) = e^{2x+5}$.

b) Show that $f \circ f^{-1} = 1_{\mathbf{R}^+}$ and $f^{-1} \circ f = 1_{\mathbf{R}}$.

10. For each of the following functions $f\colon \mathbf{R} \to \mathbf{R}$, determine whether f is invertible, and, if so, determine f^{-1}.

a) $f = \{(x, y) | 2x + 3y = 7\}$

b) $f = \{(x, y) | ax + by = c, b \neq 0\}$

c) $f = \{(x, y) | y = x^3\}$

d) $f = \{(x, y) | y = x^4 + x\}$

11. Prove Theorem 5.9.

12. If $A = \{1, 2, 3, 4, 5, 6, 7\}$, $B = \{2, 4, 6, 8, 10, 12\}$, and $f\colon A \to B$ where $f = \{(1, 2), (2, 6), (3, 6), (4, 8), (5, 6), (6, 8), (7, 12)\}$, determine the preimage of B_1 under f in each of the following cases.

a) $B_1 = \{2\}$ **b)** $B_1 = \{6\}$

c) $B_1 = \{6, 8\}$ **d)** $B_1 = \{6, 8, 10\}$

e) $B_1 = \{6, 8, 10, 12\}$ **f)** $B_1 = \{10, 12\}$

13. Let $f\colon \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) = \begin{cases} x + 7, & x \leq 0 \\ -2x + 5, & 0 < x < 3 \\ x - 1, & 3 \leq x \end{cases}$$

a) Find $f^{-1}(-10)$, $f^{-1}(0)$, $f^{-1}(4)$, $f^{-1}(6)$, $f^{-1}(7)$, and $f^{-1}(8)$.

b) Determine the preimage under f for each of the intervals (i) $[-5, -1]$; (ii) $[-5, 0]$; (iii) $[-2, 4]$; (iv) $(5, 10)$; and (v) $[11, 17)$.

14. Let $f\colon \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$. For each of the following subsets B of $\mathbf{R}$, find $f^{-1}(B)$.

a) $B = \{0, 1\}$ **b)** $B = \{-1, 0, 1\}$

c) $B = [0, 1]$ **d)** $B = [0, 1)$

e) $B = [0, 4]$ **f)** $B = (0, 1] \cup (4, 9)$

15. Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{6, 7, 8, 9, 10, 11, 12\}$. How many functions $f: A \to B$ are such that $f^{-1}(\{6, 7, 8\}) = \{1, 2\}$?

16. Let $f: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = \lfloor x \rfloor$, the greatest integer in x. Find $f^{-1}(B)$ for each of the following subsets B of $\mathbf{R}$.

a) $B = \{0, 1\}$
b) $B = \{-1, 0, 1\}$
c) $B = [0, 1)$
d) $B = [0, 2)$
e) $B = [-1, 2]$
f) $B = [-1, 0) \cup (1, 3]$

17. Let $f, g: \mathbf{Z}^+ \to \mathbf{Z}^+$ where for all $x \in \mathbf{Z}^+$, $f(x) = x + 1$ and $g(x) = \max\{1, x - 1\}$, the maximum of 1 and $x - 1$.

a) What is the range of f?
b) Is f an onto function?
c) Is the function f one-to-one?
d) What is the range of g?
e) Is g an onto function?
f) Is the function g one-to-one?
g) Show that $g \circ f = 1_{\mathbf{Z}^+}$.
h) Determine $(f \circ g)(x)$ for $x = 2, 3, 4, 7, 12$, and 25.
i) Do the answers for parts (b), (g), and (h) contradict the result in Theorem 5.8?

18. Let f, g, h denote the following closed binary operations on $\mathcal{P}(\mathbf{Z}^+)$. For $A, B \subseteq \mathbf{Z}^+$, $f(A, B) = A \cap B$, $g(A, B) = A \cup B$, $h(A, B) = A \triangle B$.

a) Are any of the functions one-to-one?
b) Are any of f, g, and h onto functions?
c) Is any one of the given functions invertible?
d) Are any of the following sets infinite?

(1) $f^{-1}(\emptyset)$
(2) $g^{-1}(\emptyset)$
(3) $h^{-1}(\emptyset)$
(4) $f^{-1}(\{1\})$
(5) $g^{-1}(\{2\})$
(6) $h^{-1}(\{3\})$
(7) $f^{-1}(\{4, 7\})$
(8) $g^{-1}(\{8, 12\})$
(9) $h^{-1}(\{5, 9\})$

e) Determine the number of elements in each of the finite sets in part (d).

19. Prove parts (a) and (c) of Theorem 5.10.

20. a) Give an example of a function $f: \mathbf{Z} \to \mathbf{Z}$ where (i) f is one-to-one but not onto; and (ii) f is onto but not one-to-one.

b) Do the examples in part (a) contradict Theorem 5.11?

21. Let $f: \mathbf{Z} \to \mathbf{N}$ be defined by

$$f(x) = \begin{cases} 2x - 1, & \text{if } x > 0 \\ -2x, & \text{for } x \leq 0. \end{cases}$$

a) Prove that f is one-to-one and onto.
b) Determine f^{-1}.

22. If $|A| = |B| = 5$, how many functions $f: A \to B$ are invertible?

23. Let $f, g, h, k: \mathbf{N} \to \mathbf{N}$ where $f(n) = 3n$, $g(n) = \lfloor n/3 \rfloor$, $h(n) = \lfloor (n + 1)/3 \rfloor$, and $k(n) = \lfloor (n + 2)/3 \rfloor$, for each $n \in \mathbf{N}$. (a) For each $n \in \mathbf{N}$ what are $(g \circ f)(n)$, $(h \circ f)(n)$, and $(k \circ f)(n)$? (b) Do the results in part (a) contradict Theorem 5.7?

5.7 Computational Complexity†

In Section 4.4 we introduced the concept of an algorithm, following the examples set forth by the division algorithm (of Section 4.3) and the Euclidean algorithm (of Section 4.4). At that time we were concerned with certain properties of a general algorithm:

- The precision of the individual step-by-step instructions
- The input provided to the algorithm, and the output the algorithm then provides
- The ability of the algorithm to solve a certain type of problem, not just specific instances of the problem
- The uniqueness of the intermediate and final results, based on the input

†The material in Sections 5.7 and 5.8 may be skipped at this point. It will not be used very much until Chapter 10. The only place where this material appears before Chapter 10 is in Example 7.13, but that example can be omitted without any loss of continuity.

- The finite nature of the algorithm in that it terminates after the execution of a finite number of instructions

When an algorithm correctly solves a certain type of problem and satisfies these five conditions, then we may find ourselves examining it further in the following ways.

1) Can we somehow measure how long it takes the algorithm to solve a problem of a certain size? Whether we can may very well depend, for example, on the compiler being used, so we want to develop a measure that doesn't actually depend on such considerations as compilers, execution speeds, or other characteristics of a given computer.

   For example, if we want to compute a^n for $a \in \mathbf{R}$ and $n \in \mathbf{Z}^+$, is there some "function of n" that can describe how fast a given algorithm for such exponentiation accomplishes this?

2) Suppose we can answer questions such as the one set forth at the start of item 1. Then if we have two (or more) algorithms that solve a given problem, is there perhaps a way to determine whether one algorithm is "better" than another?

In particular, suppose we consider the problem of determining whether a certain real number x is present in the list of n real numbers $a_1, a_2, \ldots, a_n$. Here we have a problem of size n.

If there is an algorithm that solves this problem, how long does it take to do so? To measure this we seek a function $f(n)$, called the *time-complexity function*† of the algorithm. We expect (both here and in general) that the value of $f(n)$ will increase as n increases. Also, our major concern in dealing with any algorithm is how the algorithm performs for *large* values of n.

In order to study what has now been described in a somewhat informal manner, we need to introduce the following fundamental idea.

Definition 5.23 Let $f, g\colon \mathbf{Z}^+ \to \mathbf{R}$. We say that g *dominates* f (or f is *dominated* by g) if there exist constants $m \in \mathbf{R}^+$ and $k \in \mathbf{Z}^+$ such that $|f(n)| \leq m|g(n)|$ for all $n \in \mathbf{Z}^+$, where $n \geq k$.

Note that as we consider the values of $f(1), g(1), f(2), g(2), \ldots,$ there is a point (namely, k) after which the size of $f(n)$ is bounded above by a positive multiple (m) of the size of $g(n)$. Also, when g dominates f, then $|f(n)/g(n)| \leq m$ [that is, the size of the quotient $f(n)/g(n)$ is bounded by m], for those $n \in \mathbf{Z}^+$ where $n \geq k$ and $g(n) \neq 0$.

When f is dominated by g we say that f is of *order* (*at most*) g and we use what is called "big-Oh" notation to designate this. We write $f \in O(g)$, where $O(g)$ is read "order g" or "big-Oh of g." As suggested by the notation "$f \in O(g)$," $O(g)$ represents the set of all functions with domain $\mathbf{Z}^+$ and codomain $\mathbf{R}$ that are dominated by g. These ideas are demonstrated in the following examples.

EXAMPLE 5.65 Let $f, g\colon \mathbf{Z}^+ \to \mathbf{R}$ be given by $f(n) = 5n$, $g(n) = n^2$, for $n \in \mathbf{Z}^+$. If we compute $f(n)$ and $g(n)$ for $1 \leq n \leq 4$, we find that $f(1) = 5$, $g(1) = 1$; $f(2) = 10$, $g(2) = 4$; $f(3) =$

†We could also study the *space-complexity function* of an algorithm, which we need when we attempt to measure the amount of memory required for the execution of an algorithm on a problem of size n. In this text, however, we limit our study to the time-complexity function.

15, $g(3) = 9$; and $f(4) = 20$, $g(4) = 16$. However, $n \geq 5 \Rightarrow n^2 \geq 5n$, and we have $|f(n)| = 5n \leq n^2 = |g(n)|$. So with $m = 1$ and $k = 5$, we find that for $n \geq k$, $|f(n)| \leq m|g(n)|$. Consequently, g dominates f and $f \in O(g)$. [Note that $|f(n)/g(n)|$ is bounded by 1 for all $n \geq 5$.]

We also realize that for all $n \in \mathbf{Z}^+$, $|f(n)| = 5n \leq 5n^2 = 5|g(n)|$. So the dominance of f by g is shown here with $k = 1$ and $m = 5$. This is enough to demonstrate that the constants k and m of Definition 5.23 need *not* be unique.

Furthermore, we can generalize this result if we now consider functions $f_1, g_1: \mathbf{Z}^+ \to \mathbf{R}$ defined by $f_1(n) = an$, $g_1(n) = bn^2$, where a, b are nonzero real numbers. For if $m \in \mathbf{R}^+$ with $m|b| \geq |a|$, then for all $n \geq 1 (= k)$, $|f_1(n)| = |an| = |a|n \leq m|b|n \leq m|b|n^2 = m|bn^2| = m|g_1(n)|$, and so $f_1 \in O(g_1)$.

In Example 5.65 we observed that $f \in O(g)$. Taking a second look at the functions f and g, we now want to show that $g \notin O(f)$.

EXAMPLE 5.66

Once again let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ be defined by $f(n) = 5n$, $g(n) = n^2$, for $n \in \mathbf{Z}^+$.

If $g \in O(f)$, then in terms of quantifiers, we would have

$$\exists m \in \mathbf{R}^+ \; \exists k \in \mathbf{Z}^+ \; \forall n \in \mathbf{Z}^+ \; [(n \geq k) \Rightarrow |g(n)| \leq m|f(n)|].$$

Consequently, to show that $g \notin O(f)$, we need to verify that

$$\forall m \in \mathbf{R}^+ \; \forall k \in \mathbf{Z}^+ \; \exists n \in \mathbf{Z}^+ \; [(n \geq k) \wedge (|g(n)| > m|f(n)|)].$$

To accomplish this, we first should realize that m and k are arbitrary, so we have no control over their values. The only number over which we have control is the positive integer n that we select. Now no matter what the values of m and k happen to be, we can select $n \in \mathbf{Z}^+$ such that $n > \max\{5m, k\}$. Then $n \geq k$ (actually $n > k$) and $n > 5m \Rightarrow n^2 > 5mn$, so $|g(n)| = n^2 > 5mn = m|5n| = m|f(n)|$ and $g \notin O(f)$.

For those who prefer the method of proof by contradiction, we present a second approach. If $g \in O(f)$, then we would have

$$n^2 = |g(n)| \leq m|f(n)| = mn$$

for all $n \geq k$, where k is some fixed positive integer and m is a (real) *constant*. But then from $n^2 \leq mn$ we deduce that $n \leq m$. This is impossible because $n (\in \mathbf{Z}^+)$ is a variable that can increase without bound while m is still a constant.

EXAMPLE 5.67

a) Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ with $f(n) = 5n^2 + 3n + 1$ and $g(n) = n^2$. Then $|f(n)| = |5n^2 + 3n + 1| = 5n^2 + 3n + 1 \leq 5n^2 + 3n^2 + n^2 = 9n^2 = 9|g(n)|$. Hence for all $n \geq 1$ $(= k)$, $|f(n)| \leq m|g(n)|$ for any $m \geq 9$, and $f \in O(g)$. We can also write $f \in O(n^2)$ in this case.

In addition, $|g(n)| = n^2 \leq 5n^2 \leq 5n^2 + 3n + 1 = |f(n)|$ for all $n \geq 1$. So $|g(n)| \leq m|f(n)|$ for any $m \geq 1$ and all $n \geq k \geq 1$. Consequently $g \in O(f)$. [In fact, $O(g) = O(f)$; that is, any function from $\mathbf{Z}^+$ to $\mathbf{R}$ that is dominated by one of f, g is also dominated by the other. We shall examine this result for the general case in the Section Exercises.]

b) Now consider $f, g: \mathbf{Z}^+ \to \mathbf{R}$ with $f(n) = 3n^3 + 7n^2 - 4n + 2$ and $g(n) = n^3$. Here we have $|f(n)| = |3n^3 + 7n^2 - 4n + 2| \leq |3n^3| + |7n^2| + |-4n| + |2| \leq 3n^3 + 7n^3 + 4n^3 + 2n^3 = 16n^3 = 16|g(n)|$, for all $n \geq 1$. So with $m = 16$ and $k = 1$, we find that f is dominated by g, and $f \in O(g)$, or $f \in O(n^3)$.

Since $7n - 4 > 0$ for all $n \geq 1$, we can write $n^3 \leq 3n^3 \leq 3n^3 + (7n - 4)n + 2$ whenever $n \geq 1$. Then $|g(n)| \leq |f(n)|$ for all $n \geq 1$, and $g \in O(f)$. [As in part (a), we also have $O(f) = O(g) = O(n^3)$ in this case.]

We generalize the results of Example 5.67 as follows. Let $f: \mathbf{Z}^+ \to \mathbf{R}$ be the polynomial function where $f(n) = a_t n^t + a_{t-1}n^{t-1} + \cdots + a_2 n^2 + a_1 n + a_0$, for $a_t, a_{t-1}, \ldots, a_2, a_1, a_0 \in \mathbf{R}$, $a_t \neq 0$, $t \in \mathbf{N}$. Then

$$\begin{aligned}
|f(n)| &= |a_t n^t + a_{t-1}n^{t-1} + \cdots + a_2 n^2 + a_1 n + a_0| \\
&\leq |a_t n^t| + |a_{t-1}n^{t-1}| + \cdots + |a_2 n^2| + |a_1 n| + |a_0| \\
&= |a_t| n^t + |a_{t-1}| n^{t-1} + \cdots + |a_2| n^2 + |a_1| n + |a_0| \\
&\leq |a_t| n^t + |a_{t-1}| n^t + \cdots + |a_2| n^t + |a_1| n^t + |a_0| n^t \\
&= (|a_t| + |a_{t-1}| + \cdots + |a_2| + |a_1| + |a_0|) n^t.
\end{aligned}$$

In Definition 5.23, let $m = |a_t| + |a_{t-1}| + \cdots + |a_2| + |a_1| + |a_0|$ and $k = 1$, and let $g: \mathbf{Z}^+ \to \mathbf{R}$ be given by $g(n) = n^t$. Then $|f(n)| \leq m|g(n)|$ for all $n \geq k$, so f is dominated by g, or $f \in O(n^t)$.

It is also true that $g \in O(f)$ and that $O(f) = O(g) = O(n^t)$.

This generalization provides the following special results on summations.

EXAMPLE 5.68

a) Let $f: \mathbf{Z}^+ \to \mathbf{R}$ be given by $f(n) = 1 + 2 + 3 + \cdots + n$. Then (from Examples 1.40 and 4.1) $f(n) = \left(\frac{1}{2}\right)(n)(n+1) = \left(\frac{1}{2}\right)n^2 + \left(\frac{1}{2}\right)n$, so $f \in O(n^2)$.

b) If $g: \mathbf{Z}^+ \to \mathbf{R}$ with $g(n) = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \left(\frac{1}{6}\right)(n)(n+1)(2n+1)$ (from Example 4.4), then $g(n) = \left(\frac{1}{3}\right)n^3 + \left(\frac{1}{2}\right)n^2 + \left(\frac{1}{6}\right)n$ and $g \in O(n^3)$.

c) If $t \in \mathbf{Z}^+$, and $h: \mathbf{Z}^+ \to \mathbf{R}$ is defined by $h(n) = \sum_{i=1}^{n} i^t$, then $h(n) = 1^t + 2^t + 3^t + \cdots + n^t \leq n^t + n^t + n^t + \cdots + n^t = n(n^t) = n^{t+1}$ so $h \in O(n^{t+1})$.

Now that we have examined several examples of function dominance, we shall close this section with two final observations. In the next section we shall apply the idea of function dominance in the analysis of algorithms.

1) When dealing with the concept of function dominance, we seek the best (or tightest) bound in the following sense. Suppose that $f, g, h: \mathbf{Z}^+ \to \mathbf{R}$, where $f \in O(g)$ and $g \in O(h)$. Then we also have $f \in O(h)$. (A proof for this is requested in the Section Exercises.) If $h \notin O(g)$, however, the statement $f \in O(g)$ provides a "better" bound on $|f(n)|$ than the statement $f \in O(h)$. For example, if $f(n) = 5$, $g(n) = 5n$, and $h(n) = n^2$, for all $n \in \mathbf{Z}^+$, then $f \in O(g)$, $g \in O(h)$, and $f \in O(h)$, but $h \notin O(g)$. Therefore, we are provided with more information by the statement $f \in O(g)$ than by the statement $f \in O(h)$.
2) Certain orders, such as $O(n)$ and $O(n^2)$, often occur when we deal with function dominance. Therefore they have come to be designated by special names. Some of the most important of these orders are listed in Table 5.11.

Table 5.11

Big-Oh Form	Name
$O(1)$	Constant
$O(\log_2 n)$	Logarithmic
$O(n)$	Linear
$O(n \log_2 n)$	$n \log_2 n$
$O(n^2)$	Quadratic
$O(n^3)$	Cubic
$O(n^m)$, $m = 0, 1, 2, 3, \ldots$	Polynomial
$O(c^n)$, $c > 1$	Exponential
$O(n!)$	Factorial

EXERCISES 5.7

1. Use the results of Table 5.11 to determine the best "big-Oh" form for each of the following functions $f: \mathbf{Z}^+ \to \mathbf{R}$.

a) $f(n) = 3n + 7$ b) $f(n) = 3 + \sin(1/n)$

c) $f(n) = n^3 - 5n^2 + 25n - 165$

d) $f(n) = 5n^2 + 3n \log_2 n$

e) $f(n) = n^2 + (n - 1)^3$

f) $f(n) = \dfrac{n(n+1)(n+2)}{(n+3)}$

g) $f(n) = 2 + 4 + 6 + \cdots + 2n$

2. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$, where $f(n) = n$ and $g(n) = n + (1/n)$, for $n \in \mathbf{Z}^+$. Use Definition 5.23 to show that $f \in O(g)$ and $g \in O(f)$.

3. In each of the following, $f, g: \mathbf{Z}^+ \to \mathbf{R}$. Use Definition 5.23 to show that g dominates f.

a) $f(n) = 100 \log_2 n$, $g(n) = \left(\frac{1}{2}\right) n$

b) $f(n) = 2^n$, $g(n) = 2^{2n} - 1000$

c) $f(n) = 3n^2$, $g(n) = 2^n + 2n$

4. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ be defined by $f(n) = n + 100$, $g(n) = n^2$. Use Definition 5.23 to show that $f \in O(g)$ but $g \notin O(f)$.

5. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$, where $f(n) = n^2 + n$ and $g(n) = \left(\frac{1}{2}\right) n^3$, for $n \in \mathbf{Z}^+$. Use Definition 5.23 to show that $f \in O(g)$ but $g \notin O(f)$.

6. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ be defined as follows:

$$f(n) = \begin{cases} n, & \text{for } n \text{ odd} \\ 1, & \text{for } n \text{ even} \end{cases} \qquad g(n) = \begin{cases} 1, & \text{for } n \text{ odd} \\ n, & \text{for } n \text{ even} \end{cases}$$

Verify that $f \notin O(g)$ and $g \notin O(f)$.

7. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ where $f(n) = n$ and $g(n) = \log_2 n$, for $n \in \mathbf{Z}^+$. Show that $g \in O(f)$ but $f \notin O(g)$.

(*Hint*:

$$\lim_{n \to \infty} \frac{n}{\log_2 n} = +\infty.$$

This requires the use of calculus.)

8. Let $f, g, h: \mathbf{Z}^+ \to \mathbf{R}$ where $f \in O(g)$ and $g \in O(h)$. Prove that $f \in O(h)$.

9. If $g: \mathbf{Z}^+ \to \mathbf{R}$ and $c \in \mathbf{R}$, we define the function cg: $\mathbf{Z}^+ \to \mathbf{R}$ by $(cg)(n) = c(g(n))$, for each $n \in \mathbf{Z}^+$. Prove that if $f, g: \mathbf{Z}^+ \to \mathbf{R}$ with $f \in O(g)$, then $f \in O(cg)$ for all $c \in \mathbf{R}$, $c \neq 0$.

10. a) Prove that $f \in O(f)$ for all $f: \mathbf{Z}^+ \to \mathbf{R}$.

b) Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$. If $f \in O(g)$ and $g \in O(f)$, prove that $O(f) = O(g)$. That is, prove that for all $h: \mathbf{Z}^+ \to \mathbf{R}$, if h is dominated by f, then h is dominated by g, and conversely.

c) If $f, g: \mathbf{Z}^+ \to \mathbf{R}$, prove that if $O(f) = O(g)$, then $f \in O(g)$ and $g \in O(f)$.

11. The following is analogous to the "big-Oh" notation introduced in conjunction with Definition 5.23.

For $f, g: \mathbf{Z}^+ \to \mathbf{R}$ we say that f is of *order at least* g if there exist constants $M \in \mathbf{R}^+$ and $k \in \mathbf{Z}^+$ such that $|f(n)| \geq M|g(n)|$ for all $n \in \mathbf{Z}^+$, where $n \geq k$. In this case we write $f \in \Omega(g)$ and say that f is "big Omega of g." So $\Omega(g)$ represents the set of all functions with domain $\mathbf{Z}^+$ and codomain $\mathbf{R}$ that dominate g.

Suppose that $f, g, h: \mathbf{Z}^+ \to \mathbf{R}$, where $f(n) = 5n^2 + 3n$, $g(n) = n^2$, $h(n) = n$, for all $n \in \mathbf{Z}^+$. Prove that (a) $f \in \Omega(g)$; (b) $g \in \Omega(f)$; (c) $f \in \Omega(h)$; and (d) $h \notin \Omega(f)$—that is, h is *not* "big Omega of f."

12. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$. Prove that $f \in \Omega(g)$ if and only if $g \in O(f)$.

13. a) Let $f: \mathbf{Z}^+ \to \mathbf{R}$ where $f(n) = \sum_{i=1}^{n} i$. When $n = 4$, for example, we have $f(n) = f(4) = 1 + 2 + 3 + 4 > 2 + 3 + 4 > 2 + 2 + 2 = 3 \cdot 2 = \lceil (4+1)/2 \rceil 2 = 6 >$

$(4/2)^2 = (n/2)^2$. For $n = 5$, we find $f(n) = f(5) = 1+2+3+4+5 > 3+4+5 > 3+3+3 = 3 \cdot 3 = \lceil (5+1)/2 \rceil 3 = 9 > (5/2)^2 = (n/2)^2$. In general, $f(n) = 1+2+\cdots+n > \lceil n/2 \rceil + \cdots + n > \lceil n/2 \rceil + \cdots + \lceil n/2 \rceil = \lceil (n+1)/2 \rceil \lceil n/2 \rceil > n^2/4$. Consequently, $f \in \Omega(n^2)$.

Use

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

to provide an alternative proof that $f \in \Omega(n^2)$.

b) Let $g: \mathbf{Z}^+ \to \mathbf{R}$ where $g(n) = \sum_{i=1}^{n} i^2$. Prove that $g \in \Omega(n^3)$.

c) For $t \in \mathbf{Z}^+$, let $h: \mathbf{Z}^+ \to \mathbf{R}$ where $h(n) = \sum_{i=1}^{n} i^t$. Prove that $h \in \Omega(n^{t+1})$.

14. For $f, g: \mathbf{Z}^+ \to \mathbf{R}$, we say that f is "big Theta of g," and write $f \in \Theta(g)$, when there exist constants $m_1, m_2 \in \mathbf{R}^+$ and $k \in \mathbf{Z}^+$ such that $m_1|g(n)| \leq |f(n)| \leq m_2|g(n)|$, for all $n \in \mathbf{Z}^+$, where $n \geq k$. Prove that $f \in \Theta(g)$ if and only if $f \in \Omega(g)$ and $f \in O(g)$.

15. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$. Prove that

$$f \in \Theta(g) \text{ if and only if } g \in \Theta(f).$$

16. a) Let $f: \mathbf{Z}^+ \to \mathbf{R}$ where $f(n) = \sum_{i=1}^{n} i$. Prove that $f \in \Theta(n^2)$.

b) Let $g: \mathbf{Z}^+ \to \mathbf{R}$ where $g(n) = \sum_{i=1}^{n} i^2$. Prove that $g \in \Theta(n^3)$.

c) For $t \in \mathbf{Z}^+$, let $h: \mathbf{Z}^+ \to \mathbf{R}$ where $h(n) = \sum_{i=1}^{n} i^t$. Prove that $h \in \Theta(n^{t+1})$.

5.8 Analysis of Algorithms

Now that the reader has been introduced to the concept of function dominance, it is time to see how this idea is used in the study of algorithms. In this section we present our algorithms as pseudocode procedures. (We shall also present algorithms as lists of instructions. The reader will find this to be the case in later chapters.)

We start with a procedure to determine the balance in a savings account.

EXAMPLE 5.69

In Fig. 5.12 we have a procedure (written in pseudocode) for computing the balance in a savings account n months (for $n \in \mathbf{Z}^+$) after it has been opened. (This balance is the procedure's output.) Here the user supplies the value of n, the input for the program. The variables *deposit*, *balance*, and *rate* are real variables, while i is an integer variable. (The annual interest rate is 0.06.)

```
procedure AccountBalance(n: integer)
begin
  deposit := 50.00          {The monthly deposit}
  i := 1                    {Initializes the counter}
  rate := 0.005             {The monthly interest rate}
  balance := 100.00         {Initializes the balance}
  while i ≤ n do
    begin
      balance := deposit + balance + balance * rate
      i := i + 1
    end
end
```

Figure 5.12

Consider the following specific situation. Nathan puts \$100.00 in a new account on January 1. Each month the bank adds the interest (*balance* $*$ *rate*) to Nathan's account—on the first of the month. In addition, Nathan deposits an additional \$50.00 on the first of each month (starting on February 1). This program tells Nathan the balance in his account after n months have gone by (assuming that the interest rate does not change). [*Note*: After one month, $n = 1$ and the balance is \$50.00 (new deposit) + \$100.00 (initial deposit) + (\$100.00)(0.005) (the interest) = \$150.50. When $n = 2$ the new balance is \$50.00 (new deposit) + \$150.50 (previous balance) + (\$150.50) (0.005) (new interest) = \$201.25.]

Our objective is to count (measure) the total number of operations (such as assignments, additions, multiplications, and comparisons) this program segment takes to compute the balance in Nathan's account n months after he opened it. We shall let $f(n)$ denote the total number of these operations. [Then $f: \mathbf{Z}^+ \to \mathbf{R}$. (Actually, $f(\mathbf{Z}^+) \subseteq \mathbf{Z}^+$.)]

The program segment begins with four assignment statements, where the integer variable i and the real variable *balance* are initialized, and the values of the real variables *deposit* and *rate* are declared. Then the **while** loop is executed n times. Each execution of the loop involves the following seven operations:

1) Comparing the present value of the counter i with n.
2) Increasing the present value of *balance* to *deposit* + *balance* + *balance* * *rate*; this involves one multiplication, two additions, and one assignment.
3) Incrementing the value of the counter by 1; this involves one addition and one assignment.

Finally, there is one more comparison. This is made when $i = n + 1$, so the **while** loop is terminated and the other six operations (in steps 2 and 3) are not performed.

Therefore, $f(n) = 4 + 7n + 1 = 7n + 5 \in O(n)$. Consequently, we say that $f \in O(n)$. For as n gets larger, the "order of magnitude" of $7n + 5$ depends primarily on the value n, the number of times the **while** loop is executed. Therefore, we could have obtained $f \in O(n)$ by simply counting the number of times the **while** loop was executed. Such shortcuts will be used in our calculations for the remaining examples.

Our next example introduces us to a situation where three types of complexity are determined. These measures are called the *best-case* complexity, the *worst-case* complexity, and the *average-case* complexity.

EXAMPLE 5.70

In this example we examine a typical *searching* process. Here an array of n (≥ 1) integers $a_1, a_2, a_3, \ldots, a_n$ is to be searched for the presence of an integer called *key*. If the integer is found, the value of *location* indicates its first location in the array; if it is not found the value of *location* is 0, indicating an unsuccessful search.

We cannot assume that the entries in the array are in any particular order. (If they were, the problem would be easier and a more efficient algorithm could be developed.) The input for this algorithm consists of the array (which is read in by the user or provided, perhaps, as a file from an external source), along with the number n of elements in the array, and the value of the integer *key*.

The algorithm is provided in the pseudocode procedure in Fig. 5.13.

We shall define the complexity function $f(n)$ for this algorithm to be the number of elements in the array that are examined until the value *key* is found (for the first time) or the array is exhausted (that is, the number of times the **while** loop is executed).

What is the best thing that can happen in our search for *key*? If *key* $= a_1$, we find that *key* is the first entry of the array, and we had to compare *key* with only one element of the array. In this case we have $f(n) = 1$, and we say that the *best-case complexity* for our algorithm

```
procedure LinearSearch(key, n: integer; a1,a2,a3,...,an: integers)
begin
  i := 1                          {initializes the counter}
  while (i ≤ n and key ≠ ai) do
    i := i + 1
  if i ≤ n then location := i     {successful search}
  else location := 0              {unsuccessful search}
end {location is the subscript of the first array entry that equals key;
    location is 0 if key is not found}
```

Figure 5.13

is $O(1)$ (that is, it is constant and independent of the size of the array). Unfortunately, we cannot expect such a situation to occur very often.

From the best situation we turn now to the worst. We see that we have to examine all n entries of the array if (1) the first occurrence of *key* is a_n or (2) *key* is not found in the array. In either case we have $f(n) = n$, and the *worst-case complexity* here is $O(n)$. (The worst-case complexity will typically be considered throughout the text.)

Finally, we wish to obtain an estimate of the average number of array entries examined. We shall assume that the n entries of the array are distinct and are all equally likely (with probability p) to contain the value *key*, and that the probability that *key* is not in the array is equal to q. Consequently, we have $np + q = 1$ and $p = (1 - q)/n$.

For each $1 \leq i \leq n$, if *key* equals a_i, then i elements of the array have been examined. If *key* is not in the array, then all n array elements are examined. Therefore, the *average-case complexity* is determined by the average number of array elements examined, which is

$$\begin{aligned} f(n) &= (1 \cdot p + 2 \cdot p + 3 \cdot p + \cdots + n \cdot p) + n \cdot q = p(1 + 2 + 3 + \cdots + n) + nq \\ &= \frac{pn(n+1)}{2} + nq. \end{aligned}$$

If $q = 0$, then *key* is in the array, $p = 1/n$ and $f(n) = (n+1)/2 \in O(n)$. For $q = 1/2$, we have an even chance that *key* is in the array and $f(n) = (1/(2n))[n(n+1)/2] + (n/2) = (n+1)/4 + (n/2) \in O(n)$. [In general, for all $0 \leq q \leq 1$, we have $f(n) \in O(n)$.]

EXAMPLE 5.71†

The result in Example 5.70 for the average number of array elements examined in the linear search algorithm may also be calculated using the idea of the random variable. When the algorithm is applied to the array $a_1, a_2, a_3, \ldots, a_n$ (of n distinct integers), we let the discrete random variable X count the number of array elements examined in the search for the integer *key*. Here the sample space can be considered as $\{1, 2, 3, \ldots, n, n^*\}$, where for $1 \leq i \leq n$, we have the case where *key* is found to be a_i — so that the i elements $a_1, a_2, a_3, \ldots, a_i$ have been examined. The entry n^* denotes the situation where all n elements are examined but *key* is not found among any of the array elements $a_1, a_2, a_3, \ldots, a_n$.

Once again we assume that each array entry has the same probability p of containing the value *key* and that q is the probability that *key* is not in the array. Then $np + q = 1$ and

†This example uses the concept of the discrete random variable which was introduced in the optional material in Section 3.7. It may be skipped without loss of continuity.

we have $Pr(X = i) = p$, for $1 \le i \le n$, and $Pr(X = n^*) = q$. Consequently, the average number of array elements examined during the execution of the linear search algorithm is

$$E(X) = \sum_{i=1}^{n} iPr(X = i) + nPr(X = n^*)$$
$$= \sum_{i=1}^{n} ip + np = p(1 + 2 + 3 + \cdots + n) + nq = \frac{pn(n+1)}{2} + nq.$$

Early in the discussion of the previous section, we mentioned how we might want to compare two algorithms that both correctly solve a given type of problem. Such a comparison can be accomplished by using the time-complexity functions for the algorithms. We demonstrate this in the next two examples.

EXAMPLE 5.72

The algorithm implemented in the pseudocode procedure of Fig. 5.14 outputs the value of a^n for the input a, n, where a is a real number and n is a positive integer. The real variable x is initialized as 1.0 and then used to store the values $a, a^2, a^3, \ldots, a^n$ during execution of the **for** loop. Here we define the time-complexity function $f(n)$ for the algorithm as the number of multiplications that occur in the **for** loop. Consequently, we have $f(n) = n \in O(n)$.

```
procedure Power1(a: real; n: positive integer)
begin
  x := 1.0
  for i := 1 to n do
    x := x * a
end
```

Figure 5.14

EXAMPLE 5.73

In Fig. 5.15 we have a second pseudocode procedure for evaluating a^n for all $a \in \mathbf{R}$, $n \in \mathbf{Z}^+$. Recall that $\lfloor i/2 \rfloor$ is the greatest integer in (or the floor of) $i/2$.

```
procedure Power2(a: real; n: positive integer)
begin
  x := 1.0
  i := n
  while i > 0 do
    begin
      if i ≠ 2 * ⌊i/2⌋ then    {i is odd}
        x := x * a
      i := ⌊i/2⌋
      if i > 0 then
        a := a * a
    end
end
```

Figure 5.15

For this procedure the real variable x is initialized as 1.0 and then used to store the appropriate powers of a until it contains the value of a^n. The results shown in Fig. 5.16 demonstrate what is happening to x (and a) for the cases where $n = 7$ and 8. The numbers 1, 2, 3, and 4 indicate the first, second, third, and fourth times the statements in the **while** loop (in particular, the statement $i := \lfloor i/2 \rfloor$) are executed. If $n = 7$, then because $2^2 < 7 < 2^3$, we have $2 < \log_2 7 < 3$. Here the **while** loop is executed three times and

$$3 = \lfloor \log_2 7 \rfloor + 1 < \log_2 7 + 1,$$

where $\lfloor \log_2 7 \rfloor$ denotes the greatest integer in $\log_2 7$, which is 2. Also, when $n = 8$, the number of times the **while** loop is executed is

$$4 = \lfloor \log_2 8 \rfloor + 1 = \log_2 8 + 1,$$

since $\log_2 8 = 3$.

n = 7 x := 1.0 i := 7	n = 8 x := 1.0 i := 8
1 { x := x * a {x = a} i := 3 a := a * a	1 { i := 4 a := a * a
2 { x := x * a {x = a^3} i := 1 a := a * a	2 { i := 2 a := a * a
	3 { i := 1 a := a * a
3 { x := x * a {x = a^7} i := 0 [x = $a^7 = a \cdot a^2 \cdot a^4$]	4 { x := x * a {x = a^8} i := 0 [x = $(((a)^2)^2)^2$]

Figure 5.16

We shall define the time-complexity function $g(n)$ for (the implementation of) this exponentiation algorithm as the number of times the **while** loop is executed. This is also the number of times the statement $i := \lfloor i/2 \rfloor$ is executed. (Here we assume that the time interval for the computation of each $\lfloor i/2 \rfloor$ is independent of the magnitude of i.) On the basis of the foregoing two observations, we want to establish that for all $n \geq 1$, $g(n) \leq \log_2 n + 1 \in O(\log_2 n)$. We shall establish this by the Principle of Mathematical Induction (the alternative form — Theorem 4.2) on the value of n.

When $n = 1$, we see in Fig. 5.15 that i is odd, x is assigned the value of $a = a^1$, and a^1 is determined after only $1 = \log_2 1 + 1$ execution of the **while** loop. So $g(1) = 1 \leq \log_2 1 + 1$.

Now assume that for all $1 \leq n \leq k$, $g(n) \leq \log_2 n + 1$. Then for $n = k + 1$, during the first pass through the **while** loop the value of i is changed to $\left\lfloor \frac{k+1}{2} \right\rfloor$. Since $1 \leq \left\lfloor \frac{k+1}{2} \right\rfloor \leq k$, by the induction hypothesis we shall execute the **while** loop $g\left(\left\lfloor \frac{k+1}{2} \right\rfloor\right)$ more times, where $g\left(\left\lfloor \frac{k+1}{2} \right\rfloor\right) \leq \log_2 \left\lfloor \frac{k+1}{2} \right\rfloor + 1$.

Therefore

$$g(k+1) \leq 1 + \left[\log_2 \left\lfloor \frac{k+1}{2} \right\rfloor + 1\right] \leq 1 + \left[\log_2 \left(\frac{k+1}{2}\right) + 1\right]$$
$$= 1 + [\log_2(k+1) - \log_2 2 + 1] = \log_2(k+1) + 1.$$

For the time-complexity function of Example 5.72, we found that $f(n) \in O(n)$. Here we have $g(n) \in O(\log_2 n)$. It can be verified that g is dominated by f but f is not dominated by g. Therefore, *for large n*, this second algorithm is considered more efficient than the first algorithm (of Example 5.72). (However, note how much easier the pseudocode in Fig. 5.14 is than that of the procedure in Fig. 5.15.)

In closing this section, we shall summarize what we have learned by making the following observations.

1) The results we established in Examples 5.69, 5.70, 5.72, and 5.73 are useful when we are dealing with moderate to large values of n. For small values of n, such considerations about time-complexity functions have little purpose.
2) Suppose that algorithms A_1 and A_2 have time-complexity functions $f(n)$ and $g(n)$, respectively, where $f(n) \in O(n)$ and $g(n) \in O(n^2)$. We must be cautious here. We might expect an algorithm with linear complexity to be "perhaps more efficient" than one with quadratic complexity. But we really need more information. If $f(n) = 1000n$ and $g(n) = n^2$, then algorithm A_2 is fine until the problem size n exceeds 1000. If the problem size is such that we never exceed 1000, then algorithm A_2 is the better choice. However, as we mentioned in observation 1, as n grows larger, the algorithm of linear complexity becomes the better alternative.
3) In Fig. 5.17 we have graphed a log-linear plot for the functions associated with some of the orders given in Table 5.11. [Here we have replaced the (discrete) integer variable n by the (continuous) real variable n.] This should help us to develop some feeling for their relative growth rates (especially for large values of n).

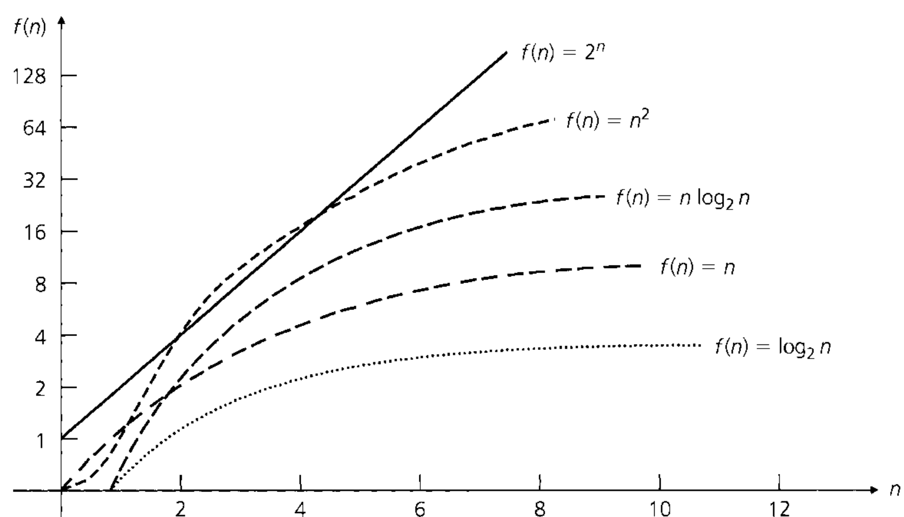


Figure 5.17

The data in Table 5.12 provide estimates of the running times of algorithms for certain orders of complexity. Here we have the problem sizes $n = 2$, 16, and 64, and we assume that the computer can perform one operation every 10^{-6} second = 1 microsecond (on the average). The entries in the table then estimate the running times in microseconds. For example, when the problem size is 16 and the order of complexity is $n \log_2 n$, then the running time is a very brief $16 \log_2 16 = 16 \cdot 4 = 64$ microseconds; for the order of complexity 2^n, the running time is 6.5×10^4 microseconds = 0.065 seconds. Since both of these time intervals are so short, it is difficult for a human to observe much of a difference in execution times. Results appear to be instantaneous in either case.

Table 5.12

	Order of Complexity					
Problem size n	$\log_2 n$	n	$n \log_2 n$	n^2	2^n	$n!$
2	1	2	2	4	4	2
16	4	16	64	256	6.5×10^4	2.1×10^{13}
64	6	64	384	4096	1.84×10^{19}	$>10^{89}$

However, such estimates can grow rather rapidly. For instance, suppose we run a program for which the input is an array A of n different integers. The results from this program are generated in two parts:

1) First the program implements an algorithm that determines the subsets of A of size 1. There are n such subsets.
2) Then a second algorithm is implemented to determine all the subsets of A. There are 2^n such subsets.

Let us assume that we have a computer that can determine each subset of A in a microsecond. For the case where $|A| = 64$, the first part of the output is executed almost instantaneously — in approximately 64 microseconds. For the second part, however, Table 5.12 indicates that the amount of time needed to determine all the subsets of A will be about 1.84×10^{19} microseconds. We cannot be too content with this result, however, since

$$1.84 \times 10^{19} \text{ microseconds} \doteq 2.14 \times 10^8 \text{ days} \doteq 5845 \text{ centuries}.$$

EXERCISES 5.8

1. In each of the following pseudocode program segments, the integer variables i, j, n, and *sum* are declared earlier in the program. The value of n (a positive integer) is supplied by the user prior to execution of the segment. In each case we define the time-complexity function $f(n)$ to be the number of times the statement *sum* := *sum* + 1 is executed. Determine the best "big-Oh" form for f.

a)
```
begin
  sum := 0
  for i := 1 to n do
    for j := 1 to n do
      sum := sum + 1
end
```

b)
```
begin
  sum := 0
  for i := 1 to n do
    for j := 1 to n * n do
      sum := sum + 1
end
```

c)
```
begin
  sum := 0;
  for i := 1 to n do
    for j := i to n do
      sum := sum + 1
end
```

d)
```
begin
  sum := 0
  i := n
```

```
    while i > 0 do
      begin
        sum := sum + 1
        i := ⌊i/2⌋
      end
  end
e) begin
    sum := 0
    for i := 1 to n do
      begin
        j := n
        while j > 0 do
          begin
            sum := sum + 1
            j := ⌊j/2⌋
          end
      end
  end
```

2. The following pseudocode procedure implements an algorithm for determining the maximum value in an array $a_1, a_2, a_3, \ldots, a_n$ of integers. Here $n \geq 2$ and the entries in the array need not be distinct.

```
procedure Maximum (n: integer;
  a1, a2, a3, ..., an: integers)
begin
  max := a1
  for i := 2 to n do
    if ai > max then
      max := ai
end
```

a) If the worst-case complexity function $f(n)$ for this segment is determined by the number of times the comparison $a_i > max$ is executed, find the appropriate "big-Oh" form for f.

b) What can we say about the best-case and average-case complexities for this implementation?

3. a) Write a computer program (or develop an algorithm) to locate the first occurrence of the maximum value in an array $a_1, a_2, a_3, \ldots, a_n$ of integers. (Here $n \in \mathbf{Z}^+$ and the entries in the array need not be distinct.)

b) Determine the worst-case complexity function for the implementation developed in part (a).

4. a) Write a computer program (or develop an algorithm) to determine the minimum and maximum values in an array $a_1, a_2, a_3, \ldots, a_n$ of integers. (Here $n \in \mathbf{Z}^+$ with $n \geq 2$, and the entries in the array need not be distinct.)

b) Determine the worst-case complexity function for the implementation developed in part (a).

5. The following pseudocode procedure can be used to evaluate the polynomial

$$8 - 10x + 7x^2 - 2x^3 + 3x^4 + 12x^5,$$

when x is replaced by an arbitrary (but fixed) real number r.

For this particular instance, $n = 5$ and $a_0 = 8$, $a_1 = -10$, $a_2 = 7$, $a_3 = -2$, $a_4 = 3$, and $a_5 = 12$.

```
procedure PolynomialEvaluation1
  (n: nonnegative integer;
  r, a0, a1, a2, ..., an: real)
begin
  product := 1.0
  value := a0
  for i := 1 to n do
    begin
      product := product * r
      value := value + ai * product
    end
end
```

a) How many additions take place in the evaluation of the given polynomial? (Do not include the $n - 1$ additions needed to increment the loop variable i.) How many multiplications?

b) Answer the questions in part (a) for the general polynomial

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_{n-1}x^{n-1} + a_nx^n,$$

where $a_0, a_1, a_2, a_3, \ldots, a_{n-1}, a_n$ are real numbers and n is a positive integer.

6. We first note how the polynomial in the previous exercise can be written in the *nested multiplication method*:

$$8 + x(-10 + x(7 + x(-2 + x(3 + 12x)))).$$

Using this representation, the following pseudocode procedure (implementing *Horner's method*) can be used to evaluate the given polynomial.

```
procedure PolynomialEvaluation2
  (n: nonnegative integer;
  r, a0, a1, a2, ..., an: real)
begin
  value := an
  for j := n - 1 down to 0 do
    value := aj + r * value
end
```

Answer the questions in parts (a) and (b) of Exercise 5 for the new procedure given here.

7. Let $a_1, a_2, a_3, \ldots$ be the integer sequence defined recursively by

1) $a_1 = 0$; and

2) For $n > 1$, $a_n = 1 + a_{\lfloor n/2 \rfloor}$.

Prove that $a_n = \lfloor \log_2 n \rfloor$ for all $n \in \mathbf{Z}^+$.

8. Let $a_1, a_2, a_3, \ldots$ be the integer sequence defined recursively by

1) $a_1 = 0$; and
2) For $n > 1$, $a_n = 1 + a_{\lceil n/2 \rceil}$.

Find an explicit formula for a_n and prove that your formula is correct.

9. Suppose the probability that the integer *key* is in the array $a_1, a_2, a_3, \ldots, a_n$ (of n distinct integers) is $3/4$ and that each array element has the same probability of containing this value. If the linear search algorithm of Example 5.70 is applied to this array and value of *key*, what is the average number of array elements that are examined?

10. When the linear search algorithm is applied to the array $a_1, a_2, a_3, \ldots, a_n$ (of n distinct integers) for the integer *key*, suppose the probability that *key* has the value a_i is $i/[n(n+1)]$, for $1 \le i \le n$. Under these circumstances, what is the average number of array elements examined?

11. a) Write a computer program (or develop an algorithm) to determine the location of the first entry in an array $a_1, a_2, a_3, \ldots, a_n$ of integers that repeats a previous entry in the array.

b) Determine the worst-case complexity for the implementation developed in part (a).

12. a) Write a computer program (or develop an algorithm) to determine the location of the first entry a_i in an array $a_1, a_2, a_3, \ldots, a_n$ of integers, where $a_i < a_{i-1}$.

b) Determine the worst-case complexity for the implementation developed in part (a).

5.9 Summary and Historical Review

In this chapter we developed the function concept, which is of great importance in all areas of mathematics. Although we were primarily concerned with finite functions, the definition applies equally well to infinite sets and includes the functions of trigonometry and calculus. However, we did emphasize the role of a finite function when we transformed a finite set into a finite set. In this setting, computer output (that terminates) can be thought of as a function of computer input, and a compiler can be regarded as a function that transforms a (source) program into a set of machine-language instructions (object program).

The actual word *function*, in its Latin form, was introduced in 1694 by Gottfried Wilhelm Leibniz (1646–1716) to denote a quantity associated with a curve (such as the slope of the curve or the coordinates of a point of the curve). By 1718, under the direction of Johann Bernoulli (1667–1748), a function was regarded as an algebraic expression made up of constants and a variable. Equations or formulas involving constants and variables

Gottfried Wilhelm Leibniz (1646–1716)

came later with Leonhard Euler (1707–1783). His is the definition of "function" generally found in high school mathematics. Also, in about 1734, we find in the work of Euler and Alexis Clairaut (1713–1765) the notation $f(x)$, which is still in use today.

Euler's idea remained intact until the time of Jean Baptiste Joseph Fourier (1768–1830), who found the need for a more general type of function in his investigation of trigonometric series. In 1837, Peter Gustav Lejeune Dirichlet (1805–1859) set down a more rigorous formulation of the concepts of variable, function, and the correspondence between the independent variable x and the dependent variable y, when $y = f(x)$. Dirichlet's work emphasized the relationship between two sets of numbers and did not call for the existence of a formula or expression connecting the two sets. With the developments in set theory during the nineteenth and twentieth centuries came the generalization of the function as a particular type of relation.

Peter Gustav Lejeune Dirichlet (1805–1859)

In addition to his fundamental work on the definition of a function, Dirichlet was also quite active in applied mathematics and in number theory, where he found need for, and was the first to formally state, the pigeonhole principle. Consequently, this principle is sometimes referred to as the Dirichlet drawer principle or the Dirichlet box principle.

The nineteenth and twentieth centuries saw the use of the special function, one-to-one correspondence, in the study of the infinite. In about 1888, Richard Dedekind (1831–1916) defined an infinite set as one that can be placed into a one-to-one correspondence with a proper subset of itself. [Galileo (1564–1642) had observed this for the set $\mathbf{Z}^+$.] Two infinite sets that could be placed in a one-to-one correspondence with each other were said to have the same *transfinite cardinal number*. In a series of articles, Georg Cantor (1845–1918) developed the idea of levels of infinity and showed that $|\mathbf{Z}| = |\mathbf{Q}|$ but $|\mathbf{Z}| < |\mathbf{R}|$. A set A with $|A| = |\mathbf{Z}|$ is called *countable*, or *denumerable*, and we write $|\mathbf{Z}| = \aleph_0$ as Cantor did, using the Hebrew letter aleph, with the subscripted 0, to denote the first level of infinity. To show that $|\mathbf{Z}| < |\mathbf{R}|$, or that the real numbers were *uncountable*, Cantor devised a technique now referred to as the Cantor diagonal method. (More about the theory of countable and uncountable sets can be found in Appendix 3.)

The Stirling numbers of the second kind (in Section 5.3) are named in honor of James Stirling (1692–1770), a pioneer in the development of generating functions, a topic we will investigate later in the text. These numbers appear in his work *Methodus Differentialis*, published in London in 1730. Stirling was an associate of Sir Isaac Newton (1642–1727) and was using the Maclaurin series in his work 25 years before Colin Maclaurin (1698–1746). However, although his name is not attached to this series, it appears in the approximation known as Stirling's formula: $n! \doteq (2\pi n)^{1/2}e^{-n}n^n$, which, as justice would have it, was actually developed by Abraham DeMoivre (1667–1754).

Using the counting principles developed in Section 5.3, the results in Table 5.13 extend the ideas that were summarized in Table 1.11. Here we count the number of ways it is possible to distribute m objects into n containers, under the conditions prescribed in the first three columns of the table. (The cases wherein neither the objects nor the containers are distinct will be covered in Chapter 9.)

Table 5.13

Objects Are Distinct	Containers Are Distinct	Some Container(s) May Be Empty	Number of Distributions
Yes	Yes	Yes	n^m
Yes	Yes	No	$n!\ S(m, n)$
Yes	No	Yes	$S(m, 1) + S(m, 2) + \cdots + S(m, n)$
Yes	No	No	$S(m, n)$
No	Yes	Yes	$\binom{n+m-1}{m}$
No	Yes	No	$\binom{n+(m-n)-1}{(m-n)} = \binom{m-1}{m-n}$ $= \binom{m-1}{n-1}$

Finally, the "big-Oh" notation of Section 5.7 was introduced by Paul Gustav Heinrich Bachmann (1837–1920) in his book *Analytische Zahlentheorie*, an important work on number theory, published in 1892. This notation has become prominent in approximation theory, in such areas as numerical analysis and the analysis of algorithms. In general, the notation $f \in O(g)$ denotes that we do not know the function f explicitly but do know an upper bound on its order of magnitude. The "big-Oh" symbol is sometimes referred to as the Landau symbol, in honor of Edmund Landau (1877–1938), who used this notation throughout his work.

Further properties of the Stirling numbers of the second kind are given in Chapter 4 of D. I. A. Cohen [3] and in Chapter 6 of the text by R. L. Graham, D. E. Knuth, and O. Patashnik [7]. The article by D. J. Velleman and G. S. Call [11] provides a very interesting introduction to the Stirling numbers of the second kind — as well as the Eulerian numbers introduced in Example 4.21. For more on infinite sets and the work of Georg Cantor, consult Chapter 8 of H. Eves and C. V. Newsom [6] or Chapter IV of R. L. Wilder [12]. The presentation in the book by J. W. Dauben [5] covers the controversy surrounding set theory at the turn of the century and shows how certain aspects of Cantor's personal life played such an integral part in his understanding and defense of set theory.

More examples that demonstrate how to apply the pigeonhole principle are given in the articles by K. R. Rebman [9] and A. Soifer and E. Lozansky [10]. Further results and extensions on problems arising from the principle are covered in the article by D. S. Clark and J. T. Lewis [2]. During the twentieth century a great deal of research has been devoted to generalizations of the pigeonhole principle, culminating in the subject of Ramsey theory, named for Frank Plumpton Ramsey (1903–1930). An interesting introduction to Ramsey theory can be found in Chapter 5 of D. I. A. Cohen [3]. The text by R. L. Graham, B. L. Rothschild, and J. H. Spencer [8] provides further worthwhile information.

Extensive coverage on the topic of relational data bases can be found in the work of C. J. Date [4]. Finally, the text by S. Baase and A. Van Gelder [1] is an excellent place to continue the study of the analysis of algorithms.

REFERENCES

1. Baase, Sara, and Van Gelder, Allen. *Computer Algorithms: Introduction to Design & Analysis*, 3rd ed. Reading, Mass.: Addison-Wesley, 2000.
2. Clark, Dean S., and Lewis, James T. "Herbert and the Hungarian Mathematician: Avoiding Certain Subsequence Sums." *The College Mathematics Journal* 21 (March 1990): pp. 100–104.
3. Cohen, Daniel I. A. *Basic Techniques of Combinatorial Theory*. New York: Wiley, 1978.
4. Date, C. J. *An Introduction to Database Systems*, 7th ed. Boston, Mass.: Addison-Wesley, 2002.
5. Dauben, Joseph Warren. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Lawrenceville, N. J.: Princeton University Press, 1990.
6. Eves, Howard, and Newsom, Carroll V. *An Introduction to the Foundations and Fundamental Concepts of Mathematics*, rev. ed. New York: Holt, 1965.
7. Graham, Ronald L., Knuth, Donald E., and Patashnik, Oren. *Concrete Mathematics*, 2nd ed. Reading, Mass.: Addison-Wesley, 1994.
8. Graham, Ronald L., Rothschild, Bruce L., and Spencer, Joel H. *Ramsey Theory*, 2nd ed. New York: Wiley, 1980.
9. Rebman, Kenneth R. "The Pigeonhole Principle (What it is, how it works, and how it applies to map coloring)." *The Two-Year College Mathematics Journal*, vol. 10, no. 1 (January 1979): pp. 3–13.
10. Soifer, Alexander, and Lozansky, Edward, "Pigeons in Every Pigeonhole." *Quantum* (January 1990): pp. 25–26, 32.
11. Velleman, Daniel J., and Call, Gregory S. "Permutations and Combination Locks." *Mathematics Magazine* 68 (October 1995): pp. 243–253.
12. Wilder, Raymond L. *Introduction to the Foundations of Mathematics*, 2nd ed. New York: Wiley, 1965.

SUPPLEMENTARY EXERCISES

1. Let $A, B \subseteq \mathscr{U}$. Prove that

a) $(A \times B) \cap (B \times A) = (A \cap B) \times (A \cap B)$; and

b) $(A \times B) \cup (B \times A) \subseteq (A \cup B) \times (A \cup B)$.

2. Determine whether each of the following statements is true or false. For each false statement give a counterexample.

a) If $f: A \to B$ and $(a, b), (a, c) \in f$, then $b = c$.

b) If $f: A \to B$ is a one-to-one correspondence and A, B are finite, then $A = B$.

c) If $f: A \to B$ is one-to-one, then f is invertible.

d) If $f: A \to B$ is invertible, then f is one-to-one.

e) If $f: A \to B$ is one-to-one and $g, h: B \to C$ with $g \circ f = h \circ f$, then $g = h$.

f) If $f: A \to B$ and $A_1, A_2 \subseteq A$, then $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$.

g) If $f: A \to B$ and $B_1, B_2 \subseteq B$, then $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

3. Let $f: \mathbf{R} \to \mathbf{R}$ where $f(ab) = af(b) + bf(a)$, for all $a, b \in \mathbf{R}$. (a) What is $f(1)$? (b) What is $f(0)$? (c) If $n \in \mathbf{Z}^+$, $a \in \mathbf{R}$, prove that $f(a^n) = na^{n-1}f(a)$.

4. Let $A, B \subseteq \mathbf{N}$ with $1 < |A| < |B|$. If there are 262,144 relations from A to B, determine all possibilities for $|A|$ and $|B|$.

5. If $\mathcal{U}_1$, $\mathcal{U}_2$ are universal sets with $A, B \subseteq \mathcal{U}_1$, and $C, D \subseteq \mathcal{U}_2$, prove that $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$.

6. Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 3, 4, 5, 6\}$. How many one-to-one functions $f: A \to B$ satisfy (a) $f(1) = 3$? (b) $f(1) = 3$, $f(2) = 6$?

7. Determine all real numbers x for which

$$x^2 - \lfloor x \rfloor = 1/2.$$

8. Let $\mathcal{R} \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+$ be the relation given by the following recursive definition.

1) $(1, 1) \in \mathcal{R}$; and

2) For all $(a, b) \in \mathcal{R}$, the three ordered pairs $(a + 1, b)$, $(a + 1, b + 1)$, and $(a + 1, b + 2)$ are also in $\mathcal{R}$.

Prove that $2a \geq b$ for all $(a, b) \in \mathcal{R}$.

9. Let a, b denote fixed real numbers and suppose that $f: \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = a(x + b) - b$, $x \in \mathbf{R}$. (a) Determine $f^2(x)$ and $f^3(x)$. (b) Conjecture a formula for $f^n(x)$, where $n \in \mathbf{Z}^+$. Now establish the validity of your conjecture.

10. Let A_1, A and B be sets with $\{1, 2, 3, 4, 5\} = A_1 \subset A$, $B = \{s, t, u, v, w, x\}$, and $f: A_1 \to B$. If f can be extended to A in 216 ways, what is $|A|$?

11. Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{t, u, v, w, x, y, z\}$. (a) If a function $f: A \to B$ is randomly generated, what is the probability that it is one-to-one? (b) Write a computer program (or develop an algorithm) to generate random functions $f: A \to B$ and have the program print out how many functions it generates until it generates one that is one-to-one.

12. Let S be a set of seven positive integers the maximum of which is at most 24. Prove that the sums of the elements in all the nonempty subsets of S cannot be distinct.

13. In a ten-day period Ms. Rosatone typed 84 letters to different clients. She typed 12 of these letters on the first day, seven on the second day, and three on the ninth day, and she finished the last eight on the tenth day. Show that for a period of three consecutive days Ms. Rosatone typed at least 25 letters.

14. If $\{x_1, x_2, \ldots, x_7\} \subseteq \mathbf{Z}^+$, show that for some $i \neq j$, either $x_i + x_j$ or $x_i - x_j$ is divisible by 10.

15. Let $n \in \mathbf{Z}^+$, n odd. If $i_1, i_2, \ldots, i_n$ is a permutation of the integers $1, 2, \ldots, n$, prove that $(1 - i_1)(2 - i_2) \cdots (n - i_n)$ is an even integer. (Which counting principle is at work here?)

16. With both of their parents working, Thomas, Stuart, and Craig must handle ten weekly chores among themselves. (a) In how many ways can they divide up the work so that everyone is responsible for at least one chore? (b) In how many ways can the chores be assigned if Thomas, as the eldest, must mow the lawn (one of the ten weekly chores) and no one is allowed to be idle?

17. Let $n \in \mathbf{N}$, $n \geq 2$. Show that $S(n, 2) = 2^{n-1} - 1$.

18. Mrs. Blasi has five sons (Michael, Rick, David, Kenneth, and Donald) who enjoy reading books about sports. With Christmas approaching, she visits a bookstore where she finds 12 different books on sports.

a) In how many ways can she select nine of these books?

b) Having made her purchase, in how many ways can she distribute the books among her sons so that each of them gets at least one book?

c) Two of the nine books Mrs. Blasi purchased deal with basketball, Donald's favorite sport. In how many ways can she distribute the books among her sons so that Donald gets at least the two books on basketball?

19. Let $m, n \in \mathbf{Z}^+$ with $n \geq m$. (a) In how many ways can one distribute n distinct objects among m different containers with no container left empty? (b) In the expansion of $(x_1 + x_2 + \cdots + x_m)^n$, what is the sum of all the multinomial coefficients $\binom{n}{n_1, n_2, \ldots, n_m}$ where $n_1 + n_2 + \cdots + n_m = n$ and $n_i > 0$ for all $1 \leq i \leq m$?

20. If $n \in \mathbf{Z}^+$ with $n \geq 4$, verify that $S(n, n-2) = \binom{n}{3} + 3\binom{n}{4}$.

21. If $f: A \to A$, prove that for all $m, n \in \mathbf{Z}^+$, $f^m \circ f^n = f^n \circ f^m$. (First let $m = 1$ and induct on n. Then induct on m. This technique is known as *double induction*.)

22. Let $f: X \to Y$, and for each $i \in I$, let $A_i \subseteq X$. Prove that

a) $f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i)$.

b) $f\left(\bigcap_{i \in I} A_i\right) \subseteq \bigcap_{i \in I} f(A_i)$.

c) $f\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f(A_i)$, for f one-to-one.

23. Given a nonempty set A, let $f: A \to A$ and $g: A \to A$ where

$$f(a) = g(f(f(a))) \quad \text{and} \quad g(a) = f(g(f(a)))$$

for all a in A. Prove that $f = g$.

24. Let A be a set with $|A| = n$.

a) How many closed binary operations are there on A?

b) A closed ternary (3-ary) operation on A is a function $f: A \times A \times A \to A$. How many closed ternary operations are there on A?

c) A closed k-ary operation on A is a function $f: A_1 \times A_2 \times \cdots \times A_k \to A$, where $A_i = A$, for all $1 \leq i \leq k$. How many closed k-ary operations are there on A?

d) A closed k-ary operation for A is called *commutative* if

$$f(a_1, a_2, \ldots, a_k) = f(\pi(a_1), \pi(a_2), \ldots, \pi(a_k)),$$

where $a_1, a_2, \ldots, a_k \in A$ (repetitions allowed), and

$\pi(a_1), \pi(a_2), \ldots, \pi(a_k)$ is any rearrangement of $a_1, a_2, \ldots, a_k$. How many of the closed k-ary operations on A are commutative?

25. a) Let $S = \{2, 16, 128, 1024, 8192, 65536\}$. If four numbers are selected from S, prove that two of them must have the product 131072.

b) Generalize the result in part (a).

26. If $\mathcal{U}$ is a universe and $A \subseteq \mathcal{U}$, we define the *characteristic function* of A by $\chi_A: \mathcal{U} \to \{0, 1\}$, where

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

For sets $A, B \subseteq \mathcal{U}$, prove each of the following:

a) $\chi_{A \cap B} = \chi_A \cdot \chi_B$, where $(\chi_A \cdot \chi_B)(x) = \chi_A(x) \cdot \chi_B(x)$

b) $\chi_{A \cup B} = \chi_A + \chi_B - \chi_{A \cap B}$

c) $\chi_{\overline{A}} = 1 - \chi_A$, where $(1 - \chi_A)(x) = 1(x) - \chi_A(x) = 1 - \chi_A(x)$

(For $\mathcal{U}$ finite, placing the elements of $\mathcal{U}$ in a fixed order results in a one-to-one correspondence between subsets A of $\mathcal{U}$ and the arrays of 0's and 1's obtained as the images of $\mathcal{U}$ under χ_A. These arrays can then be used for the computer storage and manipulation of certain subsets of $\mathcal{U}$.)

27. With $A = \{x, y, z\}$, let $f, g: A \to A$ be given by $f = \{(x, y), (y, z), (z, x)\}$, $g = \{(x, y), (y, x), (z, z)\}$. Determine each of the following: $f \circ g$, $g \circ f$, f^{-1}, g^{-1}, $(g \circ f)^{-1}$, $f^{-1} \circ g^{-1}$, and $g^{-1} \circ f^{-1}$.

28. a) If $f: \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = 5x + 3$, find $f^{-1}(8)$.

b) If $g: \mathbf{R} \to \mathbf{R}$, where $g(x) = |x^2 + 3x + 1|$, find $g^{-1}(1)$.

c) For $h: \mathbf{R} \to \mathbf{R}$, given by

$$h(x) = \left|\frac{x}{x+2}\right|,$$

find $h^{-1}(4)$.

29. If $A = \{1, 2, 3, \ldots, 10\}$, how many functions $f: A \to A$ (simultaneously) satisfy $f^{-1}(\{1, 2, 3\}) = \emptyset$, $f^{-1}(\{4, 5\}) = \{1, 3, 7\}$, and $f^{-1}(\{8, 10\}) = \{8, 10\}$?

30. Let $f: A \to A$ be an invertible function. For $n \in \mathbf{Z}^+$ prove that $(f^n)^{-1} = (f^{-1})^n$. [This result can be used to define f^{-n} as either $(f^n)^{-1}$ or $(f^{-1})^n$.]

31. In certain programming languages, the functions pred and succ (for predecessor and successor, respectively) are functions from $\mathbf{Z}$ to $\mathbf{Z}$ where $\text{pred}(x) = \pi(x) = x - 1$ and $\text{succ}(x) = \sigma(x) = x + 1$.

a) Determine $(\pi \circ \sigma)(x)$, $(\sigma \circ \pi)(x)$.

b) Determine π^2, π^3, $\pi^n (n \geq 2)$, σ^2, σ^3, $\sigma^n (n \geq 2)$.

c) Determine π^{-2}, π^{-3}, π^{-n} $(n \geq 2)$, σ^{-2}, σ^{-3}, $\sigma^{-n} (n \geq 2)$, where, for example, $\sigma^{-2} = \sigma^{-1} \circ \sigma^{-1} = (\sigma \circ \sigma)^{-1} = (\sigma^2)^{-1}$. (See Supplementary Exercise 30.)

32. For $n \in \mathbf{Z}^+$, define $\tau: \mathbf{Z}^+ \to \mathbf{Z}^+$ by $\tau(n) =$ the number of positive-integer divisors of n.

a) Let $n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \ldots p_k^{e_k}$, where $p_1, p_2, p_3, \ldots, p_k$ are distinct primes and e_i is a positive integer for all $1 \leq i \leq k$. What is $\tau(n)$?

b) Determine the three smallest values of $n \in \mathbf{Z}^+$ for which $\tau(n) = k$, where $k = 2, 3, 4, 5, 6$.

c) For all $k \in \mathbf{Z}^+$, $k > 1$, prove that $\tau^{-1}(k)$ is infinite.

d) If $a, b \in \mathbf{Z}^+$ with $\gcd(a, b) = 1$, prove that $\tau(ab) = \tau(a)\tau(b)$.

33. a) How many subsets $A = \{a, b, c, d\} \subseteq \mathbf{Z}^+$, where $a, b, c, d > 1$, satisfy the property $a \cdot b \cdot c \cdot d = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$?

b) How many subsets $A = \{a_1, a_2, \ldots, a_m\} \subseteq \mathbf{Z}^+$, where $a_i > 1$, $1 \leq i \leq m$, satisfy the property $\prod_{i=1}^{m} a_i = \prod_{j=1}^{n} p_j$, where the p_j, $1 \leq j \leq n$, are distinct primes and $n \geq m$?

34. Give an example of a function $f: \mathbf{Z}^+ \to \mathbf{R}$ where $f \in O(1)$ and f is one-to-one. (Hence f is not constant.)

35. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ where

$$f(n) = \begin{cases} 2, & \text{for } n \text{ even} \\ 1, & \text{for } n \text{ odd} \end{cases} \qquad g(n) = \begin{cases} 3, & \text{for } n \text{ even} \\ 4, & \text{for } n \text{ odd} \end{cases}$$

Prove or disprove each of the following: (a) $f \in O(g)$; and (b) $g \in O(f)$.

36. For $f, g: \mathbf{Z}^+ \to \mathbf{R}$ we define $f + g: \mathbf{Z}^+ \to \mathbf{R}$ by $(f + g)(n) = f(n) + g(n)$, for $n \in \mathbf{Z}^+$. [*Note*: The plus sign in $f + g$ is for the addition of the functions f and g, while the plus sign in $f(n) + g(n)$ is for the addition of the real numbers $f(n)$ and $g(n)$.]

a) Let $f_1, g_1: \mathbf{Z}^+ \to \mathbf{R}$ with $f \in O(f_1)$ and $g \in O(g_1)$. If $f_1(n) \geq 0$, $g_1(n) \geq 0$, for all $n \in \mathbf{Z}^+$, prove that $(f + g) \in O(f_1 + g_1)$.

b) If the conditions $f_1(n) \geq 0$, $g_1(n) \geq 0$, for all $n \in \mathbf{Z}^+$, are not satisfied, as in part (a), provide a counterexample to show that

$$f \in O(f_1),\ g \in O(g_1) \not\Rightarrow (f + g) \in O(f_1 + g_1).$$

37. Let $a, b \in \mathbf{R}^+$, with $a, b > 1$. Let $f, g: \mathbf{Z}^+ \to \mathbf{R}$ be defined by $f(n) = \log_a n$, $g(n) = \log_b n$. Prove that $f \in O(g)$ and $g \in O(f)$. [Hence $O(\log_a n) = O(\log_b n)$.]

7

Relations: The Second Time Around

In Chapter 5 we introduced the concept of a (binary) relation. Returning to relations in this chapter, we shall emphasize the study of relations on a set A — that is, subsets of $A \times A$. Within the theory of languages and finite state machines from Chapter 6, we find many examples of relations on a set A, where A represents a set of strings from a given alphabet or a set of internal states from a finite state machine. Various properties of relations are developed, along with ways to represent finite relations for computer manipulation. Directed graphs reappear as a way to represent such relations. Finally, two types of relations on a set A are especially important: equivalence relations and partial orders. Equivalence relations, in particular, arise in many areas of mathematics. For the present we shall use an equivalence relation on the set of internal states in a finite state machine M in order to find a machine M_1, with as few internal states as possible, that performs whatever tasks M is capable of performing. The procedure is known as the minimization process.

7.1 Relations Revisited: Properties of Relations

We start by recalling some fundamental ideas considered earlier.

Definition 7.1 For sets A, B, any subset of $A \times B$ is called a (*binary*) *relation* from A to B. Any subset of $A \times A$ is called a (*binary*) *relation* on A.

As mentioned in the sentence following Definition 5.2, our primary concern is with binary relations. Consequently, for us the word "relation" will once again mean binary relation, unless something otherwise is specified.

EXAMPLE 7.1

a) Define the relation $\mathcal{R}$ on the set $\mathbf{Z}$ by $a \mathcal{R} b$, or $(a, b) \in \mathcal{R}$, if $a \leq b$. This subset of $\mathbf{Z} \times \mathbf{Z}$ is the ordinary "less than or equal to" relation on the set $\mathbf{Z}$, and it can also be defined on $\mathbf{Q}$ or $\mathbf{R}$, but not on $\mathbf{C}$.

b) Let $n \in \mathbf{Z}^+$. For $x, y \in \mathbf{Z}$, the *modulo n relation* $\mathcal{R}$ is defined by $x \mathcal{R} y$ if $x - y$ is a multiple of n. With $n = 7$, we find, for instance, that $9 \mathcal{R} 2$, $-3 \mathcal{R} 11$, $(14, 0) \in \mathcal{R}$, but $3 \not\mathcal{R} 7$ (that is, 3 is *not* related to 7).

c) For the universe $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7\}$ consider the (fixed) set $C \subseteq \mathcal{U}$ where $C = \{1, 2, 3, 6\}$. Define the relation $\mathcal{R}$ on $\mathcal{P}(\mathcal{U})$ by $A \mathcal{R} B$ when $A \cap C = B \cap C$. Then the sets $\{1, 2, 4, 5\}$ and $\{1, 2, 5, 7\}$ are related since $\{1, 2, 4, 5\} \cap C = \{1, 2\} = \{1, 2, 5, 7\} \cap C$. Likewise we find that $X = \{4, 5\}$ and $Y = \{7\}$ are so related because $X \cap C = \emptyset = Y \cap C$. However, the sets $S = \{1, 2, 3, 4, 5\}$ and $T = \{1, 2, 3, 6, 7\}$ are *not* related — that is, $S \not\mathcal{R} T$ — since $S \cap C = \{1, 2, 3\} \neq \{1, 2, 3, 6\} = T \cap C$.

EXAMPLE 7.2

Let Σ be an alphabet, with language $A \subseteq \Sigma^*$. For $x, y \in A$, define $x \mathcal{R} y$ if x is a prefix of y. Other relations can be defined on A by replacing "prefix" with either "suffix" or "substring."

EXAMPLE 7.3

Consider a finite state machine $M = (S, \mathcal{I}, \mathcal{O}, \nu, \omega)$.

a) For $s_1, s_2 \in S$, define $s_1 \mathcal{R} s_2$ if $\nu(s_1, x) = s_2$, for some $x \in \mathcal{I}$. Relation $\mathcal{R}$ establishes the *first level of reachability*.

b) The relation for the *second level of reachability* can also be given for S. Here $s_1 \mathcal{R} s_2$ if $\nu(s_1, x_1x_2) = s_2$, for some $x_1x_2 \in \mathcal{I}^2$. This can be extended to higher levels if the need arises. For the general *reachability* relation we have $\nu(s_1, y) = s_2$, for some $y \in \mathcal{I}^*$.

c) Given $s_1, s_2 \in S$ the relation of *1-equivalence*, which is denoted by $s_1 \mathrm{E}_1 s_2$ and is read "s_1 is 1-equivalent to s_2", is defined when $\omega(s_1, x) = \omega(s_2, x)$ for all $x \in \mathcal{I}$. Consequently, $s_1 \mathrm{E}_1 s_2$ indicates that if machine M starts in either state s_1 or s_2, the output is the same for each element of $\mathcal{I}$. This idea can be extended to states being *k-equivalent*, where we write $s_1 \mathrm{E}_k s_2$ if $\omega(s_1, y) = \omega(s_2, y)$, for all $y \in \mathcal{I}^k$. Here the same output string is obtained for each input string in $\mathcal{I}^k$ if we start at either s_1 or s_2.

If two states are k-equivalent for all $k \in \mathbf{Z}^+$, then they are called *equivalent*. We shall look further into this idea later in the chapter.

We now start to examine some of the properties a relation can satisfy.

Definition 7.2

A relation $\mathcal{R}$ on a set A is called *reflexive* if for all $x \in A$, $(x, x) \in \mathcal{R}$.

To say that a relation $\mathcal{R}$ is reflexive simply means that each element x of A is related to itself. All the relations in Examples 7.1 and 7.2 are reflexive. The general reachability relation in Example 7.3(b) and all of the relations mentioned in part (c) of that example are also reflexive. [What goes wrong with the relations for the first and second levels of reachability given in parts (a) and (b) of Example 7.3?]

EXAMPLE 7.4

For $A = \{1, 2, 3, 4\}$, a relation $\mathcal{R} \subseteq A \times A$ will be reflexive if and only if $\mathcal{R} \supseteq \{(1, 1), (2, 2), (3, 3), (4, 4)\}$. Consequently, $\mathcal{R}_1 = \{(1, 1), (2, 2), (3, 3)\}$ is not a reflexive relation on A, whereas $\mathcal{R}_2 = \{(x, y) | x, y \in A, x \leq y\}$ is reflexive on A.

EXAMPLE 7.5

Given a finite set A with $|A| = n$, we have $|A \times A| = n^2$, so there are 2^{n^2} relations on A. How many of these are reflexive?

If $A = \{a_1, a_2, \ldots, a_n\}$, a relation $\mathcal{R}$ on A is reflexive if and only if $\{(a_i, a_i) | 1 \leq i \leq n\} \subseteq \mathcal{R}$. Considering the other $n^2 - n$ ordered pairs in $A \times A$ [those of the form (a_i, a_j), where $i \neq j$ for $1 \leq i, j \leq n$] as we construct a reflexive relation $\mathcal{R}$ on A, we either include or exclude each of these ordered pairs, so by the rule of product there are $2^{(n^2-n)}$ reflexive relations on A.

Definition 7.3

Relation $\mathcal{R}$ on set A is called *symmetric* if $(x, y) \in \mathcal{R} \Rightarrow (y, x) \in \mathcal{R}$, for all $x, y \in A$.

EXAMPLE 7.6

With $A = \{1, 2, 3\}$, we have:

a) $\mathcal{R}_1 = \{(1, 2), (2, 1), (1, 3), (3, 1)\}$, a symmetric, but not reflexive, relation on A;

b) $\mathcal{R}_2 = \{(1, 1), (2, 2), (3, 3), (2, 3)\}$, a reflexive, but not symmetric, relation on A;

c) $\mathcal{R}_3 = \{(1, 1), (2, 2), (3, 3)\}$ and $\mathcal{R}_4 = \{(1, 1), (2, 2), (3, 3), (2, 3), (3, 2)\}$, two relations on A that are both reflexive and symmetric; and

d) $\mathcal{R}_5 = \{(1, 1), (2, 3), (3, 3)\}$, a relation on A that is neither reflexive nor symmetric.

To count the symmetric relations on $A = \{a_1, a_2, \ldots, a_n\}$, we write $A \times A$ as $A_1 \cup A_2$, where $A_1 = \{(a_i, a_i) | 1 \leq i \leq n\}$ and $A_2 = \{(a_i, a_j) | 1 \leq i, j \leq n, i \neq j\}$, so that every ordered pair in $A \times A$ is in exactly one of A_1, A_2. For A_2, $|A_2| = |A \times A| - |A_1| = n^2 - n = n(n-1)$, an even integer. The set A_2 contains $(1/2)(n^2 - n)$ subsets S_{ij} of the form $\{(a_i, a_j), (a_j, a_i)\}$ where $1 \leq i < j \leq n$. In constructing a symmetric relation $\mathcal{R}$ on A, for each ordered pair in A_1 we have our usual choice of exclusion or inclusion. For each of the $(1/2)(n^2 - n)$ subsets $S_{ij} (1 \leq i < j \leq n)$ taken from A_2 we have the same two choices. So by the rule of product there are $2^n \cdot 2^{(1/2)(n^2-n)} = 2^{(1/2)(n^2+n)}$ symmetric relations on A.

In counting those relations on A that are both reflexive and symmetric, we have only one choice for each ordered pair in A_1. So we have $2^{(1/2)(n^2-n)}$ relations on A that are both reflexive and symmetric.

Definition 7.4

For a set A, a relation $\mathcal{R}$ on A is called *transitive* if, for all $x, y, z \in A$, $(x, y), (y, z) \in \mathcal{R} \Rightarrow (x, z) \in \mathcal{R}$. (So if x "is related to" y, and y "is related to" z, we want x "related to" z, with y playing the role of "intermediary.")

EXAMPLE 7.7

All the relations in Examples 7.1 and 7.2 are transitive, as are the relations in Example 7.3(c).

EXAMPLE 7.8

Define the relation $\mathcal{R}$ on the set $\mathbf{Z}^+$ by $a \mathcal{R} b$ if a (exactly) divides b—that is, $b = ca$ for some $c \in \mathbf{Z}^+$. Now if $x \mathcal{R} y$ and $y \mathcal{R} z$, do we have $x \mathcal{R} z$? We know that $x \mathcal{R} y \Rightarrow y = sx$ for some $s \in \mathbf{Z}^+$ and $y \mathcal{R} z \Rightarrow z = ty$ where $t \in \mathbf{Z}^+$. Consequently, $z = ty = t(sx) = (ts)x$ for $ts \in \mathbf{Z}^+$, so $x \mathcal{R} z$ and $\mathcal{R}$ is transitive. In addition, $\mathcal{R}$ is reflexive, but not symmetric, because, for example, $2 \mathcal{R} 6$ but $6 \not\mathcal{R} 2$.

EXAMPLE 7.9

Consider the relation $\mathcal{R}$ on the set $\mathbf{Z}$ where we define $a \mathcal{R} b$ when $ab \geq 0$. For all integers x we have $xx = x^2 \geq 0$, so $x \mathcal{R} x$ and $\mathcal{R}$ is reflexive. Also, if $x, y \in \mathbf{Z}$ and $x \mathcal{R} y$, then

$$x \mathcal{R} y \Rightarrow xy \geq 0 \Rightarrow yx \geq 0 \Rightarrow y \mathcal{R} x,$$

so the relation $\mathcal{R}$ is symmetric as well. However, here we find that $(3, 0), (0, -7) \in \mathcal{R}$—since $(3)(0) \geq 0$ and $(0)(-7) \geq 0$—but $(3, -7) \notin \mathcal{R}$ because $(3)(-7) < 0$. Consequently, this relation is *not* transitive.

EXAMPLE 7.10

If $A = \{1, 2, 3, 4\}$, then $\mathcal{R}_1 = \{(1, 1), (2, 3), (3, 4), (2, 4)\}$ is a transitive relation on A, whereas $\mathcal{R}_2 = \{(1, 3), (3, 2)\}$ is not transitive because $(1, 3), (3, 2) \in \mathcal{R}_2$ but $(1, 2) \notin \mathcal{R}_2$.

At this point the reader is probably ready to start counting the number of transitive relations on a finite set. But this is not possible here. For unlike the cases dealing with the reflexive and symmetric properties, there is no known general formula for the total number of transitive relations on a finite set. However, at a later point in this chapter we shall have the necessary ideas to count the relations $\mathcal{R}$ on a finite set, where $\mathcal{R}$ is (simultaneously) reflexive, symmetric, and transitive.

For now we consider one last property for relations.

Definition 7.5

Given a relation $\mathcal{R}$ on a set A, $\mathcal{R}$ is called *antisymmetric* if for all $a, b \in A$, $(a \mathcal{R} b$ and $b \mathcal{R} a) \Rightarrow a = b$. (Here the only way we can have both a "related to" b and b "related to" a is if a and b are one and the same element from A.)

EXAMPLE 7.11

For a given universe $\mathcal{U}$, define the relation $\mathcal{R}$ on $\mathcal{P}(\mathcal{U})$ by $(A, B) \in \mathcal{R}$ if $A \subseteq B$, for $A, B \subseteq \mathcal{U}$. So $\mathcal{R}$ is the subset relation of Chapter 3 and if $A \mathcal{R} B$ and $B \mathcal{R} A$, then we have $A \subseteq B$ and $B \subseteq A$, which gives us $A = B$. Consequently, this relation is antisymmetric, as well as reflexive and transitive, but it is not symmetric.

Before we are led astray into thinking that "not symmetric" is synonymous with "antisymmetric", let us consider the following.

EXAMPLE 7.12

For $A = \{1, 2, 3\}$, the relation $\mathcal{R}$ on A given by $\mathcal{R} = \{(1, 2), (2, 1), (2, 3)\}$ is not symmetric because $(3, 2) \notin \mathcal{R}$, and it is not antisymmetric because $(1, 2), (2, 1) \in \mathcal{R}$ but $1 \neq 2$. The relation $\mathcal{R}_1 = \{(1, 1), (2, 2)\}$ is both symmetric and antisymmetric.

How many relations on A are antisymmetric? Writing

$$A \times A = \{(1, 1), (2, 2), (3, 3)\} \cup \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\},$$

we make two observations as we try to construct an antisymmetric relation $\mathcal{R}$ on A.

1) Each element $(x, x) \in A \times A$ can be either included or excluded with no concern about whether or not $\mathcal{R}$ is antisymmetric.
2) For an element of the form (x, y), $x \neq y$, we must consider both (x, y) and (y, x) and we note that for $\mathcal{R}$ to remain antisymmetric we have three alternatives: (a) place (x, y) in $\mathcal{R}$; (b) place (y, x) in $\mathcal{R}$; or (c) place neither (x, y) nor (y, x) in $\mathcal{R}$. [What happens if we place both (x, y) and (y, x) in $\mathcal{R}$?]

So by the rule of product, the number of antisymmetric relations on A is $(2^3)(3^3) = (2^3)(3^{(3^2-3)/2})$. If $|A| = n > 0$, then there are $(2^n)(3^{(n^2-n)/2})$ antisymmetric relations on A.

For our next example we return to the concept of function dominance, which we first defined in Section 5.7.

EXAMPLE 7.13

Let $\mathcal{F}$ denote the set of all functions with domain $\mathbf{Z}^+$ and codomain $\mathbf{R}$; that is, $\mathcal{F} = \{f \mid f: \mathbf{Z}^+ \to \mathbf{R}\}$. For $f, g \in \mathcal{F}$, define the relation $\mathcal{R}$ on $\mathcal{F}$ by $f \mathcal{R} g$ if f is dominated by g (or $f \in O(g)$). Then $\mathcal{R}$ is reflexive and transitive.

If $f, g: \mathbf{Z}^+ \to \mathbf{R}$ are defined by $f(n) = n$ and $g(n) = n + 5$, then $f \mathcal{R} g$ and $g \mathcal{R} f$ but $f \neq g$, so $\mathcal{R}$ is *not antisymmetric*. In addition, if $h: \mathbf{Z}^+ \to \mathbf{R}$ is given by $h(n) = n^2$, then $(f, h), (g, h) \in \mathcal{R}$, but neither (h, f) nor (h, g) is in $\mathcal{R}$. Consequently, the relation $\mathcal{R}$ is also *not symmetric*.

At this point we have seen the four major properties that arise in the study of relations. Before closing this section we define two more notions, each of which involves three of these four properties.

Definition 7.6

A relation $\mathcal{R}$ on a set A is called a *partial order*, or a *partial ordering relation*, if $\mathcal{R}$ is reflexive, antisymmetric, and transitive.

EXAMPLE 7.14

The relation in Example 7.1(a) is a partial order, but the relation in part (b) of that example is not because it is not antisymmetric. All the relations of Example 7.2 are partial orders, as is the subset relation of Example 7.11.

Our next example provides us with the opportunity to relate this new idea of a partial order with results we studied in Chapters 1 and 4.

EXAMPLE 7.15

We start with the set $A = \{1, 2, 3, 4, 6, 12\}$ — the set of positive integer divisors of 12 — and define the relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x (exactly) divides y. As in Example 7.8 we find that $\mathcal{R}$ is reflexive and transitive. In addition, if $x, y \in A$ and we have both $x \mathcal{R} y$ and $y \mathcal{R} x$, then

$$x \mathcal{R} y \Rightarrow y = ax, \text{ for some } a \in \mathbf{Z}^+, \quad \text{and}$$
$$y \mathcal{R} x \Rightarrow x = by, \text{ for some } b \in \mathbf{Z}^+.$$

Consequently, it follows that $y = ax = a(by) = (ab)y$, and since $y \neq 0$, we have $ab = 1$. Because $a, b \in \mathbf{Z}^+$, $ab = 1 \Rightarrow a = b = 1$, so $y = x$ and $\mathcal{R}$ is antisymmetric — hence it defines a partial order for the set A.

Now suppose we wish to know how many ordered pairs occur in this relation $\mathcal{R}$. We may simply list the ordered pairs from $A \times A$ that comprise $\mathcal{R}$:

$$\mathcal{R} = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 6), (1, 12), (2, 2), (2, 4), (2, 6), (2, 12), (3, 3), (3, 6), (3, 12), (4, 4), (4, 12), (6, 6), (6, 12), (12, 12)\}$$

In this way we learn that there are 18 ordered pairs in the relation. But if we then wanted to consider the same type of partial order for the set of positive integer divisors of 1800, we should definitely be discouraged by this method of simply *listing* all the ordered pairs. So let us examine the relation $\mathcal{R}$ a little closer. By the Fundamental Theorem of Arithmetic we may write $12 = 2^2 \cdot 3$ and then realize that if $(c, d) \in \mathcal{R}$, then

$$c = 2^m \cdot 3^n \quad \text{and} \quad d = 2^p \cdot 3^q,$$

where $m, n, p, q \in \mathbf{N}$ with $0 \le m \le p \le 2$ and $0 \le n \le q \le 1$.

When we consider the fact that $0 \le m \le p \le 2$, we find that each possibility for m, p is simply a selection of size 2 from a set of size 3 — namely, the set $\{0, 1, 2\}$ — where repetitions are allowed. (In any such selection, if there is a smaller nonnegative integer, then it is assigned to m.) In Chapter 1 we learned that such a selection can be made in $\binom{3+2-1}{2} = \binom{4}{2} = 6$ ways. And, in like manner, n and q can be selected in $\binom{2+2-1}{2} = \binom{3}{2} = 3$ ways. So by the rule of product there should be $(6)(3) = 18$ ordered pairs in $\mathcal{R}$ — as we found earlier by actually listing all of them.

Now suppose we examine a similar situation, the set of positive integer divisors of $1800 = 2^3 \cdot 3^2 \cdot 5^2$. Here we are dealing with $(3+1)(2+1)(2+1) = (4)(3)(3) = 36$ divisors, and a typical ordered pair for this partial order (given by division) looks like $(2^r \cdot 3^s \cdot 5^t, 2^u \cdot 3^v \cdot 5^w)$, where $r, s, t, u, v, w \in \mathbf{N}$ with $0 \le r \le u \le 3$, $0 \le s \le v \le 2$, and $0 \le t \le w \le 2$. So the number of ordered pairs in the relation is

$$\binom{4+2-1}{2}\binom{3+2-1}{2}\binom{3+2-1}{2} = \binom{5}{2}\binom{4}{2}\binom{4}{2} = (10)(6)(6) = 360,$$

and we definitely should *not* want to have to list all of the ordered pairs in the relation in order to obtain this result.

In general, for $n \in \mathbf{Z}^+$ with $n > 1$, use the Fundamental Theorem of Arithmetic to write $n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \cdots p_k^{e_k}$, where $k \in \mathbf{Z}^+$, $p_1 < p_2 < p_3 < \cdots < p_k$, and p_i is prime and $e_i \in \mathbf{Z}^+$ for each $1 \le i \le k$. Then n has $\prod_{i=1}^{k}(e_i + 1)$ positive integer divisors. And when we consider the same type of partial order for this set (of positive integer divisors of n), we find that the number of ordered pairs in the relation is

$$\prod_{i=1}^{k}\binom{(e_i+1)+2-1}{2} = \prod_{i=1}^{k}\binom{e_i+2}{2}.$$

In closing this section we introduce the equivalence relation — a concept that is very important in the study of mathematics.

Definition 7.7 An *equivalence relation* $\mathcal{R}$ on a set A is a relation that is reflexive, symmetric, and transitive.

EXAMPLE 7.16

a) The relation in Example 7.1(b) and all the relations in Example 7.3(c) are equivalence relations.

b) If $A = \{1, 2, 3\}$, then
$\mathcal{R}_1 = \{(1, 1), (2, 2), (3, 3)\}$,
$\mathcal{R}_2 = \{(1, 1), (2, 2), (2, 3), (3, 2), (3, 3)\}$,
$\mathcal{R}_3 = \{(1, 1), (1, 3), (2, 2), (3, 1), (3, 3)\}$, and
$\mathcal{R}_4 = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\} = A \times A$
are all equivalence relations on A.

c) For a given finite set A, $A \times A$ is the largest equivalence relation on A, and if $A = \{a_1, a_2, \ldots, a_n\}$, then the equality relation $\mathcal{R} = \{(a_i, a_i) | 1 \le i \le n\}$ is the smallest equivalence relation on A.

d) Let $A = \{1, 2, 3, 4, 5, 6, 7\}$, $B = \{x, y, z\}$, and $f: A \rightarrow B$ be the onto function

$$f = \{(1, x), (2, z), (3, x), (4, y), (5, z), (6, y), (7, x)\}.$$

Define the relation $\mathcal{R}$ on A by $a \mathcal{R} b$ if $f(a) = f(b)$. Then, for instance, we find here that $1 \mathcal{R} 1$, $1 \mathcal{R} 3$, $2 \mathcal{R} 5$, $3 \mathcal{R} 1$, and $4 \mathcal{R} 6$.

For each $a \in A$, $f(a) = f(a)$ because f is a function—so $a \mathcal{R} a$, and $\mathcal{R}$ is reflexive. Now suppose that $a, b \in A$ and $a \mathcal{R} b$. Then $a \mathcal{R} b \Rightarrow f(a) = f(b) \Rightarrow f(b) = f(a) \Rightarrow b \mathcal{R} a$, so $\mathcal{R}$ is symmetric. Finally, if $a, b, c \in A$ with $a \mathcal{R} b$ and $b \mathcal{R} c$, then $f(a) = f(b)$ and $f(b) = f(c)$. Consequently, $f(a) = f(c)$, and we see that $(a \mathcal{R} b \wedge b \mathcal{R} c) \Rightarrow a \mathcal{R} c$. So $\mathcal{R}$ is transitive. Since $\mathcal{R}$ is reflexive, symmetric, and transitive, it is an equivalence relation.

Here $\mathcal{R} = \{(1, 1), (1, 3), (1, 7), (2, 2), (2, 5), (3, 1), (3, 3), (3, 7), (4, 4), (4, 6), (5, 2), (5, 5), (6, 4), (6, 6), (7, 1), (7, 3), (7, 7)\}$.

e) If $\mathcal{R}$ is a relation on a set A, then $\mathcal{R}$ is both an equivalence relation and a partial order on A if and only if $\mathcal{R}$ is the equality relation on A.

EXERCISES 7.1

1. If $A = \{1, 2, 3, 4\}$, give an example of a relation $\mathcal{R}$ on A that is

a) reflexive and symmetric, but not transitive

b) reflexive and transitive, but not symmetric

c) symmetric and transitive, but not reflexive

2. For relation (b) in Example 7.1, determine five values of x for which $(x, 5) \in \mathcal{R}$.

3. For the relation $\mathcal{R}$ in Example 7.13, let $f: \mathbf{Z}^+ \rightarrow \mathbf{R}$ where $f(n) = n$.

a) Find three elements $f_1, f_2, f_3 \in \mathcal{F}$ such that $f_i \mathcal{R} f$ and $f \mathcal{R} f_i$, for all $1 \leq i \leq 3$.

b) Find three elements $g_1, g_2, g_3 \in \mathcal{F}$ such that $g_i \mathcal{R} f$ but $f \not\mathcal{R} g_i$, for all $1 \leq i \leq 3$.

4. a) Rephrase the definitions for the reflexive, symmetric, transitive, and antisymmetric properties of a relation $\mathcal{R}$ (on a set A), using quantifiers.

b) Use the results of part (a) to specify when a relation $\mathcal{R}$ (on a set A) is (i) *not* reflexive; (ii) *not* symmetric; (iii) *not* transitive; and (iv) *not* antisymmetric.

5. For each of the following relations, determine whether the relation is reflexive, symmetric, antisymmetric, or transitive.

a) $\mathcal{R} \subseteq \mathbf{Z}^+ \times \mathbf{Z}^+$ where $a \mathcal{R} b$ if $a|b$ (read "a divides b," as defined in Section 4.3).

b) $\mathcal{R}$ is the relation on $\mathbf{Z}$ where $a \mathcal{R} b$ if $a|b$.

c) For a given universe $\mathcal{U}$ and a fixed subset C of $\mathcal{U}$, define $\mathcal{R}$ on $\mathcal{P}(\mathcal{U})$ as follows: For $A, B \subseteq \mathcal{U}$ we have $A \mathcal{R} B$ if $A \cap C = B \cap C$.

d) On the set A of all lines in $\mathbf{R}^2$, define the relation $\mathcal{R}$ for two lines ℓ_1, ℓ_2 by $\ell_1 \mathcal{R} \ell_2$ if ℓ_1 is perpendicular to ℓ_2.

e) $\mathcal{R}$ is the relation on $\mathbf{Z}$ where $x \mathcal{R} y$ if $x + y$ is odd.

f) $\mathcal{R}$ is the relation on $\mathbf{Z}$ where $x \mathcal{R} y$ if $x - y$ is even.

g) Let T be the set of all triangles in $\mathbf{R}^2$. Define $\mathcal{R}$ on T by $t_1 \mathcal{R} t_2$ if t_1 and t_2 have an angle of the same measure.

h) $\mathcal{R}$ is the relation on $\mathbf{Z} \times \mathbf{Z}$ where $(a, b)\mathcal{R}(c, d)$ if $a \leq c$. [*Note*: $\mathcal{R} \subseteq (\mathbf{Z} \times \mathbf{Z}) \times (\mathbf{Z} \times \mathbf{Z})$.]

6. Which relations in Exercise 5 are partial orders? Which are equivalence relations?

7. Let $\mathcal{R}_1, \mathcal{R}_2$ be relations on a set A. (a) Prove or disprove that $\mathcal{R}_1, \mathcal{R}_2$ reflexive $\Rightarrow \mathcal{R}_1 \cap \mathcal{R}_2$ reflexive. (b) Answer part (a) when each occurrence of "reflexive" is replaced by (i) symmetric; (ii) antisymmetric; and (iii) transitive.

8. Answer Exercise 7, replacing each occurrence of $\cap$ by $\cup$.

9. For each of the following statements about relations on a set A, where $|A| = n$, determine whether the statement is true or false. If it is false, give a counterexample.

a) If $\mathcal{R}$ is a relation on A and $|\mathcal{R}| \geq n$, then $\mathcal{R}$ is reflexive.

b) If $\mathcal{R}_1, \mathcal{R}_2$ are relations on A and $\mathcal{R}_2 \supseteq \mathcal{R}_1$, then $\mathcal{R}_1$ reflexive (symmetric, antisymmetric, transitive) $\Rightarrow \mathcal{R}_2$ reflexive (symmetric, antisymmetric, transitive).

c) If $\mathcal{R}_1, \mathcal{R}_2$ are relations on A and $\mathcal{R}_2 \supseteq \mathcal{R}_1$, then $\mathcal{R}_2$ reflexive (symmetric, antisymmetric, transitive) $\Rightarrow \mathcal{R}_1$ reflexive (symmetric, antisymmetric, transitive).

d) If $\mathcal{R}$ is an equivalence relation on A, then $n \leq |\mathcal{R}| \leq n^2$.

10. If $A = \{w, x, y, z\}$, determine the number of relations on A that are (a) reflexive; (b) symmetric; (c) reflexive and symmetric; (d) reflexive and contain (x, y); (e) symmetric and contain (x, y); (f) antisymmetric; (g) antisymmetric and contain (x, y); (h) symmetric and antisymmetric; and (i) reflexive, symmetric, and antisymmetric.

11. Let $n \in \mathbf{Z}^+$ with $n > 1$, and let A be the set of positive integer divisors of n. Define the relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x

(exactly) divides y. Determine how many ordered pairs are in the relation $\mathcal{R}$ when n is (a) 10; (b) 20; (c) 40; (d) 200; (e) 210; and (f) 13860.

12. Suppose that p_1, p_2, p_3 are distinct primes and that $n, k \in \mathbf{Z}^+$ with $n = p_1^5 p_2^3 p_3^k$. Let A be the set of positive integer divisors of n and define the relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x (exactly) divides y. If there are 5880 ordered pairs in $\mathcal{R}$, determine k and $|A|$.

13. What is wrong with the following argument?

Let A be a set with $\mathcal{R}$ a relation on A. If $\mathcal{R}$ is symmetric and transitive, then $\mathcal{R}$ is reflexive.

Proof: Let $(x, y) \in \mathcal{R}$. By the symmetric property, $(y, x) \in \mathcal{R}$. Then with $(x, y), (y, x) \in \mathcal{R}$, it follows by the transitive property that $(x, x) \in \mathcal{R}$. Consequently, $\mathcal{R}$ is reflexive.

14. Let A be a set with $|A| = n$, and let $\mathcal{R}$ be a relation on A that is antisymmetric. What is the maximum value for $|\mathcal{R}|$? How many antisymmetric relations can have this size?

15. Let A be a set with $|A| = n$, and let $\mathcal{R}$ be an equivalence relation on A with $|\mathcal{R}| = r$. Why is $r - n$ always even?

16. A relation $\mathcal{R}$ on a set A is called *irreflexive* if for all $a \in A$, $(a, a) \notin \mathcal{R}$.

a) Give an example of a relation $\mathcal{R}$ on $\mathbf{Z}$ where $\mathcal{R}$ is irreflexive and transitive but not symmetric.

b) Let $\mathcal{R}$ be a nonempty relation on a set A. Prove that if $\mathcal{R}$ satisfies any two of the following properties — irreflexive, symmetric, and transitive — then it cannot satisfy the third.

c) If $|A| = n \geq 1$, how many different relations on A are irreflexive? How many are neither reflexive nor irreflexive?

17. Let $A = \{1, 2, 3, 4, 5, 6, 7\}$. How many symmetric relations on A contain exactly (a) four ordered pairs? (b) five ordered pairs? (c) seven ordered pairs? (d) eight ordered pairs?

18. a) Let $f: A \to B$, where $|A| = 25$, $B = \{x, y, z\}$, and $|f^{-1}(x)| = 10$, $|f^{-1}(y)| = 10$, $|f^{-1}(z)| = 5$. If we define the relation $\mathcal{R}$ on A by $a \mathcal{R} b$ if $a, b \in A$ and $f(a) = f(b)$, how many ordered pairs are there in the relation $\mathcal{R}$?

b) For $n, n_1, n_2, n_3, n_4 \in \mathbf{Z}^+$, let $f: A \to B$, where $|A| = n$, $B = \{w, x, y, z\}$, $|f^{-1}(w)| = n_1$, $|f^{-1}(x)| = n_2$, $|f^{-1}(y)| = n_3$, $|f^{-1}(z)| = n_4$, and $n_1 + n_2 + n_3 + n_4 = n$. If we define the relation $\mathcal{R}$ on A by $a \mathcal{R} b$ if $a, b \in A$ and $f(a) = f(b)$, how many ordered pairs are there in the relation $\mathcal{R}$?

7.2 Computer Recognition: Zero-One Matrices and Directed Graphs

Since our interest in relations is focused on those for finite sets, we are concerned with ways of representing such relations so that the properties of Section 7.1 can be easily verified. For this reason we now develop the necessary tools: relation composition, zero-one matrices, and directed graphs.

In a manner analogous to the composition of functions, relations can be combined in the following circumstances.

Definition 7.8 If A, B, and C are sets with $\mathcal{R}_1 \subseteq A \times B$ and $\mathcal{R}_2 \subseteq B \times C$, then the *composite relation* $\mathcal{R}_1 \circ \mathcal{R}_2$ is a relation from A to C defined by $\mathcal{R}_1 \circ \mathcal{R}_2 = \{(x, z) | x \in A, z \in C$, and there exists $y \in B$ with $(x, y) \in \mathcal{R}_1, (y, z) \in \mathcal{R}_2\}$.

Beware! The composition of two relations is written in an order opposite to that for function composition. We shall see why in Example 7.21.

EXAMPLE 7.17 Let $A = \{1, 2, 3, 4\}$, $B = \{w, x, y, z\}$, and $C = \{5, 6, 7\}$. Consider $\mathcal{R}_1 = \{(1, x), (2, x), (3, y), (3, z)\}$, a relation from A to B, and $\mathcal{R}_2 = \{(w, 5), (x, 6)\}$, a relation from B to C. Then $\mathcal{R}_1 \circ \mathcal{R}_2 = \{(1, 6), (2, 6)\}$ is a relation from A to C. If $\mathcal{R}_3 = \{(w, 5), (w, 6)\}$ is another relation from B to C, then $\mathcal{R}_1 \circ \mathcal{R}_3 = \emptyset$.

EXAMPLE 7.18

Let A be the set of employees at a computing center, while B denotes a set of high-level programming languages, and C is a set of projects $\{p_1, p_2, \ldots, p_8\}$ for which managers must make work assignments using the people in A. Consider $\mathscr{R}_1 \subseteq A \times B$, where an ordered pair of the form (L. Alldredge, Java) indicates that employee L. Alldredge is proficient in Java (and perhaps other programming languages). The relation $\mathscr{R}_2 \subseteq B \times C$ consists of ordered pairs such as (Java, p_2), indicating that Java is considered an essential language needed by anyone who works on project p_2. In the composite relation $\mathscr{R}_1 \circ \mathscr{R}_2$ we find (L. Alldredge, p_2). If no other ordered pair in $\mathscr{R}_2$ has p_2 as its second component, we know that if L. Alldredge was assigned to p_2 it was solely on the basis of his proficiency in Java. (Here $\mathscr{R}_1 \circ \mathscr{R}_2$ has been used to set up a matching process between employees and projects on the basis of employee knowledge of specific programming languages.)

Comparable to the associative law for function composition, the following result holds for relations.

THEOREM 7.1

Let A, B, C, and D be sets with $\mathscr{R}_1 \subseteq A \times B$, $\mathscr{R}_2 \subseteq B \times C$, and $\mathscr{R}_3 \subseteq C \times D$. Then $\mathscr{R}_1 \circ (\mathscr{R}_2 \circ \mathscr{R}_3) = (\mathscr{R}_1 \circ \mathscr{R}_2) \circ \mathscr{R}_3$.

Proof: Since both $\mathscr{R}_1 \circ (\mathscr{R}_2 \circ \mathscr{R}_3)$ and $(\mathscr{R}_1 \circ \mathscr{R}_2) \circ \mathscr{R}_3$ are relations from A to D, there is some reason to believe they are equal. If $(a, d) \in \mathscr{R}_1 \circ (\mathscr{R}_2 \circ \mathscr{R}_3)$, then there is an element $b \in B$ with $(a, b) \in \mathscr{R}_1$ and $(b, d) \in (\mathscr{R}_2 \circ \mathscr{R}_3)$. Also, $(b, d) \in (\mathscr{R}_2 \circ \mathscr{R}_3) \Rightarrow (b, c) \in \mathscr{R}_2$ and $(c, d) \in \mathscr{R}_3$ for some $c \in C$. Then $(a, b) \in \mathscr{R}_1$ and $(b, c) \in \mathscr{R}_2 \Rightarrow (a, c) \in \mathscr{R}_1 \circ \mathscr{R}_2$. Finally, $(a, c) \in \mathscr{R}_1 \circ \mathscr{R}_2$ and $(c, d) \in \mathscr{R}_3 \Rightarrow (a, d) \in (\mathscr{R}_1 \circ \mathscr{R}_2) \circ \mathscr{R}_3$, and $\mathscr{R}_1 \circ (\mathscr{R}_2 \circ \mathscr{R}_3) \subseteq (\mathscr{R}_1 \circ \mathscr{R}_2) \circ \mathscr{R}_3$. The opposite inclusion follows by similar reasoning.

As a result of this theorem no ambiguity arises when we write $\mathscr{R}_1 \circ \mathscr{R}_2 \circ \mathscr{R}_3$ for either of the relations in Theorem 7.1. In addition, we can now define the powers of a relation $\mathscr{R}$ on a set.

Definition 7.9

Given a set A and a relation $\mathscr{R}$ on A, we define the *powers of* $\mathscr{R}$ recursively by (a) $\mathscr{R}^1 = \mathscr{R}$; and (b) for $n \in \mathbf{Z}^+$, $\mathscr{R}^{n+1} = \mathscr{R} \circ \mathscr{R}^n$.

Note that for $n \in \mathbf{Z}^+$, $\mathscr{R}^n$ is a relation on A.

EXAMPLE 7.19

If $A = \{1, 2, 3, 4\}$ and $\mathscr{R} = \{(1, 2), (1, 3), (2, 4), (3, 2)\}$, then $\mathscr{R}^2 = \{(1, 4), (1, 2), (3, 4)\}$, $\mathscr{R}^3 = \{(1, 4)\}$, and for $n \geq 4$, $\mathscr{R}^n = \emptyset$.

As the set A and the relation $\mathscr{R}$ on A grow larger, calculations such as those in Example 7.19 become tedious. To avoid this tedium, the tool we need is the computer, once a way can be found to tell the machine about the set A and the relation $\mathscr{R}$ on A.

Definition 7.10

An $m \times n$ *zero-one matrix* $E = (e_{ij})_{m \times n}$ is a rectangular array of numbers arranged in m rows and n columns, where each e_{ij}, for $1 \leq i \leq m$ and $1 \leq j \leq n$, denotes the entry in the ith row and jth column of E, and each such entry is 0 or 1. [We can also write (0, 1)-matrix for this type of matrix.]

EXAMPLE 7.20

The matrix

$$E = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

is a 3×4 (0, 1)-matrix where, for example, $e_{11} = 1$, $e_{23} = 0$, and $e_{31} = 1$.

In working with these matrices, we use the standard operations of matrix addition and multiplication *with the stipulation that* $1 + 1 = 1$. (Hence the addition is called Boolean.)

EXAMPLE 7.21

Consider the sets A, B, and C and the relations $\mathcal{R}_1$, $\mathcal{R}_2$ of Example 7.17. With the orders of the elements in A, B, and C fixed as in that example, we define the *relation matrices* for $\mathcal{R}_1$, $\mathcal{R}_2$ as follows:

$$M(\mathcal{R}_1) = \begin{matrix} \\ (1) \\ (2) \\ (3) \\ (4) \end{matrix} \begin{matrix} (w) & (x) & (y) & (z) \\ \end{matrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M(\mathcal{R}_2) = \begin{matrix} (w) \\ (x) \\ (y) \\ (z) \end{matrix} \overset{(5)\ (6)\ (7)}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}$$

In constructing $M(\mathcal{R}_1)$, we are dealing with a relation from A to B, so the elements of A are used to mark the rows of $M(\mathcal{R}_1)$ and the elements of B designate the columns. Then to denote, for example, that $(2, x) \in \mathcal{R}_1$, we place a 1 in the row marked (2) and the column marked (x). Each 0 in this matrix indicates an ordered pair in $A \times B$ that is missing from $\mathcal{R}_1$. For example, since $(3, w) \notin \mathcal{R}_1$, there is a 0 for the entry in row (3) and column (w) of the matrix $M(\mathcal{R}_1)$. The same process is used to obtain $M(\mathcal{R}_2)$.

Multiplying these matrices,† we find that

$$M(\mathcal{R}_1) \cdot M(\mathcal{R}_2) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{matrix} (1) \\ (2) \\ (3) \\ (4) \end{matrix} \overset{(5)\ (6)\ (7)}{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}} = M(\mathcal{R}_1 \circ \mathcal{R}_2),$$

where the rows of the 4×3 matrix $M(\mathcal{R}_1 \circ \mathcal{R}_2)$ are marked by the elements of A while its columns are marked by the elements of C. In general we have: If $\mathcal{R}_1$ is a relation from A to B and $\mathcal{R}_2$ is a relation from B to C, then $M(\mathcal{R}_1) \cdot M(\mathcal{R}_2) = M(\mathcal{R}_1 \circ \mathcal{R}_2)$. That is, the product of the relation matrices for $\mathcal{R}_1$, $\mathcal{R}_2$, in that order, equals the relation matrix of the composite relation $\mathcal{R}_1 \circ \mathcal{R}_2$. (This is why the composition of two relations was written in the order specified in Definition 7.8.)

The reader will be asked to prove the general result of Example 7.21, along with some results from our next example, in Exercises 11 and 12 at the end of this section.

Further properties of relation matrices are exhibited in the following example.

†The reader who is not familiar with matrix multiplication or simply wishes a brief review should consult Appendix 2.

EXAMPLE 7.22

Let $A = \{1, 2, 3, 4\}$ and $\mathcal{R} = \{(1, 2), (1, 3), (2, 4), (3, 2)\}$, as in Example 7.19. Keeping the order of the elements in A fixed, we define the *relation matrix* for $\mathcal{R}$ as follows: $M(\mathcal{R})$ is the 4×4 (0, 1)-matrix whose entries m_{ij}, for $1 \le i, j \le 4$, are given by

$$m_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{R}, \\ 0, & \text{otherwise.} \end{cases}$$

In this case we find that

$$M(\mathcal{R}) = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now how can this be of any use? If we compute $(M(\mathcal{R}))^2$ using the convention that $1 + 1 = 1$, then we find that

$$(M(\mathcal{R}))^2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which happens to be the relation matrix for $\mathcal{R} \circ \mathcal{R} = \mathcal{R}^2$. (Check Example 7.19.) Furthermore,

$$(M(\mathcal{R}))^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which is also the relation matrix for the relation $\mathcal{R}^4$—that is, $(M(\mathcal{R}))^4 = M(\mathcal{R}^4)$. Also, recall that $\mathcal{R}^4 = \emptyset$, as we learned in Example 7.19.

What has happened here carries over to the general situation. We now state some results about relation matrices and their use in studying relations.

Let A be a set with $|A| = n$ and $\mathcal{R}$ a relation on A. If $M(\mathcal{R})$ is the relation matrix for $\mathcal{R}$, then

a) $M(\mathcal{R}) = \mathbf{0}$ **(the matrix of all 0's) if and only if** $\mathcal{R} = \emptyset$

b) $M(\mathcal{R}) = \mathbf{1}$ **(the matrix of all 1's) if and only if** $\mathcal{R} = A \times A$

c) $M(\mathcal{R}^m) = [M(\mathcal{R})]^m$, **for** $m \in \mathbf{Z}^+$

Using the (0, 1)-matrix for a relation, we now turn to the recognition of the reflexive, symmetric, antisymmetric, and transitive properties. To accomplish this we need the concepts introduced in the following three definitions.

Definition 7.11

Let $E = (e_{ij})_{m \times n}$, $F = (f_{ij})_{m \times n}$ be two $m \times n$ (0, 1)-matrices. We say that E *precedes*, or *is less than*, F, and we write $E \le F$, if $e_{ij} \le f_{ij}$, for all $1 \le i \le m$, $1 \le j \le n$.

EXAMPLE 7.23

With $E = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ and $F = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, we have $E \leq F$. In fact, there are eight (0, 1)-matrices G for which $E \leq G$.

Definition 7.12

For $n \in \mathbf{Z}^+$, $I_n = (\delta_{ij})_{n \times n}$ is the $n \times n$ (0, 1)-matrix where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Definition 7.13

Let $A = (a_{ij})_{m \times n}$ be a (0, 1)-matrix. The *transpose* of A, written A^{tr}, is the matrix $(a^*_{ji})_{n \times m}$ where $a^*_{ji} = a_{ij}$, for all $1 \leq j \leq n$, $1 \leq i \leq m$.

EXAMPLE 7.24

For $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$, we find that $A^{\text{tr}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$.

As this example demonstrates, the ith row (column) of A equals the ith column (row) of A^{tr}. This indicates a method we can use in order to obtain the matrix A^{tr} from the matrix A.

THEOREM 7.2

Given a set A with $|A| = n$ and a relation $\mathcal{R}$ on A, let M denote the relation matrix for $\mathcal{R}$. Then

a) $\mathcal{R}$ is reflexive if and only if $I_n \leq M$.

b) $\mathcal{R}$ is symmetric if and only if $M = M^{\text{tr}}$.

c) $\mathcal{R}$ is transitive if and only if $M \cdot M = M^2 \leq M$.

d) $\mathcal{R}$ is antisymmetric if and only if $M \cap M^{\text{tr}} \leq I_n$. (The matrix $M \cap M^{\text{tr}}$ is formed by operating on corresponding entries in M and M^{tr} according to the rules $0 \cap 0 = 0 \cap 1 = 1 \cap 0 = 0$ and $1 \cap 1 = 1$—that is, the usual multiplication for 0's and/or 1's.)

Proof: The results follow from the definitions of the relation properties and the (0, 1)-matrix. We demonstrate this for part (c), using the elements of A to designate the rows and columns in M, as in Examples 7.21 and 7.22.

Let $M^2 \leq M$. If $(x, y), (y, z) \in \mathcal{R}$, then there are 1's in row (x), column (y) and in row (y), column (z) of M. Consequently, in row (x), column (z) of M^2 there is a 1. This 1 must also occur in row (x), column (z) of M because $M^2 \leq M$. Hence $(x, z) \in \mathcal{R}$ and $\mathcal{R}$ is transitive.

Conversely, if $\mathcal{R}$ is transitive and M is the relation matrix for $\mathcal{R}$, let s_{xz} be the entry in row (x) and column (z) of M^2, with $s_{xz} = 1$. For s_{xz} to equal 1 in M^2, there must exist at least one $y \in A$ where $m_{xy} = m_{yz} = 1$ in M. This happens only if $x \mathcal{R} y$ and $y \mathcal{R} z$. With $\mathcal{R}$ transitive, it then follows that $x \mathcal{R} z$. So $m_{xz} = 1$ and $M^2 \leq M$.

The proofs of the remaining parts are left to the reader.

The relation matrix is a useful tool for the computer recognition of certain properties of relations. Storing information as described here, this matrix is an example of a *data* *structure*. Also of interest is how the relation matrix is used in the study of graph theory† and how graph theory is used in the recognition of certain properties of relations.

At this point we shall introduce some fundamental concepts in graph theory. Often these concepts will be given within examples and not in terms of formal definitions. In Chapter 11, however, the presentation will not assume what is given here and will be more rigorous and comprehensive.

Definition 7.14

Let V be a finite nonempty set. A *directed graph* (or *digraph*) G *on* V is made up of the elements of V, called the *vertices* or *nodes* of G, and a subset E, of $V \times V$, that contains the *(directed) edges*, or *arcs*, of G. The set V is called the *vertex set* of G, and the set E is called the *edge set*. We then write $G = (V, E)$ to denote the graph.

If $a, b \in V$ and $(a, b) \in E$‡, then there is an edge *from* a *to* b. Vertex a is called the *origin* or *source* of the edge, with b the *terminus*, or *terminating vertex*, and we say that b is *adjacent from* a and that a is *adjacent to* b. In addition, if $a \neq b$, then $(a, b) \neq (b, a)$. An edge of the form (a, a) is called a *loop* (at a).

EXAMPLE 7.25

For $V = \{1, 2, 3, 4, 5\}$, the diagram in Fig. 7.1 is a directed graph G on V with edge set $\{(1, 1), (1, 2), (1, 4), (3, 2)\}$. Vertex 5 is a part of this graph even though it is not the origin or terminus of an edge. It is referred to as an *isolated* vertex. As we see here, edges need not be straight line segments, and there is no concern about the length of an edge.

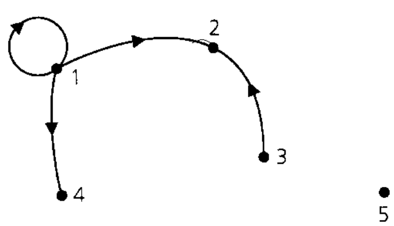


Figure 7.1

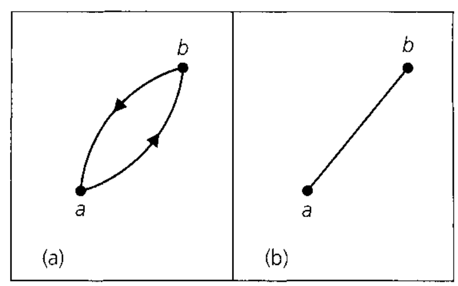


Figure 7.2

When we develop a *flowchart* to study a computer program or algorithm, we deal with a special type of directed graph where the shapes of the vertices may be important in the analysis of the algorithm. Road maps are directed graphs, where the cities and towns are represented by vertices and the highways linking any two localities are given by edges. In road maps, an edge is often directed in both directions. Consequently, if G is a directed graph and $a, b \in V$, with $a \neq b$, and both $(a, b), (b, a) \in E$, then the single undirected edge $\{a, b\} = \{b, a\}$ in Fig. 7.2(b) is used to represent the two directed edges shown in Fig. 7.2(a). In this case, a and b are called *adjacent* vertices. (Directions may also be disregarded for loops.)

†Since the terminology of graph theory is not standardized, the reader may find some differences between definitions given here and in other texts.

‡In this chapter we allow only one edge from a to b. Situations where multiple edges occur are called *multigraphs*. These are discussed in Chapter 11.

Directed graphs play an important role in many situations in computer science. The following example demonstrates one of these.

EXAMPLE 7.26

Computer programs can be processed more rapidly when certain statements in the program are executed concurrently. But in order to accomplish this we must be aware of the dependence of some statements on earlier statements in the program. For we cannot execute a statement that needs results from other statements — statements that have not yet been executed.

In Fig. 7.3(a) we have eight assignment statements that constitute the beginning of a computer program. We represent these statements by the eight corresponding vertices $s_1, s_2, s_3, \ldots, s_8$ in part (b) of the figure, where a directed edge such as (s_1, s_5) indicates that statement s_5 cannot be executed until statement s_1 has been executed. The resulting directed graph is called the *precedence graph* for the given lines of the computer program. Note how this graph indicates, for example, that statement s_7 cannot be executed until after each of the statements s_1, s_2, s_3, and s_4 has been executed. Also, we see how a statement such as s_1 must be executed before it is possible to execute any of the statements s_2, s_4, s_5, s_7, or s_8. In general, if a vertex (statement) s is adjacent *from m* other vertices (and no others), then the corresponding statements for these m vertices must be executed before statement s can be executed. Similarly, should a vertex (statement) s be adjacent *to n* other vertices, then each of the corresponding statements for these vertices requires the execution of statement s before it can be executed. Finally, from the precedence graph we see that the statements s_1, s_3, and s_6 can be processed concurrently. Following this, the statements s_2, s_4, and s_8 can be executed at the same time, and then the statements s_5 and s_7. (Or we could process statements s_2 and s_4 concurrently, and then the statements s_5, s_7, and s_8.)

```
(s1)  b := 3
(s2)  c := b + 2
(s3)  a := 1
(s4)  d := a * b + 5
(s5)  e := d - 1
(s6)  f := 7
(s7)  e := c + d
(s8)  g := b * f
```

(a)

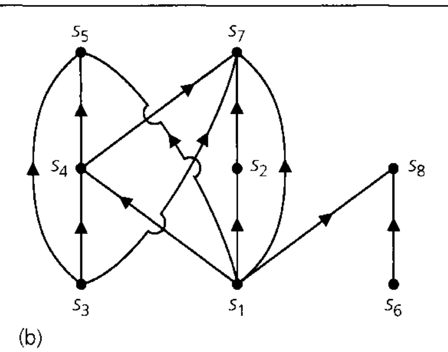


Figure 7.3

Now we want to consider how relations and directed graphs are interrelated. For a start, given a set A and a relation $\mathcal{R}$ on A, we can construct a directed graph G with vertex set A and edge set $E \subseteq A \times A$, where $(a, b) \in E$ if $a, b \in A$ and $a \mathcal{R} b$. This is demonstrated in the following example.

EXAMPLE 7.27

For $A = \{1, 2, 3, 4\}$, let $\mathcal{R} = \{(1, 1), (1, 2), (2, 3), (3, 2), (3, 3), (3, 4), (4, 2)\}$ be a relation on A. The directed graph associated with $\mathcal{R}$ is shown in Fig. 7.4(a), where the undirected edge $\{2, 3\}(= \{3, 2\})$ is used in place of the pair of distinct directed edges (2, 3) and (3, 2). If the directions in Fig. 7.4(a) are ignored, we get the *associated undirected graph* shown in part (b) of the figure. Here we see that the graph is *connected* in the sense that for any two vertices x, y, with $x \neq y$, there is a *path* starting at x and ending at y. Such a path consists of a *finite sequence of undirected edges*, so the edges $\{1, 2\}$, $\{2, 4\}$ provide a path from 1 to 4, and the edges $\{3, 4\}$, $\{4, 2\}$, and $\{2, 1\}$ provide a path from 3 to 1. The sequence of edges $\{3, 4\}$, $\{4, 2\}$, and $\{2, 3\}$ provides a path from 3 to 3. Such a *closed* path is called a *cycle*. This is an example of an undirected cycle of *length* 3, because it has three edges in it.

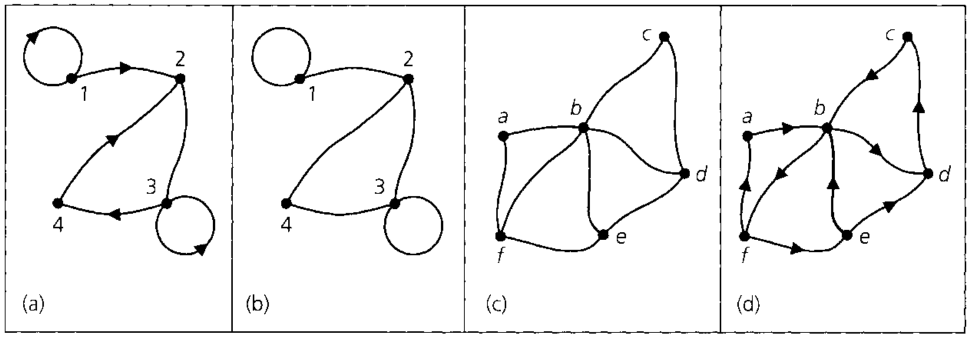


Figure 7.4

When we are dealing with paths (in both directed and undirected graphs), no vertex may be repeated. Therefore, the sequence of edges $\{a, b\}$, $\{b, e\}$, $\{e, f\}$, $\{f, b\}$, $\{b, d\}$ in Fig. 7.4(c) is *not* considered to be a path (from a to d) because we pass through the vertex b more than once. In the case of cycles, the path starts and terminates at the same vertex and has *at least three edges*. In Fig. 7.4(d) the sequence of edges (b, f), (f, e), (e, d), (d, c), (c, b) provides a *directed cycle* of length 5. The six edges (b, f), (f, e), (e, b), (b, d), (d, c), (c, b) do *not* yield a directed cycle in the figure because of the repetition of vertex b. If their directions are ignored, the corresponding six edges, in part (c) of the figure, likewise pass through vertex b more than once. Consequently, these edges are not considered to form a cycle for the undirected graph in Fig. 7.4(c).

Now since we require a cycle to have *length* at least 3, we shall not consider loops to be cycles. We also note that loops have no bearing on graph connectivity.

We choose to define the next idea formally because of its relevance to what we did earlier in Section 6.3.

Definition 7.15 A directed graph G on V is called *strongly connected* if for all $x, y \in V$, where $x \neq y$, there is a path (in G) of directed edges from x to y — that is, either the directed edge (x, y) is in G or, for some $n \in \mathbf{Z}^+$ and distinct vertices $v_1, v_2, \ldots, v_n \in V$, the directed edges $(x, v_1), (v_1, v_2), \ldots, (v_n, y)$ are in G.

It is in this sense that we talked about strongly connected machines in Chapter 6. The graph in Fig. 7.4(a) is connected but not strongly connected. For example, there is no directed path from 3 to 1. In Fig. 7.5 the directed graph on $V = \{1, 2, 3, 4\}$ is strongly connected and *loop-free*. This is also true of the directed graph in Fig. 7.4(d).

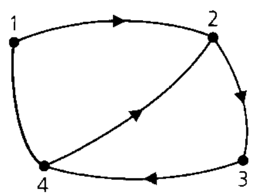


Figure 7.5

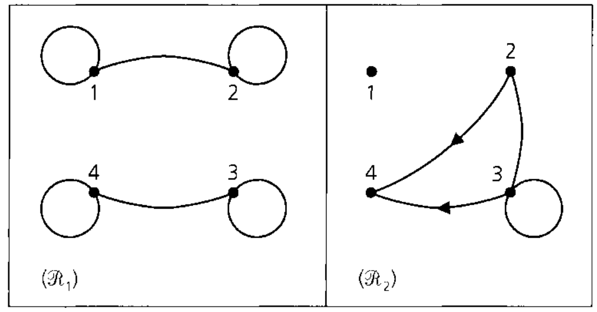


Figure 7.6

EXAMPLE 7.28

For $A = \{1, 2, 3, 4\}$, consider the relations $\mathcal{R}_1 = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3), (3, 4), (4, 3), (4, 4)\}$ and $\mathcal{R}_2 = \{(2, 4), (2, 3), (3, 2), (3, 3), (3, 4)\}$. As Fig. 7.6 illustrates, the graphs of these relations are *disconnected*. However, each graph is the union of two connected pieces called the *components* of the graph. For $\mathcal{R}_1$ the graph is made up of two strongly connected components. For $\mathcal{R}_2$, one component consists of an isolated vertex, and the other component is connected but not strongly connected.

EXAMPLE 7.29

The graphs in Fig. 7.7 are examples of undirected graphs that are loop-free and have an edge for every pair of distinct vertices. These graphs illustrate the *complete graphs* on n vertices which are denoted by K_n. In Fig. 7.7 we have examples of the complete graphs on three, four, and five vertices, respectively. The complete graph K_2 consists of two vertices x, y and an edge connecting them, whereas the complete graph K_1 consists of one vertex and no edges because loops are not allowed.

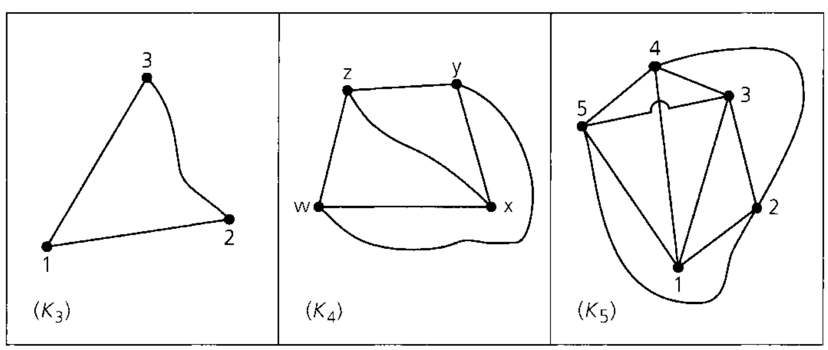


Figure 7.7

In this drawing of K_5 two edges cross, namely, $\{3, 5\}$ and $\{1, 4\}$. However, there is no point of intersection creating a new vertex. If we try to avoid the crossing of edges by drawing the graph differently, we run into the same problem all over again. This difficulty will be examined in Chapter 11 when we deal with the planarity of graphs.

A digraph G on a vertex set V gives rise to a relation $\mathcal{R}$ on V where $x \mathcal{R} y$ if (x, y) is an edge in G. Consequently, there is a (0, 1)-matrix for G, and since this relation matrix comes about from the adjacencies of pairs of vertices, it is referred to as the *adjacency matrix* for G as well as the relation matrix for $\mathcal{R}$.

At this point we tie together the properties of relations and the structure of directed graphs.

EXAMPLE 7.30

If $A = \{1, 2, 3\}$ and $\mathcal{R} = \{(1, 1), (1, 2), (2, 2), (3, 3), (3, 1)\}$, then $\mathcal{R}$ is a reflexive antisymmetric relation on A, but it is neither symmetric nor transitive. The directed graph associated with $\mathcal{R}$ consists of five edges. Three of these edges are loops that result from the reflexive property of $\mathcal{R}$. (See Fig. 7.8.) In general, if $\mathcal{R}$ is a relation on a finite set A, then $\mathcal{R}$ is reflexive if and only if its directed graph contains a loop at each vertex (element of A).

EXAMPLE 7.31

The relation $\mathcal{R} = \{(1, 1), (1, 2), (2, 1), (2, 3), (3, 2)\}$ is symmetric on $A = \{1, 2, 3\}$, but it is not reflexive, antisymmetric, or transitive. The directed graph for $\mathcal{R}$ is found in Fig. 7.9. In general, a relation $\mathcal{R}$ on a finite set A is symmetric if and only if its directed graph may be drawn so that it contains only loops and undirected edges.

EXAMPLE 7.32

For $A = \{1, 2, 3\}$, consider $\mathcal{R} = \{(1, 1), (1, 2), (2, 3), (1, 3)\}$. The directed graph for $\mathcal{R}$ is shown in Fig. 7.10. Here $\mathcal{R}$ is transitive and antisymmetric but not reflexive or symmetric. The directed graph indicates that a relation on a set A is transitive if and only if it satisfies the following: For all $x, y \in A$, if there is a (directed) path from x to y in the associated graph, then there is an edge (x, y) also. [Here (1, 2), (2, 3) is a (directed) path from 1 to 3, and we also have the edge (1, 3) for transitivity.] Notice that the directed graph in Fig. 7.3 of Example 7.26 also has this property.

The relation $\mathcal{R}$ is antisymmetric because there are no ordered pairs in $\mathcal{R}$ of the form (x, y) and (y, x) with $x \neq y$. To use the directed graph of Fig. 7.10 to characterize antisymmetry, we observe that for any two vertices x, y, with $x \neq y$, the graph contains at most one of the edges (x, y) or (y, x). Hence there are no undirected edges aside from loops.

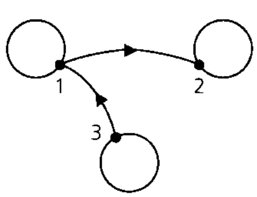


Figure 7.8

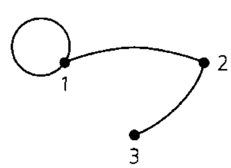


Figure 7.9

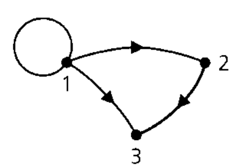


Figure 7.10

Our final example deals with equivalence relations.

EXAMPLE 7.33

For $A = \{1, 2, 3, 4, 5\}$, the following are equivalence relations on A:

$$\mathcal{R}_1 = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3), (3, 4), (4, 3), (4, 4), (5, 5)\},$$
$$\mathcal{R}_2 = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3), (4, 4), (4, 5), (5, 4), (5, 5)\}.$$

Their associated graphs are shown in Fig. 7.11. If we ignore the loops in each graph, we find the graph decomposed into components such as K_1, K_2, and K_3. In general, a relation on a finite set A is an equivalence relation if and only if its associated graph is one complete graph augmented by loops at every vertex or consists of the disjoint union of complete graphs augmented by loops at every vertex.

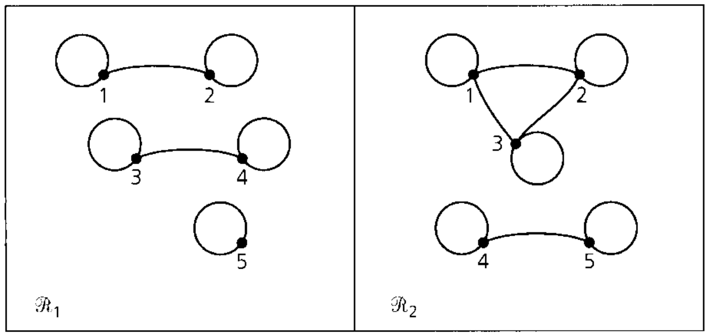


Figure 7.11

EXERCISES 7.2

1. For $A = \{1, 2, 3, 4\}$, let $\mathcal{R}$ and $\mathcal{S}$ be the relations on A defined by $\mathcal{R} = \{(1, 2), (1, 3), (2, 4), (4, 4)\}$ and $\mathcal{S} = \{(1, 1), (1, 2), (1, 3), (2, 3), (2, 4)\}$. Find $\mathcal{R} \circ \mathcal{S}$, $\mathcal{S} \circ \mathcal{R}$, $\mathcal{R}^2$, $\mathcal{R}^3$, $\mathcal{S}^2$, and $\mathcal{S}^3$.

2. If $\mathcal{R}$ is a reflexive relation on a set A, prove that $\mathcal{R}^2$ is also reflexive on A.

3. Provide a proof for the opposite inclusion in Theorem 7.1.

4. Let $A = \{1, 2, 3\}$, $B = \{w, x, y, z\}$, and $C = \{4, 5, 6\}$. Define the relations $\mathcal{R}_1 \subseteq A \times B$, $\mathcal{R}_2 \subseteq B \times C$, and $\mathcal{R}_3 \subseteq B \times C$, where $\mathcal{R}_1 = \{(1, w), (3, w), (2, x), (1, y)\}$, $\mathcal{R}_2 = \{(w, 5), (x, 6), (y, 4), (y, 6)\}$, and $\mathcal{R}_3 = \{(w, 4), (w, 5), (y, 5)\}$. (a) Determine $\mathcal{R}_1 \circ (\mathcal{R}_2 \cup \mathcal{R}_3)$ and $(\mathcal{R}_1 \circ \mathcal{R}_2) \cup (\mathcal{R}_1 \circ \mathcal{R}_3)$. (b) Determine $\mathcal{R}_1 \circ (\mathcal{R}_2 \cap \mathcal{R}_3)$ and $(\mathcal{R}_1 \circ \mathcal{R}_2) \cap (\mathcal{R}_1 \circ \mathcal{R}_3)$.

5. Let $A = \{1, 2\}$, $B = \{m, n, p\}$, and $C = \{3, 4\}$. Define the relations $\mathcal{R}_1 \subseteq A \times B$, $\mathcal{R}_2 \subseteq B \times C$, and $\mathcal{R}_3 \subseteq B \times C$ by $\mathcal{R}_1 = \{(1, m), (1, n), (1, p)\}$, $\mathcal{R}_2 = \{(m, 3), (m, 4), (p, 4)\}$, and $\mathcal{R}_3 = \{(m, 3), (m, 4), (p, 3)\}$. Determine $\mathcal{R}_1 \circ (\mathcal{R}_2 \cap \mathcal{R}_3)$ and $(\mathcal{R}_1 \circ \mathcal{R}_2) \cap (\mathcal{R}_1 \circ \mathcal{R}_3)$.

6. For sets A, B, and C, consider relations $\mathcal{R}_1 \subseteq A \times B$, $\mathcal{R}_2 \subseteq B \times C$, and $\mathcal{R}_3 \subseteq B \times C$. Prove that (a) $\mathcal{R}_1 \circ (\mathcal{R}_2 \cup \mathcal{R}_3) = (\mathcal{R}_1 \circ \mathcal{R}_2) \cup (\mathcal{R}_1 \circ \mathcal{R}_3)$; and (b) $\mathcal{R}_1 \circ (\mathcal{R}_2 \cap \mathcal{R}_3) \subseteq (\mathcal{R}_1 \circ \mathcal{R}_2) \cap (\mathcal{R}_1 \circ \mathcal{R}_3)$.

7. For a relation $\mathcal{R}$ on a set A, define $\mathcal{R}^0 = \{(a, a) | a \in A\}$. If $|A| = n$, prove that there exist $s, t \in \mathbf{N}$ with $0 \le s < t \le 2^{n^2}$ such that $\mathcal{R}^s = \mathcal{R}^t$.

8. With $A = \{1, 2, 3, 4\}$, let $\mathcal{R} = \{(1, 1), (1, 2), (2, 3), (3, 3), (3, 4), (4, 4)\}$ be a relation on A. Find two relations $\mathcal{S}$, $\mathcal{T}$ on A where $\mathcal{S} \neq \mathcal{T}$ but $\mathcal{R} \circ \mathcal{S} = \mathcal{R} \circ \mathcal{T} = \{(1, 1), (1, 2), (1, 4)\}$.

9. How many 6×6 (0, 1)-matrices A are there with $A = A^{tr}$?

10. If $E = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, how many (0, 1)-matrices F satisfy $E \le F$? How many (0, 1)-matrices G satisfy $G \le E$?

11. Consider the sets $A = \{a_1, a_2, \ldots, a_m\}$, $B = \{b_1, b_2, \ldots, b_n\}$, and $C = \{c_1, c_2, \ldots, c_p\}$, where the elements in each set remain fixed in the order given here. Let $\mathcal{R}_1$ be a relation from A to B, and let $\mathcal{R}_2$ be a relation from B to C. The relation matrix for $\mathcal{R}_i$ is $M(\mathcal{R}_i)$, where $i = 1, 2$. The rows and columns of these matrices are indexed by the elements from the appropriate sets A, B, and C according to the orders already prescribed. The matrix for $\mathcal{R}_1 \circ \mathcal{R}_2$ is the $m \times p$ matrix $M(\mathcal{R}_1 \circ \mathcal{R}_2)$, where the elements of A (in the order given) index the rows and the elements of C (also in the order given) index the columns.

Show that for all $1 \le i \le m$ and $1 \le j \le p$, the entries in the ith row and jth column of $M(\mathcal{R}_1) \cdot M(\mathcal{R}_2)$ and $M(\mathcal{R}_1 \circ \mathcal{R}_2)$ are equal. [Hence $M(\mathcal{R}_1) \cdot M(\mathcal{R}_2) = M(\mathcal{R}_1 \circ \mathcal{R}_2)$.]

12. Let A be a set with $|A| = n$, and consider the order for the listing of its elements as fixed. For $\mathcal{R} \subseteq A \times A$, let $M(\mathcal{R})$ denote the corresponding relation matrix.

a) Prove that $M(\mathcal{R}) = \mathbf{0}$ (the $n \times n$ matrix of all 0's) if and only if $\mathcal{R} = \emptyset$.

b) Prove that $M(\mathcal{R}) = \mathbf{1}$ (the $n \times n$ matrix of all 1's) if and only if $\mathcal{R} = A \times A$.

c) Use the result of Exercise 11, along with the Principle of Mathematical Induction, to prove that $M(\mathcal{R}^m) = [M(\mathcal{R})]^m$, for all $m \in \mathbf{Z}^+$.

13. Provide the proofs for Theorem 7.2(a), (b), and (d).

14. Use Theorem 7.2 to write a computer program (or to develop an algorithm) for the recognition of equivalence relations on a finite set.

15. a) Draw the digraph $G_1 = (V_1, E_1)$ where $V_1 = \{a, b, c, d, e, f\}$ and $E_1 = \{(a, b), (a, d), (b, c), (b, e), (d, b), (d, e), (e, c), (e, f), (f, d)\}$.

b) Draw the undirected graph $G_2 = (V_2, E_2)$ where $V_2 = \{s, t, u, v, w, x, y, z\}$ and $E_2 = \{\{s, t\}, \{s, u\}, \{s, x\}, \{t, u\}, \{t, w\}, \{u, w\}, \{u, x\}, \{v, w\}, \{v, x\}, \{v, y\}, \{w, z\}, \{x, y\}\}$.

16. For the directed graph $G = (V, E)$ in Fig. 7.12, classify each of the following statements as true or false.

a) Vertex c is the origin of two edges in G.

b) Vertex g is adjacent to vertex h.

c) There is a directed path in G from d to b.

d) There are two directed cycles in G.

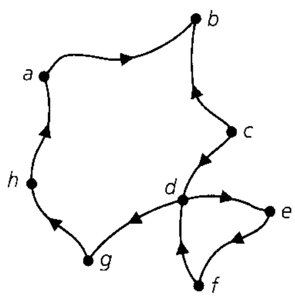


Figure 7.12

17. For $A = \{a, b, c, d, e, f\}$, each graph, or digraph, in Fig. 7.13 represents a relation $\mathcal{R}$ on A. Determine the relation $\mathcal{R} \subseteq A \times A$ in each case, as well as its associated relation matrix $M(\mathcal{R})$.

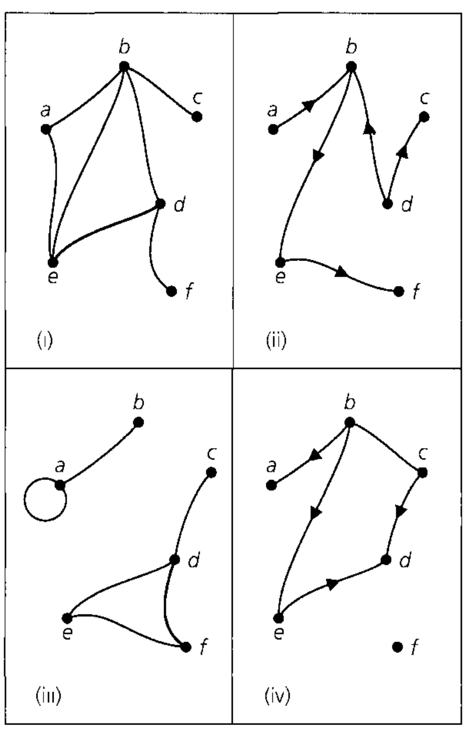


Figure 7.13

18. For $A = \{v, w, x, y, z\}$, each of the following is the (0, 1)-matrix for a relation $\mathcal{R}$ on A. Here the rows (from top to bottom) and the columns (from left to right) are indexed in the order v, w, x, y, z. Determine the relation $\mathcal{R} \subseteq A \times A$ in each case, and draw the directed graph G associated with $\mathcal{R}$.

a) $M(\mathcal{R}) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

b) $M(\mathcal{R}) = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$

19. For $A = \{1, 2, 3, 4\}$, let $\mathcal{R} = \{(1, 1), (1, 2), (2, 3), (3, 3), (3, 4)\}$ be a relation on A. Draw the directed graph G on A that is associated with $\mathcal{R}$. Do likewise for $\mathcal{R}^2$, $\mathcal{R}^3$, and $\mathcal{R}^4$.

20. a) Let $G = (V, E)$ be the directed graph where $V = \{1, 2, 3, 4, 5, 6, 7\}$ and $E = \{(i, j) | 1 \leq i < j \leq 7\}$.

i) How many edges are there for this graph?

ii) Four of the directed paths in G from 1 to 7 may be given as:

1) (1, 7);
2) (1, 3), (3, 5), (5, 6), (6, 7);
3) (1, 2), (2, 3), (3, 7); and
4) (1, 4), (4, 7).

How many directed paths (in total) exist in G from 1 to 7?

b) Now let $n \in \mathbf{Z}^+$ where $n \geq 2$, and consider the directed graph $G = (V, E)$ with $V = \{1, 2, 3, \ldots, n\}$ and $E = \{(i, j) | 1 \leq i < j \leq n\}$.

i) Determine $|E|$.

ii) How many directed paths exist in G from 1 to n?

iii) If $a, b \in \mathbf{Z}^+$ with $1 \leq a < b \leq n$, how many directed paths exist in G from a to b?

(The reader may wish to refer back to Exercise 20 in Section 3.1.)

21. Let $|A| = 5$. (a) How many directed graphs can one construct on A? (b) How many of the graphs in part (a) are actually undirected?

22. For $|A| = 5$, how many relations $\mathcal{R}$ on A are there? How many of these relations are symmetric?

23. a) Keeping the order of the elements fixed as 1, 2, 3, 4, 5, determine the (0, 1) relation matrix for each of the equivalence relations in Example 7.33.

b) Do the results of part (a) lead to any generalization?

24. How many (undirected) edges are there in the complete graphs K_6, K_7, and K_n, where $n \in \mathbf{Z}^+$?

25. Draw a precedence graph for the following segment found at the start of a computer program:

```
(s1) a := 1
(s2) b := 2
(s3) a := a + 3
(s4) c := b
(s5) a := 2 * a - 1
(s6) b := a * c
(s7) c := 7
(s8) d := c + 2
```

26. a) Let $\mathcal{R}$ be the relation on $A = \{1, 2, 3, 4, 5, 6, 7\}$, where the directed graph associated with $\mathcal{R}$ consists of the two components, each a directed cycle, shown in Fig. 7.14. Find the smallest integer $n > 1$, such that $\mathcal{R}^n = \mathcal{R}$. What is the smallest value of $n > 1$ for which the graph of $\mathcal{R}^n$ contains some loops? Does it ever happen that the graph of $\mathcal{R}^n$ consists of only loops?

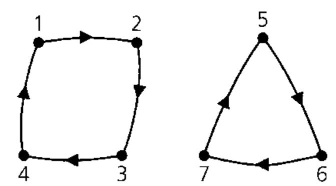


Figure 7.14

b) Answer the same questions from part (a) for the relation $\mathcal{R}$ on $A = \{1, 2, 3, \ldots, 9, 10\}$, if the directed graph associated with $\mathcal{R}$ is as shown in Fig. 7.15.

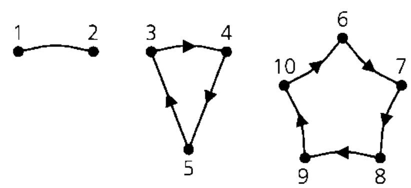


Figure 7.15

c) Do the results in parts (a) and (b) indicate anything in general?

27. If the complete graph K_n has 703 edges, how many vertices does it have?

7.3 Partial Orders: Hasse Diagrams

If you ask children to recite the numbers they know, you'll hear a uniform response of "1, 2, 3, ..." Without paying attention to it, they list these numbers in increasing order. In this section we take a closer look at this idea of order, something we may have taken for granted. We start with some observations about the sets **N**, **Z**, **Q**, **R**, and **C**.

The set **N** is closed under the binary operations of (ordinary) addition and multiplication, but if we seek an answer to the equation $x + 5 = 2$, we find that no element of **N** provides a solution. So we enlarge **N** to **Z**, where we can perform subtraction as well as addition and multiplication. However, we soon run into trouble trying to solve the equation $2x + 3 = 4$. Enlarging to **Q**, we can perform nonzero division in addition to the other operations. Yet this soon proves to be inadequate; the equation $x^2 - 2 = 0$ necessitates the introduction of the real but irrational numbers $\pm\sqrt{2}$. Even after we expand from **Q** to **R**, more trouble arises when we try to solve $x^2 + 1 = 0$. Finally we arrive at **C**, the complex numbers, where any polynomial equation of the form $c_n x^n + c_{n-1}x^{n-1} + \cdots + c_2x^2 + c_1x + c_0 = 0$, where $c_i \in \mathbf{C}$ for $0 \leq i \leq n$, $n > 0$ and $c_n \neq 0$, can be solved. (This result is known as the Fundamental Theorem of Algebra. Its proof requires material on functions of a complex variable, so no proof is given here.) As we kept building up from **N** to **C**, gaining more ability to solve polynomial equations, something was lost when we went from **R** to **C**. In **R**, given numbers r_1, r_2, with $r_1 \neq r_2$, we know that either $r_1 < r_2$ or $r_2 < r_1$. However, in **C** we have $(2 + i) \neq (1 + 2i)$, but what meaning can we attach to a statement such as "$(2 + i) < (1 + 2i)$"? We have lost the ability to "order" the elements in this number system!

As we start to take a closer look at the notion of order we proceed as in Section 7.1 and let A be a set with $\mathcal{R}$ a relation on A. The pair $(A, \mathcal{R})$ is called a *partially ordered set*, or *poset*, if relation $\mathcal{R}$ on A is a partial order, or a partial ordering relation (as given in Definition 7.6). If A is called a poset, we understand that there is a partial order $\mathcal{R}$ on A that makes A into this poset. Examples 7.1(a), 7.2, 7.11, and 7.15 are posets.

EXAMPLE 7.34

Let A be the set of courses offered at a college. Define the relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x, y are the same course or if x is a prerequisite for y. Then $\mathcal{R}$ makes A into a poset.

EXAMPLE 7.35

Define $\mathcal{R}$ on $A = \{1, 2, 3, 4\}$ by $x \mathcal{R} y$ if $x|y$ —that is, x (exactly) divides y. Then $\mathcal{R} = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (1, 3), (1, 4), (2, 4)\}$ is a partial order, and $(A, \mathcal{R})$ is a poset. (This is similar to what we learned in Example 7.15.)

EXAMPLE 7.36

In the construction of a house certain jobs, such as digging the foundation, must be performed before other phases of the construction can be undertaken. If A is a set of tasks that must be performed in building a house, we can define a relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x, y denote the same task or if task x must be performed before the start of task y. In this way we place an order on the elements of A, making it into a poset that is sometimes referred to as a PERT (Program Evaluation and Review Technique) network. (Such networks came into play during the 1950s in order to handle the complexities that arose in organizing the many individual activities required for the completion of projects on a very large scale. This technique was actually developed and first used by the U.S. Navy in order to coordinate the many projects that were necessary for the building of the Polaris submarine.)

Consider the diagrams given in Fig. 7.16. If part (a) were part of the directed graph associated with a relation $\mathcal{R}$, then because $(1, 2), (2, 1) \in \mathcal{R}$ with $1 \neq 2$, $\mathcal{R}$ could not be antisymmetric. For part (b), if the diagram were part of the graph of a transitive relation $\mathcal{R}$, then $(1, 2), (2, 3) \in \mathcal{R} \Rightarrow (1, 3) \in \mathcal{R}$. Since $(3, 1) \in \mathcal{R}$ and $1 \neq 3$, $\mathcal{R}$ is not antisymmetric, so it cannot be a partial order.

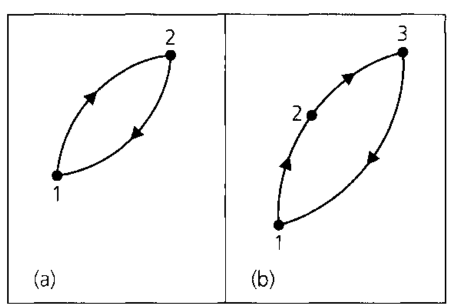


Figure 7.16

From these observations, if we are given a relation $\mathcal{R}$ on a set A, and we let G be the directed graph associated with $\mathcal{R}$, then we find that:

i) If G contains a pair of edges of the form (a, b), (b, a), for $a, b \in A$ with $a \neq b$, or

ii) If $\mathcal{R}$ is transitive and G contains a directed cycle (of length greater than or equal to three),

then the relation $\mathcal{R}$ cannot be antisymmetric, so $(A, \mathcal{R})$ fails to be a partial order.

EXAMPLE 7.37

Consider the directed graph for the partial order in Example 7.35. Figure 7.17(a) is the graphical representation of $\mathcal{R}$. In part (b) of the figure, we have a somewhat simpler diagram, which is called the *Hasse diagram* for $\mathcal{R}$.

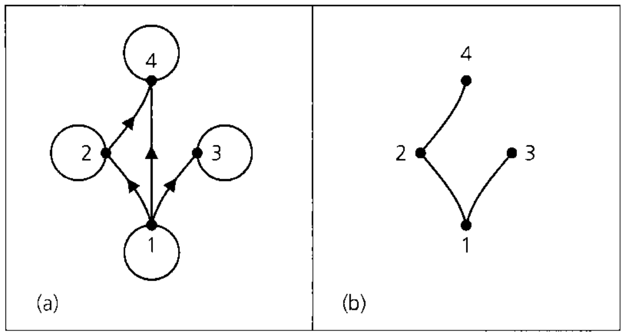


Figure 7.17

When we know that a relation $\mathcal{R}$ is a partial order on a set A, we can eliminate the loops at the vertices of its directed graph. Since $\mathcal{R}$ is also transitive, having the edges (1, 2) and (2, 4) is enough to insure the existence of edge (1, 4), so we need not include that edge. In this way we obtain the diagram in Fig. 7.17(b), where we have not lost the directions on the edges — the directions are assumed to go from the bottom to the top.

In general, if $\mathcal{R}$ is a partial order on a finite set A, we construct a Hasse diagram for $\mathcal{R}$ on A by drawing a line segment from x *up to* y, if $x, y \in A$ with $x \mathcal{R} y$ and, most important, if there is no other element $z \in A$ such that $x \mathcal{R} z$ and $z \mathcal{R} y$. (So there is nothing "in between" x and y.) If we adopt the convention of reading the diagram from bottom to top, then it is not necessary to direct any edges.

EXAMPLE 7.38

In Fig. 7.18 we have the Hasse diagrams for the following four posets. (a) With $\mathcal{U} = \{1, 2, 3\}$ and $A = \mathcal{P}(\mathcal{U})$, $\mathcal{R}$ is the subset relation on A. (b) Here $\mathcal{R}$ is the "(exactly) divides" relation

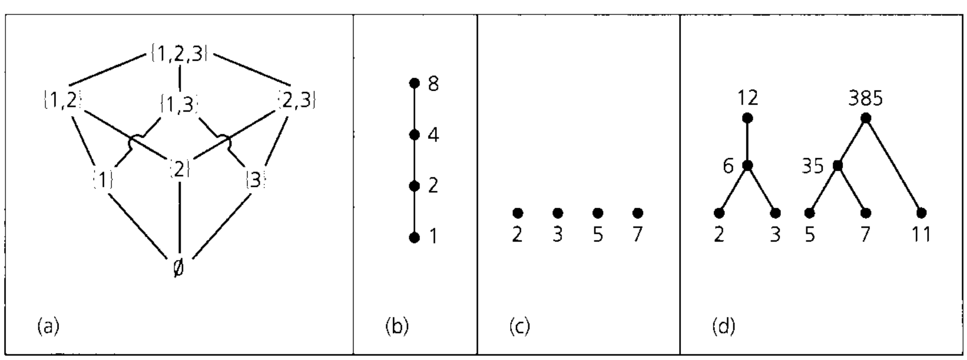


Figure 7.18

applied to $A = \{1, 2, 4, 8\}$. (c) and (d) Here the same relation as in part (b) is applied to $\{2, 3, 5, 7\}$ in part (c) and to $\{2, 3, 5, 6, 7, 11, 12, 35, 385\}$ in part (d). In part (c) we note that a Hasse diagram can have all isolated vertices; it can also have two (or more) connected pieces, as shown in part (d).

EXAMPLE 7.39

Let $A = \{1, 2, 3, 4, 5\}$. The relation $\mathcal{R}$ on A, defined by $x \mathcal{R} y$ if $x \leq y$, is a partial order. This makes A into a poset that we can denote by $(A, \leq)$. If $B = \{1, 2, 4\} \subset A$, then the set $(B \times B) \cap \mathcal{R} = \{(1, 1), (2, 2), (4, 4), (1, 2), (1, 4), (2, 4)\}$ is a partial order on B.

In general if $\mathcal{R}$ is a partial order on A, then for each subset B of A, $(B \times B) \cap \mathcal{R}$ makes B into a poset where the partial order on B is induced from $\mathcal{R}$.

We turn now to a special type of partial order.

Definition 7.16

If $(A, \mathcal{R})$ is a poset, we say that A is *totally ordered* (or, *linearly ordered*) if for all $x, y \in A$ either $x \mathcal{R} y$ or $y \mathcal{R} x$. In this case $\mathcal{R}$ is called a *total order* (or, a *linear order*).

EXAMPLE 7.40

a) On the set **N**, the relation $\mathcal{R}$ defined by $x \mathcal{R} y$ if $x \leq y$ is a total order.

b) The subset relation applied to $A = \mathcal{P}(\mathcal{U})$, where $\mathcal{U} = \{1, 2, 3\}$, is a partial, but not total, order: $\{1, 2\}, \{1, 3\} \in A$ but we have neither $\{1, 2\} \subseteq \{1, 3\}$ nor $\{1, 3\} \subseteq \{1, 2\}$.

c) The Hasse diagram in part (b) of Fig. 7.18 shows a total order. In Fig. 7.19(a) we have the directed graph for this total order—alongside its Hasse diagram in part (b).

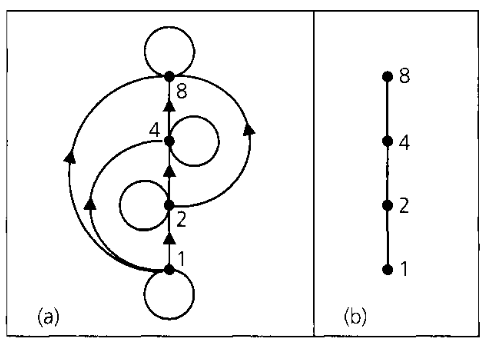


Figure 7.19

Could these notions of partial and total order ever arise in an industrial problem?

Say a toy manufacturer is about to market a new product and must include a set of instructions for its assembly. In order to assemble the new toy, there are seven tasks, denoted A, B, C, . . . , G, that one must perform in the partial order given by the Hasse diagram of Fig. 7.20. Here we see, for example, that all of the tasks B, A, and E must be completed before we can work on task C. Since the set of instructions is to consist of a listing of these tasks, numbered 1, 2, 3, . . . , 7, how can the manufacturer write the listing and make sure that the partial order of the Hasse diagram is maintained?

What we are really asking for here is whether we can take the partial order $\mathcal{R}$, given by the Hasse diagram, and find a total order $\mathcal{T}$ on these tasks for which $\mathcal{R} \subseteq \mathcal{T}$. The answer is yes, and the technique that we need is known as *topological sorting*.

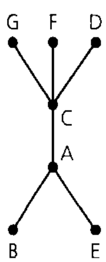


Figure 7.20

Topological Sorting Algorithm

(for a partial order $\mathcal{R}$ on a set A with $|A| = n$)

Step 1: Set $k = 1$. Let H_1 be the Hasse diagram of the partial order.

Step 2: Select a vertex v_k in H_k such that no (implicitly directed) edge in H_k starts at v_k.

Step 3: If $k = n$, the process is completed and we have a total order

$$\mathcal{T}: v_n < v_{n-1} < \cdots < v_2 < v_1$$

that contains $\mathcal{R}$.

If $k < n$, then remove from H_k the vertex v_k and all (implicitly directed) edges of H_k that terminate at v_k. Call the result H_{k+1}. Increase k by 1 and return to step (2).

Here we have presented our algorithm as a precise list of instructions, with no concern about the particulars of the pseudocode used in earlier chapters and with no reference to its implementation in a particular computer language.

Before we apply this algorithm† to the problem at hand, we should observe the deliberate use of "a" before the word "vertex" in step (2). This implies that the selection need not be unique and that we can get several different total orders $\mathcal{T}$ containing $\mathcal{R}$. Also, in step (3), for vertices v_{i-1} where $2 \leq i \leq n$, the notation $v_i < v_{i-1}$ is used because it is more suggestive of "v_i before v_{i-1}" than is the notation $v_i \, \mathcal{T} \, v_{i-1}$.

In Fig. 7.21, we show the Hasse diagrams that evolve as we apply the topological sorting algorithm to the partial order in Fig. 7.20. Below each diagram, the total order is listed as it evolves.

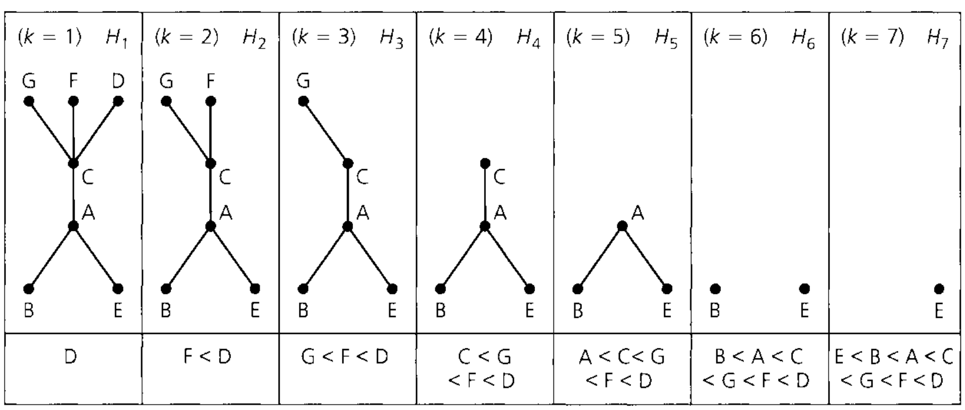


Figure 7.21

If the toy manufacturer writes the instructions in a list as 1-E, 2-B, 3-A, 4-C, 5-G, 6-F, 7-D, he or she will have a total order that preserves the partial order needed for correct assembly. This total order is one of 12 possible answers.

---

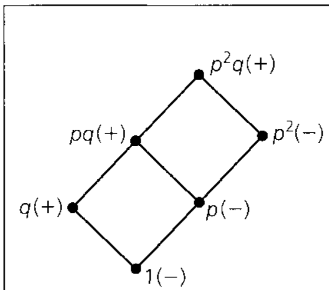
†Here we are only concerned with applying this algorithm. Hence we are assuming that it works and we shall not present a proof of that fact. Furthermore, we may operate similarly with other algorithms we encounter.

As is typical in discrete and combinatorial mathematics, this algorithm provides a procedure that reduces the size of the problem with each successive application.

The next example provides a situation where the number of distinct total orders for a particular partial order is determined.

EXAMPLE 7.41[†]

Let p, q be distinct primes. In part (a) of Fig. 7.22 we have the Hasse diagram for the partial order $\mathcal{R}$ of all positive-integer divisors of p^2q. Applying the topological sorting algorithm to this Hasse diagram, we find in Fig. 7.22(b) the five total orders $\mathcal{T}_i$, where $\mathcal{R} \subseteq \mathcal{T}_i$, for $1 \leq i \leq 5$.



(a)

$\mathcal{T}_1$: $p^2q > pq > q > p^2 > p > 1$
+, +, +, −, −, −
$\mathcal{T}_2$: $p^2q > pq > p^2 > p > q > 1$
+, +, −, −, +, −
$\mathcal{T}_3$: $p^2q > p^2 > pq > q > p > 1$
+, −, +, +, −, −
$\mathcal{T}_4$: $p^2q > pq > p^2 > q > p > 1$
+, +, −, +, −, −
$\mathcal{T}_5$: $p^2q > p^2 > pq > p > q > 1$
+, −, +, −, +, −

(b)

Figure 7.22

Now look at Fig. 7.22 again. This time focus on the three plus signs and three minus signs in part (a) of the figure and in the list below each total order in part (b). When we apply the topological sorting algorithm to the given partial order $\mathcal{R}$, step (2) of the algorithm implies that the first divisor selected is always p^2q. This accounts for the first plus sign in each $\mathcal{T}_i$, $1 \leq i \leq 5$. Continuing to apply the algorithm we get two more plus signs and the three minus signs.

Could there ever be more minus signs than plus signs in our corresponding list, as a total order is developed? For example, could we start with +, −, −,? If so, we have failed to correctly apply step (2) of the topological sorting algorithm — we should have recognized pq as the unique candidate to select after p^2q and p^2. In fact, for $0 \leq k \leq 2$, p^kq must be selected before p^k can be. Consequently, for each list of three plus signs and three minus signs, there is always at least as many plus signs as minus signs, as the list is read from left to right. Comparing now with the result in part (a) of Example 1.43, we see that the number of total orders for the given partial order is $5 = \frac{1}{3+1}\binom{2 \cdot 3}{3}$. Further, for $n \geq 1$, the topological sorting algorithm can be applied to the partial order of all positive divisors of $p^{n-1}q$ to yield $\frac{1}{n+1}\binom{2n}{n}$ total orders, another instance where the Catalan numbers arise.

In the topological sorting algorithm, we saw how the Hasse diagram was used in determining a total order containing a given poset $(A, \mathcal{R})$. This algorithm now prompts us to examine further properties of a partial order. At the start, particular emphasis will be given

[†]This example refers back to the optional material on Catalan numbers in Section 1.5. It may be skipped with no loss of continuity.

to a vertex like the vertex v_k in step (2) of the algorithm. The special property exhibited by such a vertex is now considered in the following.

Definition 7.17

If $(A, \mathcal{R})$ is a poset, then an element $x \in A$ is called a *maximal* element of A if for all $a \in A$, $a \neq x \Rightarrow x \not\mathcal{R}\, a$. An element $y \in A$ is called a *minimal* element of A if whenever $b \in A$ and $b \neq y$, then $b \not\mathcal{R}\, y$.

If we use the contrapositive of the first statement in Definition 7.17, then we can state that $x (\in A)$ is a maximal element if for each $a \in A$, $x \mathcal{R} a \Rightarrow x = a$. In a similar manner, $y \in A$ is a minimal element if for each $b \in A$, $b \mathcal{R} y \Rightarrow b = y$.

EXAMPLE 7.42

Let $\mathcal{U} = \{1, 2, 3\}$ and $A = \mathcal{P}(\mathcal{U})$.

a) Let $\mathcal{R}$ be the subset relation on A. Then $\mathcal{U}$ is maximal and $\emptyset$ is minimal for the poset $(A, \subseteq)$.

b) For B, the collection of proper subsets of $\{1, 2, 3\}$, let $\mathcal{R}$ be the subset relation on B. In the poset $(B, \subseteq)$, the sets $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$ are all maximal elements; $\emptyset$ is still the only minimal element.

EXAMPLE 7.43

With $\mathcal{R}$ the "less than or equal to" relation on the set $\mathbf{Z}$, we find that $(\mathbf{Z}, \leq)$ is a poset with neither a maximal nor a minimal element. The poset $(\mathbf{N}, \leq)$, however, has minimal element 0 but no maximal element.

EXAMPLE 7.44

When we look back at the partial orders in parts (b), (c), and (d) of Example 7.38, the following observations come to light.

1) The partial order in part (b) has the unique maximal element 8 and the unique minimal element 1.
2) Each of the four elements — 2, 3, 5, and 7 — is both a maximal element and a minimal element for the poset in part (c) of Example 7.38.
3) In part (d) the elements 12 and 385 are both maximal. Each of the elements 2, 3, 5, 7, and 11 is a minimal element for this partial order.

Are there any conditions indicating when a poset must have a maximal or minimal element?

THEOREM 7.3

If $(A, \mathcal{R})$ is a poset and A is finite, then A has both a maximal and a minimal element.

Proof: Let $a_1 \in A$. If there is no element $a \in A$ where $a \neq a_1$ and $a_1 \mathcal{R} a$, then a_1 is maximal. Otherwise there is an element $a_2 \in A$ with $a_2 \neq a_1$ and $a_1 \mathcal{R} a_2$. If no element $a \in A, a \neq a_2$, satisfies $a_2 \mathcal{R} a$, then a_2 is maximal. Otherwise we can find $a_3 \in A$ so that $a_3 \neq a_2$, $a_3 \neq a_1$ (Why?) while $a_1 \mathcal{R} a_2$ and $a_2 \mathcal{R} a_3$. Continuing in this manner, since A is finite, we get to an element $a_n \in A$ with $a_n \not\mathcal{R}\, a$ for all $a \in A$ where $a \neq a_n$, so a_n is maximal.

The proof for a minimal element follows in a similar way.

Returning now to the topological sorting algorithm, we see that in each iteration of step (2) of the algorithm, we are selecting a maximal element from the original poset $(A, \mathcal{R})$, or a poset of the form $(B, \mathcal{R}')$ where $\emptyset \neq B \subset A$ and $\mathcal{R}' = (B \times B) \cap \mathcal{R}$. At least one such element exists (in each iteration) by virtue of Theorem 7.3. Then in the second part of step (3), if x is the maximal element selected [in step (2)], we remove from the present poset all elements of the form (a, x). This results in a smaller poset.

We turn now to the study of some additional concepts involving posets.

Definition 7.18

If $(A, \mathcal{R})$ is a poset, then an element $x \in A$ is called a *least* element if $x \mathcal{R} a$ for all $a \in A$. Element $y \in A$ is called a *greatest* element if $a \mathcal{R} y$ for all $a \in A$.

EXAMPLE 7.45

Let $\mathcal{U} = \{1, 2, 3\}$, and let $\mathcal{R}$ be the subset relation.

a) With $A = \mathcal{P}(\mathcal{U})$, the poset $(A, \subseteq)$ has $\emptyset$ as a least element and $\mathcal{U}$ as a greatest element.

b) For B = the collection of nonempty subsets of $\mathcal{U}$, the poset $(B, \subseteq)$ has $\mathcal{U}$ as a greatest element. There is no least element here, but there are three minimal elements.

EXAMPLE 7.46

For the partial orders in Example 7.38, we find that

1) The partial order in part (b) has a greatest element 8 and a least element 1.
2) There is no greatest element or least element for the poset in part (c).
3) No greatest element or least element exists for the partial order in part (d).

We have seen that it is possible for a poset to have several maximal and minimal elements. What about least and greatest elements?

THEOREM 7.4

If the poset $(A, \mathcal{R})$ has a greatest (least) element, then that element is unique.

Proof: Suppose that $x, y \in A$ and that both are greatest elements. Since x is a greatest element, $y \mathcal{R} x$. Likewise, $x \mathcal{R} y$ because y is a greatest element. As $\mathcal{R}$ is antisymmetric, it follows that $x = y$.

The proof for the least element is similar.

Definition 7.19

Let $(A, \mathcal{R})$ be a poset with $B \subseteq A$. An element $x \in A$ is called a *lower bound* of B if $x \mathcal{R} b$ for all $b \in B$. Likewise, an element $y \in A$ is called an *upper bound* of B if $b \mathcal{R} y$ for all $b \in B$.

An element $x' \in A$ is called a *greatest lower bound* (glb) of B if it is a lower bound of B and if for all other lower bounds x'' of B we have $x'' \mathcal{R} x'$. Similarly $y' \in A$ is a *least upper bound* (lub) of B if it is an upper bound of B and if $y' \mathcal{R} y''$ for all other upper bounds y'' of B.

EXAMPLE 7.47

Let $\mathcal{U} = \{1, 2, 3, 4\}$, with $A = \mathcal{P}(\mathcal{U})$, and let $\mathcal{R}$ be the subset relation on A. If $B = \{\{1\}, \{2\}, \{1, 2\}\}$, then $\{1, 2\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, and $\{1, 2, 3, 4\}$ are all upper bounds for

B (in $(A, \mathcal{R})$), whereas $\{1, 2\}$ is a least upper bound (and is in B). Meanwhile, a greatest lower bound for B is $\emptyset$, which is not in B.

EXAMPLE 7.48

Let $\mathcal{R}$ be the "less than or equal to" relation for the poset $(A, \mathcal{R})$.

a) If $A = \mathbf{R}$ and $B = [0, 1]$, then B has glb 0 and lub 1. Note that $0, 1 \in B$. For $C = (0, 1]$, C has glb 0 and lub 1, and $1 \in C$ but $0 \notin C$.

b) Keeping $A = \mathbf{R}$, let $B = \{q \in \mathbf{Q} | q^2 < 2\}$. Then B has $\sqrt{2}$ as a lub and $-\sqrt{2}$ as a glb, and neither of these real numbers is in B.

c) Now let $A = \mathbf{Q}$, with B as in part (b). Here B has no lub or glb.

These examples lead us to the following result.

THEOREM 7.5

If $(A, \mathcal{R})$ is a poset and $B \subseteq A$, then B has at most one lub (glb).

Proof: We leave the proof to the reader.

We close this section with one last ordered structure.

Definition 7.20

The poset $(A, \mathcal{R})$ is called a *lattice* if for all $x, y \in A$ the elements lub$\{x, y\}$ and glb$\{x, y\}$ both exist in A.

EXAMPLE 7.49

For $A = \mathbf{N}$ and $x, y \in \mathbf{N}$, define $x \mathcal{R} y$ by $x \leq y$. Then lub$\{x, y\} = \max\{x, y\}$, glb$\{x, y\} = \min\{x, y\}$, and $(\mathbf{N}, \leq)$ is a lattice.

EXAMPLE 7.50

For the poset in Example 7.45(a), if $S, T \subseteq \mathcal{U}$, with lub$\{S, T\} = S \cup T$ and glb$\{S, T\} = S \cap T$, then $(\mathcal{P}(\mathcal{U}), \subseteq)$ is a lattice.

EXAMPLE 7.51

Consider the poset in Example 7.38(d). Here we find, for example, that

$$\text{lub}\{2, 3\} = 6,\ \text{lub}\{3, 6\} = 6,\ \text{lub}\{5, 7\} = 35,\ \text{lub}\{7, 11\} = 385,\ \text{lub}\{11, 35\} = 385,$$

and

$$\text{glb}\{3, 6\} = 3,\ \text{glb}\{2, 12\} = 2,\ \text{glb}\{35, 385\} = 35.$$

However, even though lub$\{2, 3\}$ exists, there is no glb for the elements 2 and 3. In addition, we are also lacking (among other considerations) glb$\{5, 7\}$, glb$\{11, 35\}$, glb$\{3, 35\}$, and lub$\{3, 35\}$. Consequently, this partial order is not a lattice.

EXERCISES 7.3

1. Draw the Hasse diagram for the poset $(\mathcal{P}(\mathcal{U}), \subseteq)$, where $\mathcal{U} = \{1, 2, 3, 4\}$.

2. Let $A = \{1, 2, 3, 6, 9, 18\}$, and define $\mathcal{R}$ on A by $x \mathcal{R} y$ if $x | y$. Draw the Hasse diagram for the poset $(A, \mathcal{R})$.

3. Let $(A, \mathcal{R}_1)$, $(B, \mathcal{R}_2)$ be two posets. On $A \times B$, define relation $\mathcal{R}$ by $(a, b) \mathcal{R} (x, y)$ if $a \mathcal{R}_1 x$ and $b \mathcal{R}_2 y$. Prove that $\mathcal{R}$ is a partial order.

4. If $\mathcal{R}_1$, $\mathcal{R}_2$ in Exercise 3 are total orders, is $\mathcal{R}$ a total order?

5. Topologically sort the Hasse diagram in part (a) of Example 7.38.

6. For $A = \{a, b, c, d, e\}$, the Hasse diagram for the poset $(A, \mathcal{R})$ is shown in Fig. 7.23. (a) Determine the relation matrix for $\mathcal{R}$. (b) Construct the directed graph G (on A) that is associated with $\mathcal{R}$. (c) Topologically sort the poset $(A, \mathcal{R})$.

7. The directed graph G for a relation $\mathcal{R}$ on set $A = \{1, 2, 3, 4\}$ is shown in Fig. 7.24. (a) Verify that $(A, \mathcal{R})$ is a poset and find its Hasse diagram. (b) Topologically sort $(A, \mathcal{R})$. (c) How many more directed edges are needed in Fig. 7.24 to extend $(A, \mathcal{R})$ to a total order?

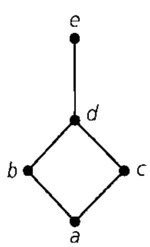


Figure 7.23

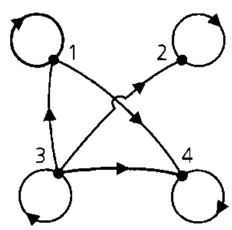


Figure 7.24

8. Prove that if a poset $(A, \mathcal{R})$ has a least element, it is unique.

9. Prove Theorem 7.5.

10. Give an example of a poset with four maximal elements but no greatest element.

11. If $(A, \mathcal{R})$ is a poset but not a total order, and $\emptyset \neq B \subset A$, does it follow that $(B \times B) \cap \mathcal{R}$ makes B into a poset but not a total order?

12. If $\mathcal{R}$ is a relation on A, and G is the associated directed graph, how can one recognize from G that $(A, \mathcal{R})$ is a total order?

13. If G is the directed graph for a relation $\mathcal{R}$ on A, with $|A| = n$, and $(A, \mathcal{R})$ is a total order, how many edges (including loops) are there in G?

14. Let $M(\mathcal{R})$ be the relation matrix for relation $\mathcal{R}$ on A, with $|A| = n$. If $(A, \mathcal{R})$ is a total order, how many 1's appear in $M(\mathcal{R})$?

15. a) Describe the structure of the Hasse diagram for a totally ordered poset $(A, \mathcal{R})$, where $|A| = n \geq 1$.

b) For a set A where $|A| = n \geq 1$, how many relations on A are total orders?

16. a) For $A = \{a_1, a_2, \ldots, a_n\}$, let $(A, \mathcal{R})$ be a poset. If $M(\mathcal{R})$ is the corresponding relation matrix, how can we recognize a maximal or minimal element of the poset from $M(\mathcal{R})$?

b) How can one recognize the existence of a greatest or least element in $(A, \mathcal{R})$ from the relation matrix $M(\mathcal{R})$?

17. Let $\mathcal{U} = \{1, 2, 3, 4\}$, with $A = \mathcal{P}(\mathcal{U})$, and let $\mathcal{R}$ be the subset relation on A. For each of the following subsets B (of A), determine the lub and glb of B.

a) $B = \{\{1\}, \{2\}\}$

b) $B = \{\{1\}, \{2\}, \{3\}, \{1, 2\}\}$

c) $B = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$

d) $B = \{\{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}\}$

e) $B = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$

18. Let $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7\}$, with $A = \mathcal{P}(\mathcal{U})$, and let $\mathcal{R}$ be the subset relation on A. For $B = \{\{1\}, \{2\}, \{2, 3\}\} \subseteq A$, determine each of the following.

a) The number of upper bounds of B that contain (i) three elements of $\mathcal{U}$; (ii) four elements of $\mathcal{U}$; (iii) five elements of $\mathcal{U}$

b) The number of upper bounds that exist for B

c) The lub for B

d) The number of lower bounds that exist for B

e) The glb for B

19. Define the relation $\mathcal{R}$ on the set $\mathbf{Z}$ by $a \mathcal{R} b$ if $a - b$ is a nonnegative even integer. Verify that $\mathcal{R}$ defines a partial order for $\mathbf{Z}$. Is this partial order a total order?

20. For $X = \{0, 1\}$, let $A = X \times X$. Define the relation $\mathcal{R}$ on A by $(a, b) \mathcal{R} (c, d)$ if (i) $a < c$; or (ii) $a = c$ and $b \leq d$. (a) Prove that $\mathcal{R}$ is a partial order for A. (b) Determine all minimal and maximal elements for this partial order. (c) Is there a least element? Is there a greatest element? (d) Is this partial order a total order?

21. Let $X = \{0, 1, 2\}$ and $A = X \times X$. Define the relation $\mathcal{R}$ on A as in Exercise 20. Answer the same questions posed in Exercise 20 for this relation $\mathcal{R}$ and set A.

22. For $n \in \mathbf{Z}^+$, let $X = \{0, 1, 2, \ldots, n-1, n\}$ and $A = X \times X$. Define the relation $\mathcal{R}$ on A as in Exercise 20. Remember that each element in this total order $\mathcal{R}$ is an ordered pair whose components are themselves ordered pairs. How many such elements are there in $\mathcal{R}$?

23. Let $(A, \mathcal{R})$ be a poset. Prove or disprove each of the following statements.

a) If $(A, \mathcal{R})$ is a lattice, then it is a total order.

b) If $(A, \mathcal{R})$ is a total order, then it is a lattice.

24. If $(A, \mathcal{R})$ is a lattice, with A finite, prove that $(A, \mathcal{R})$ has a greatest element and a least element.

25. For $A = \{a, b, c, d, e, v, w, x, y, z\}$, consider the poset $(A, \mathcal{R})$ whose Hasse diagram is shown in Fig. 7.25. Find

a) $\text{glb}\{b, c\}$

b) $\text{glb}\{b, w\}$

c) $\text{glb}\{e, x\}$

d) $\text{lub}\{c, b\}$

e) $\text{lub}\{d, x\}$

f) $\text{lub}\{c, e\}$

g) $\text{lub}\{a, v\}$

Is $(A, \mathcal{R})$ a lattice? Is there a maximal element? a minimal element? a greatest element? a least element?

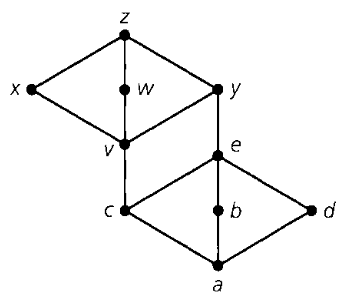


Figure 7.25

26. Given partial orders $(A, \mathcal{R})$ and $(B, \mathcal{S})$, a function $f: A \rightarrow B$ is called *order-preserving* if for all $x, y \in A$, $x \mathcal{R} y \Rightarrow f(x) \mathcal{S} f(y)$. How many such order-preserving functions are there for each of the following, where $\mathcal{R}$, $\mathcal{S}$ both denote $\leq$ (the usual "less than or equal to" relation)?

a) $A = \{1, 2, 3, 4\}$, $B = \{1, 2\}$;

b) $A = \{1, \ldots, n\}$, $n \geq 1$, $B = \{1, 2\}$;

c) $A = \{a_1, a_2, \ldots, a_n\} \subset \mathbf{Z}^+$, $n \geq 1$, $a_1 < a_2 < \cdots < a_n$, $B = \{1, 2\}$;

d) $A = \{1, 2\}$, $B = \{1, 2, 3, 4\}$;

e) $A = \{1, 2\}$, $B = \{1, \ldots, n\}$, $n \geq 1$; and

f) $A = \{1, 2\}$, $B = \{b_1, b_2, \ldots, b_n\} \subset \mathbf{Z}^+$, $n \geq 1$, $b_1 < b_2 < \cdots < b_n$.

27. Let p, q, r, s be four distinct primes and $m, n, k, \ell \in \mathbf{Z}^+$. How many edges are there in the Hasse diagram of all positive divisors of (a) p^3; (b) p^m; (c) p^3q^2; (d) p^mq^n; (e) $p^3q^2r^4$; (f) $p^mq^nr^k$; (g) $p^3q^2r^4s^7$; and (h) $p^mq^nr^ks^\ell$?

28. Find the number of ways to totally order the partial order of all positive-integer divisors of (a) 24; (b) 75; and (c) 1701.

29. Let p, q be distinct primes and $k \in \mathbf{Z}^+$. If there are 429 ways to totally order the partial order of positive-integer divisors of p^kq, how many positive-integer divisors are there for this partial order?

30. For $m, n \in \mathbf{Z}^+$, let A be the set of all $m \times n$ (0, 1)-matrices. Prove that the "precedes" relation of Definition 7.11 makes A into a poset.

7.4 Equivalence Relations and Partitions

As we noted earlier in Definition 7.7, a relation $\mathcal{R}$ on a set A is an equivalence relation if it is reflexive, symmetric, and transitive. For any set $A \neq \emptyset$, the relation of equality is an equivalence relation on A, where two elements of A are related if they are identical; equality thus establishes the property of "sameness" among the elements of A.

If we consider the relation $\mathcal{R}$ on $\mathbf{Z}$ defined by $x \mathcal{R} y$ if $x - y$ is a multiple of 2, then $\mathcal{R}$ is an equivalence relation on $\mathbf{Z}$ where all even integers are related, as are all odd integers. Here, for example, we do not have $4 = 8$, but we do have $4 \mathcal{R} 8$, for we no longer care about the size of a number but are concerned with only two properties: "evenness" and "oddness." This relation splits $\mathbf{Z}$ into two subsets consisting of the odd and even integers: $\mathbf{Z} = \{\ldots, -3, -1, 1, 3, \ldots\} \cup \{\ldots, -4, -2, 0, 2, 4, \ldots\}$. This splitting up of $\mathbf{Z}$ is an example of a partition, a concept closely related to the equivalence relation. In this section we investigate this relationship and see how it helps us count the number of equivalence relations on a finite set.

Definition 7.21 Given a set A and index set I, let $\emptyset \neq A_i \subseteq A$ for each $i \in I$. Then $\{A_i\}_{i \in I}$ is a *partition* of A if

a) $A = \bigcup_{i \in I} A_i$ and **b)** $A_i \cap A_j = \emptyset$, for all $i, j \in I$ where $i \neq j$.

Each subset A_i is called a *cell* or *block* of the partition.

EXAMPLE 7.52 If $A = \{1, 2, 3, \ldots, 10\}$, then each of the following determines a partition of A:

a) $A_1 = \{1, 2, 3, 4, 5\}$, $A_2 = \{6, 7, 8, 9, 10\}$

b) $A_1 = \{1, 2, 3\}$, $A_2 = \{4, 6, 7, 9\}$, $A_3 = \{5, 8, 10\}$

c) $A_i = \{i, i + 5\}$, $1 \le i \le 5$

In these three examples we note how each element of A belongs to *exactly one* cell in each partition.

EXAMPLE 7.53

Let $A = \mathbf{R}$ and, for each $i \in \mathbf{Z}$, let $A_i = [i, i + 1)$. Then $\{A_i\}_{i \in \mathbf{Z}}$ is a partition of $\mathbf{R}$.

Now just how do partitions come into play with equivalence relations?

Definition 7.22

Let $\mathcal{R}$ be an equivalence relation on a set A. For each $x \in A$, the *equivalence class* of x, denoted $[x]$, is defined by $[x] = \{y \in A | y \mathcal{R} x\}$.

EXAMPLE 7.54

Define the relation $\mathcal{R}$ on $\mathbf{Z}$ by $x \mathcal{R} y$ if $4|(x - y)$. Since $\mathcal{R}$ is reflexive, symmetric, and transitive, it is an equivalence relation and we find that

$$[0] = \{\ldots, -8, -4, 0, 4, 8, 12, \ldots\} = \{4k | k \in \mathbf{Z}\}$$
$$[1] = \{\ldots, -7, -3, 1, 5, 9, 13, \ldots\} = \{4k + 1 | k \in \mathbf{Z}\}$$
$$[2] = \{\ldots, -6, -2, 2, 6, 10, 14, \ldots\} = \{4k + 2 | k \in \mathbf{Z}\}$$
$$[3] = \{\ldots, -5, -1, 3, 7, 11, 15, \ldots\} = \{4k + 3 | k \in \mathbf{Z}\}.$$

But what about $[n]$, where n is an integer other than 0, 1, 2, or 3? For example, what is $[6]$? We claim that $[6] = [2]$ and to prove this we use Definition 3.2 (for the equality of sets) as follows. If $x \in [6]$, then from Definition 7.22 we know that $x \mathcal{R} 6$. Here this means that 4 divides $(x - 6)$, so $x - 6 = 4k$ for some $k \in \mathbf{Z}$. But then $x - 6 = 4k \Rightarrow x - 2 = 4(k + 1) \Rightarrow$ 4 divides $(x - 2) \Rightarrow x \mathcal{R} 2 \Rightarrow x \in [2]$, so $[6] \subseteq [2]$. For the opposite inclusion start with an element y in $[2]$. Then $y \in [2] \Rightarrow y \mathcal{R} 2 \Rightarrow$ 4 divides $(y - 2) \Rightarrow y - 2 = 4l$ for some $l \in \mathbf{Z} \Rightarrow y - 6 = 4(l - 1)$, where $l - 1 \in \mathbf{Z} \Rightarrow$ 4 divides $y - 6 \Rightarrow y \mathcal{R} 6 \Rightarrow y \in [6]$, so $[2] \subseteq [6]$. From the two inclusions it now follows that $[6] = [2]$, as claimed.

Further, we also find, for example, that $[2] = [-2] = [-6]$, $[51] = [3]$, and $[17] = [1]$. Most important, $\{[0], [1], [2], [3]\}$ provides a partition of $\mathbf{Z}$.

[*Note*: Here the index set for the partition is implicit. If, for instance, we let $A_0 = [0]$, $A_1 = [1]$, $A_2 = [2]$, and $A_3 = [3]$, then one possible index set I (as in Definition 7.21) is $\{0, 1, 2, 3\}$. When a collection of sets is called a partition (of a given set) but no index set is specified, the reader should realize that the situation is like the one given here — where the index set is implicit.]

EXAMPLE 7.55

Define the relation $\mathcal{R}$ on the set $\mathbf{Z}$ by $a \mathcal{R} b$ if $a^2 = b^2$ (or, $a = \pm b$). For all $a \in \mathbf{Z}$, we have $a^2 = a^2$ — so $a \mathcal{R} a$ and $\mathcal{R}$ is reflexive. Should $a, b \in \mathbf{Z}$ with $a \mathcal{R} b$, then $a^2 = b^2$ and it follows that $b^2 = a^2$, or $b \mathcal{R} a$. Consequently, relation $\mathcal{R}$ is symmetric. Finally, suppose that $a, b, c \in \mathbf{Z}$ with $a \mathcal{R} b$ and $b \mathcal{R} c$. Then $a^2 = b^2$ and $b^2 = c^2$, so $a^2 = c^2$ and $a \mathcal{R} c$. This makes the given relation transitive. Having established the three needed properties, we now know that $\mathcal{R}$ is an equivalence relation.

What can we say about the corresponding partition of $\mathbf{Z}$?

Here one finds that $[0] = \{0\}$, $[1] = [-1] = \{-1, 1\}$, $[2] = [-2] = \{-2, 2\}$, and, in general, for each $n \in \mathbf{Z}^+$, $[n] = [-n] = \{-n, n\}$. Furthermore, we have the *partition*

$$\mathbf{Z} = \bigcup_{n=0}^{\infty}[n] = \bigcup_{n\in\mathbf{N}}[n] = \{0\} \cup \left(\bigcup_{n=1}^{\infty}\{-n, n\}\right) = \{0\} \cup \left(\bigcup_{n\in\mathbf{Z}^+}\{-n, n\}\right).$$

These examples lead us to the following general situation.

THEOREM 7.6 If $\mathcal{R}$ is an equivalence relation on a set A, and $x, y \in A$, then (a) $x \in [x]$; (b) $x \mathcal{R} y$ if and only if $[x] = [y]$; and (c) $[x] = [y]$ or $[x] \cap [y] = \emptyset$.

Proof:

a) This result follows from the reflexive property of $\mathcal{R}$.

b) The proof here is somewhat reminiscent of what was done in Example 7.54.

If $x \mathcal{R} y$, let $w \in [x]$. Then $w \mathcal{R} x$ and because $\mathcal{R}$ is transitive, $w \mathcal{R} y$. Hence $w \in [y]$ and $[x] \subseteq [y]$. With $\mathcal{R}$ symmetric, $x \mathcal{R} y \Rightarrow y \mathcal{R} x$. So if $t \in [y]$, then $t \mathcal{R} y$ and by the transitive property, $t \mathcal{R} x$. Hence $t \in [x]$ and $[y] \subseteq [x]$. Consequently, $[x] = [y]$.

Conversely, let $[x] = [y]$. Since $x \in [x]$ by part (a), then $x \in [y]$ or $x \mathcal{R} y$.

c) This property tells us that two equivalence classes can be related in only one of two possible ways. Either they are identical or they are disjoint.

We assume that $[x] \neq [y]$ and show how it then follows that $[x] \cap [y] = \emptyset$. If $[x] \cap [y] \neq \emptyset$, then let $v \in A$ with $v \in [x]$ and $v \in [y]$. Then $v \mathcal{R} x$, $v \mathcal{R} y$, and, since $\mathcal{R}$ is symmetric, $x \mathcal{R} v$. Now $(x \mathcal{R} v$ and $v \mathcal{R} y) \Rightarrow x \mathcal{R} y$, by the transitive property. Also $x \mathcal{R} y \Rightarrow [x] = [y]$ by part (b). This contradicts the assumption that $[x] \neq [y]$, so we reject the supposition that $[x] \cap [y] \neq \emptyset$, and the result follows.

Note that if $\mathcal{R}$ is an equivalence relation on A, then by parts (a) and (c) of Theorem 7.6 the distinct equivalence classes determined by $\mathcal{R}$ provide us with a partition of A.

EXAMPLE 7.56

a) If $A = \{1, 2, 3, 4, 5\}$ and $\mathcal{R} = \{(1, 1), (2, 2), (2, 3), (3, 2), (3, 3), (4, 4), (4, 5), (5, 4), (5, 5)\}$, then $\mathcal{R}$ is an equivalence relation on A. Here $[1] = \{1\}$, $[2] = \{2, 3\} = [3]$, $[4] = \{4, 5\} = [5]$, and $A = [1] \cup [2] \cup [4]$ with $[1] \cap [2] = \emptyset$, $[1] \cap [4] = \emptyset$, and $[2] \cap [4] = \emptyset$. So $\{[1], [2], [4]\}$ determines a partition of A.

b) Consider part (d) of Example 7.16 once again. We have $A = \{1, 2, 3, 4, 5, 6, 7\}$, $B = \{x, y, z\}$, and $f: A \to B$ is the onto function

$$f = \{(1, x), (2, z), (3, x), (4, y), (5, z), (6, y), (7, x)\}.$$

The relation $\mathcal{R}$ defined on A by $a \mathcal{R} b$ if $f(a) = f(b)$ was shown to be an equivalence relation. Here

$$f^{-1}(x) = \{1, 3, 7\} = [1]\ (= [3] = [7]),$$
$$f^{-1}(y) = \{4, 6\} = [4]\ (= [6]), \quad \text{and}$$
$$f^{-1}(z) = \{2, 5\} = [2]\ (= [5]).$$

With $A = [1] \cup [4] \cup [2] = f^{-1}(x) \cup f^{-1}(y) \cup f^{-1}(z)$, we see that $\{f^{-1}(x), f^{-1}(y), f^{-1}(z)\}$ determines a partition of A.

In fact, for any nonempty sets A, B, if $f: A \to B$ is an onto function, then $A = \bigcup_{b\in B} f^{-1}(b)$ and $\{f^{-1}(b) | b \in B\}$ provides us with a partition of A.

EXAMPLE 7.57

In the programming language C++ a nonexecutable specification statement called the *union construct* allows two or more variables in a given program to refer to the same memory location.

For example, within a program the statements

```
union
{
 int a;
 int c;
 int p;
};
union
{
 int up;
 int down;
};
```

inform the C++ compiler that the integer variables a, c, and p will share one memory location while the integer variables *up* and *down* will share another. Here the set of all program variables is partitioned by the equivalence relation $\mathcal{R}$, where $v_1 \mathcal{R} v_2$ if v_1 and v_2 are program variables that share the same memory location.

EXAMPLE 7.58

Having seen examples of how an equivalence relation induces a partition of a set, we now go backward. If an equivalence relation $\mathcal{R}$ on $A = \{1, 2, 3, 4, 5, 6, 7\}$ induces the partition $A = \{1, 2\} \cup \{3\} \cup \{4, 5, 7\} \cup \{6\}$, what is $\mathcal{R}$?

Consider the cell $\{1, 2\}$ of the partition. This subset implies that $[1] = \{1, 2\} = [2]$, and so $(1, 1), (2, 2), (1, 2), (2, 1) \in \mathcal{R}$. (The first two ordered pairs are necessary for the reflexive property of $\mathcal{R}$; the others preserve symmetry.)

In like manner, the cell $\{4, 5, 7\}$ implies that under $\mathcal{R}$, $[4] = [5] = [7] = \{4, 5, 7\}$ and that, as an equivalence relation, $\mathcal{R}$ must contain $\{4, 5, 7\} \times \{4, 5, 7\}$. In fact,

$$\mathcal{R} = (\{1, 2\} \times \{1, 2\}) \cup (\{3\} \times \{3\}) \cup (\{4, 5, 7\} \times \{4, 5, 7\}) \cup (\{6\} \times \{6\}),$$

and

$$|\mathcal{R}| = 2^2 + 1^2 + 3^2 + 1^2 = 15.$$

The results in Examples 7.54, 7.55, 7.56, and 7.58 lead us to the following.

THEOREM 7.7

If A is a set, then

a) any equivalence relation $\mathcal{R}$ on A induces a partition of A, and

b) any partition of A gives rise to an equivalence relation $\mathcal{R}$ on A.

Proof: Part (a) follows from parts (a) and (c) of Theorem 7.6. For part (b), given a partition $\{A_i\}_{i\in I}$ of A, define relation $\mathcal{R}$ on A by $x \mathcal{R} y$, if x and y are in the same cell of the partition. We leave to the reader the details of verifying that $\mathcal{R}$ is an equivalence relation.

On the basis of this theorem and the examples we have examined, we state the next result. A proof for it is outlined in Exercise 16 at the end of the section.

THEOREM 7.8 For any set A, there is a one-to-one correspondence between the set of equivalence relations on A and the set of partitions of A.

We are primarily concerned with using this result for finite sets.

EXAMPLE 7.59

a) If $A = \{1, 2, 3, 4, 5, 6\}$, how many relations on A are equivalence relations?

We solve this problem by counting the partitions of A, realizing that a partition of A is a distribution of the (distinct) elements of A into identical containers, with no container left empty. From Section 5.3 we know, for example, that there are $S(6, 2)$ partitions of A into two identical nonempty containers. Using the Stirling numbers of the second kind, as the number of containers varies from 1 to 6, we have $\sum_{i=1}^{6} S(6, i) = 203$ different partitions of A. Consequently, there are 203 equivalence relations on A.

b) How many of the equivalence relations in part (a) satisfy $1, 2 \in [4]$?

Identifying 1, 2, and 4 as the "same" element under these equivalence relations, we count as in part (a) for the set $B = \{1, 3, 5, 6\}$ and find that there are $\sum_{i=1}^{4} S(4, i) = 15$ equivalence relations on A for which $[1] = [2] = [4]$.

We close by noting that if A is a finite set with $|A| = n$, then for all $n \leq r \leq n^2$, there is an equivalence relation $\mathcal{R}$ on A with $|\mathcal{R}| = r$ if and only if there exist $n_1, n_2, \ldots, n_k \in \mathbf{Z}^+$ with $\sum_{i=1}^{k} n_i = n$ and $\sum_{i=1}^{k} n_i^2 = r$.

EXERCISES 7.4

1. Determine whether each of the following collections of sets is a partition for the given set A. If the collection is not a partition, explain why it fails to be.

a) $A = \{1, 2, 3, 4, 5, 6, 7, 8\}$; $A_1 = \{4, 5, 6\}$, $A_2 = \{1, 8\}$, $A_3 = \{2, 3, 7\}$.

b) $A = \{a, b, c, d, e, f, g, h\}$; $A_1 = \{d, e\}$, $A_2 = \{a, c, d\}$, $A_3 = \{f, h\}$, $A_4 = \{b, g\}$.

2. Let $A = \{1, 2, 3, 4, 5, 6, 7, 8\}$. In how many ways can we partition A as $A_1 \cup A_2 \cup A_3$ with

a) $1, 2 \in A_1$, $3, 4 \in A_2$, and $5, 6, 7 \in A_3$?

b) $1, 2 \in A_1$, $3, 4 \in A_2$, $5, 6 \in A_3$, and $|A_1| = 3$?

c) $1, 2 \in A_1$, $3, 4 \in A_2$, and $5, 6 \in A_3$?

3. If $A = \{1, 2, 3, 4, 5\}$ and $\mathcal{R}$ is the equivalence relation on A that induces the partition $A = \{1, 2\} \cup \{3, 4\} \cup \{5\}$, what is $\mathcal{R}$?

4. For $A = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{R} = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3), (4, 4), (4, 5), (5, 4), (5, 5), (6, 6)\}$ is an equivalence relation on A. (a) What are [1], [2], and [3] under this equivalence relation? (b) What partition of A does $\mathcal{R}$ induce?

5. If $A = A_1 \cup A_2 \cup A_3$, where $A_1 = \{1, 2\}$, $A_2 = \{2, 3, 4\}$, and $A_3 = \{5\}$, define relation $\mathcal{R}$ on A by $x \mathcal{R} y$ if x and y are in the same subset A_i, for $1 \leq i \leq 3$. Is $\mathcal{R}$ an equivalence relation?

6. For $A = \mathbf{R}^2$, define $\mathcal{R}$ on A by $(x_1, y_1) \mathcal{R} (x_2, y_2)$ if $x_1 = x_2$.

a) Verify that $\mathcal{R}$ is an equivalence relation on A.

b) Describe geometrically the equivalence classes and partition of A induced by $\mathcal{R}$.

7. Let $A = \{1, 2, 3, 4, 5\} \times \{1, 2, 3, 4, 5\}$, and define $\mathcal{R}$ on A by $(x_1, y_1) \mathcal{R} (x_2, y_2)$ if $x_1 + y_1 = x_2 + y_2$.

a) Verify that $\mathcal{R}$ is an equivalence relation on A.

b) Determine the equivalence classes $[(1, 3)]$, $[(2, 4)]$, and $[(1, 1)]$.

c) Determine the partition of A induced by $\mathcal{R}$.

8. If $A = \{1, 2, 3, 4, 5, 6, 7\}$, define $\mathcal{R}$ on A by $(x, y) \in \mathcal{R}$ if $x - y$ is a multiple of 3.

a) Show that $\mathcal{R}$ is an equivalence relation on A.

b) Determine the equivalence classes and partition of A induced by $\mathcal{R}$.

9. For $A = \{(-4, -20), (-3, -9), (-2, -4), (-1, -11), (-1, -3), (1, 2), (1, 5), (2, 10), (2, 14), (3, 6), (4, 8), (4, 12)$ define the relation $\mathcal{R}$ on A by $(a, b) \mathcal{R} (c, d)$ if $ad = bc$.

a) Verify that $\mathcal{R}$ is an equivalence relation on A.

b) Find the equivalence classes $[(2, 14)]$, $[(-3, -9)]$, and $[(4, 8)]$.

c) How many cells are there in the partition of A induced by $\mathcal{R}$?

10. Let A be a nonempty set and fix the set B, where $B \subseteq A$. Define the relation $\mathcal{R}$ on $\mathcal{P}(A)$ by $X \mathcal{R} Y$, for $X, Y \subseteq A$, if $B \cap X = B \cap Y$.

a) Verify that $\mathcal{R}$ is an equivalence relation on $\mathcal{P}(A)$.

b) If $A = \{1, 2, 3\}$ and $B = \{1, 2\}$, find the partition of $\mathcal{P}(A)$ induced by $\mathcal{R}$.

c) If $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 3\}$, find $[X]$ if $X = \{1, 3, 5\}$.

d) For $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 3\}$, how many equivalence classes are in the partition induced by $\mathcal{R}$?

11. How many of the equivalence relations on $A = \{a, b, c, d, e, f\}$ have (a) exactly two equivalence classes of size 3? (b) exactly one equivalence class of size 3? (c) one equivalence class of size 4? (d) at least one equivalence class with three or more elements?

12. Let $A = \{v, w, x, y, z\}$. Determine the number of relations on A that are (a) reflexive and symmetric; (b) equivalence relations; (c) reflexive and symmetric but not transitive; (d) equivalence relations that determine exactly two equivalence classes; (e) equivalence relations where $w \in [x]$; (f) equivalence relations where $v, w \in [x]$; (g) equivalence relations where $w \in [x]$ and $y \in [z]$; and (h) equivalence relations where $w \in [x]$, $y \in [z]$, and $[x] \neq [z]$.

13. If $|A| = 30$ and the equivalence relation $\mathcal{R}$ on A partitions A into (disjoint) equivalence classes A_1, A_2, and A_3, where $|A_1| = |A_2| = |A_3|$, what is $|\mathcal{R}|$?

14. Let $A = \{1, 2, 3, 4, 5, 6, 7\}$. For each of the following values of r, determine an equivalence relation $\mathcal{R}$ on A with $|\mathcal{R}| = r$, or explain why no such relation exists. (a) $r = 6$; (b) $r = 7$; (c) $r = 8$; (d) $r = 9$; (e) $r = 11$; (f) $r = 22$; (g) $r = 23$; (h) $r = 30$; (i) $r = 31$.

15. Provide the details for the proof of part (b) of Theorem 7.7.

16. For any set $A \neq \emptyset$, let $P(A)$ denote the set of all partitions of A, and let $E(A)$ denote the set of all equivalence relations on A. Define the function $f: E(A) \to P(A)$ as follows: If $\mathcal{R}$ is an equivalence relation on A, then $f(\mathcal{R})$ is the partition of A induced by $\mathcal{R}$. Prove that f is one-to-one and onto, thus establishing Theorem 7.8.

17. Let $f: A \to B$. If $\{B_1, B_2, B_3, \ldots, B_n\}$ is a partition of B, prove that $\{f^{-1}(B_i) | 1 \leq i \leq n, f^{-1}(B_i) \neq \emptyset\}$ is a partition of A.

7.5 Finite State Machines: The Minimization Process

In Section 6.3 we encountered two finite state machines that performed the same task but had different numbers of internal states. (See Figs. 6.9 and 6.10.) The machine with the larger number of internal states contains *redundant* states — states that can be eliminated because other states will perform their functions. Since minimization of the number of states in a machine reduces its complexity and cost, we seek a process for transforming a given machine into one that has no redundant internal states. This process is known as the *minimization process*, and its development relies on the concepts of equivalence relation and partition.

Starting with a given finite state machine $M = (S, \mathcal{I}, \mathcal{O}, \nu, \omega)$, we define the relation E_1 on S by $s_1 \, E_1 \, s_2$ if $\omega(s_1, x) = \omega(s_2, x)$, for all $x \in \mathcal{I}$. This relation E_1 is an equivalence relation on S, and it partitions S into subsets such that two states are in the same subset if they produce the same output for each $x \in \mathcal{I}$. Here the states s_1, s_2 are called *1-equivalent*.

For each $k \in \mathbf{Z}^+$, we say that the states s_1, s_2 are *k-equivalent* if $\omega(s_1, x) = \omega(s_2, x)$ for all $x \in \mathcal{I}^k$. Here ω is the extension of the given output function to $S \times \mathcal{I}^*$. The relation of k-equivalence is also an equivalence relation on S; it partitions S into subsets of k-equivalent states. We write $s_1 \, E_k \, s_2$ to denote that s_1 and s_2 are k-equivalent.

Finally, if $s_1, s_2 \in S$ and s_1, s_2 are k-equivalent for all $k \geq 1$, then we call s_1 and s_2 *equivalent* and write $s_1 \, E \, s_2$. When this happens, we find that if we keep s_1 in our machine, then s_2 will be redundant and can be removed. Hence our objective is to determine the partition of S induced by E and to select one state for each equivalence class. Then we shall have a minimal realization of the given machine.

To accomplish this, let us start with the following observations.

a) If two states in a machine are not 2-equivalent, could they possibly be 3-equivalent? (or k-equivalent, for $k \geq 4$?)

The answer is no. If $s_1, s_2 \in S$ and $s_1 \not\mathrm{E}_2\, s_2$ (that is, s_1 and s_2 are not 2-equivalent), then there is at least one string $xy \in \mathscr{I}^2$ such that $\omega(s_1, xy) = v_1v_2 \neq w_1w_2 = \omega(s_2, xy)$, where $v_1, v_2, w_1, w_2 \in \mathscr{O}$. So with regard to E_3, we find that $s_1 \not\mathrm{E}_3\, s_2$ because for any $z \in \mathscr{I}$, $\omega(s_1, xyz) = v_1v_2v_3 \neq w_1w_2w_3 = \omega(s_2, xyz)$.

In general, to find states that are $(k+1)$-equivalent, we look at states that are k-equivalent.

b) Now suppose that $s_1, s_2 \in S$ and $s_1\,\mathrm{E}_2\,s_2$. We wish to determine whether $s_1\,\mathrm{E}_3\,s_2$. That is, does $\omega(s_1, x_1x_2x_3) = \omega(s_2, x_1x_2x_3)$ for all strings $x_1x_2x_3 \in \mathscr{I}^3$? Consider what happens. First we get $\omega(s_1, x_1) = \omega(s_2, x_1)$, because $s_1\,\mathrm{E}_2\,s_2 \Rightarrow s_1\,\mathrm{E}_1\,s_2$. Then there is a transition to the states $\nu(s_1, x_1)$ and $\nu(s_2, x_1)$. Consequently, $\omega(s_1, x_1x_2x_3) = \omega(s_2, x_1x_2x_3)$ if $\omega(\nu(s_1, x_1), x_2x_3) = \omega(\nu(s_2, x_1), x_2x_3)$ [that is, if $\nu(s_1, x_1)\,\mathrm{E}_2\,\nu(s_2, x_1)$].

In general, for $s_1, s_2, \in S$, where $s_1\,\mathrm{E}_k\,s_2$, we find that $s_1\,\mathrm{E}_{k+1}\,s_2$ if (and only if) $\nu(s_1, x)\,\mathrm{E}_k\,\nu(s_2, x)$ for all $x \in \mathscr{I}$.

With these observations to guide us, we now present an algorithm for the minimization of a finite state machine M.

Step 1: Set $k = 1$. We determine the states that are 1-equivalent by examining the rows in the state table for M. For $s_1, s_2 \in S$ it follows that $s_1\,\mathrm{E}_1\,s_2$ when s_1, s_2 have the same output rows.

Let P_1 be the partition of S induced by E_1.

Step 2: Having determined P_k, we obtain P_{k+1} by noting that if $s_1\,\mathrm{E}_k\,s_2$, then $s_1\,\mathrm{E}_{k+1}\,s_2$ when $\nu(s_1, x)\,\mathrm{E}_k\,\nu(s_2, x)$ for all $x \in \mathscr{I}$. We have $s_1\,\mathrm{E}_k\,s_2$ if s_1, s_2 are in the same cell of the partition P_k. Likewise, $\nu(s_1, x)\,\mathrm{E}_k\,\nu(s_2, x)$ for each $x \in \mathscr{I}$, if $\nu(s_1, x)$ and $\nu(s_2, x)$ are in the same cell of the partition P_k. In this way P_{k+1} is obtained from P_k.

Step 3: If $P_{k+1} = P_k$, the process is complete. We select one state from each equivalence class and these states yield a minimal realization of M.

If $P_{k+1} \neq P_k$, we increase k by 1 and return to step (2).

We illustrate the algorithm in the following example.

EXAMPLE 7.60

With $\mathscr{I} = \mathscr{O} = \{0, 1\}$, let M be given by the state table shown in Table 7.1. Looking at the output rows, we see that s_3 and s_4 are 1-equivalent, as are s_2, s_5, and s_6. Here E_1 partitions S as follows:

$$P_1: \{s_1\}, \{s_2, s_5, s_6\}, \{s_3, s_4\}.$$

For each $s \in S$ and each $k \in \mathbf{Z}^+$, $s\,\mathrm{E}_k\,s$, so as we continue this process to determine P_2, we shall not concern ourselves with equivalence classes of only one state.

Since $s_3\,\mathrm{E}_1\,s_4$, there is a chance that we could have $s_3\,\mathrm{E}_2\,s_4$. Here $\nu(s_3, 0) = s_2$, $\nu(s_4, 0) = s_5$ with $s_2\,\mathrm{E}_1\,s_5$, and $\nu(s_3, 1) = s_4$, $\nu(s_4, 1) = s_3$ with $s_4\,\mathrm{E}_1\,s_3$. Hence $\nu(s_3, x)\,\mathrm{E}_1\,\nu(s_4, x)$, for all $x \in \mathscr{I}$, and $s_3\,\mathrm{E}_2\,s_4$. Similarly, $\nu(s_2, 0) = s_5$, $\nu(s_5, 0) = s_2$ with $s_5\,\mathrm{E}_1\,s_2$, and $\nu(s_2, 1) = s_2$, $\nu(s_5, 1) = s_5$ with $s_2\,\mathrm{E}_1\,s_5$. Thus $s_2\,\mathrm{E}_2\,s_5$. Finally, $\nu(s_5, 0) = s_2$ and

$\nu(s_6, 0) = s_1$, but $s_2 \not\mathrm{E}_1 s_1$, so $s_5 \not\mathrm{E}_2 s_6$. (Why don't we investigate the possibility of $s_2 \mathrm{E}_2 s_6$?) Equivalence relation E_2 partitions S as follows:

$$P_2: \{s_1\}, \{s_2, s_5\}, \{s_3, s_4\}, \{s_6\}.$$

Since $P_2 \neq P_1$, we continue the process to get P_3. In determining whether $s_2 \mathrm{E}_3 s_5$, we see that $\nu(s_2, 0) = s_5$, $\nu(s_5, 0) = s_2$, and $s_5 \mathrm{E}_2 s_2$. Also, $\nu(s_2, 1) = s_2$, $\nu(s_5, 1) = s_5$, and $s_2 \mathrm{E}_2 s_5$. With $\nu(s_2, x) \mathrm{E}_2 \nu(s_5, x)$ for all $x \in \mathscr{I}$, we have $s_2 \mathrm{E}_3 s_5$. For s_3, s_4, $(\nu(s_3, 0) = s_2)$ E_2 $(s_5 = \nu(s_4, 0))$ and $(\nu(s_3, 1) = s_4)$ E_2 $(s_3 = \nu(s_4, 1))$, so $s_3 \mathrm{E}_3 s_4$ and E_3 induces the partition P_3: $\{s_1\}, \{s_2, s_5\}, \{s_3, s_4\}, \{s_6\}$.

Table 7.1

	ν		ω	
	0	1	0	1
s_1	s_4	s_3	0	1
s_2	s_5	s_2	1	0
s_3	s_2	s_4	0	0
s_4	s_5	s_3	0	0
s_5	s_2	s_5	1	0
s_6	s_1	s_6	1	0

Table 7.2

	ν		ω	
	0	1	0	1
s_1	s_3	s_3	0	1
s_2	s_2	s_2	1	0
s_3	s_2	s_3	0	0
s_6	s_1	s_6	1	0

Now $P_3 = P_2$ so the process is completed, as indicated in step (3) of the algorithm. We find that s_5 and s_4 may be regarded as redundant states. Removing them from the table, and replacing all further occurrences of them by s_2 and s_3, respectively, we arrive at Table 7.2. This is a minimal machine that performs the same tasks as the machine given in Table 7.1.

If we do not want states that skip a subscript, we can always relabel the states in this minimal machine. Here we would have s_1, s_2, s_3, s_4 $(= s_6)$, but this s_4 is not the same s_4 we started with in Table 7.1.

You may be wondering how we knew that we could stop the process when $P_3 = P_2$. For after all, couldn't it happen that perhaps $P_4 \neq P_3$, or that $P_4 = P_3$ but $P_5 \neq P_4$? To prove that this never occurs, we define the following idea.

Definition 7.23 If P_1, P_2 are partitions of a set A, then P_2 is called a *refinement* of P_1, and we write $P_2 \leq P_1$, if every cell of P_2 is contained in a cell of P_1. When $P_2 \leq P_1$ and $P_2 \neq P_1$ we write $P_2 < P_1$. This occurs when at least one cell in P_2 is properly contained in a cell in P_1.

In the minimization process of Example 7.60, we had $P_3 = P_2 < P_1$. Whenever we apply the algorithm, as we get P_{k+1} from P_k, we always find that $P_{k+1} \leq P_k$, because $(k+1)$-equivalence implies k-equivalence. So each successive partition refines the preceding partition.

THEOREM 7.9 In applying the minimization process, if $k \geq 1$ and P_k and P_{k+1} are partitions with $P_{k+1} = P_k$, then $P_{r+1} = P_r$ for all $r \geq k+1$.

Proof: If not, let r $(\geq k+1)$ be the smallest subscript such that $P_{r+1} \neq P_r$. Then $P_{r+1} < P_r$, so there exist $s_1, s_2 \in S$ with $s_1 \mathrm{E}_r s_2$ but $s_1 \not\mathrm{E}_{r+1} s_2$. But $s_1 \mathrm{E}_r s_2 \Rightarrow \nu(s_1, x) \mathrm{E}_{r-1} \nu(s_2, x)$, for all $x \in \mathscr{I}$, and with $P_r = P_{r-1}$, we then find that $\nu(s_1, x)\ \mathrm{E}_r\ \nu(s_2, x)$, for all $x \in \mathscr{I}$, so $s_1\ \mathrm{E}_{r+1}\ s_2$. Consequently, $P_{r+1} = P_r$.

We close this section with the following related idea. Let M be a finite state machine with $s_1, s_2 \in S$, and s_1, s_2 not equivalent. If $s_1\ \not\!\mathrm{E}_1\ s_2$, then these states produce different output rows in the state table for M. In this case it is easy to find an $x \in \mathscr{I}$ such that $\omega(s_1, x) \neq \omega(s_2, x)$, and this distinguishes these nonequivalent states. Otherwise, s_1 and s_2 produce the same output rows in the table but there is a smallest integer $k \geq 1$ such that $s_1\ \mathrm{E}_k\ s_2$ but $s_1\ \not\!\mathrm{E}_{k+1}\ s_2$. Now if we are to distinguish these states, we need to find a string $x = x_1x_2 \cdots x_kx_{k+1} \in \mathscr{I}^{k+1}$ such that $\omega(s_1, x) \neq \omega(s_2, x)$, even though $\omega(s_1, x_1x_2 \cdots x_k) = \omega(s_2, x_1x_2 \cdots x_k)$. Such a string x is called a *distinguishing string* for the states s_1 and s_2. There may be more than one such string, but each has the same (minimal) length $k + 1$.

Before we try to find a distinguishing string for two nonequivalent states in a specific finite state machine, let us examine the major idea at play here. So suppose that $s_1, s_2 \in S$ and that for some (fixed) $k \in \mathbf{Z}^+$ we have $s_1\ \mathrm{E}_k\ s_2$ but $s_1\ \not\!\mathrm{E}_{k+1}\ s_2$. What can we conclude? We find that

$$\begin{aligned}
s_1\ \not\!\mathrm{E}_{k+1}\ s_2 &\Rightarrow \exists x_1 \in \mathscr{I}\ [\nu(s_1, x_1)\ \not\!\mathrm{E}_k\ \nu(s_2, x_1)] \\
&\Rightarrow \exists x_1 \in \mathscr{I}\ \exists x_2 \in \mathscr{I}\ [\nu(\nu(s_1, x_1), x_2)\ \not\!\mathrm{E}_{k-1}\ \nu(\nu(s_2, x_1), x_2)], \\
\text{or}\quad &\ \exists x_1 \in \mathscr{I}\ \exists x_2 \in \mathscr{I}\ [\nu(s_1, x_1x_2)\ \not\!\mathrm{E}_{k-1}\ \nu(s_2, x_1x_2)] \\
&\Rightarrow \exists x_1, x_2, x_3 \in \mathscr{I}\ [\nu(s_1, x_1x_2x_3)\ \not\!\mathrm{E}_{k-2}\ \nu(s_2, x_1x_2x_3)] \\
&\Rightarrow \ldots \\
&\Rightarrow \exists x_1, x_2, \ldots, x_i \in \mathscr{I}\ [\nu(s_1, x_1x_2 \cdots x_i)\ \not\!\mathrm{E}_{k+1-i}\ \nu(s_2, x_1x_2 \cdots x_i)] \\
&\Rightarrow \ldots \\
&\Rightarrow \exists x_1, x_2, \ldots, x_k \in \mathscr{I}\ [\nu(s_1, x_1x_2 \cdots x_k)\ \not\!\mathrm{E}_1\ \nu(s_2, x_1x_2 \cdots x_k)].
\end{aligned}$$

This last statement about the states $\nu(s_1, x_1x_2 \cdots x_k)$, $\nu(s_2, x_1x_2 \cdots x_k)$ not being 1-equivalent implies that we can find $x_{k+1} \in \mathscr{I}$ where

$$\omega(\nu(s_1, x_1x_2 \cdots x_k), x_{k+1}) \neq \omega(\nu(s_2, x_1x_2 \cdots x_k), x_{k+1}). \tag{1}$$

That is, these *single* output symbols from $\mathscr{O}$ are different.

The result denoted by Eq. (1) also implies that

$$\omega(s_1, x) = \omega(s_1, x_1x_2 \cdots x_kx_{k+1}) \neq \omega(s_2, x_1x_2 \cdots x_kx_{k+1}) = \omega(s_2, x).$$

In this case we have two output strings of length $k + 1$ that agree for the first k symbols and differ in the $(k + 1)$st symbol.

We shall use the preceding observations, together with the partitions $P_1, P_2, \ldots, P_k, P_{k+1}$ of the minimization process, in order to deal with the following example.

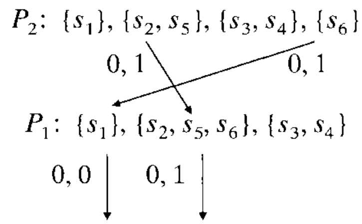
EXAMPLE 7.61

From Example 7.60 we have the partitions shown below. Here $s_2\ \mathrm{E}_1\ s_6$, but $s_2\ \not\!\mathrm{E}_2\ s_6$. So we seek an input string x of length 2 such that $\omega(s_2, x) \neq \omega(s_6, x)$.

1) We start at P_2, where for s_2, s_6, we find that $\nu(s_2, 0) = s_5$ and $\nu(s_6, 0) = s_1$ are in different cells of P_1 — that is,

$$s_5 = \nu(s_2, 0)\ \not\!\mathrm{E}_1\ \nu(s_6, 0) = s_1.$$

[The input 0 and output 1 (for $\omega(s_2, 0) = 1 = \omega(s_6, 0)$) provide the labels for the arrows going from the cells of P_2 to those of P_1.]



2) Working with s_1 and s_5 in the partition P_1 we see that

$$\omega(\nu(s_2, 0), 0) = \omega(s_5, 0) = 1 \neq 0 = \omega(s_1, 0) = \omega(\nu(s_6, 0), 0).$$

3) Hence $x = 00$ is a minimal distinguishing string for s_2 and s_6 because $\omega(s_2, 00) = 11 \neq 10 = \omega(s_6, 00)$.

EXAMPLE 7.62

Applying the minimization process to the machine given by the state table in part (a) of Table 7.3, we obtain the partitions in part (b) of the table. (Here $P_4 = P_3$.) We find that the states s_1 and s_4 are 2-equivalent but not 3-equivalent. To construct a minimal distinguishing string for these two states, we proceed as follows:

1) Since $s_1 \not\equiv_3 s_4$, we use partitions P_3 and P_2 to find $x_1 \in \mathscr{I}$ (namely, $x_1 = 1$) so that

$$(\nu(s_1, 1) = s_2) \not\equiv_2 (s_5 = \nu(s_4, 1)).$$

2) Then $\nu(s_1, 1) \not\equiv_2 \nu(s_4, 1) \Rightarrow \exists x_2 \in \mathscr{I}$ (here $x_2 = 1$) with $(\nu(s_1, 1), 1) \not\equiv_1 (\nu(s_4, 1), 1)$, or $\nu(s_1, 11) \not\equiv_1 \nu(s_4, 11)$. We used the partitions P_2 and P_1 to obtain $x_2 = 1$.
3) Now we use the partition P_1 where we find that for $x_3 = 1 \in \mathscr{I}$,

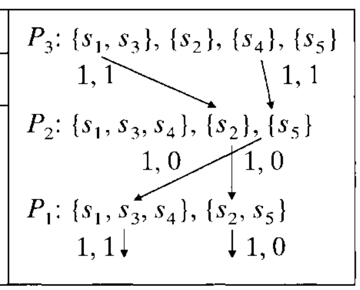
$$\omega(\nu(s_1, 11), 1) = 0 \neq 1 = \omega(\nu(s_4, 11), 1) \quad \text{or}$$
$$\omega(s_1, 111) = 100 \neq 101 = \omega(s_4, 111).$$

In part (b) of Table 7.3, we see how we arrived at the minimal distinguishing string $x = 111$ for these states. (Also note how this part of the table indicates that 11 is a minimal distinguishing string for the states s_2 and s_5, which are 1-equivalent but not 2-equivalent.)

Table 7.3

	ν		ω	
	0	1	0	1
s_1	s_4	s_2	0	1
s_2	s_5	s_2	0	0
s_3	s_4	s_2	0	1
s_4	s_3	s_5	0	1
s_5	s_2	s_3	0	0

(a)



(b)

A great deal more can be done with finite state machines. Among other omissions, we have avoided offering any rigorous explanation or proof of why the minimization process works. The interested reader should consult the chapter references for more on this topic.

EXERCISES 7.5

1. Apply the minimization process to each machine in Table 7.4.

Table 7.4

	ν		ω	
	0	1	0	1
s_1	s_4	s_1	0	1
s_2	s_3	s_3	1	0
s_3	s_1	s_4	1	0
s_4	s_1	s_3	0	1
s_5	s_3	s_3	1	0

(a)

	ν		ω	
	0	1	0	1
s_1	s_6	s_3	0	0
s_2	s_5	s_4	0	1
s_3	s_6	s_2	1	1
s_4	s_4	s_3	1	0
s_5	s_2	s_4	0	1
s_6	s_4	s_6	0	0

(b)

	ν		ω	
	0	1	0	1
s_1	s_6	s_3	0	0
s_2	s_3	s_1	0	0
s_3	s_2	s_4	0	0
s_4	s_7	s_4	0	0
s_5	s_6	s_7	0	0
s_6	s_5	s_2	1	0
s_7	s_4	s_1	0	0

(c)

2. For the machine in Table 7.4(c), find a (minimal) distinguishing string for each given pair of states: (a) s_1, s_5; (b) s_2, s_3; (c) s_5, s_7.

3. Let M be the finite state machine given in the state diagram shown in Fig. 7.26.

a) Minimize machine M.

b) Find a (minimal) distinguishing string for each given pair of states: (i) s_3, s_6; (ii) s_3, s_4; and (iii) s_1, s_2.

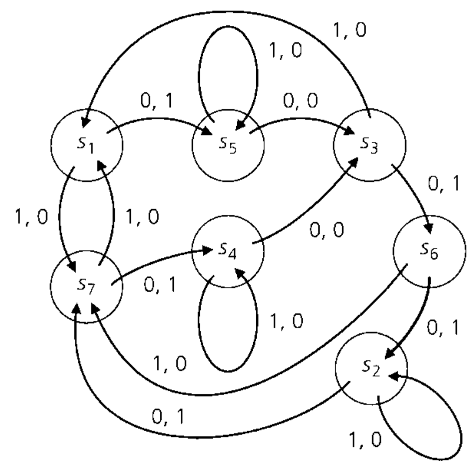


Figure 7.26

7.6 Summary and Historical Review

Once again the relation concept surfaces. In Chapter 5 this idea was introduced as a generalization of the function. Here in Chapter 7 we concentrated on relations and the special properties: reflexive, symmetric, antisymmetric, and transitive. As a result we focused on two special kinds of relations: partial orders and equivalence relations.

A relation $\mathcal{R}$ on a set A is a partial order, making A into a poset, if $\mathcal{R}$ is reflexive, antisymmetric, and transitive. Such a relation generalizes the familiar "less than or equal to" relation on the real numbers. Try to imagine calculus, or even elementary algebra, without it! Or take a simple computer program and see what happens if the program is entered into the computer haphazardly, permuting the order of the statements. Order is with us wherever we turn. We have grown so accustomed to it that we sometimes take it for granted. The origins of the subject of partially ordered sets (and lattices) came about during the nineteenth century in the work of George Boole (1815–1864), Richard Dedekind (1831–1916), Charles Sanders Peirce (1839–1914), and Ernst Schröder (1841–1902). The work of Garrett Birkhoff (1911–1996) in the 1930s, however, is where the initial work on partially ordered sets and lattices was developed to the point where these areas emerged as subjects in their own right.

For a finite poset, the Hasse diagram, a special type of directed graph, provides a pictorial representation of the order defined by the poset; it also proves useful when a total order, including the given partial order, is needed. These diagrams are named for the German number theorist Helmut Hasse (1898–1979). He introduced them in his textbook *Höhere Algebra* (published in 1926) as an aid in the study of the solutions of polynomial equations. The method we employed to derive a total order from a partial order is called topological sorting and it is used in the solution of PERT (Program Evaluation and Review Technique) networks. As mentioned earlier, this method was developed and first used by the U.S. Navy.

Although the equivalence relation differs from the partial order in only one property, it is quite different in structure and application. We make no attempt to trace the origin of the equivalence relation, but the ideas behind the reflexive, symmetric, and transitive properties can be found in *I Principii di Geometria* (1889), the work of the Italian mathematician Giuseppe Peano (1858–1932). The work of Carl Friedrich Gauss (1777–1855) on *congruence*, which he developed in the 1790s, also utilizes these ideas in spirit, if not in name.

Giuseppe Peano (1858–1932)

Carl Friedrich Gauss (1777–1855)

Basically, an equivalence relation $\mathcal{R}$ on a set A generalizes equality; it induces a characteristic of "sameness" among the elements of A. This "sameness" notion then causes the set A to be partitioned into subsets called *equivalence classes*. Conversely, we find that a partition of a set A induces an equivalence relation on A. The partition of a set arises in many places in mathematics and computer science. In computer science many searching algorithms rely on a technique that successively reduces the size of a given set A that is being searched. By partitioning A into smaller and smaller subsets, we apply the searching procedure in a more efficient manner. Each successive partition refines its predecessor, the key needed, for example, in the minimization process for finite state machines.

Throughout the chapter we emphasized the interplay between relations, directed graphs, and (0, 1)-matrices. These matrices provide a rectangular array of information about a relation, or graph, and prove useful in certain calculations. Storing information like this, in rectangular arrays and in consecutive memory locations, has been practiced in computer science since the late 1940s and early 1950s. For more on the historical background of such considerations, consult pages 456–462 of D. E. Knuth [3]. Another way to store information about a graph is the *adjacency list representation.* (See Supplementary Exercise 11.) In the study of data structures, *linked lists* and *doubly linked lists* are prominent in implementing such a representation. For more on this, consult the text by A. V. Aho, J. E. Hopcroft, and J. D. Ullman [1].

With regard to graph theory, we are in an area of mathematics that dates back to 1736 when the Swiss mathematician Leonhard Euler (1707–1783) solved the problem of the seven bridges of Königsberg. Since then, much more has evolved in this area, especially in conjunction with data structures in computer science.

For similar coverage of some of the topics in this chapter, see Chapter 3 of D. F. Stanat and D. F. McAllister [6]. An interesting presentation of the "Equivalence Problem" can be found on pages 353–355 of D. E. Knuth [3] for those wanting more information on the role of the computer in conjunction with the concept of the equivalence relation.

The early work on the development of the minimization process can be found in the paper by E. F. Moore [5], which builds upon prior ideas of D. A. Huffman [2]. Chapter 10 of Z. Kohavi [4] covers the minimization process for different types of finite state machines and includes some hardware considerations in their design.

REFERENCES

1. Aho, Alfred V., Hopcroft, John E., and Ullman, Jeffrey D. *Data Structures and Algorithms.* Reading, Mass.: Addison-Wesley, 1983.
2. Huffman, David A. "The Synthesis of Sequential Switching Circuits." *Journal of the Franklin Institute* 257, no. 3: pp. 161–190; no. 4: pp. 275–303, 1954.
3. Knuth, Donald E. *The Art of Computer Programming*, 2nd ed., Volume 1, *Fundamental Algorithms.* Reading, Mass.: Addison-Wesley, 1973.
4. Kohavi, Zvi. *Switching and Finite Automata Theory*, 2nd ed. New York: McGraw-Hill, 1978.
5. Moore, E. F. "Gedanken-experiments on Sequential Machines." *Automata Studies, Annals of Mathematical Studies*, no. 34: pp. 129–153. Princeton, N.J.: Princeton University Press, 1956.
6. Stanat, Donald F., and McAllister, David F. *Discrete Mathematics in Computer Science.* Englewood Cliffs, N.J.: Prentice-Hall, 1977.

SUPPLEMENTARY EXERCISES

1. Let A be a set and I an index set where, for each $i \in I$, $\mathcal{R}_i$ is a relation on A. Prove or disprove each of the following.

a) $\bigcup_{i \in I} \mathcal{R}_i$ is reflexive on A if and only if each $\mathcal{R}_i$ is reflexive on A.

b) $\bigcap_{i \in I} \mathcal{R}_i$ is reflexive on A if and only if each $\mathcal{R}_i$ is reflexive on A.

2. Repeat Exercise 1 with "reflexive" replaced by (i) symmetric; (ii) antisymmetric; (iii) transitive.

3. For a set A, let $\mathcal{R}_1$ and $\mathcal{R}_2$ be symmetric relations on A. If $\mathcal{R}_1 \circ \mathcal{R}_2 \subseteq \mathcal{R}_2 \circ \mathcal{R}_1$, prove that $\mathcal{R}_1 \circ \mathcal{R}_2 = \mathcal{R}_2 \circ \mathcal{R}_1$.

4. For each of the following relations on the set specified, determine whether the relation is reflexive, symmetric, antisymmetric, or transitive. Also determine whether it is a partial order or an equivalence relation, and, if the latter, describe the partition induced by the relation.

a) $\mathcal{R}$ is the relation on $\mathbf{Q}$ where $a \mathcal{R} b$ if $|a - b| < 1$.

b) Let T be the set of all triangles in the plane. For $t_1, t_2 \in T$, define $t_1 \mathcal{R} t_2$ if t_1, t_2 have the same area.

c) For T as in part (b), define $\mathcal{R}$ by $t_1 \mathcal{R} t_2$ if at least two sides of t_1 are contained within the perimeter of t_2.

d) Let $A = \{1, 2, 3, 4, 5, 6, 7\}$. Define $\mathcal{R}$ on A by $x \mathcal{R} y$ if $xy \geq 10$.

5. For sets A, B, and C with relations $\mathcal{R}_1 \subseteq A \times B$ and $\mathcal{R}_2 \subseteq B \times C$, prove or disprove that $(\mathcal{R}_1 \circ \mathcal{R}_2)^c = \mathcal{R}_2^c \circ \mathcal{R}_1^c$.

6. For a set A, let $C = \{P_i | P_i$ is a partition of $A\}$. Define relation $\mathcal{R}$ on C by $P_i \mathcal{R} P_j$ if $P_i \leq P_j$ — that is, P_i is a refinement of P_j.

a) Verify that $\mathcal{R}$ is a partial order on C.

b) For $A = \{1, 2, 3, 4, 5\}$, let P_i, $1 \leq i \leq 4$, be the following partitions: P_1: $\{1, 2\}, \{3, 4, 5\}$; P_2: $\{1, 2\}, \{3, 4\}, \{5\}$; P_3: $\{1\}, \{2\}, \{3, 4, 5\}$; P_4: $\{1, 2\}, \{3\}, \{4\}, \{5\}$. Draw the Hasse diagram for $C = \{P_i | 1 \leq i \leq 4\}$, where C is partially ordered by refinement.

7. Give an example of a poset with 5 minimal (maximal) elements but no least (greatest) element.

8. Let $A = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. Define $\mathcal{R}$ on A by $(x_1, y_1) \mathcal{R} (x_2, y_2)$, if $x_1 y_1 = x_2 y_2$.

a) Verify that $\mathcal{R}$ is an equivalence relation on A.

b) Determine the equivalence classes $[(1, 1)]$, $[(2, 2)]$, $[(3, 2)]$, and $[(4, 3)]$.

9. If the complete graph K_n has 45 edges, what is n?

10. Let $\mathcal{F} = \{f: \mathbf{Z}^+ \to \mathbf{R}\}$ — that is, $\mathcal{F}$ is the set of all functions with domain $\mathbf{Z}^+$ and codomain $\mathbf{R}$.

a) Define the relation $\mathcal{R}$ on $\mathcal{F}$ by $g \mathcal{R} h$, for $g, h \in \mathcal{F}$, if g is dominated by h and h is dominated by g — that is, $g \in \Theta(h)$. (See Exercises 14, 15 for Section 5.7.) Prove that $\mathcal{R}$ is an equivalence relation on $\mathcal{F}$.

b) For $f \in \mathcal{F}$, let $[f]$ denote the equivalence class of f for the relation $\mathcal{R}$ of part (a). Let $\mathcal{F}'$ be the set of equivalence classes induced by $\mathcal{R}$. Define the relation $\mathcal{S}$ on $\mathcal{F}'$ by $[g] \mathcal{S} [h]$, for $[g], [h] \in \mathcal{F}'$, if g is dominated by h. Verify that $\mathcal{S}$ is a partial order.

c) For $\mathcal{R}$ in part (a), let $f, f_1, f_2 \in \mathcal{F}$ with $f_1, f_2 \in [f]$. If $f_1 + f_2: \mathbf{Z}^+ \to \mathbf{R}$ is defined by $(f_1 + f_2)(n) = f_1(n) + f_2(n)$, for $n \in \mathbf{Z}^+$, prove or disprove that $f_1 + f_2 \in [f]$.

11. We have seen that the adjacency matrix can be used to represent a graph. However, this method proves to be rather inefficient when there are many 0's (that is, few edges) present. A better method uses the *adjacency list representation*, which is made up of an *adjacency list* for each vertex v and an *index list*. For the graph shown in Fig. 7.27, the representation is given by the two lists in Table 7.5.

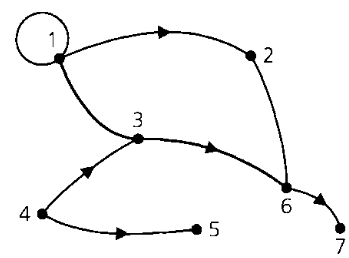


Figure 7.27

Table 7.5

Adjacency List		Index List	
1	1	1	1
2	2	2	4
3	3	3	5
4	6	4	7
5	1	5	9
6	6	6	9
7	3	7	11
8	5	8	11
9	2		
10	7		

For each vertex v in the graph, we list, preferably in numerical order, each vertex w that is adjacent from v. Hence for 1, we list 1, 2, 3 as the first three adjacencies in our adjacency list. Next to 2 in the index list we place a 4, which tells us where to start looking in the adjacency list for the adjacencies from 2. Since there is a 5 to the right of 3 in the index list, we know that the only adjacency from 2 is 6. Likewise, the 7 to the right of 4 in the index list directs us to the seventh entry in the adjacency list — namely, 3 — and we find that vertex 4 is adjacent to vertices 3 (the seventh vertex in the adjacency list) and 5 (the eighth vertex in the adjacency list). We stop at vertex 5 because of the 9 to the right of vertex 5 in the index list. The 9's in the index list next to 5 and 6 indicate that no vertex is adjacent from vertex 5. In a similar way, the 11's next to 7 and 8 in the index list tell us that vertex 7 is not adjacent to any vertex in the given directed graph.

In general, this method provides an easy way to determine the vertices adjacent from a vertex v. They are listed in the positions index(v), index(v) + 1, ..., index(v + 1) − 1 of the adjacency list.

Finally, the last pair of entries in the index list — namely, 8 and 11 — is a "phantom" that indicates where the adjacency list would pick up from *if* there were an eighth vertex in the graph.

Represent each of the graphs in Fig. 7.28 in this manner.

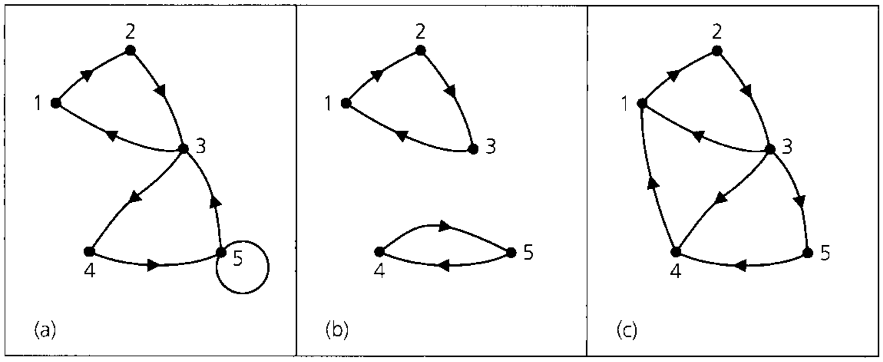


Figure 7.28

12. The adjacency list representation of a directed graph G is given by the lists in Table 7.6. Construct G from this representation.

Table 7.6

Adjacency List		Index List	
1	2	1	1
2	3	2	4
3	6	3	5
4	3	4	5
5	3	5	8
6	4	6	10
7	5	7	10
8	3	8	10
9	6		

13. Let G be an undirected graph with vertex set V. Define the relation $\mathcal{R}$ on V by $v \mathcal{R} w$ if $v = w$ or if there is a path from v to w (or from w to v since G is undirected). (a) Prove that $\mathcal{R}$ is an equivalence relation on V. (b) What can we say about the associated partition?

14. a) For the finite state machine given in Table 7.7, determine a minimal machine that is equivalent to it.

b) Find a minimal string that distinguishes states s_4 and s_6.

15. At the computer center Maria is faced with running 10 computer programs which, because of priorities, are restricted by the following conditions: (a) $10 > 8, 3$; (b) $8 > 7$; (c) $7 > 5$; (d) $3 > 9, 6$; (e) $6 > 4, 1$; (f) $9 > 4, 5$; (g) $4, 5, 1 > 2$; where, for example, $10 > 8, 3$ means that program number 10 must be run before programs 8 and 3. Determine an order for running these programs so that the priorities are satisfied.

16. a) Draw the Hasse diagram for the set of positive integer divisors of (i) 2; (ii) 4; (iii) 6; (iv) 8; (v) 12; (vi) 16; (vii) 24; (viii) 30; (ix) 32.

Table 7.7

	ν		ω	
	0	1	0	1
s_1	s_7	s_6	1	0
s_2	s_7	s_7	0	0
s_3	s_7	s_2	1	0
s_4	s_2	s_3	0	0
s_5	s_3	s_7	0	0
s_6	s_4	s_1	0	0
s_7	s_3	s_5	1	0
s_8	s_7	s_3	0	0

b) For all $2 \leq n \leq 35$, show that the Hasse diagram for the set of positive-integer divisors of n looks like one of the nine diagrams in part (a). (Ignore the numbers at the vertices and concentrate on the structure given by the vertices and edges.) What happens for $n = 36$?

c) For $n \in \mathbf{Z}^+$, $\tau(n)$ = the number of positive-integer divisors of n. (See Supplementary Exercise 32 in Chapter 5.) Let $m, n \in \mathbf{Z}^+$ and S, T be the sets of all positive-integer divisors of m, n, respectively. The results of parts (a) and (b) imply that if the Hasse diagrams of S, T are structurally the same, then $\tau(m) = \tau(n)$. But is the converse true?

d) Show that each Hasse diagram in part (a) is a lattice if we define $\text{glb}\{x, y\} = \gcd(x, y)$ and $\text{lub}\{x, y\} = \text{lcm}(x, y)$.

17. Let U denote the set of all points in and on the unit square shown in Fig. 7.29. That is, $U = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Define the relation $\mathcal{R}$ on U by $(a, b) \mathcal{R} (c, d)$ if (1) $(a, b) = (c, d)$, or (2) $b = d$ and $a = 0$ and $c = 1$, or (3) $b = d$ and $a = 1$ and $c = 0$.

a) Verify that $\mathcal{R}$ is an equivalence relation on U.

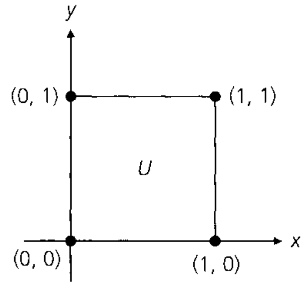


Figure 7.29

b) List the ordered pairs in the equivalence classes $[(0.3, 0.7)]$, $[(0.5, 0)]$, $[(0.4, 1)]$, $[(0, 0.6)]$, $[(1, 0.2)]$. For $0 \le a \le 1$, $0 \le b \le 1$, how many ordered pairs are in $[(a, b)]$?

c) If we "glue together" the ordered pairs in each equivalence class, what type of surface comes about?

18. a) For $\mathcal{U} = \{1, 2, 3\}$, let $A = \mathcal{P}(\mathcal{U})$. Define the relation $\mathcal{R}$ on A by $B \mathcal{R} C$ if $B \subseteq C$. How many ordered pairs are there in the relation $\mathcal{R}$?

b) Answer part (a) for $\mathcal{U} = \{1, 2, 3, 4\}$.

c) Generalize the results of parts (a) and (b).

19. For $n \in \mathbf{Z}^+$, let $\mathcal{U} = \{1, 2, 3, \ldots, n\}$. Define the relation $\mathcal{R}$ on $\mathcal{P}(\mathcal{U})$ by $A \mathcal{R} B$ if $A \nsubseteq B$ and $B \nsubseteq A$. How many ordered pairs are there in this relation?

20. Let A be a finite nonempty set with $B \subseteq A$ (B fixed), and $|A| = n$, $|B| = m$. Define the relation $\mathcal{R}$ on $\mathcal{P}(A)$ by $X \mathcal{R} Y$, for $X, Y \subseteq A$, if $X \cap B = Y \cap B$. Then $\mathcal{R}$ is an equivalence relation, as verified in Exercise 10 of Section 7.4. (a) How many equivalence classes are in the partition of $\mathcal{P}(A)$ induced by $\mathcal{R}$? (b) How many subsets of A are in each equivalence class of the partition induced by $\mathcal{R}$?

21. For $A \neq \emptyset$, let $(A, \mathcal{R})$ be a poset, and let $\emptyset \neq B \subseteq A$ such that $\mathcal{R}' = (B \times B) \cap \mathcal{R}$. If $(B, \mathcal{R}')$ is totally ordered, we call $(B, \mathcal{R}')$ a *chain* in $(A, \mathcal{R})$. In the case where B is finite, we may order the elements of B by $b_1 \mathcal{R}' b_2 \mathcal{R}' b_3 \mathcal{R}' \cdots \mathcal{R}' b_{n-1} \mathcal{R}' b_n$ and say that the chain has *length n*. A chain (of length n) is called *maximal* if there is no element $a \in A$ where $a \notin \{b_1, b_2, b_3, \ldots, b_n\}$ and $a \mathcal{R} b_1$, $b_n \mathcal{R} a$, or $b_i \mathcal{R} a \mathcal{R} b_{i+1}$, for some $1 \le i \le n-1$.

a) Find two chains of length 3 for the poset given by the Hasse diagram in Fig. 7.20. Find a maximal chain for this poset. How many such maximal chains does it have?

b) For the poset given by the Hasse diagram in Fig. 7.18(d), find two maximal chains of different lengths. What is the length of a longest (maximal) chain for this poset?

c) Let $\mathcal{U} = \{1, 2, 3, 4\}$ and $A = \mathcal{P}(\mathcal{U})$. For the poset $(A, \subseteq)$, find two maximal chains. How many such maximal chains are there for this poset?

d) If $\mathcal{U} = \{1, 2, 3, \ldots, n\}$, how many maximal chains are there in the poset $(\mathcal{P}(\mathcal{U}), \subseteq)$?

22. For $\emptyset \neq C \subseteq A$, let $(C, \mathcal{R}')$ be a maximal chain in the poset $(A, \mathcal{R})$, where $\mathcal{R}' = (C \times C) \cap \mathcal{R}$. If the elements of C are ordered as $c_1 \mathcal{R}' c_2 \mathcal{R}' \cdots \mathcal{R}' c_n$, prove that c_1 is a minimal element in $(A, \mathcal{R})$ and that c_n is maximal in $(A, \mathcal{R})$.

23. Let $(A, \mathcal{R})$ be a poset in which the length of a longest (maximal) chain is $n \ge 2$. Let M be the set of all maximal elements in $(A, \mathcal{R})$, and let $B = A - M$. If $\mathcal{R}' = (B \times B) \cap \mathcal{R}$, prove that the length of a longest chain in $(B, \mathcal{R}')$ is $n - 1$.

24. Let $(A, \mathcal{R})$ be a poset, and let $\emptyset \neq C \subseteq A$. If $(C \times C) \cap \mathcal{R} = \emptyset$, then for all distinct $x, y \in C$ we have $x \not\mathcal{R}\, y$ and $y \not\mathcal{R}\, x$. The elements of C are said to form an *antichain* in the poset $(A, \mathcal{R})$.

a) Find an antichain with three elements for the poset given in the Hasse diagram of Fig. 7.18(d). Determine a largest antichain containing the element 6. Determine a largest antichain for this poset.

b) If $\mathcal{U} = \{1, 2, 3, 4\}$, let $A = \mathcal{P}(\mathcal{U})$. Find two different antichains for the poset $(A, \subseteq)$. How many elements occur in a largest antichain for this poset?

c) Prove that in any poset $(A, \mathcal{R})$, the set of all maximal elements and the set of all minimal elements are antichains.

25. Let $(A, \mathcal{R})$ be a poset in which the length of a longest chain is n. Use mathematical induction to prove that the elements of A can be partitioned into n antichains $C_1, C_2, \ldots, C_n$ (where $C_i \cap C_j = \emptyset$, for $1 \le i < j \le n$).

26. a) In how many ways can one totally order the partial order of positive-integer divisors of 96?

b) How many of the total orders in part (a) start with $96 > 32$?

c) How many of the total orders in part (a) end with $3 > 1$?

d) How many of the total orders in part (a) start with $96 > 32$ and end with $3 > 1$?

e) How many of the total orders in part (a) start with $96 > 48 > 32 > 16$?

27. Let n be a fixed positive integer and let $A_n = \{0, 1, \ldots, n\} \subseteq \mathbf{N}$. (a) How many edges are there in the Hasse diagram for the total order $(A_n, \le)$, where "$\le$" is the ordinary "less than or equal to" relation? (b) In how many ways can the edges in the Hasse diagram of part (a) be partitioned so that the edges in each cell (of the partition) provide a path (of one or more edges)? (c) In how many ways can the edges in the Hasse diagram for $(A_{12}, \le)$ be partitioned so that the edges in each cell (of the partition) provide a path (of one or more edges) and one of the cells is $\{(3, 4), (4, 5), (5, 6), (6, 7)\}$?